Universidade do Minho
Instituto de Letras e Ciências Humanas

Anabela Alves dos Santos Rato

Cross-language Perception and Production of English Vowels by Portuguese Learners: The Effects of Perceptual Training

Doctoral Dissertation in Language Sciences
Specialization in English Linguistics

Supervised by
Professor Isabel Ermida
Professor Andreia Rauber

October 2013

To Andreia

ACKNOWLEDGEMENTS

First, I would like to thank my advisor, Professor Isabel Ermida, for accepting to supervise my PhD project, and for providing valuable advice, thorough comments and unwavering support.

I would also like to thank my co-advisor, Professor Andreia Rauber, for constant encouragement and guidance throughout this project. I am also grateful for her generosity in sharing her vast knowledge, and for her detailed and exhaustive feedback.

I thank Professor Ocke-Schwen Bohn for welcoming me at the University of Aarhus, Denmark, as a visiting student. Our monthly meetings were of vital importance at the beginning of this study, and attending his classes was enlightening and inspiring.

I would also like to thank Professor Susana Fernández for the opportunity to attend the master lectures on Second Language Acquisition. Thanks also to Professor Vinicius Carvalho, Dr. Mette Sørensen, and Fernanda Gláucia, for their hospitality during my stay at Aarhus University.

My thanks to Professor Denise Kluge for receiving me at the Federal University of Paraná, Brazil, and for discussing relevant research issues with me.

I would like to express my gratitude to Professor Ana Gabriela Macedo, Director of CEHUM, for her encouragement and support.

I would like to thank the Fundação para a Ciência e Tecnologia (FCT), Portugal, for the PhD grant SFRH/BD/63541/2009.

I am very grateful to Professor Henrique Barroso for introducing me to the thrilling science of speech, and for his friendship.

I would like to thank Professor Celeste Rodrigues for her interest in my study.

I also thank Adelina, Paulo, Ana Maria, and Vera, from CEHUM, for their availability to help me with practical matters whenever necessary.

Special thanks to my students for their willingness to participate in this study and, most importantly, for always motivating me.

My gratitude is also extended to Micaela, who helped me segment and annotate words of the production *corpora*.

Thanks to my dearest friends, Maria Alice and Salomé, for permanent care and moral support.

Habiba, Isabel, Joana, Sofia, and Tânia thank you not only for the many discussions about academic matters but also for the small talk that helped me cope with the daily work.

I would also like to thank Diana for her friendship and for proofreading the dissertation.

Thanks to the three most important women in my life, Anabela, Susete and Margarida for constant motivation, and for making me smile in the most difficult moments.

Thank you, Matilde, for your simple and pragmatic perspective on the study of speech sounds.

To my grandfather, Custódio, I would like to thank for inspiring me *in absentia* and for making me pursue my dreams.

Foremost, I would like to thank André for his endless patience, support and understanding, and for his love and care.

Braga, 4 October 2013.

ABSTRACT

Cross-language Perception and Production of English vowels by Portuguese Learners: The Effects of Perceptual Training

Several studies have demonstrated that second/foreign language (L2/FL) speech learning is a challenge to late learners (i.e., adolescents or adults) in terms of perception and production of certain non-native phonemic and phonetic contrasts (Moyer, 2013). The interaction of different factors might explain learners' difficulties, namely age of onset of learning (AOL), amount of native (L1) and non-native language (NNL) use over time, quantity and quality of NNL input, and the interference from the L1 phonological system (Piske, 2007). The *Speech Learning Model* (SLM), proposed by Flege (1995), hypothesizes that difficulties in perceiving and, consequently, in producing non-native contrasts are due to the (dis)similarities between the L1 and the NNL phonological systems. The L1 sound system is likely to hinder the formation of new non-native (L2/FL) phonological categories. However, a considerable number of cross-language studies has revealed that phonological learning is attainable for late learners, and their abilities in perceiving and producing segmental and suprasegmental non-native contrasts can improve, since the mechanisms used in the acquisition of the L1 sound system remain intact over the lifespan and can be applied to L2/FL learning (Flege, 1995). Experimental studies that investigated the effects of perceptual training on non-native speech sound perception and production reported its success not only in the modification of adult learners' perceptual patterns, but also in the improvement of their pronunciation accuracy, confirming, thereby, the plasticity of L2/FL learners' mature perceptual system (e.g., Aliaga-Garcia, 2013; Pereira & Hazan, 2013; Wang, 2008; Wang et al., 2003). Difficulties in the perception of non-native vowel contrasts have been widely described as a significant part of the problems learners have in L2/FL phonological acquisition/learning (Strange, 2007). Therefore, the present study investigated the effects of perceptual training on the learning of three English contrasts (/i/-/ɪ/; /ɛ/-/æ/; /u/-/ʊ/) by a group of EFL (English as a Foreign Language) learners. This set of vowel contrasts was selected due to reported difficulties European Portuguese native speakers have in perceiving and producing them (Flege, 1994, as cited in Flege, 1995; Rato et al., 2013). The English phonological categories /ɪ/, /æ/ and /ʊ/ tend to be assimilated to the Portuguese vowel sounds /i/, /ɛ/ and /u/, respectively, and no distinction between the two vowels of each pair is made, due to their acoustic and articulatory proximity. Specifically, this study investigated (i) whether a high variability perceptual training, which included stimuli with different phonemic contexts produced by multiple native talkers, had a positive effect on the perception of the English target segments; (ii) if transfer of improvement to oral production was observed; (iii) whether perceptual learning generalized to identification of new words produced by novel talkers; and (iv) if long-term training effects remained. The participants' perception was assessed three times with an identification test designed with natural stimuli: (1) before the auditory training – *pretest*; (2) immediately after the training was over – *posttest*; and (3) two months later – *delayed posttest*. The perceptual training program consisted of five sessions divided into two blocks, which included discrimination tasks and identification sequences followed by immediate feedback. Production was tested simultaneously in the three phases by means of a sentence-reading task with the target vowel segments. The results show that the Portuguese learners' performance in the identification of the English vowels improved significantly, and perceptual gains were retained two months after completion of the training sessions. Moreover, the results of the generalization test indicate that there was robust learning of the two front vowel pairs. Acoustic analyses of spoken data revealed that phonological learning transferred to production. In sum, these results support the claim that perceptual learning can occur in a formal non-naturalistic environment within a short period of time and corroborate previous findings on the malleability of L2/FL adult learners' perceptual systems.

RESUMO

Perceção e Produção de Vogais do Inglês por Aprendentes Portugueses: Os Efeitos do Treino Percetivo

Vários estudos têm demonstrado que, na aprendizagem de uma língua não materna (LNM), a perceção e a produção de determinados contrastes fonológicos e fonéticos não nativos são um desafio para aprendentes tardios (adolescentes ou adultos) (Moyer, 2013). Diversos fatores podem explicar essas dificuldades, tais como a idade do início da aprendizagem, a frequência de uso, a quantidade e a qualidade de exposição à LNM e a influência do sistema fonológico da língua materna (L1) (Piske, 2007). O *Speech Learning Model* (SLM), desenvolvido por Flege (1995), explica que as dificuldades percetivas e, consequentemente, produtivas se devem ao facto de o sistema de sons linguísticos da L1 impedir a formação de novas categorias fonológicas para a segunda língua ou língua estrangeira (L2/LE). No entanto, um número considerável de estudos tem revelado que aprendentes tardios podem aprender a perceber e a produzir contrastes segmentais e suprassegmentais não nativos, uma vez que os mecanismos usados para aprender o sistema de sons da L1 são ativados na aprendizagem de uma LNM e permanecem intactos durante toda a vida (Flege, 1995). Estes estudos empíricos, que investigaram os efeitos do treino percetivo na perceção e produção de sons não nativos, reportaram a sua eficácia não somente na modificação de padrões percetivos, mas também na melhoria da capacidade produtiva dos mesmos, confirmando assim a plasticidade do sistema percetivo dos aprendentes de LNMs (por exemplo, Aliaga-Garcia, 2013; Pereira & Hazan, 2013; Wang, 2008; Wang et al., 2003). Os resultados de extensa investigação indicam que as dificuldades na perceção de contrastes vocálicos não nativos são uma parte significativa dos problemas que os aprendentes revelam na aquisição/aprendizagem fonológica de uma L2/LE (Strange, 2007). Portanto, no presente estudo, investigaram-se os efeitos do treino percetivo na aprendizagem de três contrastes vocálicos da língua inglesa (/i/-/ɪ/; /ɛ/-/æ/; /u/-/ʊ/) por um grupo de aprendentes de inglês como LE. Estes três contrastes foram escolhidos devido às dificuldades percetivas e produtivas que falantes nativos de português europeu revelam na sua aprendizagem (Flege, 1994, citado em Flege, 1995; Rato et al., 2013). As categorias fonológicas /ɪ/, /æ/ e /ʊ/ da L2/LE tendem a ser assimiladas como sons da L1, /i/, /ɛ/ e /u/, respetivamente, não se verificando qualquer distinção entre as vogais dos três contrastes, devido à sua proximidade acústica e articulatória. Especificamente, pretendeu-se (i) observar o efeito de um treino percetivo de alta variabilidade, que incluiu estímulos produzidos por vários locutores nativos em diferentes contextos fonológicos, na melhoria da capacidade percetiva dos segmentos-alvo; (ii) averiguar a transferência da melhoria para a produção oral; (iii) verificar a generalização para novos contextos e novos falantes; e (iv) analisar os efeitos do treino a longo prazo. A perceção dos participantes foi testada três vezes com uma tarefa auditiva de identificação com estímulos naturais: (1) antes do treino – *pré-teste*; (2) imediatamente depois do treino – *pós-teste*; e (3) dois meses mais tarde – *teste de retenção*. O programa de treino consistiu em cinco sessões, divididas em dois blocos, que incluíram tarefas de identificação e de discriminação auditivas seguidas de correção imediata. A produção foi testada, igualmente, em três fases, através da leitura de frases veículo, contendo palavras com os segmentos vocálicos. Os resultados demonstram que os aprendentes portugueses melhoraram significativamente na identificação das vogais-alvo e essa melhoria da sua competência percetiva manteve-se dois meses após o término do treino. Para além disso, os resultados do teste de generalização indicam que houve uma aprendizagem robusta dos dois contrastes vocálicos anteriores. As análises acústicas das produções dos informantes revelaram também uma transferência da aprendizagem para a produção oral. Estes resultados suportam a afirmação de que a aprendizagem ao nível da perceção de fala pode ocorrer em contextos formais, num curto período de tempo, e corroboram resultados anteriores sobre a maleabilidade dos sistemas percetuais fonológicos de aprendentes adultos de uma LNM.

TABLE OF CONTENTS

List of Abbreviations

2L1 – Two native languages

A – Audio

AE – American English

AFC – Alternative forced-choice

AOA – Age of arrival

AOL – Age of onset of learning

ART – Articulatory

ASR – Automatic speech recognition

AV – Audiovisual

BP – Brazilian Portuguese

CEF – Common European framework of reference for languages

CG – Control group

CLT – Communicative language teaching

DISC – Discrimination

EFL – English as a foreign language

EG – Experimental group

ELF – English as a *lingua franca*

ELL – European languages and literatures

ENL – English as a native language

EP – European Portuguese

ESL – English as a second language

FA – Foreign accent

FL – Foreign language

GA – General American

GT – Generalization test

HVPT – High variability phonetic training

ID – Identification

IPA – International phonetic alphabet

ISI – Interstimulus interval

L1 – Native language

L2 – Second language

LFI – Length of formal instruction

LOR – Length of residence

LVC – Lax vowel constraint

LVPT – Low variability phonetic training

NCCS – Northern cities chain shift

NES – Native English speaker

NLM – Native language magnet

NNL – Non-native language

NNS – Non-native speaker

NS – Native speaker

NVR – Natural referent vowel

PAM – Perceptual assimilation model

RP – Received pronunciation

RT – Response time

SLA – Second language acquisition

SLM – Speech learning model

SVS – Southern vowel shift

UM – University of Minho

V – Visual

List of Figures

List of Tables

INTRODUCTION

The emergence of economic, social and cultural international networks, a phenomenon known as *globalization*, boosted substantially by the advent of communication technologies, and in particular by the Internet, has undoubtedly contributed to the spread of English around the world, and to uphold its status as a global language. English is, currently, the chief language, either with an official or working role, of international communication in several different settings that range from business meetings, political gatherings, academic conferences, international conventions, sports events, military occupations, transport operations to online community rallies. In addition, English is the main language of popular culture, including cinema, pop music, satellite broadcasting, video games, and the Internet, and it is the medium to publish a great deal of the world's academic, scientific and technological knowledge. The importance of English as an international *lingua franca* has made it the language most widely taught as a foreign language (FL) in over 100 countries (Crystal, 2003), and Portugal is among these countries in which English is a priority in FL teaching. Despite the fast growing interest and integration of Spanish in the FL formal instruction settings, which has been observed over the last years, English is still the major foreign language to be taught to children and adult learners.[1] Taking into account the spread of English around the world and its users, it is generally accepted that there are three distinct classifications of world English(es), namely English as a native language (ENL), a second language (ESL) and a foreign language (EFL). Jenkins (2009) clarifies:

> English as a native language (ENL) is the language of those born and raised in one of the countries where English is historically the first language to be spoken (mainly the UK, USA, Canada, Australia, and New Zealand). Their English speakers are thought to be around 350 million. (…) English as a Second Language refers to the language spoken in a large number of territories such as India, Bangladesh, Nigeria and Singapore, which were colonized by the English. These speakers are also thought to be around 350 million. English as a Foreign Language is the English of those for whom the language serves no purposes within their own countries. (…) the number is likely to be around 1 billion (…). (pp. 15-16)

[1] This seems to be, to some extent, a replication of the global situation. Although English is still the main language of international communication, there is some debate about the possibility of favoring Spanish in detriment of the former language. According to Jenkins (2009, p. 52), Spanish has less complex spelling, grammar, and phonology and a less colonialist discourse, and its influence has been increasing in both the European Union and the United States.

This three-way categorization is thoroughly discussed by Jenkins (2009), but some aspects should be emphasized, namely the fact that the three categories do not fully account for todays' bi- and multilingual reality, nor for the particular case of EFL countries, such as The Netherlands and Scandinavian countries where English is being used for country internal (i.e., intranational) purposes rather than solely as a foreign language. Although Portugal seems to be moving towards the latter trend,[2] the EFL classification is still the most adequate to describe the current status of English in Portugal, because it is not actually used or spoken very much in the normal course of daily life, but it is learned at school.

An alternative classification advanced by Kachru (1992) is the model of the spread of English. The author divides the world English(es) into three concentric circles, the Inner Circle, the Outer Circle and the Expanding Circle, and describes the English spoken in each circle as "norm-providing", "norm-developing", and "norm-dependent", respectively (Kachru, 1992). In this model, the EFL varieties are regarded as "performance" varieties with no official status used in EFL contexts, and thus dependent on language standards set by native speakers in the Inner Circle. Although the model provides a framework for understanding the spread of English, it has some limitations. For a detailed discussion of the model, see Jenkins (2009).

The previous classifications are closely related to the distinction between native and non-native speakers of English, that is, between those born to the language and those who learnt it through education. This distinction has been questioned within the world English research field, and there seems to be some consensus that it is only suitable for EFL, but not for English as a *lingua franca* (ELF).[3] Henceforth, the distinction native/non-native English speaker is adopted in this study.

The importance of English as an international language and the high number of non-native speakers[4] has raised the debate about the relevance of pronunciation instruction in formal EFL environments. Furthermore, the recently published work on

[2] Since 2008, English learning is part of the primary instructional curriculum as an extracurricular (i.e., non-compulsory) activity (DL 14460/2008, 26 May). Moreover, at tertiary educational level, English is occasionally used as the working language in lectures of some graduate and postgraduate courses. For example, at the University of Minho, some of the classes in the International Relations graduate course are taught in English (http://www.eeg.uminho.pt/). Finally, the importance of English as a foreign language has been recently acknowledged and reinforced by the Ministry of Education with the introduction of a compulsory national exam at the 9th grade (Desp. 11838-A/2013, 11 September). Until the current academic (2013-2014) year, only Mathematics and Portuguese were assessed by means of a national exam.

[3] Some alternatives to the native/non-native distinction have been suggested: "expert/non-expert" (with a similar connotation to that of native/non-native), and "monolingual English speaker" (MES), "bilingual English speaker (BES) (for proficient speakers in English and in another language), and "non-bilingual English speaker" (NBES) (for reasonably competent speakers, who are potential BES) (Jenkins, 2009, p. 90).

[4] Non-native English speakers greatly outnumber native speakers (Jenkins, 2007, 2009; Gonçalves, 2007).

foreign accent, written by Moyer (2013), shows that the phenomenon of non-native speech is presently a “matter of great public interest given the impact of migration on national and global affairs” (foreword). Every act of international speech communication depends on mutual intelligibility, that is, on the understanding of utterances at the acoustic-phonetic level and on expectations about the speakers’ intended meaning (Moyer, 2013, p. 180). Thus, accurate pronunciation of non-native sounds is essential for effective international spoken communication, and for a successful integration into the wider society. Conversely, a strong foreign accent can affect intelligibility and cause a breakdown in communication. Accent is fundamental to communication, for without a reasonable degree of phonological fluency, spoken interaction will falter. However, second language phonology acquisition or learning might be challenging to L2 users, because, as Moyer (2013) concisely explains:

> They must learn to perceive fine phonemic differences and establish a new system of phonological rules; produce sounds and sound sequences that often contradict the rules of their native languages; and replicate the patterns of stress, rhythm, intonation that carry implicit as well as explicit meaning. (p.1)

The range of potential difficulties is very wide, varying from phonetic contrasts to phonotactic constraints and prosody. As cited previously, non-native speakers may have difficulties with segments (viz. contrastive phonological features and allophonic realizations), syllable structure (viz., phonotactic rules), and suprasegments (viz. stress, segment duration, intonation, pitch, speech rate, rhythm and timing). These errors might occur not only at the level of production, but also, and primarily, at the level of perception. Production difficulties consist mainly in the inability to accurately articulate specific sounds, and perceptual errors in a failure to discriminate contrasting phonemes and identify phonetic segments. Thus, adult L2/FL learners are frequently characterized as having not only a foreign pronunciation, but also accented perception (Strange, 1995, p. 2).

Within the range of potential segmental errors, non-native vowel acquisition has been particularly highlighted as problematic. The findings of extensive cross-language speech research have led Strange (2007) to emphasize that “difficulties in perception of non-native vowel contrasts are a significant part of the problems many L2 learners have in mastering the L2 phonology” (p. 36). Daniel Jones (1935, as cited in Collins & Mees, 2008) also states that it is “easier to change the pronunciation of consonants than that of vowels” (p. 229).

In sum, foreign-accented speech is detected when there are divergences between the L1 and L2 phonemic/phonetic systems both at segmental and suprasegmental levels (Flege, 1995). The implications of sounding foreign, that is, of having a foreign accent can be various (see Moyer, 2013, for a thorough discussion). On the one hand, it has an immediate impact on intelligibility (i.e., on the negotiation of meaning), and, on the other hand, a long term influence on L2 speakers' social integration (e.g., Bresnahan, Ohashi, Nebashi, Liu, & Shearman, 2002; Giles, Williams, Mackie, & Rosselli, 1995).

Age of first exposure to the non-native language (NNL), that is, age of onset of learning (AOL) is widely cited as the main neurobiological predictor of success in L2 phonology learning, and there is widespread agreement on the premise "the earlier, the better". Related with this premise is Lenneberg's (1967) critical period hypothesis,[5] which predicted that neural plasticity[6] declines around puberty, when neurological maturation achieves its developmental peak. This would explain adult learners' phonological difficulties. However, regardless of the difficulties L2/FL learners' may face, Flege (1995) claims that the perceptual ability to learn non-native sounds is available throughout the lifespan. Several studies have provided evidence of the plasticity of adult learners' perceptual systems by showing that auditory training can improve both perception and production of non-native phonemic/phonetic segmental and suprasegmental contrasts. Research in speech perception and production has led Flege (1995) to suggest that "if a critical period exists, it does not result in a sharp discontinuity in L2 pronunciation ability at around puberty" (p. 234). Despite advances in second language acquisition (SLA) research, age-related cessation of neural plasticity is still a theme of debate among researchers (Moyer, 2013, p. 182).

Various other factors determine the outcomes of non-native phonological development in second/foreign language. Hence, difficulties with L2/FL sounds result from the interaction of diverse factors related to both the learner and the learning context. Extrinsic factors include learners' L2 background and experience, viz. age of onset of L2 learning (AOL), length of residence (LOR) in an L2-speaking environment, quantity and quality of L2 exposure, amount of L1 and L2 use; and intrinsic factors consist of individual differences such as motivation, memory, and language learning aptitude (Munro & Bohn, 2007; Moyer, 2013; Piske, 2007). Another main factor,

[5] Lenneberg (1967) predicted that accent in a foreign language would be especially difficult to acquire beyond age 9 or 10 years.
[6] Plasticity is the flexibility of neural structures associated with lateralization, the assignment of specific functions, such as auditory and visual processing, to a specific neural area located either in the right or left hemisphere (Lenneberg, 1967; Penfield, 1965).

mentioned earlier, that accounts for phonological acquisition is L1 attrition, that is, the (dis)similarities between the L1 and the NNL phonemic and phonetic systems.

Research investigating the influence of formal L2/FL classroom instruction on degree of foreign accent has revealed that instructional factors do not or hardly affect non-native pronunciation (Piske, MacKay, & Flege, 2001; Piske, 2007). One of the reasons that might explain this is the minor role of L2/FL phonological learning in most foreign language classrooms, being often considered as a “low priority area of study” (Hewings, 2004). The findings of these studies seem to indicate that amount of L2/FL classroom instruction does not lead to any significant decrease in degree of non-native accent (e.g., Rauber, Rato, & Silva, 2010). However, the interaction of AOL and other variables might prove that the previous claim is not well founded (see Piske, 2007, for a discussion on this issue). On the one hand, there is little evidence that amount of formal instruction affects degree of L2 foreign accent, but, on the other hand, research suggests that if L2 classroom teaching involves intensive and adequate training in the perception and production of non-native sounds, it will have a larger effect on L2 pronunciation accuracy. Moyer (2013) argues that the L2/FL classroom is an adequate context for interactive practice and targeted feedback, thus, offering some advantages over informal immersion experiences.

Classroom approaches to phonological instruction include, for example, Audiolingualism in the 1950s and 1960s, which focused on pronunciation training consisting of a set of drills for imitation and repetition, aimed at eradicating and preventing errors so as to sound native-like. Some audiolingual techniques can still be found in today’s classroom books, such as imitating recorded native speakers as models; reading aloud and reciting words, sentences, and longer passages to focus on prosodic features; practicing tongue twisters; practicing transcription with the International Phonetic Alphabet (IPA), among others (Moyer, 2013, p. 149). According to Moyer (2013) the same type of exercises has not changed over the years, their only innovation being the administration through new technologies. This approach was replaced by Cognitivism, in the 1970s, according to which errors were considered as a natural outcome of the learning process, inherently surpassed over time as long as sufficient input was provided. Experimental methods (e.g., Total Physic Response, Community Language Learning) were followed by the Communicative Language Teaching (CLT) in the 1980s, which considered accent as a feature of linguistic competence, not of communicative competence. This is still the dominant approach today, and it emphasizes the importance of global fluency rather than discrete-segmental accuracy. Therefore, accent is not considered as essential to convey meaning, but grammar and vocabulary are. This explains, to some extent, the little attention that focused pronunciation training receives nowadays in the L2/FL classroom, and also some teachers' lack of a formal phonetics background.

As aforementioned, the EFL classroom offers an adequate scenario for phonological learning and for improvement of accurate pronunciation and fluency. Phonology instruction in an L2/FL classroom context provides a unique opportunity to raise learners' phonological awareness and to practice target areas of difficulty, with teachers' feedback and correction. Studies further reported provide evidence that adequate phonetic training can lead to both short- and long-term gains in the two speech dimensions, namely perception and production.

Another question that is raised in phonological learning and teaching in the L2/FL classroom is related to its role among other communicative skills and to the guidelines for different levels of learners' proficiency. The need to establish standards in today's bilingual and multilingual reality has led the Council of Europe to establish guidelines for L2/FL learning, in 2008. This set of standards, known as the *Common European Framework of Reference for Languages* (CEF), details reading, speaking, listening, and writing skills. Within the broad area of communicative language competence, linguistic competence is one subdomain (alongside sociolinguistic and pragmatic competencies), and phonology is one of six other parameters. Skills are divided into three proficiency levels ("basic", "independent" and "proficient"), and in terms of phonological abilities, six levels are described (see Appendix A). The descriptors emphasize the communicative importance of accent, though there are some gaps between the different levels, which give the impression that phonological fluency is simultaneous to overall fluency (Moyer, 2013).

In sum, many scholars still debate the relevance of pronunciation instruction in formal environments. Some argue in favour of communicative skills disregarding the role of phonological information in foreign language learning. However, even though communicative competence should indeed be seen as a primary goal, and not all EFL learners aim at nativeness,[7] that is, at sounding native-like, intelligibility should be emphasized and that can be quite challenging when pronunciation is extremely faulty

[7] Nativeness is defined by Moyer (2013) as "a speaker's ability to produce or perceive aspects of a language on par with its native speakers" (p. 181).

(Becker, 2013). Another aspect that must also be taken into account is that different learners have different learning goals (Celce-Murcia, Brinton, & Goodwin, 1996; Morley, 1991). Some might need to acquire a more flawless pronunciation than others, which is the case, for instance, of prospective teachers of English, the target participants of the present study. Thus, this experimental study aims at investigating how effective phonological instruction will be for this specific group of learners.

Several studies have shown that the L1 phonological system has direct influence on the perception and production of the sounds of an L2/FL (e.g., Major, 1987; Flege, 1987, 1995; Flege, Schirru & Mackay, 2003; Rochet, 1995), which makes the learning of a non-native phonemic/phonetic system a challenge. Among these, research has been carried out focusing on the perception and production of English vowels by learners with different L1 vowel systems (e.g. Bohn & Flege, 1990, 1992; Cebrian, 2009; Escudero & Chládková, 2010; Flege, Munro, & Fox, 1994; Flege, Bohn, & Jang, 1997; Flege, Mackay, & Meador, 1999; Iverson & Evans, 2007a, Kim, 2003; Lengeris & Hazan, 2007; Lengeris, 2009; Munro, 1993; Tsukada et al., 2003; Rauber, Rato, & Silva, 2010), including the Brazilian Portuguese vowel inventory (Bion, Escudero, Rauber, & Baptista, 2006; Rauber, Escudero, Bion, & Baptista, 2005; Rauber, 2010). These studies have provided insightful information about the perceptual and production difficulties adult L2/FL users might have when acquiring/learning non-native speech sounds. Concerning the perception-production relationship, research has proved perceptual training to be effective in the improvement of non-native speakers' ability to perceive and to produce foreign language sounds (e.g., Bradlow, 2008; Bradlow, Yamada, Pisoni, & Tohkura, 1999; Hardison, 2003; Hazan, Sennema, Iba, & Faulkner, 2005; Jamieson & Morosan, 1986; Logan & Pruitt, 1995; Ortega-Llebaria, Faulkner, & Hazan, 2001; Pisoni, Aslin, Perey, & Hennessy, 1982; Strange & Dittmann, 1984; Wang, Jongman, & Sereno; 2003), and a study, particularly interesting to formulate the hypotheses of the present study, is that by Nobre-Oliveira (2007, 2008) on the perception and production of English vowels by native Brazilian Portuguese speakers.

The relevance of the present study is discussed next. First, to our knowledge, research on the perception and production of English vowels by European Portuguese native speakers is scarce (e.g., Flege, 1994, as cited in Flege, 1995; Rato, Rauber, Soares, & Lucas, 2013), and there has been no research that analyzes the effects of perceptual training on pronunciation of English vowels by Portuguese learners, although there are some studies that investigated the perceptual training of English vowels by learners with different L1 vowel systems (e.g., Aliaga-Garcia, 2013; Aliaga-Garcia & Mora, 2009; Iverson & Evans, 2007a, 2007b, 2009; Lambacher et al., 2005; Lengeris & Hazan, 2010; Lengeris, 2008; Pereira & Hazan, 2013; Munro, 1993; Nobre-Oliveira, 2007, 2008; Wang, 2008; Wang & Munro, 2004). Nonetheless, not many have investigated transfer of learning to production, in particular by means of quantitative analyses of L1/L2 acoustic similarity, that is, acoustic measurements of duration and formant frequency. Previous findings, reporting positive effects of certain training methods, contributed to the development of the experimental design of this study. Therefore, a high variability phonetic training (HVPT) approach will be adopted (Logan, Lively, & Pisoni, 1991; Lively, Logan, & Pisoni, 1993), within an EFL classroom context. The HVPT that will be carried out in this study will be a perceptual audio-only training, which will include natural stimuli produced by multiple talkers and will consist of discrimination and identification tasks. In addition, to guarantee that potential phonological learning is promoted by vowel-centered training tasks, and not by task familiarization, the control group, to which no training is normally administered, will undertake a similar training program, but focused on consonants. Therefore, the present study is novel in the choice of the L1-L2 pairing, in the type of production analyses to be made, and in the active participation of the control group of EFL learners.

The research questions are based on the findings of two studies about English vowel acquisition by European Portuguese speakers. Flege (1994, as cited in Flege, 1995) reported discrimination failure of vowel contrasts /æ/-/ɛ/, /ɛ/-/ɪ/, /u-/ʊ/, and /ʌ/-/ɑ/ (p. 249), and Rato et al. (2013) observed difficulties in the identification and production of the two front vowel contrasts /i/-/ɪ/ and /æ/-/ɛ/. Thus, considering the documented difficulties that Portuguese EFL learners have in perceiving and producing English vowels (Flege, 1995; Rato et al., 2013), we will now focus on the research objectives and hypotheses of the study.

In 1935, Daniel Jones (as cited in Collins & Mees, 2008, p. 229) claimed that one of the two aims of a phonetician is “to cause his pupils to perform unaccustomed movements with their organs of speech; in other words, to pronounce new sounds or new combinations of sounds”. In line with this statement, the main pedagogical objective of the present study, conducted in an EFL formal instructional context, is to raise Portuguese learners’ awareness of L1-L2 perceptual and articulatory (dis)similarities and to improve their abilities to perceive and pronounce English vowel sounds.

By combining the areas of phonetics and second language acquisition (SLA), this cross-language study aims at investigating the efficiency of perceptual training in the improvement of perception and production of English vowels by Portuguese EFL learners. The specific research objectives are: (1) investigate whether a high variability perceptual training has a positive effect on the perception of the English vowel contrasts /i/-/ɪ/, /æ/-/ɛ/, and /u/-/ʊ/ by native European Portuguese speakers; (2) examine if long-term potential effects of auditory training will remain after training is over; (3) by means of a generalization test, observe whether perceptual learning will transfer to identification of new words produced by novel talkers; (4) explore the relationship between L2 perception and production by investigating transfer to production improvement. Furthermore, (5) an examination of perceptual assimilation patterns and phonetic realizations of the target English vowel sounds is carried out, in three testing moments, to find which target vowel pairs will be more easily perceived and produced by the EFL listeners/speakers before, immediately after training, and two months later; and (6) which acoustic cues (duration or spectral quality) EFL speakers rely on to produce non-native vowels. To our knowledge, only a few studies about native Portuguese learners' perception and production of English vowels have been carried out (Flege, 1995; Rato et al., 2013). Thus, as aforementioned, the hypotheses are mostly based on these studies and on previous research with Brazilian EFL learners.

The hypotheses for each research question are the following: (1) with a limited number of training sessions, perceptual training will have a positive effect on the perception of English vowel contrasts, that is, the ability to identify non-native vowels will improve (e.g., Aliaga-Garcia, 2013; Nobre-Oliveira, 2007; Pereira & Hazan, 2013; Wang, 2008); (2) long-term positive effects of auditory training will still be observed two months after its completion (e.g., Nishi & Kewley-Port, 2007; Nobre-Oliveira, 2007; Wang, 2008); (3) generalization of perceptual learning will transfer to new talkers (e.g., Aliaga-Garcia, 2010; Nobre-Oliveira, 2007; Wang, 2008), and to new tokens (e.g., Aliaga-Garcia, 2010; Lacabex, Lecumberri, & Cooke, 2009; Nobre-Oliveira, 2007); (4) high variability perceptual training will also lead to production improvement even without any specific production training, as found in previous studies (e.g., Aliaga-Garcia, 2013; Lacabex and Lecumberri, 2010; Lambacher, 2005; Lengeris, 2008; Pereira & Hazan, 2013), showing there is a link between perception and production; (5) based on previous studies (Bion et al., 2006; Nobre-Oliveira, 2007; Rauber, 2010), the least difficult pair to perceive and produce will be /i-ɪ/, followed by /ʊ-u/, and then /ɛ-æ/; (6) EFL learners will rely on both vowel duration and quality to produce the English vowels as observed in previous studies (e.g., Rato et al., 2013; Rauber, 2010), but cue weighting will be used differently from native speakers of American English; that is, durational and spectral distinctions between vowels of the target contrasts will be less salient.

As far as the structure of this study is concerned, the present dissertation is divided into four chapters, which are briefly summarized as follows.

Chapter 1 outlines the acoustic principles of speech production and summarizes the articulatory and acoustic features of vowel sounds. In particular, it describes the American English and the European Portuguese vowel systems and reports experimental studies on the production of vowels by native speakers of both languages. Finally, it reviews a few studies on L2 speech production.

Chapter 2 describes the human hearing process, and characterizes vowels as perceptual phonemic categories. Speech perception theoretical models are summarized and complemented with a review of studies on cross-language speech perception. Moreover, it discusses the interface between L2 speech perception and production, and the effects of perceptual training on both dimensions.

The design and method of the study are described in *Chapter 3*. Information about the participants, the testing and training materials, and the procedures adopted to collect and analyze the perception and production data is provided.

Chapter 4 describes the acoustic and statistical data analyses and presents their results. Furthermore, it discusses the effects of perceptual training on vowel perception and production, and examines the interrelation between the two speech domains.

Finally, conclusions about the main findings are provided and related to the research hypotheses formulated for the present study. The limitations of the study are acknowledged, and suggestions for further research are presented.

CHAPTER 1

SPEECH PRODUCTION

Every speech that is spoken (…) in his substance is but air.
- Chaucer

This chapter provides some background to speech production by briefly explaining its acoustic theory and by introducing some of the properties of speech sounds, in particular the acoustic characteristics of vowels. Moreover, it describes the American English and the European Portuguese vowel inventories in articulatory and acoustic terms and reviews some experimental studies on the production of vowels by native speakers of both languages. Finally, it summarizes a few contributions to the investigation of L2 speech production.

1.1 The Acoustics of Speech Production

The production of speech is a complex process that consists of different stages, including *conceptualization*, *formulation* and *articulation* (Levelt, 1989; Levelt, Roelofs, & Meyer, 1999). The *conceptualization* phase involves determining what to say, *formulation* implies translating the conceptual representation into a linguistic form,[8] and *articulation* implicates motor execution of phonetic gestures. The description of the functioning of these cognitive processes is beyond the scope of this chapter, as well as the study of the nervous system that controls the articulation of speech. Nonetheless, the present study aims at analyzing the output (i.e., the sound waves of speech sounds) of the *articulation* processing stage (see Figure 1). Therefore, the focus of this chapter is on the physiological production of speech and on the physical properties of speech sounds, represented by the left and central parts of Figure 2.

[8] The *formulation* phase includes (1) lexical selection, (2) morphological and phonological encoding (syllabification), and (3) phonetic encoding (Levelt et al., 1999).

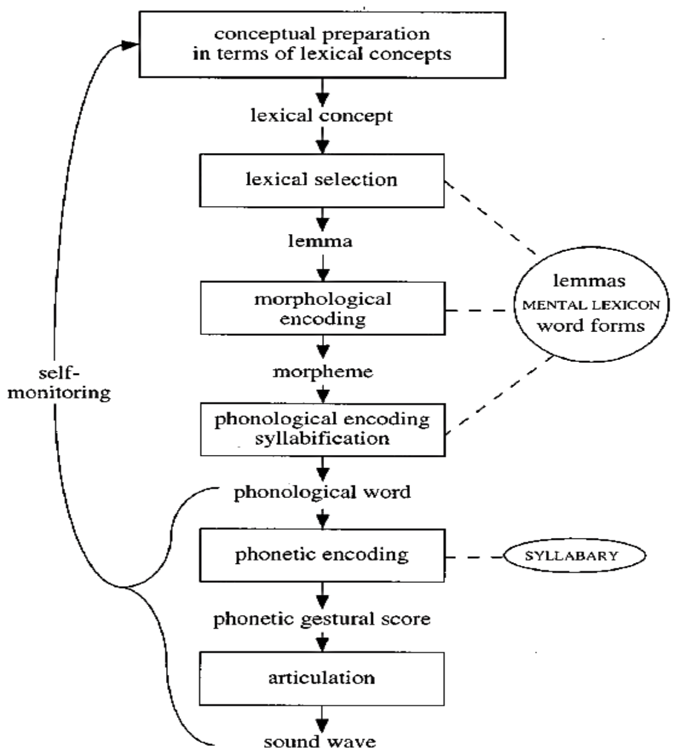


Figure 1. Flow diagram of the theory explaining the underlying cognitive processes involved in speech production (Levelt et al., 1999, p.3).

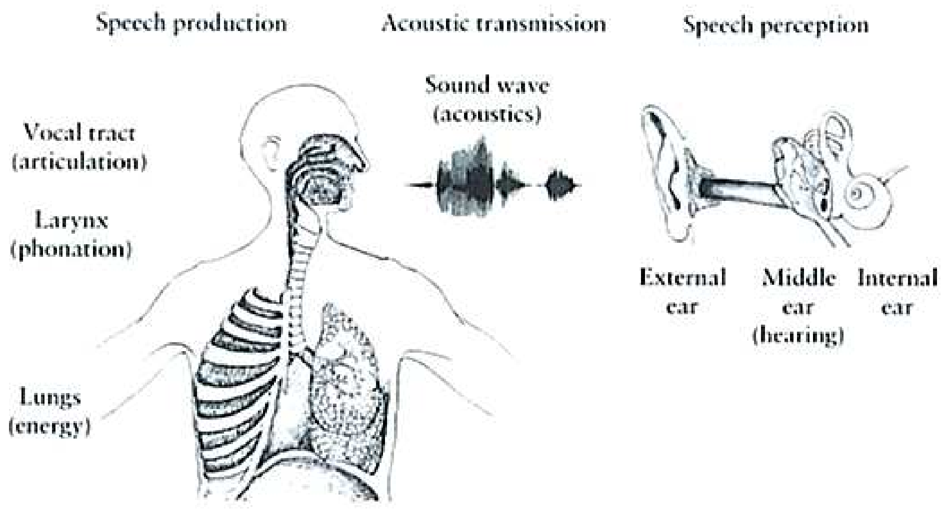


Figure 2. The speech chain (Reetz & Jongman, 2009, p. 2).

In order to understand the physiology of speech, we need to divide the speech production system into three components that correspond to anatomical structures: (1) the subglottal structure, which includes the lungs and the respiratory muscles, located below the larynx; (2) the larynx; and (3) the supralaryngeal vocal tract above the larynx, which consists of the pharynx and the nasal and oral airways (Lieberman & Blumstein, 1988; Reetz & Jongman, 2009; Stevens, 1998). The three anatomical structures are represented in the left part of Figure 2: (1) the lungs, which generate the energy necessary for phonation[9] occurring at (2) the larynx, and (3) the vocal tract, where speech sounds are articulated.

The main source of acoustic energy[10] in the production of phones is the airflow generated by the subglottal components. This egressive (outward) pulmonic airstream is converted into a sequence of periodic pulses of air produced by the rapid inward and outward movement of the vocal folds, located at the larynx. The quick movements of the vocal folds open[11] and close the glottis (the space between the vocal folds), and it is when they briefly draw apart (i.e., abduct) that air flows into the supraglottal cavities. The source of acoustic energy is lastly modified at the supralaryngeal vocal tract, which functions as a filter that helps modulate sounds (Lieberman & Blumstein, 1988). This physiological description of speech production is related to the acoustic principles that underlie the articulation of speech sounds, as explained next.

Most speech sounds, viz. voiced sounds, are produced by the vibration of the vocal folds, but others, viz. voiceless sounds, can be generated either by burst noises (e.g., stop consonants) or turbulent noises (e.g., fricatives),[12] or they can also be the result of a combination of the two, such as voiced fricatives and voiced stops[13] (Hayward, 2000; Mateus, Falé, & Freitas, 2005). In terms of acoustics, these different sound sources produce either periodic sound waves,[14] with regular repetition of

[9] Phonation (or voicing) is the vibration of the vocal folds.
[10] Hayward (2000) calls it the "power supply for speech" (p. 210).
[11] For the vocal folds to open, an amount of 5 to 10 cm H_2O of subglottal air pressure must build up below them (Lieberman & Blumstein, 1988, p. 92; Stevens, 1998, p. 4).
[12] Noise is caused by a constriction somewhere along the vocal tract. The burst noise occurs at the moment of stop consonant release, and the turbulent noise is produced in the articulation of fricatives, as airflow escapes from a narrow constriction in the vocal tract. Reetz and Jongman (2009, p. 162) add another possible type of sound source, namely the vibration of the tip of the tongue in the production of a trilled /r/.
[13] In the case of voiced fricatives, turbulent noise is superimposed on the periodic waveform associated with vocal fold vibration, whereas in the case of voiced stops, a periodic wave can be observed during the closure interval, when the vocal folds are vibrating, but this will be followed by a transient, corresponding to the release of the oral closure. Voiced stops combine periodic and aperiodic sound sources, which occur in sequence (periodicity followed by aperiodicity), rather than simultaneously as in voiced fricatives (Hayward, 2000, p. 30).
[14] Complex periodic waveforms are characteristic of voiced sonorants, i.e., vowels and sonorant consonants (Hayward, 2000, p. 28).

frequency[15] patterns, or aperiodic waveforms, with no repeating pattern, or a combination of both. Within the aperiodic waveforms there are continuous (noisy) aperiodic, in which there is no interruption of the airflow, and non-continuous (transient) aperiodic, in which there is a complete obstruction of the airflow[16] (Hayward, 2000, p. 28).

The vibration of the vocal folds generates complex periodic sound waves, whose first frequency is called the *fundamental frequency* (F0). A complex periodic waveform includes a set of frequencies multiple[17] of F0, which are called *harmonics*. The fundamental frequency is also referred to as *first harmonic*, because it is the first regular frequency vibration of a sound. The rate of the abduction-adduction movement of the vocal folds determines the period[18] and, consequently, the fundamental frequency of the glottal airflow. The F0 rate also depends on speakers' physical features associated with age and gender, such as size, length, and density of the vocal folds. Therefore, typical F0 values are 120 Hz for men, 210 Hz for women, and over 300 Hz for children. The higher F0, which is perceived as a higher pitch, is correlated with a faster vibration rate that results from the smaller vocal folds of women and children when compared to those of men (Mateus et al., 2005; Traunmüller & Eriksson, 1995).

The complex periodic wave forms, generated by the vibration of the focal folds (F0 and harmonics) are modified and filtered when passing through the supraglottal cavities. The vocal tract reduces airflow energy at certain frequencies and amplifies energy at other frequencies, which are the resonance frequencies, also called *formant frequencies*. Therefore, different length and shape configurations of the supralaryngeal vocal tract result in particular formant patterns. Formants are typical of sounds produced with resonance in the oral cavities, namely vowels, semi-vowels, liquids, and nasals (Mateus et al., 2005).

The above description of speech production as a system of sound sources (periodic and aperiodic), which generate sound, and filters (supraglottal cavities), which modify the sound produced by these sound sources, is a summary of the acoustic theory of speech production, also known as the *source-filter theory* (Fant, 1960). This theory,

[15] The frequency of a waveform is the number of times a complete cycle of vibration is repeated per second. The commonly used unit for frequency is the Hertz (abbreviated Hz), where 1 Hertz=1 vibration/second. For example, in a 300 Hz waveform, 300 complete cycles of vibration repeat in a second (Delgado-Martins, 1998; Mateus, Falé, & Freitas, 2005).

[16] As earlier mentioned, the aperiodic noisy waveform is typical of voiceless fricatives, and the transient aperiodic of stop consonants (Hayward, 2000, pp. 28-29).

[17] For example, if F0 has a frequency of 100 Hz, the second harmonic will have a frequency of 200 Hz, and the third harmonic a frequency of 300 Hz.

[18] The period is the time to make one complete vibrational cycle, and it is measured in seconds.

based on the relation between the articulatory gestures of speech sounds and their specific acoustic characteristics, can predict the resonance frequencies of the vocal tract when unconstricted, as in the production of schwa (/ə/). The vocal tract configuration to produce schwa can be seen as a uniform tube open at one end (lips) and closed at the glottis. The airstream flowing in the vocal tract resonates in contact with the tube, producing regular frequency responses (i.e., peaks), regularly spaced and separated by valleys. The frequencies of the peaks are the formant frequencies.

The resonance frequencies of a tube open at one end, such as in the production of schwa, can be calculated, as indicated next. The resonance frequencies are odd multiples resulting from the formula $f=c/4l$ (where f is frequency, c is velocity of sound in air, and l is length of the tube). If c=340meters/second (34000cm/s), and l=17 cm (the length of a male vocal tract), the result of f is 500=34000/(4x17). Hence, if F1 is 500 Hz, and the other formants are multiples of F1, F2=1500 Hz, and F3=2500 Hz (Hayward, 2000, p. 83; Johnson, 2003, pp. 95-96; Reetz & Jongman, 2009, p. 164). However, to explain the acoustic properties of other vowels, it is necessary to include constrictions at different locations (e.g., lip protrusion, larynx lowering) to resemble the configuration of the vocal tract. Therefore, to calculate the resonance frequencies of other vowels it would be necessary to add more tubes and, consequently, more calculations; however, this is not our purpose. Conversely, describing the acoustic correlates of different vocal tract configurations is important to understand the next sections of this chapter. As previously stated, the articulatory and acoustic features of speech sounds are correlated, and, therefore, a brief explanation of vowel phonetic characteristics follows.

Vowels are produced with vibration of the vocal folds and with almost no constriction of the vocal tract. Therefore, they are very resonant and thus salient in visual representations, such as oscillograms and spectrograms[19] (see Figure 3).

[19]The oscillogram (or waveform) is a graphical representation of the sound wave. The horizontal x-axis represents the time dimension, while the vertical y-axis represents the sound wave pressure. The spectrogram is a graph that displays the variation of spectral shape as a function of time. It has three axes: the horizontal x-axis represents time, the vertical y-axis represents frequency, and the darkness of marking represents the intensity of the signal at a particular time-frequency slot. Spectrograms are divided into two types, depending on the bandwidth of the analyzing filter: wideband, which gives good resolution in the time domain, and narrow band, which gives good resolution in the frequency domain (Reetz & Jongman, 2009).

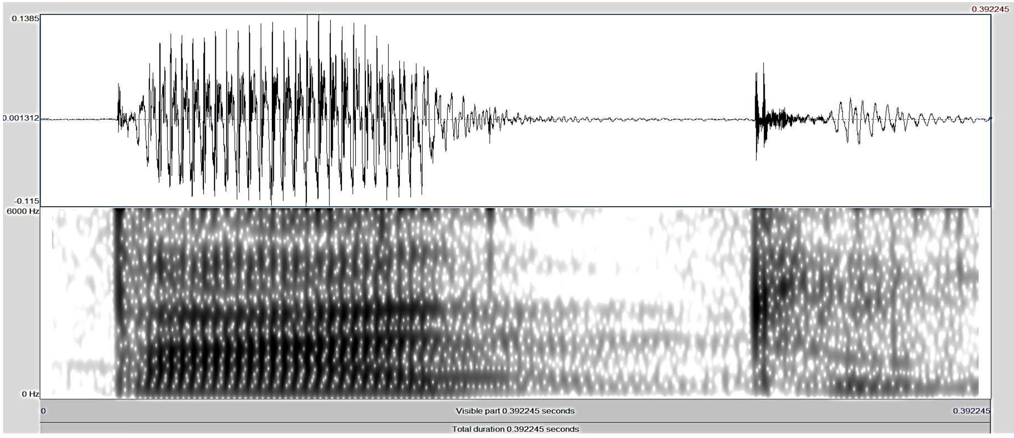


Figure 3. Oscillogram and wideband spectrogram of the Portuguese word *pato* (/ˈpatu/).

The different configurations of the vocal tract intensify or weaken distinct parts of the acoustic signal, and, as a result, each vowel presents a different acoustic image, that is, a particular formant structure. The first two formants (counted from below in the spectrogram) are the most important for vowel identification (Mateus et al., 2005), so they are summarized next.

The articulatory parameter of vowel height is related to first formant (F1) frequency, and the dimension of frontness/backness to second formant (F2) frequency (Hayward, 2000; Ladefoged, 1993). Therefore, a constriction between the tongue surface (displaced upwards) and the palate results in low F1 frequency, as in the case of the prototypical high vowels /i/ and /u/. When there is constriction at the larynx end,[20] F1 frequency is high. Low vowels, such as /a/, have high F1 frequencies. Vowel height is negatively correlated with F1 frequency; thus, the greater the degree of constriction, the lower the F1. In terms of F2 frequencies, vowels have higher frequency values the more constricted the oral tract is when the tongue is displaced toward the palate, and the frequencies are lower when the tongue is raised toward the velum. Thereby, front vowels (e.g., /i/) have the highest F2 frequencies and back vowels (e.g., /u/) have the lowest frequencies (Stevens, 1998; Reetz & Jongman, 2009). Vowel backness is, thus, correlated with F2 frequency, that is, with the distance between the first and second formant frequencies (Reetz & Jongman, 2009, p. 184). Figure 4 shows that the difference between F1 and F2 is larger for the front vowels than for the back vowels.

[20] The lowering of the tongue results in the narrowing of the vocal tract in the lower pharyngeal region near the root of the tongue (Stevens, 1998).

Acoustic differences in terms of lip rounding are found in the complex relationship between F2 and third formant (F3) frequencies.[21] Hayward (2000) states that, despite being an independent articulatory dimension, “it is less clear that (lip rounding) has a corresponding acoustic correlate” (p. 149). Most phoneticians agree that the rounding gesture tends to lower formant frequencies, most commonly F2 frequencies. For example, back and front unrounded vowels have higher F2 than their rounded counterparts.

Figure 4 displays the frequencies of the first three formants for American English monophthongs produced by 50 male speakers (Hillenbrand et al., 1995) and illustrates the correlation between the two dimensions - height and backness - for describing vowels and their acoustic parameters - F1 and F2 -, as previously described. The vowels in the horizontal axis are organized from front to back.

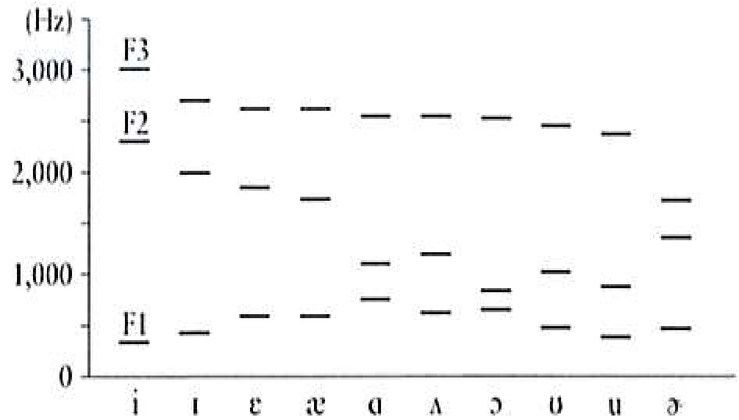


Figure 4. Average formant frequencies (F1, F2, and F3) of the monophthongal vowels of American English (Hillenbrand et al., 1995).

In the following sections, the focus is on the description of the American English and the European Portuguese vowel systems. The choice of the American English variety will be justified in *Chapter 3* and related to the data obtained from the background questionnaire.

[21] As Reetz and Jongman (2009) explain, “the lengthening of the vocal tract due to lip rounding lowers all formants, and moves F3 close to F2, which distinguishes between front unrounded and front rounded vowels” (p. 184). Therefore, in languages such as Dutch, French, German and Swedish, which have both rounded and unrounded vowels, F3 also provides an acoustic cue to vowel quality.

1.2 The American English Vowel System

In American English (AE), there are twelve monophthongs (/i/, /ɪ/, /e/, /ɛ/, /æ/, /ɜ/ or /ə/,[22] /ʌ/, /ɑ/, /ɔ/, /o/, /ʊ/, /u/) and three diphthongs (/aɪ/, /aʊ/, /ɔɪ/), as illustrated in Figures 5 and 6. Monophthongs /e/ and /o/ are mostly produced as diphthongal realizations[23] (/eɪ/ and /oʊ/) (Bohn & Caudery, 2007; Giegerich, 1992; Ladefoged, 2005; Reetz & Jongman, 2009).

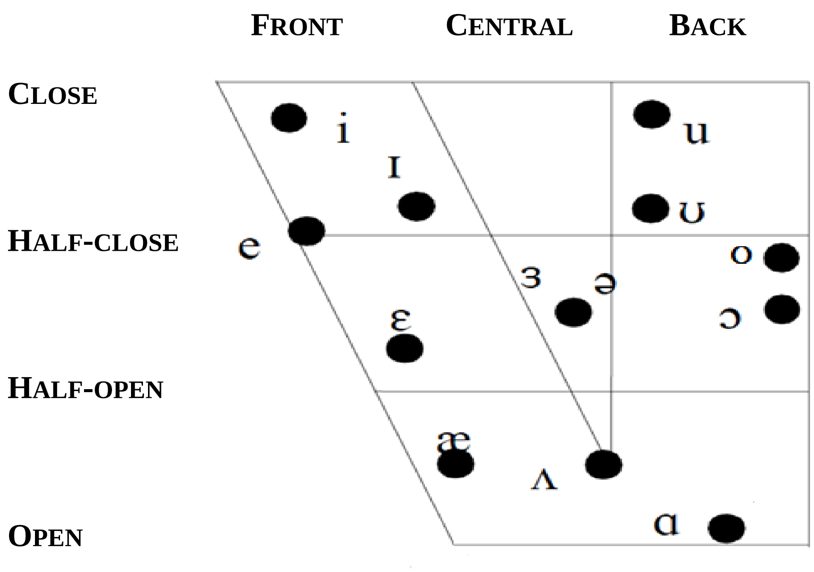


Figure 5. The American English vowel chart (adapted from Bohn & Caudery, 2007).

[22] In General American English (GA, the standard variety of AE), the production of /ɜ/ is always combined with retroflexion (i.e., articulated with the tip/blade of the tongue curled back, which forms a constriction with the passive articulator). Thus, the phonetic transcription of this sound is /ɜɻ/. The unstressed vowel /ə/ is also frequently "r-colored", i.e. rhotacized, and transcribed as /ɚ/ (Bohn & Caudery, 2007, p. 67).

[23] Giegerich (1992), Bohn and Caudery (2007), and Reetz and Jongman (2009) refer to /aɪ/, /aʊ/, /ɔɪ/ as "true diphthongs" to distinguish them from the diphthongized vowels /eɪ/ and /oʊ/. Giegerich (1992, p. 50) explains that the reason for not treating the latter as diphthongs on a phonemic level is the fact that /e/ and /o/ can have monophthongal realizations in some accents of English, such as in GA, whereas true diphthongs are of a diphthongal quality in the three main accents, the Southern British Standard English (RP), the Scottish Standard English (SSE), and the GA. Roca and Johnson (as cited in Rauber, 2010, p. 18) differentiate these two groups of diphthongs in terms of articulatory gestures. Diphthongs /aɪ/, /aʊ/, and /ɔɪ/ are considered heterogeneous because the two vowels that form the diphthong are not close in articulatory position and do not share lip gesture, whereas /eɪ/ and /oʊ/ are homogeneous diphthongs because both phases of the diphthongs are close in articulatory position and share lip gesture. Bohn and Caudery (2007) add that the true diphthongs "travel quite a distance from onglide to offglide, whereas /eɪ/ and /oʊ/ have a much shorter trajectory" (p. 69).

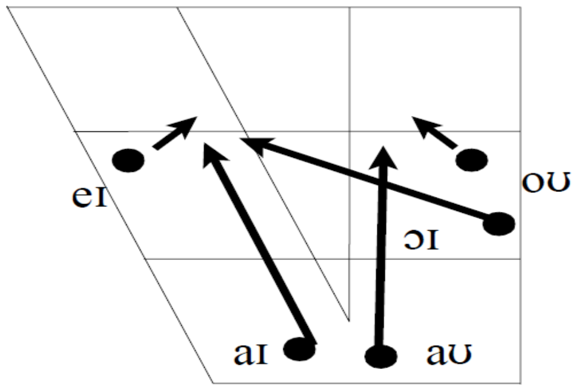


Figure 6. The American English diphthongs' trajectories in the vowel space (Bohn & Caudery, 2007).

The articulation of American English vowels can be described in relation to three main parameters, viz. tongue height, tongue position and lips position,[24] as summarized in Table 1 (Bohn & Caudery, 2007; Giegerich, 1992).

Table 1

Articulatory Classification of AE Vowels

vowel	tongue height	tongue position	lips position	examples
/i/	below close	front	unrounded	*beat*
/ɪ/	above half-close	centralized front	unrounded	*bit*
/e/	half-close	front	unrounded	*bait*
/ɛ/	above half-open	front	unrounded	*bet*
/æ/	between half-open and open	front	unrounded	*bat*
/ɜ/;/ə/	between half-open and half-close	central	unrounded	*Bert*
/ʌ/	between open and half-open	central	unrounded	*but*
/ɑ/	open	between center and back	unrounded	*bot(tom)*
/ɔ/	between half-close and half-open	back	rounded	*bought*
/o/	half-close	back	rounded	*boat*
/ʊ/	above half-close	centralized back	rounded	*book*
/u/	below close	centralized back	rounded	*boot*

Traditional articulatory descriptions of English vowels also include duration as a distinctive feature that contrasts long (e.g., /iː/ in *beat*) and short (e.g., /ɪ/ in *bit*) vowels. However, Bohn and Caudery (2007, p. 62) believe that this distinction is "misleading" because vowel length is not the main phonological feature that differentiates vowel pairs

[24] In American English, lip rounding is a redundant articulatory feature, i.e., it is not a distinctive phonological feature, because there are no vowels that differ only in lip rounding (Reetz & Jongman, 2009).

in English. Rather, it is vowel quality, which is the result of differences in terms of articulation (tongue height and tongue position), that primarily distinguishes English vowels and is used by native English speakers to differentiate their vowels. Furthermore, vowel duration depends primarily, amongst other factors,[25] on the phonetic contexts in which they occur. As the authors state, "a so-called 'long' vowel which is followed by a voiceless stop may be shorter than a so-called 'short' vowel if the 'short' vowel is followed by a (…) voiced fricative" (Bohn & Caudery, 2007, p. 62). From their point of view, the long-short distinction is more important in identifying a following consonant than the vowel itself. Giegerich (1992) also treats the quantity difference as redundant. In line with the previous statements, we also do not consider vowel duration as the main distinctive feature in the analysis of the vowel pairs under study.

Finally, English vowels can be divided in relation to tenseness. The distinction between tense and lax vowels is mostly related to their different articulatory characteristics and to syllabification. In terms of articulatory gestures, tense[26] vowels are described as having greater force of constriction than their lax counterparts, and they also have a tendency to be longer. In terms of syllabic structure, tense vowels can occur in open and closed syllables,[27] but lax vowels are restricted to closed syllables, which is known as the lax vowel constraint (LVC) (Cebrian, 2009). The tense vowels of English are /i/, /e/, /ɜ/, /ɑ/, /o/, /ɔ/, /u/ and the lax vowels are /ɪ/, /ɛ/, /æ/, /ʌ/, /ʊ/, /ə/.[28] In sum, the main difference between English tense and lax vowels is quality (i.e., formant pattern) rather than duration. As Reetz and Jongman (2009) explain, "A shortened /i/ still sounds like an /i/ just as a lengthened /ɪ/ still sounds like an /ɪ/" (p.186). Next, a few studies on the acoustic characteristics of American English vowels will be reviewed in chronological order, including the widely cited study by Peterson and Barney (1952).

Although the main purpose of Peterson and Barney's (1952) study was to discuss the use of spectrographic analysis[29] as an adequate method to make acoustic measurements of vowels, it provided important information regarding formant frequency values of vowel productions. In order to obtain these acoustic values, 76

[25] These include, for instance, speaking rate and syllable prominence.
[26] The definition of [tense] as a binary feature is given by Giegerich (1992, p. 98): "Tense sounds are produced with a deliberate, accurate, maximally distinct gesture that involves considerable muscular effort; nontense sounds are produced rapidly and somewhat indistinctly."
[27] Open syllables end in a vowel, whereas closed syllables end in a consonant.
[28] Vowel /ə/ is included in the group of lax vowels, though it occurs in open syllables.
[29] When the study was conducted, the sound spectrograph had been recently (around 1945) developed by Bell Telephone Laboratories.

speakers (33 men, 28 women and 15 children) read twice a list of monosyllabic /hVd/ words with ten vowels (/i/, /ɪ/, /ɛ/, /æ/, /ʌ/, /ɑ/, /ɔ/, /ʊ/, /u/, /ɚ/,). Acoustic measurements of fundamental frequency (F0), formant frequencies (F1, F2, and F3), and formant amplitudes were made on the steady-state[30] portion of each vowel. Randomized /hVd/ tokens, produced by ten speakers, were also presented for identification to a group of 70 adult speakers over eight sessions. One of the relevant results of this study includes the finding of a strong relation between the measured acoustic values and the identification of the intended vowel. The identification test revealed that vowels were accurately categorized, since the overall error rate was 5.6%, and all misidentifications occurred between adjacent vowels. However, high formant frequency cross-speaker variability was found, as well as considerable degree of overlap of formant frequency values amongst adjacent vowels. The reported variability of formant values may be explained by the fact that dialect was not a statistically controlled variable (see Table B1). Other cited limitations of the study (cf. Ladefoged, 2003) are related to the fact that the number of children was unbalanced in comparison to that of adults, and no background information was provided about their age and gender. In addition, identification results were not described separately for the three different groups of listeners.

Forty-three years later, Hillenbrand, Getty, Clark, and Wheeler (1995) replicated Peterson and Barney’s (1952) study and measured twelve American English vowels (/i/, /ɪ/, /e/, /ɛ/, /æ/, /ɝ/ /ʌ/, /ɑ/, /ɔ/, /o/, /ʊ/, /u/). Not only did their study include another two vowels (/i/ and /e/), but the number of participants also increased considerably. A total of 139 speakers (45 men, 48 women, and 46 children) from southeastern and southwestern parts of the state of Michigan (i.e., northern Midwesterns) were selected through a screening procedure that consisted of dialect assessment focused mainly on the /ɑ/- /ɔ/ distinction that speakers were expected to make either in spontaneous speech or in a 128-word-text reading task, and in a discrimination task. Similarly to Peterson and Barney’s (1952), the production test included 12 lists of randomized /hVd/ words that each participant was asked to read three times. Together with formant frequencies and fundamental frequency, Hillenbrand et al. (1995) also measured vowel duration. Acoustic measurements for the 12 vowels are listed in Table B1. Results regarding vowel duration showed that there were differences amongst the three groups of

[30] The steady-state portion of a vowel is the stretch of the vowel during which the formant frequencies are relatively stable, usually in the middle of the vowel, away from the influence of surrounding segments (Reetz & Jongman, 2009, p. 185).

speakers, with men producing significantly shorter vowels when compared to women and children. In terms of F0 values of men and women, they differed by only a few Hz in comparison to Peterson and Barney's (1952) reported values. While there were similarities between both studies, many differences were also found, namely the different vowel locations in the F1 and F2 space, in particular in the case of vowels /ɛ/ and /æ/. The findings by Hillenbrand et al.'s (1995) study contrasted with those of Peterson and Barney's (1952), since they reported higher F2 values for /æ/ than for /ɛ/ and slightly lower F1 values for /æ/ than for /ɛ/. In the review of this study, Clopper, Pisoni, and Jong (2005) emphasize that Hillenbrand et al. (1995) found evidence of the Northern Cities Chain Shift[31] (NCCS) in their data, including the backing of /ʌ/. The authors also consider that the dissimilarities between the two abovementioned studies are mainly a result of the regional dialect differences between the two groups of speakers and of diachronic change, since there is a time gap of about 40 years between both studies. The study by Hillenbrand et al. (1995) also included an identification task in which 20 native speakers listened to one presentation of each of the 1668 /hVd/ tokens over two one-hour sessions. Despite the inconsistencies in terms of acoustic measurements, the two identification experiments carried out by Peterson and Barney (1952) and Hillenbrand et al. (1995) had similar results, that is, the identification rates were very similar.

The acoustic properties (the first three formant frequencies and F0) of 11 English vowels (/i/, /ɪ/, /e/, /ɛ/, /æ/, /ɚ/, /ʌ/, /ɑ/, /o/, /ʊ/, /u/) produced by southern Californian English-speaking monolinguals were examined by Hagiwara (1997) in order to illustrate cross-variability amongst American English dialects. The researcher aimed at comparing vowel production data from his group of speakers with data from the two aforementioned studies. Nine men and six women, whose ages ranged from 18 to 26 years, produced 33 real words with three consonant contexts (/bVt/, /tVk/, and /hVd/), inserted in the frame "Cite (word) twice", three times in random order. Although Hagiwara measured the mean F1, F2 and F3 values for each vowel produced by each gender, the influence of consonantal context in vowel production was not analyzed. Hence, the formant values reported are not, to some extent, comparable to the previous studies. In comparison to Peterson and Barney's (1952) results, Hagiwara

[31] The NCCS is characterized by raising and fronting of /æ/, backing of /ɛ/ and /ʌ/, lowering of /ɔ/, and lowering and fronting of /ɑ/ (Clopper et al., 2005, p. 1661).

(1997) reported that the main differences were the F2 frequencies of the back and central vowels: (1) the back vowels /ʊ/ and /u/ were less rounded; thus, more centralized; (2) the central vowel /ʌ/ had a much higher F2 value; and (3) the low vowels /æ/ and /ɑ/ were 200 Hz higher in southern Californian women's productions.

Each of the earlier cited studies focused on a single regional variety. Peterson and Barney (1952) intended to describe General American English,[32] Hillenbrand et al. (1995) replicated the earlier study with speakers from northern Midwest, and Hagiwara (1997) provided data for speakers of southern California. A subsequent study by Clopper et al. (2005) contributed enormously to the description of American English vowels by providing acoustic data from six regional varieties of American English. In their study, acoustic measurements of duration and first and second formant frequencies were obtained from five repetitions of 11 vowels (/i/, /ɪ/, /e/, /ɛ/, /æ/, /ɑ/, /ɔ/, /ʌ/, /o/, /ʊ/, /u/) produced by 48 monolingual native speakers, between the ages of 18 and 25 years. Speakers included four men and four women from six dialect regions of the United States, namely New England, Mid-Atlantic, North, Midland, South, and West. Ten vowels were inserted in /hVd/ tokens, and vowel /ɔ/ was inserted in the words *frogs* and *logs*, yielding 56 tokens per speaker. A brief summary of the main findings is presented in Table 2 (see Clopper et al., 2005, for detailed descriptions of each regional dialect).

[32] Clopper et al. (2005) remark that Peterson and Barney's (1952) data are closer to "American eastern seaboard" dialect than to GA.

Table 2

Dialect-specific Features of American English Vowels (Clopper et al., 2005)

AE dialect	Dialect-specific features	Comments
New England	(1) /æ/ raising (2) merger of /ɑ/ and /ɔ/	
Mid-Atlantic North	(1) merger of /ɑ/ and /ɔ/* (1) /ɑ/ lowering and fronting* (2) /æ/ raising and fronting* (3) /ɛ/ and /ʌ/ backing (females)*	The 3 features are typical of the NCCS
Midland	(1) merger of /ɑ/ and /ɔ/* (2) /ɛ/ raising (males)* (3) Centralized /e/ (females)* (4) /u/ fronting (5) /æ/ raising	Features (2), (3) and (4) are characteristic of the SVS. Feature (5) is found in the NCCS.
South	(1) /o/ fronting* (2) /e/ centralization* (3) /u/ fronting (males)* (4) /ɛ/ raising (males)* (5) /æ/ raising (males)* (6) /ʊ/ and /u/ raising (7) longer lax vowels than other dialects	Features (1) to (4) are typical of the SVS.
West	(1) merger of /ɑ/ and /ɔ/* (2) /u/ fronting (males)*	

Note: * indicates consistent (i.e., statistically significant) dialect-specific features across all speakers.

The analysis of vowel duration and formant frequency measures confirmed the presence of the Northern Cities Chain Shift, and the Southern Vowel Shift[33] (SVS). Some Midland speakers also exhibited features of the Southern dialect, while others showed Northern characteristics. According to Clopper and colleagues, these results corroborate earlier claims that the Midland dialect region is not a single dialect, but a transition area between the North and the South (e.g. Davis & Houck, 1992, cited in Clopper et al., 2005). Frequency values measured by Clopper et al. (2005) were not published in their paper, but the authors gave permission for them to be reported in Rauber (2010), and these values are presented in Table B2.

Finally, the acoustic measurements collected by Rauber (2010) and Rauber, Rato and Silva (2010) will be reported, because the stimuli used to obtain perceptual data in the present study were produced by speakers of the Western and the Midland dialects. The native Western American English speakers, who produced the tokens for the

[33] This shift is characterized by the fronting of the back vowels /u/ and /o/, the raising and fronting of /ɪ/ and /æ/, and the lowering and backing of /i/ and /e/ (Clopper et al., 2005, p. 1662).

perception test, belong to a subset of the larger group of monolinguals recorded by Rauber (2010) and the native Midland AE speakers, who participated in the study by Rauber et al. (2010), produced tokens that were also used in this study.

Although the main objective of Rauber's (2010) study was to investigate the pronunciation and perception of American English vowels by Brazilian advanced learners of English as a Foreign Language (EFL), only the production results of the American English monolinguals will be reported so as to add more information to the previous acoustic measurements of vowels. In this study, nine American English monolingual speakers (five men and four women), from Sacramento, California, produced eleven vowels (/i/, /ɪ/, /eɪ/, /ɛ/, /æ/, /ɑ/, /ɔ/, /ʌ/, /oʊ/, /ʊ/, /u/) in monosyllabic words with the following consonantal frames: /bVt/, /pVt/, /sVt/, /tVt/, /tVk/, and /kVp/. Sixty-six words, six for each vowel, were inserted in the structure "CVC. CVC and CVC sound like CVC" (e.g., *Beat. Beat and Pete sound like seat*) and presented three times in a random order to native speakers. The results indicate that: (1) /i/ and /u/ were undoubtedly produced as high vowels; (2) /ɛ/, /ʌ/, /ɑ/, and /ɔ/ comprised the group of low vowels; (3) /e/ was higher than the other mid vowels /ɪ/ and /oʊ/ for women; and (4) men produced /ɪ/, /eɪ/ and /oʊ/ with similar degrees of height. Rauber (2010) compared her measurement results (see Table B1) with the previous aforementioned findings and concluded that, despite having similar results, differences were also found. The production results showed that vowel /ɛ/ was more fronted and higher than /æ/, contrary to Hillenbrand et al.'s (1995) finding of /ɛ/ as lower and further back than /æ/. A similar inconsistency was found for vowel /u/, which was more fronted and higher than /ʊ/, contrariwise to what Hillenbrand et al. (1995) reported, that is, a higher and further back /u/ in comparison to /ʊ/. The low vowel /ɑ/ was lower and more fronted than /ɔ/ for female participants than in the two earlier studies by Peterson and Barney (1952) and Hillenbrand et al. (1995), who reported /ɑ/ as being more fronted and lower than /ɔ/. Vowel /ɪ/ was further back than /eɪ/ in all the previous studies, but Rauber (2010) reported it as being also lower than /eɪ/, whereas earlier research reported it as being higher than /eɪ/.The same pattern was found for the first element of the diphthong /oʊ/. Rauber (2010) described it as being higher and further back than /ʊ/, and the studies by Hagiwara (1997) and Hillenbrand et al. (1995) reported it as being slightly lower than /ʊ/.

Rauber et al. (2010) investigated the production and identification of American English front vowels by Chinese learners of English as a foreign language, but for the purpose of this chapter only the production results of the American English monolinguals are concisely summarized in Table B1, because a set of the tokens produced by these native speakers were used in the present study as perceptual stimuli. In this study, seven female native speakers of American English (with ages ranging from 22 to 39 years), from Davenport, Iowa, were asked to read CVC (/bVt/, /tVk/, and /hVd/) words twice with four English monophthongs /i/, /ɪ/, /ɛ/, /æ/ embedded in the carrier sentence "Say CVC now.". A total of 164 vowels were produced in order to measure F1 and F2 frequency values and vowel duration. In comparison to Clopper et al. (2005), vowel /ɛ/ had higher F1 values and vowel /æ/ had lower F1 values, resulting in a smaller Euclidean distance between the vowels of this pair than between the vowel productions of the female monolinguals analyzed by Clopper et al. (2005).

1.3 The European Portuguese Vowel System

The European Portuguese (EP) vowel system comprises nine oral vowels and five nasal vowels, presented in Table 3. All vowels, except /ɨ/, occur in stressed position. In pretonic position, all oral vowels occur (/i/, /e/, /ɛ/, /ɨ/, /ɐ/, /a/, /u/, /o/, /ɔ/), whereas in posttonic position they are reduced to four (/i/, /ɨ/, /ɐ/, /u/), and in word-final position to three (/ɨ /, /ɐ /, /u/). The five nasal vowels occur in pretonic position, but only two (/ɐ̃/, /ũ/) in posttonic position (Barroso, 1999; Mateus, 1990). [34] Authors generally acknowledge that there is internal symmetry in the EP vowel inventory, similarly to what happens in other languages. The three unrounded front vowels can be paired with the three rounded back vowels, resulting in a high vowel pair (/i/-/u/), a mid-high pair (/e/-/o/) and a mid-low pair (/ɛ/-/ɔ/).[35]

[34] Examples of the EP vowel inventory: (a) Oral vowels in stressed position: [ˈbiku]-[ˈbeku],[ˈsedɨ]- [ˈsɛdɨ], [ˈtɐʎɐ]-[ˈtaʎɐ],[ˈbolɐ]-[ˈbɔlɐ]-[ˈbulɐ]; (b) Nasal vowels in stressed position: [ˈpĩtɨ]-[ˈpẽtɨ]-[ˈpõtɨ], [ˈmɐ̃du]-[ˈmõdu]-[ˈmũdu]; (c) Oral vowels in pretonic position [fiˈnal]-[fɐwˈdal], [mɛˈziɲɐ]-[maˈziɲɐ]-[mɨˈziɲɐ], [kɔˈradɐ]-[kuˈradɐ], [toˈradɐ]-[tɐˈradɐ]; (d) Nasal vowels in pretonic position [sĩˈtar]-[sẽˈtar], [mɐ̃ˈdar]-[mõˈdar], [fĩˈdar]-[fũˈdar]; (e) Oral vowels in posttonic position [ˈsatiru], [iˈliɐku], [ˈlatɨgu], [ˈmaʃkulu]; (f) Oral vowels in word-final position [ˈabrɨ]-[ˈabrɐ]-[ˈabru]; (g) Nasal vowels in posttonic position [ˈfɔrũ], [ˈɔrfɐ̃] (Mateus, 1990, p. 306).

[35] Mateus et al. (as cited in Escudero et al., 2009) includes vowel /a/ in the /ɛ/ and /ɔ/ group, and refers to them as low vowels.

Table 3

EP Vowels (adapted from Mateus, Falé, & Freitas, 2005)

		Front	Central	Back
Oral vowels	**High**	/i/	/ɨ/	/u/
	Mid-high	/e/	/ɐ/	/o/
	Mid-low	/ɛ/		/ɔ/
	Low		/a/	
Nasal vowels	**High**	/ĩ/		/ũ/
	Mid-high	/ẽ/	/ɐ̃/	/õ/

European Portuguese vowels are described in terms of four articulatory parameters: velum position (oral, nasal), tongue position (front/palatal, central, back/velar), tongue height (high/close, mid-high/half-close, mid-low/half-open, low/open), and lip position (rounded, unrounded). From this point forward, only the oral vowels will be described, as this study does not include production of nasal vowels.

Mateus (1990, p. 52) describes the articulation of EP oral vowels in stressed position according to tongue position, tongue height and lip position. Therefore, the author explains that in the production of /ɪ/, /e/, and /ɛ/ the front of the tongue raises forward to the palate, so these vowels can be called palatal or anterior; in the articulation of central /a/ and /ɐ/ the tongue is in the center of the oral cavity in a low position, though /a/ is slightly more palatal than /ɐ/; in the production of /ɔ/, /o/, and/u/ the tongue dorsum raises towards the velum, so they can be called velar or posterior vowels. Regarding tongue height, Mateus (1990) divides the eight vowels into three groups, viz. low or open vowels /ɛ/, /a/, and /ɔ/, mid vowels /e/, /ɐ/, and /o/, and high or close vowels /i/ and /u/. According to lip position, only three vowels (/ɔ/, /o/, and/u/) are articulated with lip rounding, so they are designated rounded. However, as in English, rounding is a redundant articulatory feature, that is, it is not a distinctive phonological feature of EP vowel pairs.

Martins, Carbone, Pinto, Silva, and Teixeira (2008) conducted the first magnetic resonance imaging (MRI) study of European Portuguese sound production, but only results relative to oral vowels will be briefly presented and compared with traditional articulatory descriptions. The researchers reviewed traditional articulatory descriptions of the nine oral vowels and noticed that the most inconsistent is the description of the central /ɨ/, which includes the schwa (/ə/) (c.f. Mateus, 1990), the high central vowel /ɨ/ (c.f. Barroso, 1999), or, as suggested by Cruz-Ferreira (1999), an articulation close to

/u/, viz. /ɯ/. To verify to what extent previous articulatory descriptions of European Portuguese oral vowels were accurate, the group of researchers obtained magnetic resonance (MRI) static images of the vowel tract during the production of EP vowels by a 25-year-old male speaker from the north of Portugal. Each vowel was pronounced and artificially sustained during 5.6 s, and to help the speaker a reference word, containing the target phone, was presented before the sequence (e.g., "Please say /a/ as pronounced in /ˈpatu/"). One of the most relevant findings of this study is related to central vowel height. Contrary to traditional descriptions, in which /ɨ/ is considered as high as /i/ and /u/, Martins and her colleagues (2008) found that /ɨ/ is produced with the highest tongue point, that is, with the highest position of tongue dorsum among the central vowels, but not so high as to be considered a high central vowel. They add that the articulatory differences between the three central vowels are mainly related to tongue dorsum (TD) position and shape, jaw height and pharyngeal cavity dimensions. Though /ɨ/ is not as high as traditionally described, it is the one that is produced with the highest TD position, followed by /ɐ/ and /a/. The researchers also reported that vowel /a/ is produced with very low jaw, high lip aperture and posterior position of tongue (TD) and tongue root (TR) position, and further explain that the last characteristic goes against the traditional classification of /a/ as a central vowel. Thus, the researchers suggest that this vowel is better described as a low pharyngeal vowel. As far as anterior oral vowels are concerned, results were in line with earlier articulatory descriptions. Vowel /ɛ/ was produced with the lowest position of TD; /i/ with the most raised in anterior position, and /e/ in an intermediate position, being closer to /ɛ/ in the anterior-posterior axis. As for posterior vowels, /u/ was produced with the highest TD position amongst the three posterior vowels, followed by /o/ and /ɔ/, with the lowest and more posterior position. Compared to anterior and central vowels, posterior segments were produced with lower TD than anterior vowel segments, being only /a/ produced with lower TD than the lowest posterior /ɔ/. Despite being a pioneering study of European Portuguese vowel production, their findings are limited by the fact that data was produced by only one male speaker of a dialect different from standard EP[36] (see Figure C1 for MRI images of EP vowels).

[36] The standard variety of European Portuguese is described as a central-southern variety located mainly in Lisbon (Emiliano, 2009, p. 4) and in Coimbra (Cruz-Ferreira, 1999, p.126).

To our knowledge, only a few studies have described European Portuguese vowels acoustically (Delgado-Martins, 1973, 1975,[37] and Escudero, Boersma, Rauber, & Bion, 2009), and all of them restricted the analyses to the Lisbon dialect. A review of these studies will be reported below.

Delgado-Martins (1973) measured the first three formants and duration of eight oral vowels (/i/, /e/, /ɛ/, /ɐ/, /a/, /o/, /ɔ/, /u/) in word-medial stressed position produced by eight adult male speakers from Lisbon. Participants read 69 disyllabic words (stressed on the penultimate syllable) with the syllabic frameworks CVC, CVCV, CVVCV, CVCCV and CVCVC inserted in the carrier sentence "Digo a palavra… outra vez" (I say the word… again). The vowels were inserted in different consonantal contexts that varied in place and manner of articulation, and voicing. To get the acoustic EP vowel triangle (Figure 7), the researcher obtained the mean F1 and F2 values, presented in Table 4.

Table 4

Mean Values and Standard Deviation Values (SD) of F1, F2 and Duration of EP Vowels in Stressed Position (adapted from Delgado-Martins, 1973)

Vowel	**F1(Hz)**	**(SD)**	**F2 (Hz)**	**(SD)**	**Duration (ms)**
i	294	(37)	2343	(139)	84.8
e	403	(40)	2084	(187)	94.8
ɛ	501	(46)	1893	(155)	111.5
ɐ	511	(56)	1602	(105)	86.2
a	626	(78)	1326	(157)	109.4
ɔ	531	(57)	994	(81)	109.3
o	426	(46)	864	(111)	102.1
u	315	(45)	678	(124)	89.4

Note. Hz=Hertz; SD=Standard deviation; ms=milliseconds.

Similarly to Hagiwara's (1997) study, the limitation of this study was the fact that the influence of consonantal context in vowel production was not reported.

[37] Both studies were republished in Delgado-Martins (2002).

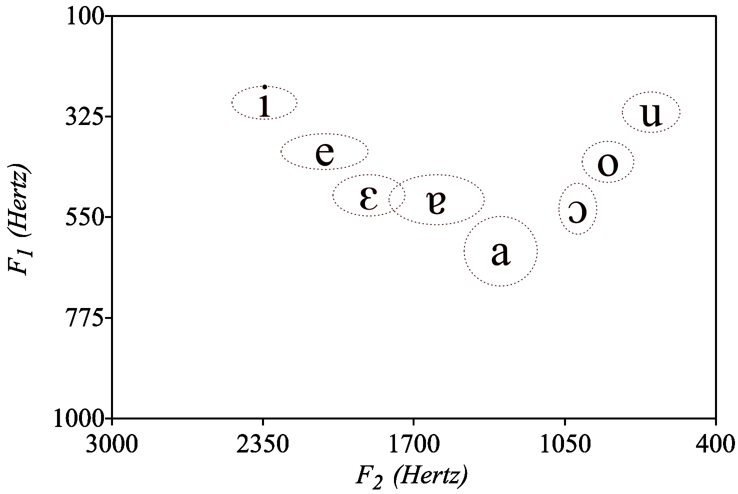


Figure 7. European Portuguese vowel triangle (Delgado-Martins, 1973).

In a subsequent study, Delgado-Martins (1975) investigated duration and intensity of nine oral vowels produced by a 25-year-old male speaker from Lisbon (see Table 5). Vowels were produced in 40 sentences with an average of 20 phonemes each. Findings revealed that the stressed vowels with more duration and intensity were the [+low, -high] vowels (/ɛ/,/ɔ/, and /a/), followed by the [- high, -low] vowels (/e/ and /o/). Finally, with less duration and intensity were the [+ high, -low] vowels (/i/ and /u/) (see Table 5). Results provided evidence that duration and intensity vary according to vowel quality and stress position.

Table 5

Mean Values of Duration and Intensity of EP Vowels (adapted from Delgado-Martins, 1975)

Vowel	Duration (ms)	Intensity (mm^2)
i	68	245
e	80	299
ɛ	106	425
ɐ	68	228
a	93	326
ɔ	97	355
o	96	320
u	69	208
ɨ	49	165

Escudero, Boersma, Rauber, and Bion (2009) measured five acoustic parameters of Brazilian Portuguese (BP) and European Portuguese (EP) vowels, viz. first formant (F1), second formant (F2), and third formant (F3) frequencies, duration, and fundamental frequency (F0). Regarding the EP vowels, data were produced by 20 young adults from Lisbon (10 men, with the mean age of 18.7 years, and 10 women, with a mean age of 19.8 years). The target vowels were the seven oral vowels that BP and EP have in common in stressed position (/i/, /e/, /ɛ/, /a/, /o/, /ɔ/, /u/). They occurred in the first syllable of CVCV words, which were embedded in a carrier sentence such as "*Pêpe*. Em *pêpe* e *pêpo* temos *ê*" (*Pêpe*. In *pêpe* and *pêpo* we have *ê*). The total number of vowel tokens analyzed per dialect was 2800. The results of this study revealed important findings about the production of European Portuguese oral vowels (see Table 6). Therefore, some of the main results are summarized below.

Table 6

Averages of Vowel Duration, F1, F2, F3, and Formant Ceiling for Female and Male Speakers of EP (Escudero et al., 2009)

vowel		i	e	ɛ	a	ɔ	o	u
Duration (ms)	F	92 (1.15)	106 (1.15)	115 (1.14)	122 (1.14)	118 (1.14)	110 (1.16)	94 (1.21)
	M	84 (1.14)	97 (1.15)	106 (1.16)	108 (1.18)	104 (1.15)	99 (1.14)	83 (1.15)
F0 (Hz)	F	216 (1.08)	211 (1.08)	204 (1.08)	201 (1.09)	204 (1.08)	211 (1.08)	222 (1.09)
	M	126 (1.18)	122 (1.17)	117 (1.16)	115 (1.15)	117 (1.15)	123 (1.17)	127 (1.19)
F1 (Hz)	F	313 (1.24)	402 (1.13)	511 (1.15)	781 (1.19)	592 (1.27)	422 (1.15)	335 (1.23)
	M	284 (1.09)	355 (1.09)	455 (1.13)	661 (1.08)	491 (1.11)	363 (1.11)	303 (1.09)
F2 (Hz)	F	2760 (1.03)	2508 (1.04)	2360 (1.03)	1662 (1.08)	1118 (1.09)	921 (1.18)	862 (1.14)
	M	2161 (1.05)	1987 (1.06)	1836 (1.07)	1365 (1.06)	934 (1.08)	843 (1.09)	814 (1.13)
F3 (Hz)	F	3283 (1.05)	3007 (1.04)	2943 (1.04)	2535 (1.17)	2729 (1.09)	2636 (1.19)	2458 (1.20)
	M	2774 (1.06)	2559 (1.06)	2475 (1.05)	2333 (1.04)	2414 (1.078)	2429 (1.07)	2315 (1.04)
Ceiling (Hz)	F	5875 (1.09)	5734 (1.09)	5662 (1.10)	5278 (1.09)	5259 (1.13)	5165 (1.12)	5066 (1.12)
	M	4570 (1.15)	4733 (1.15)	4792 (1.10)	4523 (1.12)	4537 (1.14)	4512 (1.11)	4366 (1.07)

Note. The standard deviations, converted to ratios, are between parentheses.

The researchers concluded that the seven Portuguese (EP and BP) vowels divide up into four F1 regions, and each back vowel has an F1 similar to its front counterpart, which revealed internal symmetry, as expected. As can be observed in Table 6, in terms of F1 values, amongst the EP vowels, low /a/ had the highest F1 value, followed by the mid-low vowels /ɛ/ and /ɔ/, then the mid-high vowels /e/ and /o/, and finally /i/ and /u/, which had the lowest F1 values. Portuguese women tended to have higher F1 values (478 Hz) than men (409 Hz), and they used a larger part of the F1 space than men. Furthermore, though the structure of the vowel inventory is symmetric, back vowels had slightly higher F1 values than their front counterparts, following the tendency found in languages such as American English, Parisian French, Northern German (Escudero et al., 2009, p. 1379). In relation to F2 values, women also had higher values than men, thus the size of the F2 space was greater for Portuguese women than for men. As reported above, female informants had larger F1 and F2 space sizes than male participants, which is explained by the fact that women and men have different vocal tract sizes. Moreover, women produced longer vowels than men. In terms of duration, results showed that the lower vowels were longer than the higher vowels, and that the vowel intrinsic duration[38] effect (the duration ratio of low and high vowels) was strong, and, compared with other languages without a phonological length contrast, such as Iberian Spanish, Peruvian Spanish, and European French, this effect was even stronger. According to Escudero et al. (2009),

> this suggests that in Portuguese the effect is not only of an automatic articulatory nature, it seems that Portuguese has turned duration into a language-specific (minor) cue[39] for phonological vowel identity analogously to how English vowel duration has become a cue for the phonological voicing of a following obstruent both in production and in perception. (p. 1390)

Some of the results regarding the comparison between EP and BP vowels are also mentioned further, since studies on English vowel perception and production by Brazilian Portuguese speakers will be reported in *Chapter 2*.

[38] Intrinsic vowel duration differences derive from physiology, since low vowels require longer time to be articulated due to more jaw lowering (Lehiste, 1970, p. 18-19).
[39] A cue is an acoustic feature that is considered to provide unique information about the identity of a particular segment such as vowel height, frontness, or rounding (Reetz & Jongman, 2009, p. 185).

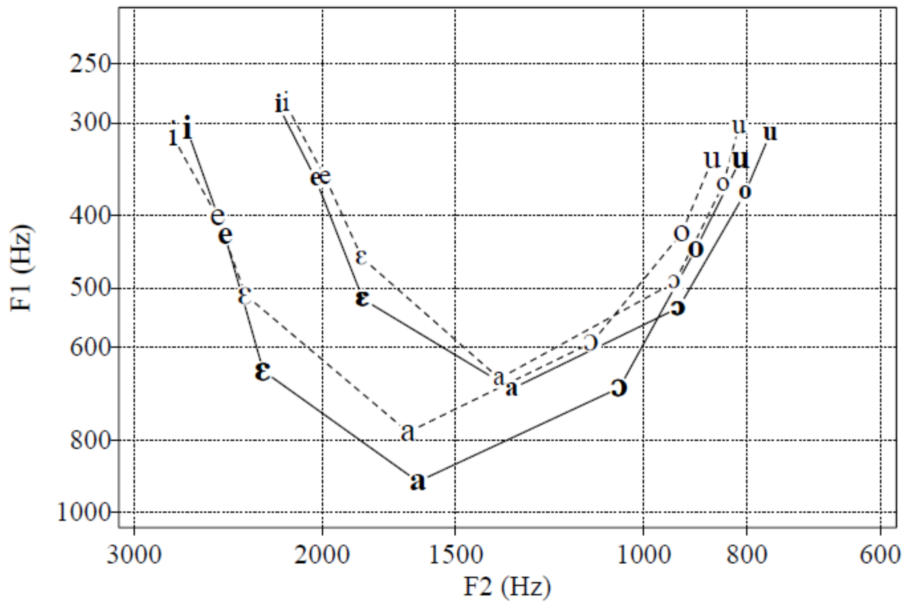


Figure 8. The vowel spaces of EP and BP. Solid lines and bold symbols=BP; dashed lines=EP; large font=women; small font=men (Escudero et al., 2009).

As can be observed in Figure 8, mid-low and low vowels /ɛ/, /ɔ/, and /a/ were higher in European Portuguese than in Brazilian Portuguese productions. In addition, Portuguese speakers tended to have slightly higher F2 values than the Brazilian speakers. The acoustic distance between mid-low and mid-high vowels was larger in BP than in EP. Escudero and her colleagues explain that these differences are related to the fact that /ɛ/ and /ɔ/ are lower in BP than in EP and add that the smaller /ɛ/-/e/ distance in EP as compared to BP is due to a more raised /ɛ/ than to a lowered /e/. Vowel /ɛ/ is higher, that is, less open in EP than in BP, which may indicate, according to the authors, an impending merger of EP /ɛ/ into /e/ (Escudero et al., 2009, p. 1391). In terms of duration, their findings are in line with previous research on the comparison between European Spanish and Peruvian Spanish (Morrison & Escudero, 2007), since vowels in South American varieties seem to be longer than in European varieties. In Figure 9, the mean duration of the target vowels is displayed, and we can observe that vowels are longer in Brazilian Portuguese than in European Portuguese.

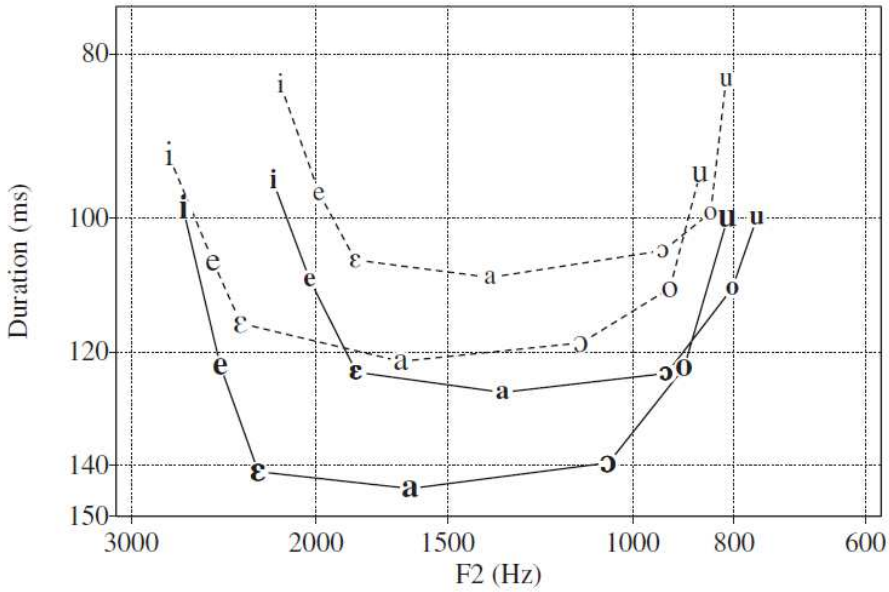


Figure 9. Mean duration as a function of vowel category. Solid lines and bold symbols=BP; dashed lines=EP; large font=women; small font=men (Escudero et al., 2009, p. 1387).

1.4 Second Language Speech Production

Second language (L2) speech production is as complex a process as L1 speech production, if not more. There are several theoretical models that propose explanations for the cognitive processes underlying L2 speech production based on models of L1 production (e.g., De Bot, 1992; Green, 1986; Poulisse & Bongaerts, 1994).[40] Since differences amongst these models revolve around the conceptualization level, that is, on the cognitive processes implied in the activation of two languages, they are not discussed here. However, De Bot's (1992) proposal that the articulator is one for both languages, where a large set of L1 and L2 segments and suprasegments are stored together, that is, located in a common phonological space, is important to highlight because it is this line of thought that most likely explains the L1 phonological interference in L2 acquisition (Mota, 2010). This section, which is focused on the production of overt L2 speech, that is, on the output of the articulation stage, reviews some cross-language speech production studies.

Second language speech production has been the focus of many research studies conducted over the last few decades. Researchers have investigated different types of

[40] See Mota (2010) for a review of L1 and L2 speech production models.

pronunciation difficulties, at the segmental, suprasegmental (e.g., Albini, 2012; Lacabex, Lecumberri, & Cooke, 2008a, 2008b 2009; Lee & Cho, 2010; Wang, Jongman, & Sereno, 2003, 2006) and non-segmental (i.e. phonotactic level, e.g., Cebrian, 2006, 2009) levels,[41] to understand the causes of inaccurate production,[42] and the influence of diverse variables on the degree of foreign accent. Variables that have been investigated range from differences in terms of linguistic experience, such as age of L2 learning, amount of L2 exposure (e.g., Bohn & Flege, 1992; Flege, Munro, & MacKay, 1995a, 1995b; Flege, Schirru, & MacKay, 2003; Fullana & Mora, 2009; MacKay & Fullana, 2009), individual differences in acoustic and phonological memory (e.g., Aliaga-Garcia, Mora & Cerviño-Povedano, 2010; Safronova, 2013), phonological attention control (e.g., Darcy, Mora, & Daidone, 2013; Kim & Hazan, 2010), musical training (e.g., Gottfried, 2007) to L1 orthography (e.g., Silveira, 2009), L1 phonetic inventory size (e.g., Souza & Carlet, 2013; Horslund & Bohn, 2013) and L1 vocabulary size (e.g., Fullana & MacKay, 2013). Some of these studies are reviewed in this section in line with the contents of this chapter and according to the main research objectives of the present study. In addition, research on the effects of production training on the learning of non-native sounds is summarized.

One of the major factors that influences L2 speech production and has been extensively investigated is the age of onset of learning (AOL), that is, the age of first exposure to the target L2 language. Age of L2 learning is an important determinant of success in learning to accurately produce non-native sounds, and a major predictor of degree of foreign accent. Many studies examining L2 speech production have provided evidence that suggests "the earlier, the better" when acquiring a new sound system, and that the correlation between AOL and accent is mostly a negative one (Moyer, 2013; Piske, 2007). Considering the focus of this study, the next section focuses on studies that examine how speakers from different first languages produce non-native segments.

Flege (1987) investigated the influence of language experience on the production of vowels by measuring their formants (F1, F2, F3) in French and English words produced by native French participants, who were highly experienced in English, and by three groups of native English speakers, differing according to French-language experience. The researcher also tested the hypothesis of whether equivalence

[41] Major (2001, as cited in Moyer, 2013) states that potential production difficulties for L2 learners may include: (1) segments; (2) syllable structure; (3) prosody (stress, segment length, or duration); (4) intonation; (5) pitch; (6) rhythm and timing; and (7) speech rate.

[42] It is now widely acknowledged that production inaccuracy is mainly attributable to perceptual biases, caused by L1 interference rather than by articulatory problems (Best & Tyler, 2007; Flege, 1995; Rochet, 1995).

classification[43] between L1 and L2 phonological systems limits the learning of a vowel similar to an L1 category (e.g., /u/ of French and English), but not the acquisition of a new vowel, that is, a vowel which has no L1 counterpart (e.g., French /y/ for native English speakers). Results showed that the three groups of English participants, who differed in amount of exposure to French, produced /y/ with formant values that approximated the French speakers' /y/. However, the least experienced group produced this vowel with F2 values considerably lower than those of native French speakers. In the case of vowel /u/, the least experienced English speakers produced a more anterior /u/ (closer to the French /y/) than their English /u/. The other more experienced native English speakers produced /u/ with F2 values that differed significantly from those of French native speakers. These results suggest that amount of exposure to the L2 determines production accuracy and seem to support the hypothesis that equivalence classification prevents experienced learners from producing similar non-native phones, but not new phones.

Flege and Bohn (1992) further examined the effect of L2 experience on production accuracy of new (/æ/) and similar (/i/, /ɪ/, and /ɛ/) English vowels by native speakers of German. Two experiments compared the production of English vowels by two groups of L1 German speakers, differing in L2 experience (inexperienced vs. experienced), namely an acoustic experiment, which analyzed the spectral and temporal characteristics of the English vowels produced in /bVt/ words, and an identification test, in which the same tokens were assessed for intelligibility. The results showed that amount of L2 experience did not affect the production of the similar vowels /i/ and /ɪ/, but production accuracy of vowel /ɛ/ was, to some extent, influenced. The group of inexperienced German speakers did not produce /ɛ/ differently from native English speakers, but the group of experienced speakers produced this vowel shorter than the native English and the inexperienced German speakers. This seems to indicate that a similar sound is learned in the early stages of L2 learning, and amount of experience does not influence production accuracy, in this case. However, the new vowel /æ/ was produced by the experienced German group, but not by the inexperienced group, with acoustic values that approximated those of native speakers. This finding corroborated the hypothesis previously tested by Flege (1987) that L2 experience improves production accuracy of a new sound.

[43] Flege (1987) describes equivalence classification as "a basic cognitive mechanism 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" (p. 49).

In the aforementioned studies, researchers were interested in comparing the learning of sounds differing in degree of equivalence when considering L1 and L2 interphonology, namely the learning of new and similar L2 vowels, and in examining the influence of L2 experience. The next studies focus mainly on the importance of age in L2 learning.

Flege (1991) examined whether age of learning (early vs. late) determined Spanish-English bilinguals learning of the different Spanish and English phonetic realizations of /t/, which differ in voice onset time[44] (VOT). Spanish /t/ has shorter VOT than English /t/. The results showed that the late L2 learners produced English /t/ with VOT values that were intermediate to Spanish short-lag values and to English long-lag values. The early learners' VOT values for English /t/ did not differ from English VOT. These findings seem to indicate that non-native speakers who learn an L2 in early childhood are able to establish phonetic categories for sounds in the L2 that differ acoustically from corresponding sounds in the native language.

Flege, Munro, and MacKay (1995b) investigated the production of English consonants by native speakers of Italian, who differed in age of learning (AOL) from two to 23 years. Word-initial, word-medial and word-final tokens of English stops and fricatives were acoustically analyzed and assessed through forced-choice judgments made by native English listeners. Data was obtained through a reading task, which consisted of 14 words embedded in a carrier sentence. Age of learning (AOL) affected significantly the productions of English consonants by ten subgroups of native Italian participants, although informants had lived in Canada for an average of 32 years, and reported speaking English more than Italian. Strong effects of AOL were noted for nearly every consonant examined and for the ten subgroups of native Italian participants.

To test the hypothesis whether L2 vowels produced by bilinguals differ from vowels produced by monolingual speakers of the L2, Flege, Schirru, and MacKay (2003) tested four groups of Italian English bilinguals, who differed in age of arrival (AOA) in Canada (early vs. late) and frequency of Italian use (low L1-use vs. high L1-use). According to the authors, native speakers of Italian tend to identify English /eɪ/ as instances of the Italian vowel category /e/, even though English /e/ is produced with more tongue movement than is Italian /e/. Acoustic analyses revealed that English /e/

[44] Voice onset time (VOT) is the interval between the release of stop closure and the onset of voicing. VOT is the most important cue for the voicing contrast.

was produced by early bilinguals, who seldom use Italian (low L1-use), with significantly more movement than native English speakers, but more accurately than late bilinguals did. The researchers explained that, although the tongue movement was overall exaggerated, early bilinguals were able to form a phonetic category for English /e/ different from Italian /e/, whereas late bilinguals did not establish a new category for English /e/, because they merged the phonetic categories of English /e/ and Italian /e/.

The context of L2 learning in the aforementioned studies is a naturalistic one, given that Flege and his colleagues have mainly examined immigrants living in a country where the L2 is the predominant language. However, most foreign language students do not live in an L2-speaking environment, and learn the non-native language (NNL) in a classroom through formal instruction. Therefore, a few studies that were conducted in a formal instructional setting in the learners' native country are briefly described next. It is not our aim to discuss the (different) conditions under which immigrants and students in a foreign language classroom learn a NNL, but to understand if the same factors, viz. age of learning and exposure to the target language, influence L2 production in a formal learning setting.

MacKay and Fullana (2009) investigated the effects of age of learning (AOL) and amount of formal exposure to the target language on the production of English vowels /i/, /ɪ/, /æ/, and /u/ by Spanish learners of English as a foreign language (EFL) in a classroom setting. Learners varying in AOL (8, and 11 years) and L2 exposure (2.5, 4.5, and 7.5 years) produced 34 English words, and seven native speakers (NS) both rated their productions in a 9-point foreign accent (FA) scale, and performed an identification task. No significant effects for the two factors were found on the accent ratings and vowel identification scores. Neither AOL nor exposure to the foreign language explained the degree of foreign accent in Catalan/Spanish speakers in the production of the English vowels. Therefore, the authors suggest than an AOL of less than 8 years and more than 7.5 years of exposure are needed to produce English vowel sounds accurately in a formal learning context.

Fullana and Mora (2009) further investigated the effects of AOL (before vs. after 8 years) and amount of exposure to the FL (school exposure vs. extra exposure) on the perception and production of voicing contrasts (/s/-/z/, /p/-/b/, /t/-/d/) in English word-final position by Catalan/Spanish advanced EFL learners in a formal learning setting. The voicing contrasts /s/-/z/, /p/-/b/, /t/-/d/ in word-final position do not occur, at the phonetic level, either in Catalan or in Spanish. Results revealed that neither starting age nor exposure had a significant effect on perception and production. However, they also demonstrated that learners who started learning English before eight years (i.e., the earlier starters) approximated English native speakers' productions by articulating slightly shorter voiced and longer voiceless consonants in word-final position, and greater vowel duration differences in some word pairs, thereby confirming the advantage of earlier starting age in naturalistic settings. Nonetheless, an increase in extra exposure by means of extracurricular language courses or/and stays abroad did not result in learners' using vowel duration differences as a cue to voicing in word-final positions. In sum, these results provided evidence that starting age (AOL) and exposure were not significant in determining Catalan/Spanish bilinguals' production of voicing contrasts in English word-final position.

The studies reviewed earlier focused mainly on L1 interference at the segmental level. However, it is important to remark that, although it is not the focus of the present research, differences in syllabification between the L1 and L2 have also been found to result in inaccurate production of non-native sounds. Therefore, an example of a study, which examined this relation between non-segmental information and L2 sound acquisition, is reported below.

In order to investigate the extent to which L1 syllable structure and phonotactics are carried over to the L2, Cebrian (2009) focused on the acquisition of L2 phonotactic constraints and syllabification processes to assess how this knowledge influences the acquisition of the English tense-lax contrast by Catalan/Spanish speakers. The English lax vowel constraint (LVC) restricts lax vowels, but not tense vowels, to closed syllables. The segmental tense-lax contrast and syllabic constraints on vowels do not exist either in Spanish or Catalan. Therefore, the LVC is a new non-segmental feature for Catalan/Spanish learners of English. Moreover, English and Catalan differ in syllable structure. The learners' knowledge of the LVC and the syllabification patterns was tested in a series of production tasks with the English vowel contrasts /i/-/ɪ/ and /eɪ/-/ɛ/ in CVC and CVCVC words. The results revealed that the L2 learners showed some knowledge of the LVC, though to a lesser extent than native English speakers. However, knowledge of the LVC did not result in target-like syllabification of medial consonants, since Catalan/Spanish learners syllabified English words following L1 principles. In addition, the greater tendency to mispronounce /ɪ/ as /i/ in CVCVC (disyllabic words) than in CVC (monosyllabic) words can be interpreted, according to

Cebrian (2009, p. 241), "as a 'tensing' strategy effect in bisyllables". The interference from L1 syllabification patterns restricted the ability to apply the LVC in bisyllables.

These findings indicate that a specific type of segmental error can be the result of non-segmental factors such as the effect of L1 and L2 phonotactics and syllable structure. Cebrian (2009) emphasizes that the assessment of L2 segmental learning needs to include the study of non-native contrasts in different prosodic and syllabic contexts, since segmental errors depend on the context they occur, that is, analyses of segments should be complemented with analyses at the suprasegmental and phonotactic levels (Cebrian, 2009, p. 242).

A few studies have also investigated the effect of production training on the learning of non-native sounds. They are briefly described in the next section, because their research design includes production training tasks, but more studies regarding this theme are reviewed in *Chapter 2*.

Flege (1988) examined the use of visual information specifying the tongue position of non-native vowels for training vowel production. Tongue-palate distances were measured at four locations along the hard palate with an optoelectronic glossometer to provide visual feedback in terms of tongue targets for English /i/, /ɪ/, /æ/, and /ɑ/. Before training, the Spanish participant, who had three years of exposure to the target language in an English-speaking country, neutralized the tongue position difference between English /i/-/ɪ/ and /æ/-/ɑ/. Physiological (viz. tongue position and tongue differences), acoustic (viz. first three formants), and identification tests performed by a native English speaker showed that the participant produced a difference between /i/-/ɪ/ after only 10 minutes of visual articulatory training. The researcher explains that a similar improvement was not observed for /æ/-/ɑ/, because the tongue differences between the latter could not be shown adequately by the glossometer. This study provided evidence that visual articulatory information shown during production training can improve production accuracy.

A more complex study by Aliaga-Garcia (2010) investigated whether native-like cue weighting is best promoted by identification (ID) or articulatory (ART) audiovisual phonetic training. Separate groups of bilingual Catalan/Spanish learners of English received ID training (i.e., perceptual training) and ART training (i.e., production training) on 11 English RP monothongs. In the identification training, participants listened to CVC words with the target vowels and had to label them from three or four alternatives, and received immediate feedback. The production training consisted of an imitation training task, in which learners were first presented with words audiovisually and then recorded the same words. At the end of each session, they heard their own productions and compared them to those produced by a native speaker, as many times as they wanted. The pre- and posttests included a four-choice categorization task with natural CVC words and a forced-choice categorization task based on eight synthetic /hVd/ continua. A significant main effect of training type was not found, meaning that ID and ART types of training did not produce significantly different effects on L2 vowel categorization. After ID and ART training, Catalan/Spanish learners of English were more reliably able to distinguish two similar vowels based on spectral cues, but no further significant differences were found in the performance of the two learners' groups.

Lacabex, Lecumberri, and Cooke (2009) also compared the effects of auditory (i.e., perceptual) training with articulatory (i.e., production) training on English vowel reduction (/ə/) in lexical words by Spanish learners of English. Perceptual training was based on discrimination exercises, and production training provided learners with articulatory and visual information, which was based on their productions of the tokens with /ə/, and individual feedback. Both experimental groups improved after training, but results revealed no differences between them, that is, the two training modalities did not result in significant differences in the perception of vowel reduction before and after training. Perception was assessed by means of a two-alternative forced-choice task (full vowels vs. schwa). Although no significant effect of training type was found, authors suggest that production training influences perceptual abilities positively, and it might be as efficient as perceptual training. Similar findings had been previously reported by Leather (1990), who examined the effect of training on the production and perception of Chinese lexical tones. The training design was similar to Lacabex et al. (2009) in that one group received training in perception, and the other was trained to use visual feedback in production. In Leather's (1990) study, the perceptual and production abilities of both groups were tested, and results suggested that both training types had positive effects, since training in one modality tended to be sufficient to enable a learner to perform well in the other.

Having reviewed some studies on L1 and L2 speech production, the next chapter focuses on the interrelation between perceptual and productive performance in L2 speech learning in more detail. It discusses some L2 speech perception theories and reviews various studies that investigated not only the relationship between perception and production, but also the effects of perceptual training on the acquisition of L2 sounds.

CHAPTER 2

SPEECH PERCEPTION

Take care of the sense, and the sounds will take care of themselves.
- Lewis Carroll

In the first section of this chapter, the physiology of the hearing process is briefly described, and vowels are characterized as perceptual categories to complement the acoustic and articulatory information provided in the previous chapter. Afterwards, some theoretical frameworks on the perception of native speech are summarized so as to facilitate the understanding of this process. Two paramount models that explain how non-native speech sounds are perceived by L2 learners are also discussed and complemented with a review of studies on cross-language speech perception. In the last sections, the relationship between L2 speech perception and production is analyzed, and the effects of perceptual training on both dimensions are exemplified with findings of several experimental studies.

2.1 The Hearing Process

Speech perception is a complex process that is difficult to define and to explain.[45] Moreover, in comparison with the study of speech production, the investigation of the hearing process is recent and, consequently, further knowledge of all the implied domains is still needed (Reetz & Jongman, 2009). The complexity of speech perception underlies the first part of this chapter, where several theories proposing different views of this process are discussed.

Despite the inherently complex nature of the speech perception process, it has been generally defined as the ability to recognize and identify auditorily and visually phonetic segments as speech sounds in a given language (Fowler & Magnuson, 2012;

[45] Hayward (2000) suggests that "the first difficulty confronting any (…) researcher into speech perception is deciding just what speech perception is" (p. 105), and Ladefoged (2005) adds that "there is no one way of perceiving speech" (p. 110).

Hayward, 2000).[46] Therefore, from this point forward, speech perception refers solely to the recognition of phonetic segments.

More specifically, speech perception is "the conversion of continuous acoustic signals into a set of discrete units", that is, into words and their constituent consonants and vowels (Reetz & Jongman, 2009, p. 251). At the acoustic level, speech is a complex, flowing and continuous[47] (i.e., non-segmented) signal. Conversely, at the perceptual level, speech signals consist of individual phonemic segments, that is, vowels and consonants. However, speech perception is not linear, and phonetic segments "are not like beads strung on a string, one segment after another" (Yeni-Komshian, 1998, p. 111). Rather, speech is like "a braid in which the properties that help us identify phonetic segments are tightly intertwined and overlap greatly" (*ibidem*). This notion will be further addressed in Section 2.1.1.

Contrary to the analysis of speech production carried out in this study, which is focused on acoustic measurements of vocalic segments, that is, on the output of cognitive production processes, our investigation of speech perception aims at understanding how acoustic properties of signals are cognitively processed (i.e., perceived) as speech sounds. Therefore, several theories that have been developed to explain how this process occurs will be further reviewed, but, firstly, the physiology of hearing is described to provide some background for the understanding of the auditory processing of speech.

The hearing process can be divided into two parts: the peripheral auditory system and the central auditory system. The peripheral system includes the external, middle and internal ear parts, and the central system consists of the neural system with the auditory nerve, the auditory cortex, and other parts of the brain involved in speech perception. In the auditory periphery (represented in Figure 10), the mechanical vibrations of the sound wave (i.e., the acoustic signal) are converted into neural impulses (i.e., neural signals), and in the central part, neural signals are decoded in the auditory cortex as speech (Hayward, 2000; Reetz & Jongman, 2009; Johnson, 2003). This process performed by the peripheral auditory system is described next.

[46] Hayward (2000, p. 105) adds that the term *speech perception* can be also applied to the study of word recognition. Hence, as an alternative to the former generic term the author suggests the designation *phonetic perception* to refer to the study of the recognition of individual phonemes. Since the present study focuses on phonetic segments exclusively, the terms *speech perception and phonetic perception* are used interchangeably.

[47] There are no pauses at the start and end points of words and the production of each phoneme is affected by the surrounding (preceding and following) phonetic segments due to the effect of coarticulation.

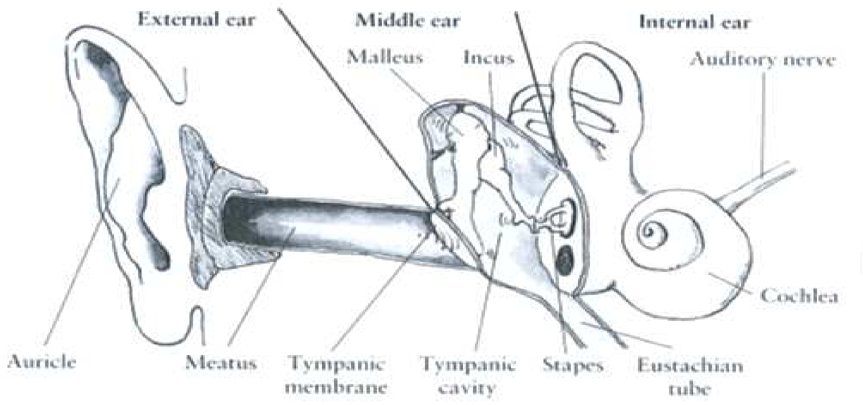


Figure 10. The hearing organ: external ear, middle ear, and internal ear (Reetz & Jongman, 2009, p. 227).

Firstly, the sound wave passes the auricle (i.e., the pinna, which is the visible part of the external ear); secondly, it flows into the ear canal (i.e., the meatus, which is about 2.5 cm long, and 0.8 cm wide), and travels down to the eardrum[48] (i.e., the tympanum). The tympanic membrane, located at the boundary between external and middle ear (more precisely, at the eardrum), is set into vibration and these vibrating movements are transmitted to a chain of three bones in the middle ear (i.e., the ossicles).[49] The bone chain transmits the sound waves to the basilar membrane (i.e., cochlear) at the end of the internal ear,[50] and the basilar membrane causes the surrounding (perilymphatic) fluid inside the internal ear to vibrate. These vibrations are, then, transformed into neural impulses that are transmitted through the auditory nerve into the brain (Delgado-Martins, 1998; Reetz & Jongman, 2009).

In acoustic terms: (1) the ear canal functions as a filter that amplifies the frequencies of the sound wave; (2) the tympanic membrane protects the inner parts of the ear from high sound pressure levels through the mechanism of partial oscillations, which reduces the total force of transmission of high sounds; (3) the ossicles increase the pressure variations of the sound wave; (4) the cochlea amplifies sound pressure when going from air into a liquid and additionally attenuates sound frequencies, in

[48] The eardrum is a thin elastic membrane of skin that moves due to air pressure variations of sound waves (Reetz & Jongman, 2009, p. 228).
[49] The bone chain consists of the malleus (hammer) that touches the tympanic membrane and transmits its movements into the incus (anvil), which in turn transmits them to the stapes (stirrup). The ossicles convert the pressure changes from an elastic medium (air) to pressure changes in a liquid (water) (Reetzer & Jongman, 2009, p. 228).
[50] In the internal ear there is a bony spiral tube (about 3.2-3.5 cm long) - the cochlea - filled with (perilymphatic) liquid. Internally, the cochlea is divided into two passages: the upper scala vestibuli and the lower scala tympani, which are separated by the basilar membrane.

particular lower frequencies; (5) the basilar membrane resonates in response to the frequencies[51] of the sound waves, and it is this oscillation of the basilar membrane[52] that is essential for the transformation of pressure waves into neural impulses in the cochlea; and (6) the nerve cells determine the amplitude of a sound signal and which frequencies it contains by measuring where the membrane oscillates.

The transformations on the acoustic signal performed by the auditory system are so many that the representation of the signal by the electric nerve impulses may be quite different from the initial acoustic input signal. Therefore, it is important to make a distinction between the physical stimulus (waveform) and the auditory sensation (perception) of speech sounds. More detailed descriptions of the auditory system can be found in Hayward (2000), Johnson (2003), Delgado-Martins (1998), and Reetz & Jongman (2009).

Taking into account the importance of complementing articulatory and acoustic descriptions of speech sounds with auditory and perceptual information, the segments under study, that is, vowels are described as perceptual categories in the following subsection.

2.1.1 The perception of vowels

As exemplified earlier by the “braid” comparison, the term *speech perception* can refer not only to the recognition of phonetic segments but also to the processing of other linguistic factors, such as lexical, syntactic, semantic and pragmatic information. In this sense, it would be a synonym for interpretation of spoken speech. However, since the focus of this study is on how listeners analyze the acoustic patterns of non-native contrasts and perceive them as speech sounds, other linguistic factors are not discussed. Thereby, this subsection describes the acoustic properties, that is, the perceptual cues that are essential to perceiving English vowels.

As mentioned in *Chapter 1*, the acoustic signal of vowels is very loud because they are produced with an unobstructed airstream that flows through a relatively open vocal tract. Perceptually, vowels are also very salient, because they are voiced and rather long when compared with other speech sounds. As explained in the previous chapter, the most important acoustic cue to vowel quality is the location of the formant

[51] Higher frequencies cause the membrane to oscillate more prominently at the base, where it is thicker and narrower, than at the apex, where the membrane is thinner and wider (Reetz & Jongman, 2009).
[52] The basilar membrane has the organ of Corti that is linked to the auditory nerve (Reetz & Jongman, 2009).

frequencies. Each vowel is produced with a distinct location of the formant frequencies that correspond to particular vocal tract configurations (see, for example, Figure 4). When hearing a vowel, the listener extracts information of the first two (or three) formants, and maps the formantic pattern into a phonemic category. For example, if F1 has a frequency of 300 Hz and F2 a frequency of 2300 Hz, an English listener may perceive the formant frequencies as being the vowel /i/. However, some experiments[53] have shown that not only formant frequencies measured in the steady-state portion[54] of the vowel are important cues to vowel quality, but also dynamic properties, in particular vowel duration and formant transitions, provide essential perceptual information to the recognition of vocalic segments (Lieberman & Blumstein, 1988; Reetz & Jongman, 2009; Yeni-Komshian, 1998).

Moreover, several factors influence vowels' formant frequency patterns, including phonetic context (e.g. Van Son & Pols, 1995; Treiman, Kessler, & Bick, 2003; Treiman, Kessler, Devin, Bick, & Davis, 2006), speaking rate (e.g., Gottfried, Miller, & Payton, 1990; Kessinger & Blumstein, 1998) and speakers' vocal tract size (adults vs. children; women vs. men). Specifically, variations in phonetic context, that is, in the segments that flank the vowel, and in conversational speaking rate contribute to vowel undershoot[55] (Reetz & Jongman, 2009). Moreover, vowel perception is even more challenging when produced by a variety of speakers, because vowels do not have the same formant frequencies across a group of speakers (compare, for example, the sex differences in terms of F1 and F2 frequencies in Table B1). Despite these variations, the listener must perceive the different acoustic patterns as pertaining to the same phonemic category. In order to do so, that is, to remove speaker variation, listeners perform a normalization process. Reetz and Jongman (2009) define normalization as a "procedure for factoring out systematic covariation in acoustic properties so that the apparent overlap among vowel categories can be reduced or eliminated" (p. 254). The ability to perceive vowels produced by multiple speakers implies, therefore, the activation of this pattern recognition system, that is, *normalization* (Reetz & Jongman, 2009; Yeni-Komshian, 1998).

[53] For example, Hillebrand and Nearey (1999); Macchi (1980); and Strange et al. (1976), as cited in Reetz and Jongman (2009, p. 256); and Jenkins et al. (1986), as cited in Yeni-Komshian (1998, p. 126).

[54] It is the middle portion of the vowel that is distant from flanking segments and is considered minimally or not affected by contextual effects, such as coarticulation.

[55] Reetz and Jongman (2009, p. 252) explain that vowel undershoot occurs when formant frequencies do not reach their "optimal" values, i.e., those values as measured in vowels produced in isolation. This happens, for example, in the production of CVC words in which the articulators do not reach their optimal position for the vowel due to the articulatory configuration of the initial and final consonants and/or the speed of production. The articulators undershoot their target positions and, hence, the formant frequencies undershoot their optimal values. Yeni-Komshian (1998) calls this phenomenon *vowel under-articulation*.

Although there are a few models that explain (speaker or vocal tract) normalization, their discussion is beyond the scope of this subsection. See, for example, Reetz and Jongman (2009) for a detailed description of normalization models.

2.2 Theories of Speech Perception

As aforementioned, in terms of production, speech sounds are highly variable acoustic signals that differ according to several factors, including phonetic context, speakers' characteristics (age, sex, accent), and speaking rate, but phonetic perception is "resilient" to all this variation (Fowler & Magnuson, 2012, p. 13). The phonetic constancy of speech perception despite the lack of acoustic invariance poses challenges to theories of speech perception that try to explain this process. In order to understand how listeners map the extracted acoustic information from the speech signal into cognitive representations, we will summarize four major theoretical frameworks of the perception of speech, in particular the motor theory, the direct realist theory, the fuzzy logical model, and the analysis by synthesis approach. These models are primarily concerned with perception of phonetic segments, and thus do not explain the effects of lexical and other higher-level cognitive knowledge into the process of speech perception.

The motor theory was initially proposed by Liberman and Mattingly (1985) to explain the problem of lack of invariance.[56] The lack of acoustic invariance raised the question of which kind of information listeners use to perceive speech sounds, and researchers departed from the belief that, if there is invariance in the acoustic domain, the solution would be to find invariance in the articulatory domain. Therefore, this theory claims that the objects of speech perception are articulatory gestures rather than acoustic or auditory events. This view emerged from the finding that perceived phonemes have a simpler (one-to-one) relationship to articulation than to acoustics. More specifically, the motor theory hypothesizes that articulatory events perceived by listeners are neuromotor commands to the articulators (e.g., tongue, lips and vocal folds) rather than more peripheral events such as actual articulatory movements. The neuromotor commands are also referred to as intended (i.e., abstract) gestures, which exist in the speakers's mind.

[56] The lack of invariance is the lack of stable acoustic features of a given segment in a variety of contexts. Some of the variability is due to the immediate phonetic context (coarticulation), but it can also arise from a change in the overall context, relating to such factors as speaking rate and speakers' sex, age, and dialect (Hayward, 2000; Fowler & Magnuson, 2012).

Both the motor theory and the direct realist theory posit that the immediate objects of speech perception are articulatory gestures, rather than acoustic events (i.e., the physical changes in frequency, amplitude and duration of the speech signal). However, while for the motor theory the objects of perception are intended gestures, or the neuromotor commands that generate those gestures, the direct realist framework advances that these objects of speech perception are actual vocal tract movements. According to this theory, developed by Fowler (1986, 2003), the listener uses information from the acoustic signal to recover actual vocal tract gestures.

The fuzzy logical model, proposed by Dominic Massaro (1987), tries to explain the complex mapping of acoustic features onto cognitive representations by viewing speech perception as a probabilistic process of matching features to prototype representations in memory. According to this model, listeners have a set of prototypes stored in memory, corresponding to the numerous perceptual units of language (viz. V, VC and VC syllables), which contain a number of distinctive features. The features of the prototype correspond to the ideal values that a segment should have to belong to that category. For example, the prototype for Portuguese /pa/ is characterized as having initial closure of the lips. To decide whether or not a given CV syllable is or is not /pa/, the listener has to integrate information form a variety of sources. Each source of information may or may not provide a clear answer; thus, answers can be not only “yes” or “not” but also “possibly”. By combining definite and fuzzy truth values from many available sources of information, listeners evaluate whether the consonant /p/ is labial (Hayward, 2000, p. 126). In sum, feature information is evaluated, integrated and matched against prototype descriptions in memory, and recognition is made on the basis of the relative goodness of match of the segmental information with the relevant prototype descriptions.

The analysis by the synthesis approach advanced by Stevens and Blumstein (1978) hypothesizes that listeners perceive (i.e., analyze) speech by implicitly generating (i.e., synthesizing) speech from what they hear and then compare the “synthesized” version with the auditory speech sound. According to this model, perceptual mapping involves the extraction of abstract (i.e., invariant and distinctive) phonemic features from the acoustic signal. In addition, this approach suggests that various acoustic properties increase the auditory salience of phonological contrasts.

More detailed accounts of these models can be found in Diehl, Lotto and Holt (2004), Fowler and Magnuson (2012), Reetz and Jongman (2009), and Yeni-Komshian (1998).

In the remaining part of this section, three characteristics of speech perception are briefly explained because they contribute to the understanding of some key concepts later mentioned in this dissertation.

The first is the concept of *categorical perception*, which is the ability to perceive "stimuli, equally spaced along some physical continuum, as belonging to one or another perceptual category instead of as varying as a function of their physical values" (Reetz & Jongman, 2009, p. 265). As aforesaid, when listening to speech, perceivers hear speech sounds not as a continuum but as a number of discrete categories. This can be observed when listeners are asked to identify (i.e., label) stimuli varying along the continuum, and they switch abruptly rather than gradually, between categories, and when listeners perform better at discriminating between stimuli (i.e., hearing the difference between a pair of sounds) that belong to different categories than they do between stimuli which pertain to the same category. In short, categorical perception implies that "listeners perceive the stimuli not as sounds pure and simple, but rather as members of categories" (Hayward, 2000, p. 117).

The second feature of speech perception is that it is a multi or bimodal process. This means that perception of speech involves the integration of not only auditory information but also visual information conveyed by facial gestures such as lip movements. The extent to which these multiple cues are integrated in the recognition of all phonetic segments is not discussed here, but it seems clear that the combination of both visual and auditory cues facilitates the identification of certain segments (e.g., labials) to a greater extent than others (e.g., velars). The most cited study that investigated the role of visual information in speech perception is that of McGurk and MacDonald (1976). In their study, the production of four CVCV nonsense words (*baba*, *gaga*, *papa* and *kaka*) by a woman was filmed. When participants watched the videos or listened to the stimuli they identified the four words correctly. When the videos were dubbed in such a way that the visual labials (*baba* and *papa*) were combined with audio velars (*gaga* and *kaka*), they heard *dada* or *tata*. In the reverse situation, specifically audio bilabials with visual velars, they tended to hear words that included both types of consonant, such as *bagba* or *paka*. This showed that what the participants heard was influenced by what they saw, and this phenomenon of visual information influencing what listeners hear is known as the McGurk effect (Hayward, 2000).

Another important aspect regarding speech perception is that it becomes language-specific at a very early stage of human development. How listeners come to perceive sounds in a manner that is particular to their L1 is briefly explained as follows. Although children and adults show some malleability in their phonetic categories, infants as young as two to four months are sensitive to a great variety of subtle acoustic differences, regardless of these differences serving to differentiate phonemes in their ambient language. That is, infants categorically discriminate among phonemes of their native language as well as many other non-native phonemes. However, this sensitivity weakens at the end of the first year of life (e.g., Werker, 1989; Werker & Tees, 1984, 1999). Hence, in the first 12 months the first language acts as a filter that emphasizes native sound contrasts while de-emphasizes non-native contrasts. By the end of the first year, infants start to become perceptually tuned to their L1 (Diehl, Lotto, & Holt, 2003), that is, they respond to the speech sounds in a language-specific manner, discriminating acoustic differences between phoneme categories of their language, but no longer distinguishing sounds within those categories (e.g., Werker & Tees, 1984, 1999). This means that there is a shift from a language-general to a language-specific pattern during the first year of life. For example, Polka and Werker (1994) observed the same pattern of change for English-speaking infants' discrimination of two non-native contrasts. The study by Polka and Bohn (1996), which assessed language-specific influences on infant vowel perception, concluded that English and German infants (6-8 and 10-12 months) did not differ in their discrimination of either German /dut/-/dyt/ and English /dæt/-/dɛt/ contrasts, and there were no age differences in the discrimination of either contrasts for English and German infants. However, in both language groups at both ages, there were clear differences in performance related to the direction in which the vowel change was presented to infants. For the German contrast, discrimination was poorer when presented from /dut/ to /dyt/ and for the English contrast discrimination was poorer when the contrast changed from /dæt/ to /dɛt/. According to the researchers, these directional asymmetries seem to point to a language-universal perceptual pattern, which suggests that vowels produced with extreme articulatory gestures function as perceptual attractors in infant vowel perception.

This finding of an apparent language-universal perceptual pattern was further tested (see Polka and Bohn, 2003, for a review of data on infant vowel perception asymmetries) and, consequently, the same researchers proposed a framework for understanding early phonetic development, viz. the Natural Referent Vowel (NRV) framework (2011). The NRV was initially supported by the finding that directional asymmetries tend to occur in infant vowel discrimination, which seems to indicate an underlying perceptual bias favoring vowels that are located closer to the periphery of the F1/F2 vowel space. As a result of this evidence, the researchers propose that vowels with extreme articulatory-acoustic properties (i.e., peripheral in the vowel space) act as natural referent vowels, which influence and guide the development of vowel perception. Moreover, they add that language experience influences the default pattern established by the initial vowel perception biases, which means that through experience with a specific language learners access other vowel categories and organize their vowel perception to optimize native language perception.

Directional asymmetries in vowel perception studies, reported by Kuhl and colleagues, led to the development of the Native Language Magnet (NLM) model (Kuhl & Iverson, 1995). The NLM model argues that exposure to language early in life produces a change in perceived distances in the acoustic space underlying phonetic distinctions, and this subsequently alters both the perception and the production of spoken language. Infants and adults were tested on sets of synthetic vowels that fall within the same vowel category but vary in their category goodness-of-fit to include prototypic and non-prototypic instances of the same vowel. Results revealed that both adults and infants find more difficult to distinguish a prototypical example of /i/ from variants that surround it in the acoustic space as compared with distinguishing a non-prototypical instance of vowel /i/ from its variants, even though the auditory distance between (non)prototype and variants is the same for both conditions. The prototypical vowels in the native language act like perceptual magnets, attracting other members of the category and, thereby, reshaping the perceptual space to facilitate access to native categories. According to the researchers, this magnet effect also contributes to the difficulty that listeners have perceiving non-native phonetic distinctions in a new language. Native language magnets warp the perceptual space, resulting in the attraction of similar sounds, which, in turn, makes certain non-native contrasts difficult to distinguish. The prediction of the NLM model is that perceptual difficulty depends on the proximity of a given L2 sound to an L1 magnet. The nearer it is to a magnet, the more it will be assimilated to the L1 category. The idea is that good instances of native-language categories act as magnets that filter the new language's phonetic units. Hence, experience with L1 speech sounds leads to the creation of category prototypes or high-density representations of exemplars that act as perceptual magnets, which distort the perceptual space.

2.3 Second Language Speech Perception

The research area that investigates the acquisition/learning[57] of non-native sounds is commonly referred to as the study of cross-language speech perception and production, which is also known as phonetics of second language acquisition (Reetz & Jongman, 2009). A question that this research field has extensively addressed is the extent to which the phonological system of the L1 influences the perception and production of an L2. Findings have shown that the relation between the sound inventories of the L1 and L2 is a major predictor of the way in which non-native phonemes will be perceived and pronounced.

Although there are several theoretical frameworks,[58] such as the BiPhon model (Boersma, 2007), the Second Language Linguistic Perception (L2LP) model (Escudero, 2005), and the Automatic Selective Perception (ASP) working model (Strange, 2011) that explain how non-native sounds are perceived by L2 learners, in this section we only briefly review two of the main speech perception models that explain L2 phonetic category acquisition, namely the Speech Learning Model (SLM) proposed by Flege (1995) and the Perceptual Learning Model (PAM-L2) advanced by Best and Tyler (2007), and developed from Best's (1995) Perceptual Assimilation Model (PAM). Both models depart from cross-language (L1-L2) phonetic similarity to predict learners' success or failure in the acquisition of non-native segmental contrasts.

In L1 acquisition research, the hypothesis that production problems are caused by inaccurate perception has been explored. However, studies that investigated the link between these two dimensions, namely the ability to perceive speech sounds and their misarticulation, concluded that "the cause of many L1 segmental production errors are to be found at a 'motor level rather than at a 'mentalistic' (level) of linguistic organization" (Locke, 1980, as cited in Flege, 1995, p. 237). According to Flege (1995), this conclusion drawn for L1 acquisition cannot be applied to L2 learning in the same

[57] The debate on the use of the terms "learning" and "acquisition" of a second language (L2) is not discussed, but the former will be understood generally as L2 learning in formal instructional contexts, whereas the latter will be used scarcely to refer to L2 learning in naturalistic environments, i.e., in a foreign language-speaking country.
[58] For a review of L2 Speech Perception models, see, for example, Boersma and Hamann (2009).

way, because speakers tend to perceive the non-native sounds in relation to their L1 phonology and, thus, differently from native speakers.

The Speech Learning Model (SLM) is concerned with the ultimate attainment of L2 pronunciation, that is, the final developmental state of learning,[59] and posits that many L2 production errors have a perceptual basis, that is, are perceptually motivated and happen through the filter of the L1. Therefore, it hypothesizes about how L2 sounds are perceived by non-native speakers and predicts their production accuracy based on their perceptual performance. The SLM departs from the distinction between *new* and *similar* phones, advanced by Flege (1987) in an earlier study about the production of non-native sounds by L2 learners. According to the researcher, an L2 sound is perceived as *new*, when it is not equivalent to any L1 sound (i.e., when it has no counterpart in the L1, and therefore differs from all L1 phones), or *similar*, when it resembles (but is not identical) to an established native category. An L2 sound can be also perceived as being *identical*, that is, equivalent to an L1 category. Flege (1987, 1995) predicts that if an L2 sound is perceived as new (or sufficiently dissimilar from L1 categories), the learner tends to create a new category for that sound, whereas if it is perceived as similar, that is, as an allophone of an L1 sound (i.e., perceptually equivalent to an L1 phone), the learner is not able to establish a new category for that sound. In sum, the SLM proposes that the process of equivalence classification prevents learners from creating a new category for a similar, but not for a new L2 sound. This process makes learners' perceive a non-native phone 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 an L2 phonetic category as an allophone of an L1 sound; conversely, the greater the perceived phonetic dissimilarity between L1 and L2 phones, the more likely it is that phonetic differences between the sounds will be perceived, and a new category will be created. This process of perceptual equivalence classification helps determining L2 learners' production accuracy.

The four postulates of Flege's SLM (1995) model are presented and briefly summarized below.

[59]According to Birdsong (2004, p. 82), "ultimate" is not to be thought of as a synonym of "native-like," although native-likeness is one of the observed outcomes of L2 acquisition.

> P1. The mechanisms and processes used in learning the L1 sound system, including category formation, remain intact over the life span, and can be applied to L2 learning.
> P2. Language-specific aspects of speech sounds are specified in long-term memory representations called phonetic categories.
> P3. Phonetic categories established in childhood for L1 sounds evolve over the life span to reflect the properties of all L1 or L2 phones identified as a realization of each category.
> P4. Bilinguals strive to maintain contrast between L1 and L2 phonetic categories, which exist in a common phonological space. (p. 239)

The first postulate refers to the plasticity[60] of the perceptual system, that is, the innate potential to modify perceptual patterns throughout the lifespan. It states that the ability to perceptually learn new L2 phonetic categories is available to adults learning a non-native language. The second postulate entails the idea that phonetic categories are mental representations of the distinctive features of speech sounds. The third premise suggests that listeners continuously modify their perception of L1 and L2 speech gestures, and the last postulate suggests that L1 and L2 phonetic categories exist in a common phonological space. Although the SLM hypothesizes that the ability to create new non-native sound categories continues over the lifespan, it also suggests that L2 speech learning is limited by learners' age and native language (L1). Age of learning (AOL) is, according to Flege (1995), a very influential factor on phonetic perception.

Since the SLM claims that production accuracy is related to perceptual phonetic category representation, inaccurate perception will cause inaccurate production, and any age-perceptual constraints will also apply to the production domain. However, the author suggests that the perceived relation of L1 and L2 sounds and language-specific perceptual patterns may change if extensive naturalistic L2 input and exposure are provided. If these conditions are met, new non-native sound categories can be established and near native-like perception[61] can be achieved.

Best (1995) proposes the Perceptual Assimilation Model (PAM) that establishes correspondences between L1 and non-native phonological categories and predicts how listeners of an unfamiliar foreign language will assimilate non-native phones with respect to their L1 phonological categories and, specifically, how they will discriminate non-native contrasts. The PAM focuses on non-native speech perception by naïve

[60] This notion of plasticity is associated with the critical period hypothesis (Lenneberg, 1967) that suggests that there is an ideal time window during which native proficiency with a language can be obtained. Although there is not yet consensus about the existence and duration of this critical period, after which language acquisition becomes more difficult, it is now well documented (e.g., Werker & Tees, 1984) that there is a high degree of plasticity in the perceptual system not only in infants and children, but also in adults.

[61] The model predicts that even when a category is established for an L2 sound, it might not be produced exactly as it is produced by native speakers (Flege, 1995, p. 243).

speakers (i.e., unfamiliar with the target language), and tries to predict non-native sound categorization at initial stages of learning. Best and Tyler (2007) adapted the PAM to predict the learnability of non-native sounds by experienced L2 learners (i.e., familiarized with the L2) from the category assimilation types proposed in the PAM. In order to present the models chronologically, the PAM is summarized first and, then, the PAM-L2.

According to the PAM, which is grounded on a direct-realist approach to perception, non-native contrasts are perceived in terms of their (gestural) phonetic similarity to the phonological categories of the L1, that is, articulatory gestures are assumed to be perceptual objects. The PAM makes predictions about discrimination performance taking into consideration the way two segments of a contrastive pair are perceptually assimilated. The model posits that when hearing an unfamiliar non-native speech sound naïve listeners are likely to perceptually assimilate the non-native phone to the most articulatorily similar L1 sound. The PAM predicts the occurrence of six patterns in the perceptual assimilation of non-native contrasts: (1) if two non-native phones are perceived as exemplars of two different native phonemes (TC – *two category assimilation*), discrimination is expected to be excellent; (2) poor discrimination is predicted if two non-native sounds are perceived as equally good or poor instances of the same native phoneme (SG – *single category assimilation*); and (3) when two non-native phones are heard as instances of the same native phoneme but with different category goodness ratings intermediate discrimination is predicted (CG – *category-goodness assimilation*). If one or both members of an L2 sound pair are very dissimilar from any L1 category, that is, non-existing in the L1, they are uncategorized in the L1 system. The PAM predicts (4) very good discrimination for categorized-uncategorized (UC) contrasts, (5) while discrimination of two uncategorized sounds (UU) depends on the extent to which each phone maps onto a distinct L1category. Non-native sounds can also be perceived (6) as non-speech sounds and, thus, are nonassimilable (NA).

The PAM-L2 (Best & Tyler, 2007) posits hypotheses on L2 listeners that are acquiring experience with non-native contrasts. According to the researchers, this model is compatible with the SLM equivalence classification, a process by which an L2 phonological category is perceptually assimilated to the L1 phonological entity.

The four perceptual assimilation patterns predicted by the PAM-L2 are described below and compared to the PAM.

1) *Only one L2 phonological category is perceived as equivalent (perceptually assimilated) to a given L1 phonological category.*

In this case, the learner perceives an L1 and an L2 phonological category as being equivalent. This may happen in an uncategorized-categorized (UC) assimilation or in a two-category (TC) assimilation type.

2) *Both L2 phonological categories are perceived as equivalent to the same L1 phonological category, but one is perceived as being more deviant than the other.*

This case is the same as the category goodness (CG) assimilation. Learners are expected to discriminate these L2 phonemes well, but not as well as in two-category (TC) assimilation types. Therefore, a new L2 phonemic category is likely to be established for the deviant L2 phone, while the L2 phone that is perceived as a better exemplar is likely to be perceived as equivalent to the L1 category, and hence no new category is formed.

3) *Both L2 phonological categories are perceived as equivalent to the same L1 phonological category, but as equally good or poor instances of that category.*

This is a case of single-category (SG) assimilation. The authors add that the L2 learner may initially have difficulty discriminating these L2 sounds, which are likely to be assimilated to the single L1 category, and minimally contrasting L2 words are expected to be heard as homophones. In SLM terms, both L2 speech sounds would be merged with the L1 phonetic category (cf. Flege, 1995; Flege, Shirru, & Mackay, 2003). Whether or not L2 listeners can learn to perceive a difference between single-category assimilated L2 phones depend on whether they are perceived as good or poor exemplars of the L1 phoneme.

4) *No L1-L2 assimilation.*

If the naïve listener does not perceive either of the contrasting L2 sounds as belonging to any single phonological L1 category, but rather as having a mixture of more modest similarities to several L1 phonological categories (uncategorized, in the PAM terms), then one or two new phonological categories may be relatively easy to learn perceptually. This is similar to the SLM's concept of *new* phone. In the PAM's formulation it is not only the similarity or dissimilarity of a given L2 phone to the closest individual L1 phonetic category that is crucial to perceptual learning, but also its comparative relationships within the interlanguage phonological system. These four perceptual patterns consider non-native sounds that are perceived as speech, that is, that are assimilated into the L1 phonological space of the L2 learner. However, as the PAM also predicts, there are speech sounds, such as Zulu clicks, which are perceived as non-sounds by non-click L1 speakers. In these cases, two situations may occur: they might be perceptually assimilated into the phonological space of the L2 listener as uncategorized speech sounds, possibly resulting in the learning of one or two new phonological categories, or L2 learners may never assimilate these sounds into their phonological space.

The two models previously summarized depart from cross-language (L1-L2) phonetic similarity to predict learners' success or failure in the acquisition of non-native segmental contrasts. However, unlike the SLM that considers both the production and perception aspects of language processing when proposing that pronunciation problems have a perceptual basis, the PAM-L2 focuses exclusively on L2 speech perception. Flege's model predicts that L2 speech sounds are easier to acquire if they differ phonetically from those in the L1, that is, the process of equivalence is addressed only at a phonetic level, whereas the PAM-L2 considers equivalence both at the phonetic and the phonological levels. Finally, the PAM focuses on the perception of L2 contrasts by naïve listeners, whereas the SLM focuses on the production and perception of L2 contrasts by late experienced learners. The PAM-L2, however, was developed to also predict L2 perceptual assimilation by L2 adult learners.

In conclusion, although adult listeners have difficulty perceiving many non-native contrasts, perception is not uniformly poor, as demonstrated by the SLM and the PAM-L2. Furthermore, the perceptual system exhibits a considerable plasticity such that phonetic categories can be modified even in adults. However, the perceptual difficulty that adult learners may have cannot be predicted solely on the basis of a comparison between the L1 and L2 phonemic inventories. Instead, research indicates that several other factors need to be taken into account, including the specific acoustic cues that distinguish a phonemic contrast, the allophonic distribution of phonetic segments in the L1 and L2, the age of the learner, the amount of exposure to the L2, the quality of the input, and the extent of usage of both the L1 and L2. Therefore, some studies on L2 speech perception are discussed next to exemplify the previous observation.

Due to the great number of cross-language studies on L2 speech perception that have been conducted since the 1960s, just a few experiments are summarized in this section in order to point out some aspects involved in L2 vowel categorization, namely the weighting of perceptual cues, the role of L2 experience, and the influence of consonantal context in the perception of vowels. The importance of examining the perceived distance of L1 and L2 sounds is also briefly discussed.

The degree of difficulty in perceiving non-native contrasts varies significantly, depending on various factors, viz. the psychophysical salience of the acoustic parameters differentiating phonetic contrasts, similarities and differences in the phonetic structure of L1 and L2 categories, and the phonetic and phonotactic contexts in which contrasts occur. In addition, though L2 learners may learn to perceptually differentiate non-native contrasts, they may use different acoustic cues from those used by native speakers and when this happens L2 speakers seem to have perceptual difficulties.

A number of studies have demonstrated that, although L2 adult speakers can learn to differentiate a non-native contrast, they may still use different perceptually acoustic parameters from those used by native speakers, that is, their selective perceptual strategies may differ. We will refer to three studies that investigated the weighting of acoustic cues used to differentiate non-native vowel contrasts, whether or not they are used in L2 learners' native languages. These studies assessed the weighting of temporal (i.e., vowel duration) and spectral (i.e., degree of openness and frontness/backness of the vowel) cues in the perception of English vowel contrasts.

For instance, Bohn and Flege (1990) investigated the perception of the English front vowels /i/, /ɪ/, /ɛ/, /æ/ by two groups of adult native speakers of German differing in language experience (experienced vs. inexperienced). Researchers considered vowels /i/, /ɪ/, and /ɛ/ perceptually similar, if not identical, to German vocalic phonemes, and /æ/ a new vowel for German learners of English. In the synthetic continua (*beat-bit*, *bet-bat*), in which vowel formants and duration varied, both groups of German learners of English perceived the similar English contrast /ɪ/-/i/ as native speakers did, and both groups of speakers used spectral cues to do so. However, inexperienced German speakers used durational cues to distinguish the English /ɛ/-/æ/ contrast, instead of the spectral cues native speakers use. This finding suggests that when spectral cues are insufficient to differentiate L2 vowel contrasts, durational cues are used instead.

A study by Bohn (1995) further investigated cue weighting by three groups of young adult learners of English by examining which acoustic cues (spectral or/and duration) native speakers of German, Mandarin and Spanish use to identify the English front vowel contrasts /i/-/ɪ/, and /ɛ/-/æ/, which differ in both vowel spectrum and duration. The three experiments followed the same design as Bohn and Flege (1990).

The target vowels were identified from two continua, one ranging from *beat* to *bit*, and the other from *bet* to *bat*. The results provided evidence that German speakers used both spectral and durational cues to identify the L2 /ɛ/-/æ/ contrast, which was predicted since vowel pairs in their L1 differ in both spectrum and duration. Moreover, duration was a more important cue to the /ɛ/-/æ/ contrast for them than for native English speakers, as expected because duration is an important cue used by German speakers to differentiate vowels. However, in Spanish, duration is not used contrastively, and in Mandarin there is only a duration difference between two of the four tones. In spite of this, Spanish and Mandarin listeners relied on the temporal cue to differentiate the English vowel contrasts, rather than on spectral cues. According to Bohn (1995), these findings suggest that L1 experience does not entirely determine how non-native vowel contrasts are perceived. The results from Spanish and Mandarin speakers indicate that duration cues in vowel perception are easy to access whether or not listeners have had linguistic experience with them. This perceptual strategy used by L2 learners is designated by Bohn (1995) as the *desensitization hypothesis.* This principle states that "whenever spectral differences are insufficient to differentiate vowel contrasts because previous linguistic experience did not sensitize listeners to these spectral differences, duration differences will be used to differentiate the non-native vowel contrast" (Bohn, 1995, pp. 294-5).

Cebrian (2006) also assessed the relative weighting of quality and duration cues in the identification of the English vowel contrast /i/-/ɪ/ by adult Catalan/Spanish speakers, whose L1s have no durational contrasts. This was evaluated by means of an identification test with a synthetic vowel continuum from /i/ to /ɪ/ to /ɛ/. Results for /i/ and /ɪ/ showed that L2 learners made a greater use of temporal cues than native English speakers did, corroborating Bohn's (1995) desensitization hypothesis. This study also examined the role of L2 experience (length of residence in an L2-speaking country) in non-native vowel categorization by means of a perceptual assimilation task, but no positive effect of experience was found. Both groups of L2 learners relied on duration regardless of amount of L2 experience.

The use of different perceptual cues to distinguish non-native contrasts has also been documented in studies about L2 consonant perception. For example, Yamada (1995) compared native speakers of Japanese, native speakers of English and Japanese-English bilinguals' perceptual performance by asking listeners to identify stimuli in which F2 frequency, and F3 and F1 transitional duration varied factorially (F2 and F3 onset and transition frequencies, and F1 onset duration) of synthetic /r/ and /l/ continua. The researcher found that native speakers of Japanese tended to focus more on F2 than on F1/F3, while English speakers predominantly focused on the F1/F3 cues to distinguish the liquid contrast. The bilingual speakers did not focus on these cues as much as native Japanese speakers did.

Learners also seem to have perceptual difficulties when L1 uses only one cue and L2 uses multiple cues to perceive a certain segment. For instance, Flege (1984) investigated the perception of English final fricative /s/ and /z/ by experienced and inexperienced Arabic speakers. There is a "trading relation" between final fricative duration and preceding vowel duration in English. Native speakers of English lengthen a vowel if a following final segment is short (voiced) and shorten a vowel if the following final segment is long (voiceless). The authors wanted to examine if the Arabic speakers could identify final /s/ and /z/ by controlling consonant and vowel duration. The experienced Arabic speakers did not show any difference in identifying the two segments, whereas the inexperienced speakers were different in that they did not use consonant durations, but rather used vowel duration to distinguish final /s/ and /z/. Since Arabic has phonemic vowel length, the L2 speakers used that L1 perceptual cue to identify /z/. Flege and Hillenbrand's (1986) study further examined the differential use of temporal cues to the English /s/-/z/ contrast in word-final position by native and non-native speakers of English. The results showed that the French, Swedish and Finnish speakers used cues established for the perception of phonetic contrasts in their native language to identify fricatives /s/ and /z/ instead of simultaneously using (i.e., integrating) both vowel and fricative temporal cues that are used by native speakers to distinguish the English /s/-/z/ contrast. Their perceptual voicing judgments relied predominantly on vowel duration than on fricative duration.

In terms of research on the influence of phonetic context in vowel perception, three studies are reviewed next. In 1995, a study by Van Son and Pols emphasized that vowel identification improved significantly when context beyond the CV and VC transitions was added, which indicates that vowel perception is context-dependent. The researchers compared vowel identification in three contexts, namely in isolation (V) (only the central 50 ms of the vowel), in CV and VC contexts, and in CVC words. The presentation of the vowel in isolation proved inadequate for determining vowel identity, but adding a flanking consonant CV, but not VC, decreased incorrect identification. In CVC contexts vowel identification scores increased considerably.

To investigate whether and how cross-language perception of vowels is affected by the consonantal context in which vowels occur, Bohn and Steinlen (2003) conducted a study with native Danish learners of English. Native Danish listeners identified the eleven Southern British English (SBE) monophthongs produced in three consonantal contexts (/hVt/, /dVt/, and /gVk/) using Danish response categories, and provided goodness ratings of their identifications. Results indicated that perceptual assimilation of some vowels is highly affected by consonantal context. The location of the Danish vowels in the acoustic F1/F2 space was only slightly affected by flanking consonants, whereas SBE vowels were strongly affected by consonantal context. It seems, thus, that Danish learners transferred their L1 expectations about lack of vowel coarticulation to SBE. According to the researchers, these results seem to indicate that patterns of perceptual assimilation and degrees of learning difficulty of non-native vowels cannot be predicted exclusively from comparisons between L1 and L2 phoneme inventories, because vowel perception depends on the phonetic contexts in which segments occur. This supports earlier findings such as Strange, Yamada, Fitzgerald, and Kubo's (1998), whose study on perceptual assimilation of American English vowels by native Japanese speakers concluded that cross-language perceptual patterns are not context-independent.

However, Strange, Bohn, Nishi, and Trent (2005) did not find any evidence of context-dependent vowel assimilation when investigating perceptual assimilation of North German and American English vowels in CVC syllables (C=labial, alveolar, velar stops) in sentences. Results showed that perceptual assimilation patterns did not vary with context, which suggests that listeners adopt a context-independent strategy when evaluating cross-language similarity of vowels produced and presented in continuous speech contexts. According to the researchers, training materials that include vowel contrasts in contexts graded from easier to more difficult might optimize performance and lead to robust L2 phonetic categorization.

Flege, Munro, and Fox (1994) investigated the relation between vowel distance in an F1-F2 acoustic space and perceived vowel dissimilarity in two experiments with English monolinguals and native Spanish learners of English. These groups of listeners rated the dissimilarity of two Spanish vowel categories, two English vowels, or one Spanish and one English vowel category. L2 experience was also examined, but no differences were found in dissimilarity ratings between experienced and inexperienced Spanish participants, which suggested that naturalistic exposure to an L2 is unlikely to result in changes in the perception of vowel dissimilarity. Results showed that for both native English and Spanish speakers perceived dissimilarity increased as the distance between vowels in the F1-F2 acoustic space increased, which demonstrates the importance of F1 and F2 frequency on the perception of vowel quality and vowel dissimilarity.

Guion, Flege, Akahane-Yamada, and Pruitt (2000) found that the perceived phonetic distance of L1 and L2 sounds predicts learning effects in the discrimination of L2 sounds by testing native Japanese speakers' perception of English consonants. The first experiment was a cross-language mapping experiment in which Japanese listeners identified English and Japanese consonants in terms of a Japanese category and then rated the identifications for goodness-of-fit to that Japanese category. The second experiment was a categorial discrimination test with the same stimuli. Three groups of Japanese speakers, varying in English-language experience, and one group of native English speakers participated, and the contrastive pairs tested included two English consonants, two Japanese consonants, and one English and one Japanese consonant. The results showed that the perceived phonetic distance of L2 consonants from the closest L1 consonant predicted the discrimination of L2 sounds. In terms of L2 experience, some of the consonant contrasts showed evidence of learning, with experienced speakers obtaining significantly higher scores than the inexperienced Japanese groups. Cebrian, Mora, and Aliaga-Garcia (2010) propose that a combination of two methods of perceptual assessment, viz. a perceptual assimilation task and a rated dissimilarity task, is advantageous to obtain more reliable measures of cross-language perceptual similarity.

A final aspect of L2 vowel perception is now mentioned, because it supports the Natural Referent Vowel (NRV) framework proposed by Polka and Bohn (2011), previously cited. Kerschhofer-Puhalo (2010) investigated the perception of German vowels by a group of learners of German with five different L1s (Albanian, Polish, Romanian, SerboCroatian, and Turkish) by means of a cross-language categorization test. Results showed that peripheral vowels were perceived more accurately and were often used as perceptual targets and substitutes. Asymmetries in vowel perception suggest that in the acquisition of L2 vowels there are language-universal perceptual preferences. The NRV model postulates language-universal biases towards more peripheral vowels in contrast pairs. The more peripheral vowels seem to facilitate the formation of language-specific vowel categories, whereas the less peripheral vowels tend to be assimilated to the reference vowel. The results of the study on German vowel categorization showed that incorrect categorizations occurred more often with universally disfavored vowels, while peripheral vowels /i/, /a/, and /u/ were better categorized.

For a detailed historical review of cross-language studies on speech perception see, for example, Strange (1995).

2.4 The Interface between L2 Speech Perception and Production

Several hypotheses have been suggested to explain the cause of foreign accent in L2 speech production, including that neural maturation might reduce neural plasticity (Penfield, 1965; Lenneberg, 1967) leading to a diminished ability to add or modify sensorimotor programs for producing sounds in an L2. Another explanation suggests (Flege, 1995; Rochet, 1995) that foreign accents are caused by the inaccurate perception of sounds in an L2. To explore this further, the interface between L2 speech perception and production is discussed in this section.

The relationship between perception and production has been the focus of several studies for the last few decades, but the nature of this link still remains relatively unclear (Rochet, 1995). Research has provided consistent empirical evidence for the existence of a link between perception and production. However, the question of which speech domain comes first (if perception precedes production or production precedes perception) is still debatable, especially when considering the abilities of second language learners. Although studies on L1 speech acquisition have consistently supported that acquisition of perception precedes that of production,[62] L2 research has only been able to show that there is a tendency for perception to precede production. Moreover, it seems that L2 learners tend to reach higher ultimate attainment in production than in perception, that is, learners' perceptual difficulties may remain after production has been mastered (e.g., Sheldon & Strange, 1982). In this sense, L2 learning follows a different pattern from L1 acquisition. Besides, L2 learning by adults and children also differs, being the latter often more successful than the former (e.g., Baker, Trofimovich, Flege, Mack, & Halter, 2008). Both perceptual and productive skills appear to be associated with the age at which children and adults are first exposed to the L2, indicating that age may mediate the relationship between these two abilities.

[62] Strange (1995) points out that "native language patterns of perception are well established long before children have mastered the production of the phonetic segments and sequences of that language" (p. 40).

Furthermore, it also seems that L2 learners who are able to perceive a non-native contrast may accomplish the task in a different way than native speakers do, that is, their selective perceptual strategies may differ, as explained previously. Next, we will briefly review some studies on the relationship between perception and production in L2 acquisition to clarify the previous observations.

Numerous studies have provided empirical evidence that show that there is a link between perception and production inaccuracy. Some studies that investigated the relationship between vowel perception and production are reviewed next, but a few studies that examined L2 consonant acquisition are also mentioned to provide evidence for the earlier statements. There are various studies that investigated perception and production of non-native vowels, but only an exemplary set is summarized.

Rochet (1995) analyzed Portuguese and English speakers' perception and production of the French vowels /u/, /i/ and /y/ and found a correlation between inaccurate perception and production of vowel /y/. When /y/ was produced incorrectly, Portuguese speakers tended to produce it more like /i/, whereas English speakers produced it more like /u/. In perceptual identification tests, Portuguese and English speakers also performed differently from native French speakers. Portuguese speakers identified /y/ as /i/ and English speakers as /u/, which followed the same pattern as that observed in production. These findings indicate that there is a strong correlation between perception and production.

Flege, Bohn, and Jang (1997) investigated the effect of L2 experience in the perception and production of the English front vowels /i/, /ɪ/, /ɛ/, and /æ/ by 80 native speakers of German, Spanish, Mandarin and Korean. Significant differences between experienced and inexperienced non-native speakers of English were found in terms of production and perception accuracy, with both dimensions being related. Differences in degrees of accuracy in producing and perceiving the English vowels depended on L1 background, possibly because of the differences in the perceived relation between the L1 and L2 vowel inventories. Another study by Flege, Mackay, and Meador (1999) assessed the relation between perception and production of a larger set of English vowels by native Italian speakers, whose age of arrival (AOA) in Canada differed. As in Flege et al. 1997, the results showed that L2 experience influenced both production and perception, since speakers' accuracy in producing and perceiving English vowels diminished as AOA increased. Moreover, a significant correlation between vowel production and perception was found, and the authors concluded that L2 vowel production accuracy was limited by how accurately they were perceived.

Rauber, Rato and Silva (2010) carried out two experiments to test the perception and production of English front vowels by native speakers of Mandarin. The findings showed that perception accuracy outperformed production in the case of the L2 similar vowels /ɪ/ and /æ/, which were better identified than produced. They found that perception and production were interrelated: higher identification accuracy rates were related to better production results, whereas lower identification rates were related to poorer production.

Considering the cross-linguistic similarity between European and Brazilian Portuguese, the results of three studies on English vowel perception and production by Brazilian learners of English as a foreign language (EFL) will be reported. Rauber, Escudero, Bion, and Baptista (2005) investigated the identification and production of English vowels, and their findings confirmed that inaccurate production was related to poor categorical discrimination. Bion, Escudero, Rauber, and Baptista (2006) further investigated this relation by examining the perception and production of a smaller subset of English front vowels (/i/, /ɪ/, /ɛ/, /æ/) and provided evidence for a strong link between perception and production, since greater discrimination in the perception test was related to better production results. Rauber (2010) also examined whether there was a correlation between the perception and production of /i/, /ɪ/, /ɛ/, /æ/, /u/, and /ʊ/ by proficient English learners. Her findings showed that the rate of accurate perception of L2 vowels was higher than that of production, indicating that the vowels which were better perceived were also the ones produced more accurately by L2 learners, which in turn confirmed the interrelation between perceptual and productive abilities.

Overall, the previous vowel studies provided evidence of a relationship between L2 sound perception and production, and a precedence of perceptual categorization over production learning. Flege, Munro, and Mackay (1995b) hypothesize that “an L2 phone must be perceived in a fully native-like fashion if it is to be produced in a fully native-like fashion” (p. 22). However, a few studies on L2 consonant categorization also provided evidence that production accuracy may precede perceptual learning.

For example, Sheldon and Strange (1982) investigated the relationship between the production and perception of the English liquids /r/ and /l/ by native speakers of Japanese learning English. The results showed that the production of the liquid contrast was more accurate than their perception of the pair. The difficulty in perceiving the contrast varied with its position in the word, namely prevocalic /r/ and /l/ in consonant clusters caused the highest percentage of perceptual errors, while word-final liquids were accurately perceived. The authors suggest that the pattern of errors might be the result of acoustic phonetic factors, namely difficulty in using the native distinctive spectral cue, the transition of third formant (F3), that distinguishes English /r/ and /l/. However, as Flege (1991) emphasizes, this result should be interpreted carefully, because "the Japanese speakers might have formal English education to direct them to use articulatory strategies such as 'to produce /l/, put the tongue on…'" (p. 265).

Finally, Wilson, Saygin, Sereno, and Iacoboni (2004) provided cortical evidence for the link between perception and production by using functional magnetic resonance imaging (fMRI). The researchers showed that acoustic input activates specific areas in the brain that are responsible for articulatory representation. Although no claim is made on which ability comes first, the results suggest that perception precedes production.

After the review of studies relating L2 speech perception and production, the next section will summarize research on perceptual training to improve L2 perception and production skills.

2.5 Perceptual Training

Research on phonetic training started in the early 1960s as a means to assess L1 performance of people with speech disorders, such as dyslexia and aphasia. As a consequence of the positive results in the field of speech therapy, training started to be applied to L2 learning as a way to improve phonetic abilities of non-native speakers and, consequently, to enhance intelligibility and reduce foreign accent (Nobre-Oliveira, 2010).

A considerable number of empirical studies have shown that L2 speakers can learn to perceive and produce non-native contrasts in a native-like manner, mainly because the perceptual processes used in learning the L1 sound system can be applied to L2 learning and "remain intact over the life span" (Flege, 1995, p. 239).

Substantial research has been conducted to investigate whether and how perceptual training on non-native segments and suprasegments can modify L1 perceptual patterns. Results have revealed complex findings in which several variables interact to determine the extent to which language-specific perceptual patterns can be adapted and modified. Bohn (2000) identifies three types of variables:

> subject variables, which define what a listener brings to the task of perceptually organizing non-native contrasts (e.g., L1 background, L2 experience, age, etc); task variables, which define the experimental procedures used to assess changes in perceptual patterns (e.g., identification or discrimination, memory load in different tasks, etc); and stimulus variables, which define what the listener is trying to organize perceptually (e.g. consonant or vowel contrasts, temporal or spectral cues, etc). (p. 11)

Over the last three decades, numerous studies have focused on the effects of perceptual training on the perception and production of non-native segmental and suprasegmental contrasts, and have reported its positive effects both on modification of sound perceptual patterns and on improvement of production accuracy, thus supporting the claim that there is a close relationship between production and perception. Although these studies have, to a certain extent, similar objectives, they have tested different hypotheses and have followed different procedures.

The scope of research in this field is very wide insofar as not only perceptual training of segments (consonants and vowels) but also suprasegments (stress, tone) have been the focus of study. The immediate effects of training on both perceptive and productive abilities of L2 learners have been investigated, as well as long-term effects, and generalization[63] of learning. Therefore, we will review several perceptual training studies following, as far as possible, a chronological order. Due to the emergence of a series of cross-language perceptual training studies in the last years, a selection of the most relevant studies to the present research has been made. Recent studies of vowel perceptual training are also summarized.

Two of the first pioneering studies investigating the effects of perceptual training focused on the ability of adults to identify and discriminate stop voice onset time (VOT) contrasts that are not phonologically distinctive in their native language. Pisoni, Aslin, Perey, and Hennessy (1982) trained 12 monolingual native English speakers to perceive three voicing categories, namely synthesized prevoiced, voiceless unaspirated, and voiceless aspirated stops differing in VOT (-70 ms, 0 ms, and + 70 ms, respectively). The training consisted of an ABX discrimination task[64] with immediate feedback[65] and

[63] Generalization is the ability to transfer the acquired knowledge to multiple dimensions, such as to novel productions of new talkers, to new productions of the same talker, to new tasks. Generalization has been used as a measure of robustness of learning (Hardison, 2004; Lively et al., 1994; Logan & Pruitt, 1995).

[64] ABX discrimination task is one of the three categories of discrimination tasks (the others are the AX task, and the category change task). In an ABX task three stimuli are presented, one of which is the same of the other two, and the listener indicates which two are the same, or alternatively, which is the different one (Logan & Pruitt, 1995, p. 355).

[65] Feedback is the information provided to learners about their performance in the training tasks. It enables participants to determine whether what they are doing is appropriate or not and, thus, whether they have to modify their response or not. Feedback can be immediately given to listeners on a trial-by-trial basis, or after longer periods, such as blocks of trials or an entire training session (Logan & Pruitt, 1995, p. 362).

it involved four one-hour sessions. The results demonstrated that the adult perceptual mechanisms used to categorize stop consonants can be modified with simple laboratory training techniques in a short period of time (after only one hour of training, half of the participants reached 85% accuracy in identifying the three stops differing in VOT). A subsequent study by McClaskey, Pisoni, and Carrell (1983) aimed at training English speakers to discriminate synthesized voiced and voiceless stops (VOT from – 70 to +70 ms). One group of participants was trained with labial stops, and the other group with velar stops. The researchers reported that after discrimination training of VOT, most of the participants could discriminate both labial and velar stops, and this perceptual ability was transferred to another place of articulation (alveolar), which was not part of the training. Another study which examined the effects of laboratory training on stop consonants was conducted by Werker and Tees (1984). The researchers investigated whether English-speaking adults' perception of non-native contrasts could be modified using laboratory training involving multiple natural stimuli and continuous feedback. Thirty English speakers were trained to discriminate voicing and place contrasts, viz. the Hindi breathy voiced-voiceless dental aspirated stops (voicing contrast) and the dental-retroflex stops (place contrast). The results showed that, after a short period of training, participants could perceive the voicing contrast, but not the place contrast.

Jamieson and Morosan (1986) conducted a perceptual training study consisting solely of an identification task[66] (to avoid calling attention to between-category distinctions) using synthesized stimuli with immediate feedback. Native speakers of Canadian French were trained on the voicing distinction between the English interdental fricatives /θ/ and /ð/ in synthetic CV syllables. The researchers used the perceptual fading technique[67] by manipulating the formants of the natural stimuli gradually. The contrast enhancement was obtained by an increase in the fricative duration in the synthetic CV tokens. The contrast was perceived more easily in the enhanced condition, and the duration of the fricative was gradually reduced during the training program until the normal duration was presented. The results showed that after the 90-minute identification training, which consisted of 12 sessions of three to eight blocks each, the participants improved their within-category perceptual ability and transferred it to untrained natural stimuli in both identification and discrimination posttests.

[66] In an identification or labeling task a stimulus is presented and the listener has to give a label to the segment (Logan & Pruitt, 1995, p. 357).
[67] The fading technique consists of enhancing the perceptual contrast between stimuli at the beginning of the training and gradually reducing it towards the end of the training (Logan & Pruitt, 1995).

Rochet and Chen (1992) also used a perceptual fading technique to train Mandarin speakers to identify a voicing contrast between French word-initial labial stops (/p/ and /b/, starting from VOT – 60 to + 130 ms). The aims of this study were to (1) examine whether perceptual training with synthetic stimuli promoted the learning of the French stop voicing contrast, (2) investigate whether improvement in perception transferred to production, and (3) observe whether learning of the labial stop voicing contrast was carried over to stops with a different place of articulation, and to different phonetic contexts. Mandarin speakers were trained on the French voicing contrast with identification tasks over six 30-minute sessions divided into nine blocks. The researchers used the fading technique by manipulating the VOT values of /p/ and /b/ so that the participants could gradually become aware of the distinction. In addition, participants had to achieve a minimum of 95% of identification accuracy in a block to advance to the following block.[68] The speakers showed significant improvement in the identification of the trained voicing contrast and generalization to untrained phonetic contexts and to the voiceless stops /t/ and /k/. However, as to generalization of learning to natural stimuli, only generalization to the voiceless targets was found, with no significant change on the perception of natural tokens of voiced stops. The results of the production data revealed transfer of improved perceptual performance to the production level. These findings suggest that adults can learn to perceive and produce non-native contrasts with limited perceptual training.

In one of three experiments conducted in 1989, Flege also focused on the training of a voicing contrast, namely the word-final English stops /t/ and /d/ (e.g. *beat* vs. *bead*) by Mandarin and Shanghainese speakers. The researcher reported that after training, which included multiple natural stimuli with immediate feedback, there was improvement in the perception of trained words, and generalization to untrained words. However, the Chinese speakers of Shanghainese, whose dialect has words ending in a glottal stop, seemed to benefit more from the training than Mandarin speakers whose L1 has no word-final stops. This latter finding was important because it suggested that L1 syllabification might also influence L2 learning.

Several perceptual training studies have also given attention to the English /r/-/l/ contrast but one of the first studies that investigated the effects of perceptual training on this contrast was the one conducted by Strange and Dittman in 1984. The aim of this

[68] This method is called *adaptive training*. Participants have to master a structure to proceed to a more difficult one (Jamieson & Morosan, 1986).

study was to improve native Japanese speakers' perception of English r/ and /l/, which are not contrastive sounds in their native language. Training involved AX discrimination tasks[69] with synthesized stimuli of the target contrast in minimal pairs such as "rock-lock" with immediate feedback. The researchers used synthesized speech, that is, manipulated stimuli with gradually increasing between- and within-category differences, namely a synthetic "rake-lake" continuum, varying in F1, F2, and F3 frequencies. After 18 training sessions over three months, Japanese speakers showed significant improvement in synthetic stimuli identification and oddity discrimination in the "lake-rake" continua. The discrimination task was found to direct attention to between-category distinctions. However, generalization to the identification of natural tokens did not occur.

In the 1990s, various studies followed Strange and Dittman's (1984) research on the effects of perceptual training on the /r/-/l/ contrast with Japanese speakers, namely Pisoni and colleagues' series of studies (e.g., Bradlow, Pisoni, Yamada, & Tokhura, 1997; Bradlow, Yamada, Pisoni, & Tohkura, 1999; Lively, Logan, & Pisoni, 1993; Lively, Yamada, Tokhura, & Pisoni, 1994; Logan, Lively, & Pisoni, 1991; Yamada, Tohkura, Bradlow, & Pisoni, 1996). The first of these studies (Logan, Lively, & Pisoni 1991) trained six Japanese speakers to identify the liquid contrast /r/-/l/ in 15 forty-minute sessions of 272 trials each over three weeks. The stimuli used in the identification task were recorded by various talkers. This method of input recording, called high-variability identification training procedure, which aimed at preventing the participants from relying on specific vocal characteristics of the talkers, rather than on the target sounds, proved to be a successful procedure to promote robust learning. The results indicated not only improvement in perception, but also transfer of learning to novel stimuli, and to new talkers. The second study by Lively, Logan and Pisoni (1993) investigated the effects of two different training procedures, namely identification tasks (1) with stimuli in five phonetic contexts and recorded by a single talker, and (2) with stimuli in three phonetic contexts (initial position, initial consonant clusters, and intervocalic position) and recorded by five talkers. The results demonstrated that variability of contexts and multiple talkers play a significant role in promoting perceptual learning and robust category formation. Lively, Pisoni, Yamada, Tohkura,

[69] AX discrimination task is one of the three categories of discrimination tasks (ABX task, category change task). In an AX task a pair of stimuli is presented and the listener indicates whether the stimuli are the same or different (Logan & Pruitt, 1995, p. 355).

and Yamada (1994) used the same high-variability[70] approach to investigate the long-term effects of training, that is, whether there would be retention of improvement, by perceptually testing the trained Japanese speakers (Lively et al., 1993) three and six months after the training was over. The posttest results had revealed an immediate perceptual improvement of 16%. Three months later, identification accuracy reduced only 2% from the posttest; however, a higher reduction of improvement was observed six months later, but still with a retention of 4.5% of accuracy in relation to the pretest. The results suggested that the high-variability training method promotes a long-term modification of speakers' phonetic perception. Bradlow, Pisoni, Yamada, and Tohkura (1997) conducted the fourth study and adopted the same high-variability approach of Logan et al. (1991). The main aim of the study was to verify if there would be transfer of perceptual improvement to speech production. The perceptual training consisted of 45 sessions of identification training with natural stimuli recorded by five talkers over a period of three to four weeks. The production data of eleven Japanese speakers was assessed through a word-reading task and an imitation task with naturally spoken /r/-/l/ minimal word pairs. The results showed improvement in perception with generalization to new words and to new talkers and, the most important finding, transfer to production, without any specific production training. As a follow-up to the previous study, Bradlow, Yamada, Pisoni, and Tohkura (1999) investigated long-term retention of improvement in both speech perception and production. This study on the /r/-/l/ contrast followed the same testing and high-variability training procedures previously used in Bradlow et al. (1997). The perceptual training consisted of fifteen 20-30-minute sessions over the same period of time (three to four weeks). The results of the perceptual tests showed that the trained Japanese speakers improved significantly from the pretest to the posttest and maintained improved levels of perceptual performance in the three-month retention test. Perceptual evaluations by native American English listeners of the Japanese speakers' productions showed that not only did the participants improve from the pretest to the posttest, but the effects of training were retained three months after the training was over. Another study by the same researchers, Yamada, Tokhura, Bradlow, and Pisoni (1996), used the same method to investigate further whether training in the perception domain would transfer to improvement in production and whether there would be retention of learning three and six months after training was over. Native

[70] The high variability phonetic training (HVPT) is a method which consists of presenting L2 learners with natural stimuli produced by multiple talkers in various phonetic contexts.

speakers of Japanese received 45 sessions of perception training over 15 days (three sessions per day). The results of the tests showed significant improvements from pretest to posttest in both perception and production. Moreover, participants retained improvement in perceptual and production performance three and six months after the conclusion of the training. These results show that training in the perception domain can produce long-term modifications in both speech perception and production.

The findings of this series of studies were fundamental for cross-language speech research because they provided evidence that there is a close relationship between speech production and perception in the development of non-native speech contrasts, and that perceptual training promotes robust improvement of both perceptual and productive abilities of learners. Furthermore, in terms of training procedures, the aforementioned studies enhanced the importance of stimulus and talker variability, that is, of a high variability approach, to achieve significant and robust improvement in the identification and pronunciation of non-native sound contrasts.

In the following decade, the effects of perceptual training on L2 speech perception and production were extensively investigated, and previous findings were systematically tested, most of which were replicated, but a few were not observed. For example, Trapp and Bohn (2000) observed training effects in the perception of the English word-final /s/-/z/ voicing contrast by Danish speakers, and generalization to new phonetic environments and untrained talkers, but they found no transfer of learning to production. Reis and Nobre-Oliveira (2008) observed perceptual training effects in the production of only two of the three trained English voiceless stops /p/, /t/, and /k/ by Brazilian learners of English, and identification and discrimination tests revealed contradictory results. The identification accuracy of the English voiceless stops improved, but results of the discrimination test showed that there was a significant negative effect of training.

Therefore, given the considerable number of cross-language studies conducted in recent years, only some of the research that provided relevant new findings and applied innovative training procedures are reported below. For example, in the 2000s, several studies (e.g., Wang, Spence, Jongman, & Sereno, 1999; Wang, Jongman, & Sereno, 2003; Wayland & Li, 2008) showed that the positive effects of perceptual training are not restricted to the segmental level. Wang et al. (1999) conducted research that indicates that perceptual training is also effective in the acquisition of non-native suprasegmental contrasts, namely Mandarin tones. Using the high-variability training method, American learners of Mandarin were trained in increasing difficulty in eight 45-minute sessions to identify four tones in natural monosyllables produced by four native Mandarin speakers. The results indicated improvement in identification accuracy with transfer to new stimuli and novel talkers, and retention of the perceptual learning six months after the training was over.

Wang et al. (2003) further investigated whether the Mandarin tone contrasts learned perceptually would transfer to production. The same trained speakers of American English who participated in the previous perceptual study were recorded producing a list of Mandarin words, and their productions were acoustically analyzed and judged by native Mandarin speakers. Results revealed significant tone production improvement after perceptual training. Wayland and Guion (2004) focused on training tones as well. Native Chinese speakers (five Taiwanese speakers and one Mandarin speaker) and native English speakers were trained on the identification and discrimination of mid-low tone contrasts in Thai. The identification and discrimination training with immediate feedback consisted of five 30-minute sessions. The results showed that in discrimination both groups of participants improved, but the native Chinese speakers outperformed the native English speakers, and in identification a significant improvement was observed in the Chinese speakers' group, but not in the English speakers' group. The researchers concluded that previous experience with tones in one language facilitates the learning of tones from other languages. In 2008, Wayland and Li further evaluated the effects of two perceptual training procedures on the discrimination of the mid-low tone contrast in Thai. Native English speakers and native Chinese speakers received training for two sessions of 60 minutes using either an identification task or a categorial (same/different) discrimination task to improve their ability to discriminate the Thai mid-low tone contrast. The results suggested that both training procedures were equally effective in improving the participants' perceptual discrimination of tone contrast in Thai, although, as in the study by Wayland and Guion (2004), native Chinese speakers outperformed native English speakers, which suggests that prior experience with tone contrasts may benefit the learning of a tone language.

The introduction of visual cues[71] in perceptual training studies was examined in a series of studies by Hardison (2003, 2004) and Sueyoshi and Hardison (2005), who added visual information (viz. the talker's face) in training tasks. Hardison (2003)

[71] Cue is an acoustic or visual property of a speech sound that is phonetically contrastive.

focused on the perception of the widely investigated /ɹ/ - /l/ contrast by native Japanese and Korean speakers, and compared the effects of audiovisual (AV) training to those of auditory training. Both kinds of training consisted of identification tasks, in which the participants would hear/see a stimulus, label it, receive feedback, and hear/see the same stimulus again (reinforcement). The results showed that the two trained groups (native speakers of Japanese and Korean) 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. Generalization to novel stimuli and to a new talker was also found, as well as improvement in production accuracy. The researcher suggested that the audiovisual training provided extra information that contributed to the formation of new phonetic categories.

Hardison (2004) also included the visual modality in the training of suprasegmentals. Native English speakers were trained on French pitch contours over thirteen 40-minute sessions with 90 short French sentences with feedback corresponding to real-time visual displays of pitch contour. In the training, participants read each sentence, and its pitch contour appeared in the computer screen simultaneously to reading. Then, participants heard the same sentence produced by a native French speaker, and the corresponding pitch contour was displayed on the screen. After that, the two pitch contours were overlaid in a contrasting color in an additional window. After receiving this feedback, participants then reread the sentence. Results showed a significant improvement in prosody accuracy. Similarly to Hardison (2003), findings revealed greater improvement for the group trained with the audiovisual modality than for the audio training group. Hazan, Sennema, Iba, and Faulkner (2005) also focused on the use of visual cues for speech perception by investigating the effects of audiovisual perceptual training on L2 learners' perception and production of a new phonetic contrast. The use of visual cues by native Japanese speakers was evaluated for two English contrasts, the labial-labiodental contrast /v/-/b/-/p/ and the /r/-/l/ contrast. In the first experiment, Japanese learners of English participated in ten sessions of either auditory (A) or audiovisual (AV) training on the labial-labiodental contrast, and results indicated that the audiovisual training was more effective than the auditory training. In the second experiment, participants undertook ten sessions of perceptual training of the /r/-/l/ contrast with either auditory stimuli, natural audiovisual stimuli or audiovisual stimuli with a synthetic face synchronized to natural speech. The results showed that perception of the /r/-/l/ contrast improved in all groups, but learners trained audiovisually did not improve more than those trained auditorily. The learners' production of /r/-/l/ improved significantly after perceptual training, and a greater improvement was obtained for the AV natural training group. This study shows that sensitivity to visual cues for non-native phonemic contrasts can be enhanced through audiovisual perceptual training. The researchers conclude that audiovisual training is more effective than auditory training when the visual cues to the phonemic contrast are sufficiently salient, and it also leads to a greater improvement in production, even for contrasts with relatively low visual salience.

The manipulation of acoustic cues has been another focus of some perceptual training studies. Iverson, Hazan, and Bannister (2005) investigated whether cue weighting can be altered by perceptual training and, thus, compared the effectiveness of three different cue manipulation techniques in perceptual training: (1) all enhancement (enhancing the contrasts by manipulating F3); (2) fading technique (F3 enhancement gradually reduced during training); and (3) secondary cues (manipulating F2). Japanese speakers were assigned to four training conditions – one natural and three with cue manipulation – and received high variability training (with natural words from multiple talkers) on the identification of the English /r/-/l/ contrast over two to three weeks. The results showed that all the training techniques improved /r/-/l/ identification, but there were no significant differences between them. Training also changed the use of secondary acoustic cues.

Pruitt, Jenkins and Strange (2006) manipulated the acoustic cue - vowel duration - in the stimuli used for training. Native American English speakers and native Japanese speakers were trained on the identification of the Hindi place contrast, namely the dental stop-retroflex stop, which does not exist in the phonemic inventories of English or Japanese. However, English and Japanese differ in their phonetic relationship to Hindi regarding this contrast. English has both dental and retroflex allophones of alveolar stops, whereas Japanese, unlike English, has a contrast similar to Hindi, specifically, the Japanese /d/ versus the flapped /r/ which is sometimes produced as a retroflex. The manipulation consisted of four levels of shortening the duration of the vowel in the stimuli's CV syllables. The training consisted of twelve 30- to 50-minute sessions of ten listening blocks, which gradually increased in difficulty by decreasing vowel duration in stimuli and by adding new talkers. Vowel shortening in CV syllables increased discrimination accuracy of the target consonants and, thus, all participants improved significantly in distinguishing the contrast. However, Japanese listeners outperformed the American participants, showing that phonemic experience with a place of articulation contrast similar to the trained Hindi dental-retroflex contrast influences perceptual performance. Learning was transferred to three untrained consonantal contexts, to a new vowel context and to a new talker's productions, but not to natural stimuli.

In order to study speech perception there are not only indirect methods that allow inferences on how native and non-native speech sounds are perceived, such as identification and discrimination tasks, but also more direct procedures as brain imaging techniques[72] that allow researchers to observe, to a certain extent, brain activity when a listener hears speech sounds. Recent studies have demonstrated that the effects of perceptual training observed in participants' performance can also be assessed by brain imaging.

Callan et al. (2003) investigated whether perceptual training can induce neural plasticity with the use of functional brain imaging (fMRI). The researchers examined localized changes in brain activity to observe the effects of one-month extensive perceptual identification training on the English /r/-/l/ contrast with nine Japanese speakers. Before and after training, functional brain images were obtained for identification of the English /r/-/l/ contrast (difficult for Japanese speakers), /b/-/g/ contrast (easy) and /b/-/v/ contrast (difficult) to observe whether there was modification in areas of brain activation. Identification tests, which evaluated behavioral performance, were conducted inside the scanner before and after training, and the training procedure was the same followed by Bradlow et al. (1997). The behavioral results showed an improvement in the identification of the /r/-/l/ contrast. Concerning brain imaging, the results indicated that improved identification performance may be due to the acquisition of perceptual-articulatory (i.e., motor) mappings. Speech perception of a difficult non-native phonetic contrast appears to be facilitated by learning-induced plasticity in cortical as well as subcortical regions that are potentially involved with formation of perceptual–motor mappings between auditory and articulatory representations of speech. The researchers concluded that using perceptual training for acquisition of difficult L2 phonetic contrasts may be beneficial.

[72] These include, for example electro-encephalography (EEG), magneto-encephalography (MEG), and functional magnetic resonance imaging (fMRI) (Reetz & Jongman, 2009, p. 226).

A study of both behavioral and cortical changes resulting from learning a new L2 contrast, viz. the Mandarin tone, was conducted by Sereno and Wang (2007). In a first experiment, eight speakers of English, beginning learners of Mandarin Chinese, were trained to identify Mandarin tones in a two-week high-variability training program, consisting of eight 40-minute sessions with an identification task followed by feedback. Perception was assessed via an identification test with 100 stimuli, and production was evaluated by native speakers' judgments and acoustic analyses. Both perception and production accuracy improved substantially after training, learning of tonal distinctions generalized to new talkers and new stimuli, and six months after training perceptual improvement was still observed. In sum, behavioral improvements were observed, and the HVPT promoted consistent and robust learning of the Mandarin tones. To test whether these behavioral changes could be observed at the cortical level, participants were evaluated prior to training and immediately after training using two fMRI scans, during which participants performed a tone identification task with 40 words. The fMRI data revealed cortical changes associated with tone training, including an increase in volume of activation (in the Wernicke's area), as well as the involvement of neighboring neural areas (viz. the left hemisphere superior temporal gyrus, and the right hemisphere inferior frontal gyrus), showing a cortical recruitment of resources needed to identify unfamiliar tones (Sereno & Wang, 2007, p. 252).

Ylinen et al. (2009) investigated whether phonetic training can modify non-native cue weighting by using behavioral and electrophysiological methods, specifically by measuring the mismatch negativity (MMN) brain response that has been used to examine long-term memory representations for speech sounds. The researchers compared the use of spectral and duration cues of English tense /i/ and lax /ɪ/ vowels (e.g., *beat* vs. *bit)* between native Finnish and English speakers before and after training. Although there is a duration difference between /i/ and /ɪ/, native speakers of English rely primarily on spectral cues or on the integration of spectral and duration cues to identify both phonemes. In contrast, Finnish L2 users of English seem to base their identification on vowel duration, because duration is used phonologically in their L1 to separate short and long phonemes. Therefore, Finns were expected to weigh duration cues more than native English speakers. The behavioral results suggested that, before training, the Finns relied more on duration in vowel recognition than the native speakers of English did. However, after training, the Finns were able to use the spectral cues of the vowels more reliably than before. Accordingly, the MMN brain responses revealed that training had enhanced the Finns' ability to process the spectral cues of the English vowels. This suggests that, as a result of training, plastic changes can occur in the weighting of phonetic cues at early processing stages in the cortex.

All the studies described in this section illustrate the overall positive effects of perceptual training of non-native segmental and suprasegmental contrasts. The most relevant aspects reviewed will be summarized next, namely (1) the relation between perceptual training and both perception and production improvement; (2) generalization; and (3) the retention effects of learning.

The first studies conducted in the 1980s focused primarily on the role of perceptual training in cross-language speech perception. For example, Pisoni et al. (1982), McClaskey et al. (1983), Strange and Dittman (1984), Werker and Tees (1984), Jamieson and Morosan (1986), and Flege (1989) all found evidence of perceptual learning. Subsequent studies showed that perception training can also lead to production improvement without any articulatory training. Several studies, such as Rochet and Chen (1992), Yamada et al. (1996), Bradlow et al. (1997), Wang et al. (1999), Hardison (2003), Wang et al. (2003), and Hazan et al. (2005) have provided empirical evidence of the effects of perceptual training on speech production. For instance, Rochet and Chen (1992) reported a positive correlation between perceptual and productive improvement after training. However, this parallel degree of learning was not found in other studies, which have shown improvement in production after perceptual training, but not to the same extent as in perception. Studies by Bradlow et al. (1997), Wang et al. (2003), and Hazan et al. (2005) have shown this tendency.

Several studies (e.g., Bradlow et al., 1997; Pruitt et al., 1996; Hazan et al., 2005; McClaskey et al., 1983; Rochet & Chen, 1992; Yamada et al., 1996) reported generalization of learning to multiple dimensions. For example, McClaskey et al. (1983), whose training study focused on the voicing contrast between voiceless and voiced labial and velar stops, found generalization of learning to an untrained (viz. alveolar) place of articulation. In the study by Rochet and Chen (1992), in which the word-initial voicing contrast /b/-/p/ was trained, the perceptual improvement of synthesized tokens was transferred to new tokens and to natural stimuli. Yamada et al. (1996) reported high accuracy rates in generalization tests with new (untrained) words after training of the /r/-/l/ contrast with Japanese learners. In a replication of this study, Bradlow et al. (1997) achieved the same results. Pruitt et al. (2006) also found evidence for generalization of learning of the Hindi consonants to novel vocalic and consonantal contexts produced by a different speaker. Generalization was also reported in a study with audiovisual training. Hazan and her colleagues (2005) compared the degree of generalization in two different groups, one that received auditory training and the other that undertook audiovisual training. Although the training modalities differed, the results demonstrated that there was transfer of learning to new stimuli, to the same extent, in both groups. Wang et al. (1999) reported high levels of learning transfer at the suprasegmental level in a study with American English speakers learning Mandarin tones. The generalization tests confirmed that learning was transferred to novel talkers and tokens.

Some of the aforementioned research also showed the influence of stimulus and task variables by examining the use of synthetic vs. natural stimuli and by comparing discrimination to identification tasks. For example, Strange and Dittman (1984) reported that participants' performance in the generalization test was better in tasks with synthesized stimuli than with natural stimuli. Similar findings were reported by Pruitt et al. (2006) that found no transfer of learning with synthesized stimuli to natural tokens. Rochet and Chen (1992), who trained participants on voicing stop contrasts, found generalization to natural tokens, but it was limited to voiceless segments. In terms of perceptual training tasks, Jamieson and Morosan (1986) concluded that discrimination tasks were not the most appropriate to promote generalization. This finding was supported by other researchers, such as Logan et al., (1991) and Logan and Pruitt (1995), who also suggested that discrimination tasks were not the most adequate to use in training listeners to perceive new phonetic categories because they tend to focus listeners' attention on between-category rather than on within-category acoustic differences. Rather, they state that identification tasks (with feedback) seem to be more effective than discrimination tasks in promoting learning and in improving perceptual abilities to map non-native sounds onto L2 phonemic categories. Nevertheless, according to Logan and Pruitt (1995), discrimination training can be very useful, especially in the very beginning of training to show learners that two sounds are different (even though they cannot say which sound is which), or in more advanced stages to train between-category variability. A final consideration about the training procedures is related to the adoption of a high-variability approach, which includes stimulus and talker variability. Since the 1990s, a number of studies (e.g., Bradlow et al., 1997, 1999; Iverson et al., 2005; Lively et al., 1994; Logan et al., 1991; Wang et al.,

1999) have adopted a HVPT method because this procedure proved to promote robust learning of non-native perceptual categories, viz. segments and suprasegments.

Some of the studies reported earlier, such as Bradlow et al., 1999, Lively et al., 1994, Wang et al., 1999, and Yamada et al., 1996, also aimed at investigating whether the knowledge acquired by means of perceptual training would be retained after training. Bradlow et al. (1999), for instance, reported that perceptual learning was retained three months after the training was over, and Lively et al. (1994) and Yamada et al. (1996) reported that the degree of perceptual accuracy was maintained not only in the following three months, but also six months after the conclusion of the training. This long-term retention was found not only at a segmental level but also at a suprasegmental level in the study by Wang et al. (1999). The results of Yamada et al. (1996) also showed that retention of learning acquired through perceptual training may go beyond the domain of speech perception. Three months and six months after the training was over, L2 learners performed significantly better in the production retention tests than the learners in the control group (who did not undertake any training), whose performance was worse than in the pretest.

In sum, the impact and efficacy of perceptual training on perceptual and production learning of non-native phonemic contrasts depends on a number of variables that have to be decided by the researcher(s): (1) what kind of training tasks are more suitable to promote learning (hence, categorization) of a given non-native contrast; (2) whether feedback is provided immediately on a trial-by-trial basis or cumulatively after a training block or session; (3) what type of testing procedure is more appropriate to assess perception and production accuracy; (4) whether the sequencing of stimuli presentation focuses on critical acoustic parameters to be trained; (5) whether stimuli include synthesized or natural tokens; (6) if stimuli are recorded by multiple talkers or not; (7) whether the contrast to be trained is presented in various phonetic contexts and syllable positions or not; (8) which training modality is more suitable to enhance salience of non-native contrasts; and (9) whether the training program is adaptive or not.[73]

Therefore, different perceptual training procedures have been tested regarding perceptual training tasks (discrimination and identification tasks); stimulus presentation (e.g., fading technique); stimulus type (natural and synthesized stimuli; real words and pseudo words); stimulus context (one phonetic context vs. multiple contexts; stimulus in

[73] Some of these variables are described by Bohn (2000, p. 14).

isolation, or embedded in syllable, word, phrase or sentence); input (one talker or multiple talkers); feedback (trial-by-trial or cumulative feedback); duration of training (short-term or long-term training); training design (adaptive training vs. non-adaptive); and type of training (auditory, visual, and audiovisual training). The assessment of perceptual training efficiency has looked at (1) learning effects both in perception and production accuracy; (2) generalization of improvement to new contexts, new talkers, and new segments; and (3) long-term effects. It has been evaluated mainly through indirect behavioral methods as those applied in this study but also by direct observing methods, such as brain images.

The studies reviewed earlier have provided empirical evidence that successful training of non-native contrasts can improve L2 learners' perception and production performance. Next, the focus will be on perceptual studies that trained vowels specifically.

Some perceptual training studies were summarized to provide background knowledge of the main findings and to highlight the diverse training methods that have been applied in research over the last few years. However, since the focus of this study is on the learning of non-native vowels, a chronological presentation of several studies that were conducted to assess the effectiveness of phonetic training on L2 vowel perception and production are summarized next so as to show the various methods that have been used and the main findings. For this purpose, the review of these studies is slightly more detailed. Moreover, all of them focus exclusively on English vowel training with native speakers from different L1s, including Mandarin, Cantonese, Japanese, German, Greek, Spanish (Chilean-Spanish monolinguals and Catalan/Spanish bilinguals), and Brazilian Portuguese.

The study by Wang (2008), and Wang and Munro (2004) tested the effects of computer-assisted perceptual training on the learning of three English vowel contrasts (/i/-/ɪ/, /ɛ/-/æ/, and /u/-/ʊ/) by native speakers of Mandarin and Cantonese living in Canada. Sixteen participants were trained during two months with identification tasks with immediate feedback. The training sessions began with synthesized tokens, followed by fading sessions, and the final stages included variable natural tokens produced by multiple native English speakers. Participants had some control over the training sessions insofar as they could take the training tasks at a self-determined pace. Identification tests with synthesized "heed-hid", "who'd-hood" and "had-head" continua and with natural minimal pair words contrasting the three vowel pairs were used to assess perceptual learning. The effect of training on production was assessed through an intelligibility test in which participants' productions, which were obtained with a word-reading task containing the target vowel contrasts, were identified by native English speakers and through acoustic measurements of vowel durations. The participants' identification accuracy improved significantly from pretest to posttest, and perceptual learning was generalized to new talkers. Moreover, perceptual improvement was retained three months after training was over. In contrast, production accuracy did not improve significantly after training, and this finding was supported by both assessment methods, namely by the identification test rates and the vowel duration measurements. Wang (2008) proposes that perceptual training might not be sufficient for significant improvement in production and suggests combining auditory training modalities with articulatory training tasks to promote robust learning of L2 vowels. Wang and Munro (2004) concluded that computer-assisted perceptual training was an efficient means to deal with learners' individual differences, especially in terms of amount of training time needed. Nevertheless, this was a limitation of the study because, since the individual amounts of training time were not controlled, researchers could not claim that the participants who got the best results were the ones who trained longer.

Lambacher, Martens, Kakehi, Marasinghe, and Molholt (2005) investigated the effects of a high variability identification perceptual training on the perception and production of the mid and low American English vowels /æ/, /ɑ/, /ʌ/, /ɔ/ and /ɝ/. The study consisted of two experiments. Thirty-four Japanese speakers participated in a perception experiment and a subset of 20 participants also took part in a production experiment. The perception pre- and posttests consisted of a five-alternative forced-choice (AFC) task[74] with 150 CVC tokens. Production data were collected with a reading task of 20 CVC minimal pairs containing the five vowels and were evaluated aurally by a group of native American English listeners and acoustically by the researchers. The training consisted of the identification of 75 CVC tokens (15 phonetic contexts for each of the five vowels), and immediate feedback was provided. The six 20-minute sessions of the training program were administered on a weekly basis over a period of six weeks. The overall results revealed that both perception and production

[74] Forced-choice identification tasks are labeling tasks in which the listener has to choose a label from a given number of choices. For example, in a two-alternative forced-choice task, there are only two response options.

improved considerably after training, without any explicit instruction on vowel articulation.

Nobre-Oliveira (2007, 2008) studied the effect of perceptual training on the learning of English vowels /i/-/ɪ/, /ɛ/-/æ/, and /u/-/ʊ/ by Brazilian Portuguese speakers, and compared training with synthesized speech stimuli with cue enhancement[75] (SynS) to training with natural tokens (NatS). The experimental group of 29 undergraduate students of English was divided into two sub-groups, according to the type of training undertaken, and both groups were tested three times, viz. before (pretest), immediately after (posttest), and one month after (retention test) the pronunciation training with an AFC identification task with natural stimuli. Moreover, acoustic measurements of F1 and F2 for each vowel were first collected by means of a reading task containing 116 monosyllabic English words, and then analyzed to verify whether there were also effects of training on production. After a three-week training period, which consisted of combined identification and discrimination tasks with immediate feedback, learners' perceptual competence improved in the identification of the three vowel contrasts, and production of the high front vowel contrast (/i/-/ɪ/) also showed significant improvement. Although there was no significant difference between the two groups of trainees, results indicate that the performance of the group trained with synthesized stimuli (SynS) was slightly better in the posttest than the performance of the other group (NatS), which, according to the researcher, suggests that training with enhanced stimuli may be more effective than with natural stimuli. Moreover, perceptual learning of the three vowel pairs was transferred to new talkers for the NatS group and to new tokens for the SynS group. Finally, the researcher also reported that perceptual improvement was maintained one month after the training was over, as well as the productive ability to distinguish the high front vowels.

The role of auditory training of English vowels on L2 learning by speakers with different L1 vowel inventories was investigated by Iverson and Evans (2007a), who compared how native Spanish speakers (with a five-vowel inventory) and German speakers (with an 18-vowel system) learn English vowels (viz. /aɪ/, /eɪ /, /ɛ/, /i/, /ɪ /, /ɑ/, /ɜ/, /aʊ/, /əʊ/, /ɒ/, /ɔ/, /u/, /æ/, and /ʌ/) by means of a high-variability perceptual training. Twenty-six participants (13 native Spanish and 13 native German speakers) undertook five 45-minute training sessions, which comprised 225 trials of identification with

[75] The synthesized stimuli consisted of computer-generated utterances with enhanced spectral cues and no variation in duration, whereas the natural stimuli were recorded by native speakers of American English (Nobre-Oliveira, 2007).

immediate feedback. The training stimuli were recorded by five native talkers of Standard Southern British English (SSBE), but none of these were the same as the talkers of the test stimuli. Before and after training, participants were tested in their identification of 56 isolated /bVt/ words with a closed-set identification task (with all 14 vowels as response options). Results demonstrated that Germans improved perception accuracy twice as much (20%) as Spanish speakers (10%) following the high variability identification training with feedback. The results suggest that the dense L1 vowel space of German speakers facilitates (i.e., makes L2 learners more sensitive to gradient categorical differences between vowels) rather than interferes with new learning.

In a follow-up experiment, Iverson and Evans (2009) revealed that Spanish speakers were able to improve as much as the German group after undertaking an additional ten-session training program, consisting of identification tasks with stimuli recorded by five English talkers, and that both groups retained perceptual learning. The findings suggest that a larger vowel category inventory may facilitate new learning, and support the hypothesis that high-variability phonetic training improves identification performance of non-native phonemes.

In order to evaluate whether L2 learning by experienced L1 French speakers (who lived, on average for 18 months, in an English-speaking country) differs from learning by inexperienced L1 French learners (who learned English at school), Iverson, Pinet, and Evans (2012) used a high-variability auditory training procedure, and a battery of perception and production tests to evaluate their performance. The eight 45-minute sessions of high-variability identification training with 225 trials each included the same stimuli as Iverson and Evans (2009), and the pre- and posttest stimuli comprised recordings of English /bVt/ words from ten speakers of British English. Perceptual performance was assessed by means of a vowel closed-set identification test and a category discrimination test (i.e., an oddity task), and pre- and posttest production data, collected with a word reading task, were given to a group of four British English listeners for identification judgments. The experiments revealed that both groups of experienced and inexperienced English speakers equally benefited from training, which suggests that laboratory training promotes a type of learning that is different from that obtained in more naturalistic settings. The researchers concluded that auditory training improves the efficiency of categorization, and provides additional learning to L2 naturalistic experience.

Lengeris (2008) studied the effectiveness of auditory phonetic training on native Greek speakers' perception and production of Southern British English vowels (viz. /aɪ/, /eɪ /, /ɛ/, /i/, /ɪ /, /ɑ/, /ɜ/, /aʊ/, /əʊ/, /ɒ/, /ɔ/, /u/, /æ/, and /ʌ/). The training procedures and stimuli were the same as in Iverson and Evans (2007a). The trainees identified and produced English /bVt/ words before and after five sessions of auditory training with feedback. All trainees improved in their overall identification of English vowels, and their post-training productions were more accurately identified by native English speakers than their pre-training productions. In sum, findings seem to indicate that speakers with a small L1 vowel inventory (viz. five vowels) can improve perception and production of non-native vowels within a larger vowel system, by means of intensive laboratory training.

Lengeris and Hazan (2010) further examined whether success in learning L2 vowels is related to learners' L1 vowel processing ability or their frequency discrimination acuity. A group of 18 Greek learners of English received five sessions of HVPT, and pre- and posttests assessed different aspects of participants' L2 and L1 vowel processing and frequency acuity. The training software, stimuli and procedures were the same as Iverson and Evans (2009), but assessment methods differed. L2 and L1 vowel processing were assessed via: (1) natural English vowel identification in quiet and in multi-talker babble (noise), and natural Greek vowel identification in babble; (2) categorization of synthetic English and Greek vowel continua; and (3) discrimination of the same continua. Frequency discrimination acuity was assessed for a non-speech continuum. English vowel production was judged by two Southern British English speakers in a forced-choice identification task, and measured acoustically. The results replicated the finding that HVPT significantly improved the identification of L2 vowels, and that learning transfers to vowel production as judged by native English speakers and confirmed by acoustic analyses of the English vowels. This study, however, provided new information concerning the effect of phonetic training in speech-in-noise L2 perception. Results revealed, as expected, that perception was significantly poorer in noise than in quiet. Nonetheless, training in the quiet condition significantly improved perception in noise. Frequency discrimination acuity was related to measures of both L1 and L2 vowel processing, a finding that supports an auditory processing hypothesis over the L1 phonetic hypothesis (speech-specific explanation) on individual variability in L2 vowel learning. Participants with better frequency discrimination acuity at pretest were also the most accurate in both English vowel perception and production after training.

According to the researchers, it seems that while L1 experience affects L2 processing, some individuals are better at using acoustic information to overcome L1 perceptual biases.

Aliaga-Garcia and Mora (2007) investigated the effects of six one-hour phonetic training sessions on the perception and production of two English vowel contrasts (/i/-/ɪ/ and /æ/-/ʌ/) by 36 advanced Catalan/Spanish bilingual learners of English in a formal instructional context. The method used was HVPT, which included practice on perception and production tasks[76] followed by feedback. Learners' accuracy in vowel perception was tested by means of an AX discrimination task, and production was assessed by measuring formant frequency and length of the target vowels in both the pretest and posttest phases, which were collected by means of an imitation task using a delayed repetition technique.[77] Results showed that phonetic training significantly affected L2 learners' ability to discriminate the vowel contrasts, but no significant differences were found in vowel production.

In a follow-up study, summarized in *Chapter 1*, Aliaga-Garcia (2010) found that both identification (ID) and articulatory (ART) audiovisual high-variability training were equally effective in promoting a significant improvement in the identification of English natural and synthetized (i.e., with modified duration) vowels, and in generalizing learning to untrained tokens and novel talkers. In 2013, the researcher presented further results regarding production. Production of eleven RP vowels was pre- and post-tested with a delayed repetition task, including trained and untrained tokens in auditory and audiovisual conditions, and pre- and posttest acoustic measures of height, frontness, duration, spectral distance and overall vowel dispersion scores were used to assess vowel production. The results revealed significant training effects on vowel production, namely on vowel height and frontness, and higher spectral distances scores for English front and low vowels.

The study by Wong (2012) examined the effectiveness of two perceptual training methods, specifically low (LVPT) and high variability phonetic training (HVPT), on the perception and production of the English vowel contrast /e/-/æ/ by Cantonese ESL learners. Both training procedures were based on two-AFC

[76] The training sessions included identification and discrimination tasks, phonetic transcription, articulatory (visual) description of sounds, imitation practice, reading aloud, and tongue-twisters. Feedback was given on a trial-by-trial basis and weekly during the training (Aliaga-García & Mora, 2007, p. 12).

[77] A delayed repetition technique consists of an elicitation procedure, in which tokens are elicited using orthographic and aural cues. Each token is produced after the participant hears the token spoken by a native speaker. This procedure ensures that orthography (i.e., difficulty in reading) does not interfere in a participant's performance (Flege et al., 1995).

identification tasks with immediate feedback, but the HVPT included stimuli produced by seven native speakers of four different English varieties (viz. General American accent, Northern American accent, Northern British accent, and Southern East British accent), whereas the tokens of the LVPT were produced by only one native talker. The training program consisted of ten sessions, spread over three weeks, and learning was tested with a word-list-reading task (production) and an identification task (perception). Generalization was tested with a sentence-reading task and an identification test with untrained words. After perceptual training, the two trained groups (22 participants in the LVPT and 19 in the HVPT) revealed significant perceptual improvement in the identification of the two vowels after training and generalization of learning to new words and new speakers, with the HVPT group demonstrating more significant improvement cases than the LVPT group of trainees. Robust transfer of perceptual learning of vowel /e/ to production was observed in both groups, with the HVPT group outperforming the LVPT group. However, generalization to production at sentence-level was not observed. Wong (2013) further examined the proficiency level of participants in the HVPT group to verify whether language proficiency influenced perceptual and productive performance. After training, both low and high proficiency groups improved significantly in the identification of the vowel contrast, thus, proficiency was not a significant factor. Moreover, perceptual learning was generalized to new words, regardless of proficiency level. In conclusion, proficiency did not affect the learning of the two English vowels, suggesting that the HVPT procedure is beneficial to learners with both low and high proficiency levels.

Nishi and Kewley-Port (2007) investigated the influence of training set sizes by training native Japanese listeners to identify American English (AE) vowels. Two groups of Japanese learners of English were either trained on nine AE vowels (/i/, /ɪ/, /ɛ/, /æ/, /ɑ/, /ʌ/, /ɔ/, /ʊ/, and /u/) (fullset training group) or on three (difficult) vowels (/ɑ/, / ʌ /, and /ʊ/) (subset training group). Performance of listeners was assessed with identification tests before and after training, as well as three months after the training was over. The training consisted of nine 90-minute sessions, which included identification tasks with interactive feedback. Results indicated that both training groups significantly improved the perception of AE trained vowels and maintained improvement three months after the completion of training. In addition, both groups generalized improvement to untrained words and tokens spoken by novel talkers. However, the subset group did not transfer learning to untrained vowels, which suggests that training programs for learning non-native vowels should present a full set of vowels instead of focusing only on a subset of difficult segments. The researchers (2008) extended the target participants to Koreans and examined whether training programs combining the two stimulus sets provided more effective training. Three groups of five Korean speakers were trained on American English vowels over a period of nine days, using one of three procedures: (a) fullset only; (b) first three days on subset and then six days on fullset; (c) six days on fullset and then three days on subset. Results revealed that fullset training was effective but no advantage was found for the two combined procedures over the fullset.

Pereira and Hazan (2013) conducted a research study that investigated the impact of three different modalities of high variability vowel training on L2 learners' perception and production. Forty-seven Chilean-Spanish speakers (divided into three groups), beginner learners of English, were trained on eleven British English vowels. Five HVPT sessions of auditory (A), audio-visual (AV) or video-only (V) training were attended by each group of learners. Pre- and posttests were used to evaluate perception and production of the eleven English monophthongs. Significant improvement in vowel identification was found in the word-level perceptual test, but was not transferred to a sentence-level context, regardless of the training mode. Production competence, which was measured by means of native English speakers (NES) ratings and acoustic analysis, revealed that there was improvement and change in spectral measures after training. Findings also indicated a lack of sensitivity to visual cues for English vowels, since participants did not seem to benefit from the visual information available prior to training nor after it.

Lacabex, Lecumberri and Cooke (2008a, 2008b, 2009) analyzed auditory (i.e., perceptual) versus articulatory (i.e., production) training effects in the identification of the English full vowel-schwa contrast in foreign language (FL) formal instruction contexts by 34 Spanish learners of English, as described in the previous chapter. Spanish learners' perceptual performance was assessed in isolated words to verify the effect of training, and in sentences to investigate the robustness of learning in generalizing to an untrained context. Both trained groups improved their performance on the perception of the reduced vowel after training and were able to generalize learning to an untrained context (viz. sentences). A follow-up experiment by Lacabex and Lecumberri (2010) investigated the production of English weak forms before and after perceptual and production training. The productions of 34 Spanish EFL learners were tested with a reading-aloud task and an imitation task. Results showed a significant improvement in the production of weak forms after training in both production tests, though posttest production was only judged correct 50% of the time. This latter result seems to indicate that training should be extended. Moreover, these findings seem to support that training in one speech modality (perception or production) facilitates the other modality, pointing to a facilitating relationship between perception and production training. Lacabex and Lecumberri (2012) further observed overgeneralization effects in the posttest, namely in the production of the weak vowel in words with a strong vowel in unstressed syllable.

Hirata and Kelly (2010) investigated the effects of lips and hand gestures on auditory learning of Japanese phonemic vowel length contrasts by native speakers of English. Sixty native English speakers participated in one of four types of training: audio-only; audio-mouth; audio-hands; and audio-mouth-hands. Participants were pre- and post-tested with perception tasks that assessed their ability to identify short and long vowels in Japanese. Although the four groups improved from pre- to posttest, the participants in the audio-mouth condition improved more than those in the audio-only condition, whereas the two conditions involving hand gestures did not. The researchers concluded that seeing lip movements during training significantly helped learners perceive difficult phonemic contrasts, but seeing hand gestures did not.

The use of acoustic spectrograms and waveforms in vowel pronunciation training has recently been investigated. For example, Okuno (2013) examined the benefits of waveform displays as visual cues on the acquisition of L2 vowel duration in Japanese by native English speakers. The researcher compared identification and production accuracy among three groups of 64 participants: (1) a control group (with no training); (2) an audio-visual (AV) group; and (3) a visual (V) group. Findings revealed that segmental identification improved for both AV and V groups, but improvement was greater for the group trained with visual waveforms. Quintana-Lara (2012) investigated the effects of acoustic spectrographic instruction on the production of the English high front contrast. This type of instruction is based on the assumption that physical representations of speech sounds (spectrograms) help learners see and modify pronunciation features. Twenty-six EFL teachers participated, and during two weeks, the 16 participants in the experimental group received acoustic spectrographic instruction. Production accuracy was tested by two production tasks and a perceptual identification task. Acoustic measurements and data from the perception test indicated that pronunciation of both high front vowels improved.

Overall, perceptual training studies on English vowels showed significant improvement in perceptual performance immediately after training. Generalization of perceptual learning to new talkers (e.g., Aliaga-Garcia, 2010; Nobre-Oliveira, 2007; Wang, 2008), to new tokens (e.g., Aliaga-Garcia, 2010; Lacabex et al., 2009; Nobre-Oliveira, 2007) and to new contexts (i.e., to embedding sentence) (e.g., Lacabex et al., 2007, 2008a, 2008b) was reported, and long-term retention was also observed one month (Nobre-Oliveira, 2007) and three months (e.g., Nishi-Kewley, 2007; Wang, 2008) after completion of training. Several studies found significant training effects on production (e.g., Aliaga-Garcia, 2013; Lacabex and Lecumberri, 2010; Lambacher, 2005; Lengeris, 2008; Pereira & Hazan, 2013), but a few also reported no improvement in production accuracy, such as Aliaga-Garcia and Mora (2007), and Wang (2008). The procedures to assess perceptual performance do not differ much amongst studies, varying from discrimination (e.g. Aliaga-Garcia & Mora, 2007) to identification tests. However, in terms of production assessment, two methods have been used either in combination or individually. For example, Aliaga-Garcia (2010) and Nobre-Oliveira (2007) measured vowel production acoustically, whereas Lengeris (2008) and Iverson et al. (2012) used native speakers' (NS) identification rating judgments. Lambacher (2005), Wang (2008), and Pereira-Hazan (2013), for instance, combined both acoustic analysis and NSs identification ratings to assess L2 learners' production accuracy.

Several experiments were conducted to investigate which training procedures are more effective in promoting L2 categorical learning. For example, Nobre-Oliveira (2007) compared training with natural stimuli to synthesized stimuli and found that, though there was a tendency for the group trained with synthesized tokens to have better results, there was no significant difference between them. Wang (2008) also included synthesized and natural tokens in the training, but did not compare stimulus type. A comparison between auditory (perceptual) and articulatory (production) training was also made, for example by Lacabex et al. (2009) and Aliaga-Garcia (2010), who concluded that both methods are effective. Preliminary results of Pereira and Hazan's (2013) comparative study between visual (V), audio (A), and audiovisual (AV) training modalities seem to indicate that adding visual cues to training stimuli (i.e., audiovisual stimuli) does not facilitate the learning of English vowels. The influence of other variables has also been examined, namely L2 experience or language proficiency

(Iverson et al., 2012, and Wong, 2013, respectively), L1 and L2 vowel inventory size (Iverson & Evans, 2009; Lengeris, 2008), degree of stimuli variability (low vs. high, as in Wong, 2012), and number of vowels included in the training (fullset vs. subset, as in Nishi & Kewley-Port, 2007, 2008). Regarding L2 experience and proficiency, no significant differences were found, which seems to indicate that laboratory training equally promotes learning in groups with diverse L2 backgrounds.

Two very important findings are the following: (1) although large L1 vowel inventories, as German, seem to facilitate learning (e.g., Iverson & Evans, 2009), learners, whose L1s have a small vowel space, such as Greek, can successfully learn non-native vowels within larger inventories (e.g., Lengeris, 2008); and (2) a high-variability training method (with multiple talkers and stimuli) is very effective. Moreover, Lengeris and Hazan (2010) found that though L1 phonetic experience affects L2 speech processing, individual differences may be related to a better ability to process acoustic information in order to overcome L1 perceptual biases.

The review of training studies is fundamental to understand the method of the present study, which is described in the next chapter. Several decisions about the aforementioned training variables were made based on the findings of these studies.

CHAPTER 3

METHOD

The true method of knowledge is experiment.
- William Blake

This chapter describes the design of the study by providing detailed information about the participants, the testing and training materials, the procedures followed to administer both the tests and the training program, and the analyses carried out to assess production and perception of the target American English vowels in the three testing moments - pretest, posttest and delayed posttest - of the experiment.

3.1 Participants

In this first section, the biographical characteristics of the two groups of participants will be described. First, a detailed account of the adult learners' of English (L2) background will be provided, and then a brief summary about the L1 informants will be presented. Furthermore, the eligibility and exclusion criteria underlying the selection of the participants will be reported.

3.1.1 L2 participants

The L2 participants of this study were 34 undergraduate students of the European Languages and Literatures (ELL) degree course (English Major and English Monolingual) enrolled in *English Linguistics 1 – Phonetics and Phonology* in the second semester of the first year.

The Portuguese students agreed to participate in the research study voluntarily[78] and signed a consent form written in their native language (Appendix D). The selection of the participants was based on the data collected by means of a background questionnaire (Appendix E), whose relevant results are described in the next section.

[78] Due to the fact that the study was conducted within the *English Phonetics and Phonology* course, 49 students volunteered to participate and, thus, filled in the background questionnaire. However, 15 students were excluded from the experiment, because they did not meet the following criteria: (1) be native speakers of European Portuguese (L1) who were first exposed to English (L2) around the age of 8 years old; (2) had lived in an English-speaking country for no more than three months, and (3) have no visual, hearing, or speech-related impairments.

Participants were informed about the procedures, the potential benefits, and the main purposes of the research. However, detailed information about the specific aims and focus of the experiment was not disclosed so as not to affect participants' performance, especially in the pretest phase. In order to avoid participant attrition,[79] students were given an extra mark[80] for taking part in the four phases of the experiment: (1) pretest, (2) training, (3) posttest, and (4) delayed posttest. This strategy aimed at keeping students motivated and committed to the research experiment.

This group of Portuguese learners of English had no previous explicit phonetic training, and had been exposed to English mainly through formal classroom instruction in a Foreign Language (FL) context.[81] All reported having no hearing or speech-related impairments.

The cohort was divided into an experimental group (*n*=22) and a control group (*n*=12), according to the following criteria: (1) their scores in the identification pretest, so that the American English vowel identification accuracy rates would be matched between both groups to facilitate the interpretation of results, and (2) their timetable.[82] In these matched groups,[83] the range of identification accuracy scores was 53%-80% (mean 66%) for the experimental group and 52%-86% (mean 69%) for the control group (results are displayed in Table 7). The difference in the number of participants in the two groups is due to two reasons, namely their timetable constraints, and the potential benefits of the perceptual training on the learning of the target vowel contrasts.

---

[79] In a longitudinal study with a pretest, posttest and delayed posttest design, participants may drop out before they complete the experiment, due to many reasons, such as being no longer available or willing to take part. This is known as "participant attrition" or "mortality" (Mackey & Grass, 2005; Monette, Sullivan & DeJong, 2011).

[80] In order to give all volunteers the opportunity to be granted an extra mark and to attend the training sessions, the 15 students who were excluded (see note 78) took the same tests and training tasks as the group of participants, and were not initially told that their data were not considered for analysis. They were informed about the selection only after the research was completed, i.e., at the end of the semester so as not to influence their performance and commitment during the experiment, as well as their classmates'.

[81] In this learning context, input tends to be limited and very likely foreign-accented. In Portuguese state schools, the majority of teachers of EFL (English as Foreign Language) are native speakers (L1) of European Portuguese, the only exception being in private language schools, whose teachers and tutors are mainly native speakers of the foreign language.

[82] Although participants were assigned to an experimental group and a control group, they were not told this information. They were explained that the division into two groups was related to their timetables (the two classes enrolled in the course – English Monolingual and English Major – had slightly different timetables). They were also told that, although the auditory training sessions would differ in terms of phonetic segments to be focused on, at the end of the semester, the two groups would have access to all the training tasks.

[83] The researcher followed the same procedure as Iverson and Evans (2009, p. 868) to match the two groups of informants.

Table 7

Individual Identification Test Scores in the Pretest

Experimental Group		Control Group	
Participant	**Pretest (%)**	**Participant**	**Pretest (%)**
E1	72.38	C23	85.71
E2	58.57	C24	60.95
E3	75.24	C25	60.95
E4	66.67	C26	66.19
E5	62.38	C27	65.23
E6	70.0	C28	73.8
E7	69.52	C29	80.47
E8	55.71	C30	75.23
E9	52.86	C31	74.28
E10	65.23	C32	60.47
E11	55.23	C33	52.38
E12	60.47	C34	66.66
E13	75.23		
E14	57.61		
E15	53.33		
E16	74.29		
E17	63.81		
E18	65.71		
E19	70.48		
E20	69.04		
E21	69.04		
E22	80.47		
Mean (M)	**65.6** (*SD* = 7.81)		**68.5** (*SD* = 9.55)

Note. E=experimental; C=control.

The experimental group consisted of 12 male (54.5%) and 10 female participants (45.5%), whose ages ranged from 18 to 42 years (mean=22.32 years, standard deviation (SD)=6.69 years); and the control group included four men (33.3%) and eight women (66.7%) between 18 and 40 years old (mean=24.33 years, SD= 6.99 years). The differences in age were not expected to influence the participants' performance during the experiment, since no association was found between age and the identification pretest results. For more detailed information, see Table F1.

The two groups suffered attrition from the posttest to the delayed posttest. There were three dropouts[84] (13.34%) in the experimental group, and one dropout (8.34%) in the control group, as shown in Table 8.

[84] Although we tried to avoid this situation by contacting these students personally, they did not want to come back two months after the training was over, i.e., at the end of the semester, due to lack of interest in both their commitment to the subject and the course

Table 8

Number of Participants in Each Phase of the Experiment

	EG	CG	Total
Pretest	*n*=22	*n*=12	*n*= 34
Training _Vowels	*n*=22	-	*n*= 22
Training _Consonants	-	*n*=12	*n*= 12
Posttest	*n*=22	*n*=12	*n*= 34
Generalization Test	*n*= 22	*n*= 12	*n*= 34
Delayed Posttest	*n*=19	*n*=11	*n*= 30

Note. EG=experimental group; CG=control group.

Since the group of selected participants was homogenous, the following description refers to the whole cohort. Individual answers regarding participants' language experience are shown in Table F2.

The level of language proficiency of each individual was assessed according to the CEF,[85] and based on the required levels of English proficiency in each semester of the undergraduate course. Thirty first-year students (88.2%) had a B1 level at the end of the first semester,[86] two second-year students had a B2 level of English (5.9%), and two other students[87] (5.9%) had a C1 level of English. Apart from a third-year student, who was not attending any English Language class, 33 students reported having English Language lessons[88] four hours per week, and studying English, on average, 2.75 hours (*SD*= 2.28 h) weekly.

The mean age of learning (AOL), that is, the age at which participants started learning English, was 9.71 years (range= 6-11 years, SD=.94), which means that the majority of students started their English formal instruction at middle school, in the 5th grade,[89] and learned English formally for eight years (mean=8.18 years, range=5-12, SD=1.70).

evaluation. They failed the continuous assessment of the course, and since the extra mark would not make any difference in their final evaluation, they dropped out.

[85] The *Common European Framework of Reference for Languages* (CEF) was created by the Council of Europe, in 2008, to standardize foreign language proficiency levels. There are six reference levels: A1, A2 (Basic User); B1, B2 (Independent User); C1, and C2 (Proficient User). In terms of phonological competence, the CEF descriptors for each level are presented in Appendix A. /http://www.coe.int/t/dg4/linguistic/Source/Framework_EN.pdf/.

[86] Although the assessment of students' language proficiency in the English Language classes comprises different competences, such as comprehension, speaking, and writing, which are graded from 1 to 20 marks, they all achieved the B1 level in the first semester. At the time of the study (the second semester) they were attending B1+ level classes.

[87] One was a third-year student, and the other was a psychologist, who was majoring English Language and Literature.

[88] It is important to emphasize that the language of teaching in this undergraduate course is the target foreign language. Therefore, the English Literature, English Culture, and English Linguistics classes are taught in English by native speakers of English and bilingual speakers (2L1 and L2). In some other Portuguese state universities, these courses are taught in Portuguese.

[89] Mandatory English formal instruction is part of the National Learning Curriculum, and it starts at middle school in the 5th grade.

In the background questionnaire, apart from filling in biographical data, participants had to answer several questions about their use of English. To facilitate further analyses, we consider the frequency of English use from two perspectives: active and passive. The active use of English includes mainly oral production, that is, speaking the L2 to communicate with Portuguese speakers, native English speakers and other non-native speakers, and the passive use of English consists of watching movies, and TV programs, listening to music, playing video games, and reading. With reference to the active use, 14 students (41. 2%) reported speaking English frequently, that is, at least once a week, and the other students (58.8%) stated not using English to communicate on a weekly basis. Regarding the passive use, 28 students (80.9%) reported contacting with English daily. The majority of students specified contacting with English one hour per day, on average.

To sum up, participants tended to use English more passively than actively, which seems to indicate that, although having a certain amount of daily input through the mass media, the frequency with which these individuals maintained oral interactions in English was low, mainly because the opportunities to speak the L2 were scarce.

Other relevant information about the participants' experience with English includes the learning of the L2 in other formal contexts, such as private language schools, interruption of formal instruction of English, experiences abroad (more than one month but less than three months in English-speaking countries), and their awareness of the variety of English they speak. As far as the learning of English in other contexts is concerned, only five students (14.7%) stated having studied English in private schools. Twenty-two (64.7%) participants interrupted the formal learning of English at some point in their academic life,[90] before starting their undergraduate course. Only one student reported having an experience abroad. This informant lived in England for three months at the age of 20 and spoke both Portuguese and English on a daily basis. When asked about the variety of English they use, the majority of students (76.5%) claimed being more familiar with American English, four students (11.8%) believed they speak the British English variety, three (8.8 %) explained they speak a hybrid variety of American and British English, and a student (2.9%) could not identify any variety. The choice to focus on American English phonetic segments in this study is closely related to the above results. The majority of students are more familiarized with the American variety of English because the (outside school) input they are mostly

[90] This can be explained by the fact that many state high schools do not offer the English language subject on the 12th grade.

exposed to is American English.[91] A previous study (Rato et al., 2013) conducted with a group of participants from the same university and with identical backgrounds also contributed to this choice. A summary of the aforementioned statistical results of the background questionnaire is provided in Appendix G.

3.1.2 L1 participants

Seven native speakers of American English also participated as a baseline group. They validated the perception testing and training materials by performing the perception tasks that were used in the experiment,[92] thus providing baseline data. This group included five women (71.4%) and two men (28.6%), whose ages ranged from 26 to 55 (mean=39.71 years, SD=11.28). See Table 9 for a summary of biodata about this group of AE NSs.

Table 9

Background Information of the American English NSs

Part	Sex	Age	Place of Birth	Current residence	Occupation
1	F	33	Fairfield, CT	Fredonia, NY	Assistant Professor of TESOL
2	F	43	Detroit, MI	Ermesinde, PT	Theology teacher
3	M	32	Artesia, CA	New York, NY	Finance analyst
4	F	26	Ean Claire, WI	Porto, PT	Translator and proofreader
5	F	35	Minneapolis, MN	Columbus, OH	PhD student
6	M	54	Berwick, PA	Maia, PT	Evangelical missionary
7	F	55	Niagara Falls, NY	Maia, PT	Evangelical missionary

Note. CT=Connecticut; MI=Michigan; CA=California; WI=Wisconsin; MN=Minnesota; PA=Pennsylvania; NY=New York; OH=Ohio; PT= Portugal; TESOL=Teachers of English to Speakers of Other Languages.

Since the aim of including this baseline group was to validate the experiments and check if all the stimuli were accurately identified and discriminated, and since there was a battery of thirteen perception tasks, repetitions of tokens were not added so as to reduce the amount of stimuli and time. Moreover, native listeners could hear each stimulus only once before choosing an answer. These American English NSs

[91] In formal FL teaching contexts, as explained previously, teachers and tutors have different native language backgrounds, and learning materials, such as listening activities (audio and audiovisual), include recordings of NSs of different varieties of English.
[92] The training tasks that were prepared for the control group and used as “distractor” exercises were not administered to the L1 participants.

participated voluntarily and were not paid. Four of them took the perception tests individually in a quiet room of the University of Minho using a laptop computer and Sennheiser HD202 headphones, and three of them performed the tests at home. These were sent by email along with detailed instructions on how to install and run the computer program with the perception tasks, and how to perform them. The data obtained from the L1 group were analyzed to check if there were tokens which needed to be replaced or deleted. After careful observation, we did not find any stimulus that was systematically misidentified and, thus, no task or test was changed.

3.2 Materials

This section describes the materials used to conduct the study. It is subdivided into two main sections. In the first section, the materials used to elicit the production data are described, and, in the second section, the tools to collect the perception data are presented. In the latter section, the first part is dedicated to the description of the testing materials and the second part to the training tasks.

3.2.1 Questionnaires

This study included two questionnaires written in Portuguese:[93] a background questionnaire and a follow-up questionnaire on the learners' assessment of the training program.The background questionnaire (Appendix E) consisted of 18 closed and open questions[94] and aimed at collecting background information that could influence the participants' performance (Dörnyei & Csizér, 2012). These comprised not only questions about biographic information, such as age, place of birth, and residence, but also questions about their language learning history, such as age at which learning of English began (AOL), number of years, that is, length of L2 formal instruction (LFI), and awareness of different varieties of English. This questionnaire was administered to L2 participants individually, before the production pretest. The researcher was present to clarify any questions that could come up.

[93] The questionnaires were written in Portuguese, because it is a general belief, with which we agree, that the quality of the obtained data improves if presented in the participants' mother tongue (Mackey & Gass, 2005, p. 96).

[94] Mackey and Gass (2005) explain that "a closed-item question is one for which the researcher determines the possible answers, whereas an open-ended question allows respondents to answer in the manner they see fit" (p. 93). The questionnaires included more closed-ended than open-ended questions because they offer a greater uniformity of measurement, which facilitates statistical analysis.

The follow-up questionnaire on the learners' assessment of the training sessions included ten closed-ended questions and an open question (Appendix H). The aim of this questionnaire was to collect information about students' opinions and attitudes concerning the learning process, namely the type of perception task which was easier to do, the vowel pair which was more difficult to learn, the level of motivation, concentration and tiredness they felt throughout the sessions, and some general comments. This questionnaire was filled in only by the experimental group, immediately after the training program was completed. The data obtained from this questionnaire will be discussed in the next chapter.

3.2.2 Testing Materials

The production and perception tests included in the present study are described in the following subsections.

3.2.2.1 Production

The L2 participants' performance on the production of the American English vowels was measured three times by means of a production test: (1) before the training (*pretest*); (2) immediately after the five-week training period (*posttest*); and (3) two months after the training program was over (*delayed posttest*). The production of the European Portuguese vowels was collected once, in the pretest phase, by means of a similar production test.

3.2.2.1.1 European Portuguese vowels

The European Portuguese corpus consisted of 42 disyllabic words, listed in Table 10. Each word had one of the following seven oral vowels, /i/, /e/, /ɛ/, /a/, /u/, /o/, and /ɔ/, in stressed position.

Table 10

EP Words Read by the L2 Participants

Picture	Vowel	pVpV	pVtV	pVkV	tVpV	tVtV	tVkV
bico (beak)	/i/	pipo	pito	pico	tipo	tito	tico
dedo (finger)	/e/	pêpo[a]	pêto	pêco	têpo[a]	têto	têco[a]
lego	/ɛ/	pépo[a]	péto	péco	tépo[a]	této	téco
pato (duck)	/a/	papo	pato	paco	tapo	tato	taco
copo (glass)	/ɔ/	pópo[a]	póto[a]	póco[a]	tópo	tóto[a]	tóco
côco (coconut)	/o/	pôpo[a]	pôto	pôco[a]	tôpo	tôto[a]	tôco
cubo (cube)	/u/	pupo	puto	puco	tupo	tuto	tuco

Note. See Table I1 for phonetic transcriptions of EP words.
[a] Pseudo words.

The 30 words and 12 pseudo words[95] were embedded medially in the carrier sentence "Digo (CVCo word) novamente" ("I say (CVCo word) again") (C=consonant, V=vowel). This was preceded by a picture illustrating a word that rhymed with the CVCo word in the sentence so as to facilitate the reading of the words, and elicit the accurate production of the target vowels (see Figure 11).

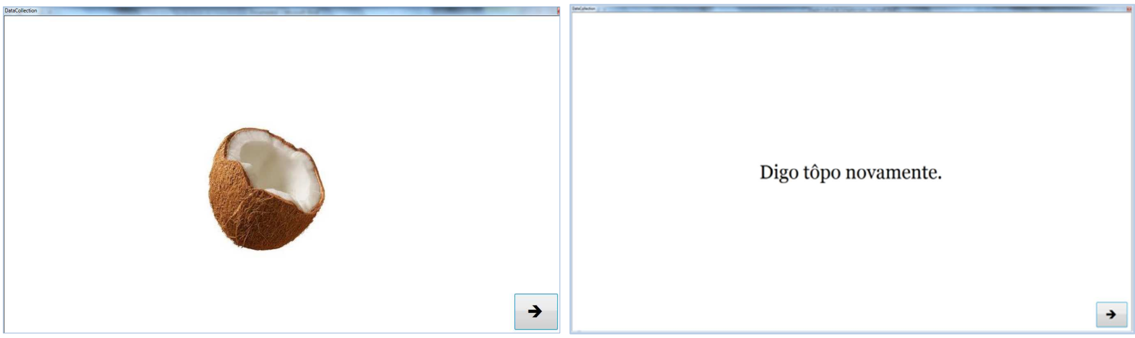


Figure 11. The computer screens of the EP production test. The screen on the left shows the picture of the key word used to elicit the accurate pronunciation of the target word inserted in the carrier sentence, shown on the right.

Furthermore, orthographic accents were added in stressed syllables with vowels <e> and <o>, for example, *têpo, tépo, tôpo, tópo*, to help participants distinguish between /e/ and /ɛ/, and /o/ and /ɔ/, respectively. Moreover, words with orthographic diphthongs, such as <poupo> or <pouco> were not included to avoid diphthongization.

[95] Pseudo words were considered to be words that do not violate phonological constraints and are not indexed in the first edition of the *Vocabulário Ortográfico do Português* (Portuguese Orthographic Vocabulary) (2010). The VOP has 210 000 lexical entries, which include onomastic vocabulary, gentiles, toponyms, and (foreign) loan words. /http://www.portaldalinguaportuguesa.org/vop.html./

As mentioned previously, words had a CVCV syllabic structure, and the seven stressed vowels were inserted in six phonological contexts: /pVpu/, /pVtu/, /pVku/, /tVpu/, /tVtu/, and /tVku/. These vowels were flanked by voiceless plosive and fricative consonants so as to facilitate their segmentation and minimize vowel duration variation. The words illustrated by a picture were not considered for analysis. Each individual read the 42 sentences, listed in Appendix J, twice. However, only one production of each context was selected for analysis,[96] resulting in 42 tokens (7 vowels x 6 contexts). The total number of European Portuguese vowels analyzed acoustically was 1428 (42 tokens x 34 participants).

3.2.2.1.2 American English vowels

The American English corpus consisted of 58 monosyllabic words. The six target AE vowels /i/, /ɪ/, /ɛ/, /æ/, /u/, /ʊ/, and the distractor vowel /ʌ/ were inserted in CVC words with the phonological frames: /pVt/, /pVk/, /tVt/, /tVk/, /kVp/, /kVt/, /bVt/, /dVk/, /fVt/, /sVt/, listed in Table 11.

Table 11

AE Words Read by the L2 Participants

picture	vowel	pVt	pVk	tVt	tVk	kVp	kVt	bvt	dVk	fVt	sVt
feet	**/i/**	Pete	peak	teat	teak	keep	keat	beat		feet	seat
pig	**/ɪ/**	pit	pick	tit	tick	kip	kit	bit		fit	sit
bed	**/ɛ/**	pet	peck	Tet	tech	kept[b]	Ket	bet	deck		set
cat	**/æ/**	pat	pack	tat	tack	cap	cat	bat		fat	sat
boot	**/u/**	poop[b]		toot	tuke[a]	coop[a]	coot	boot (2x)	duke		suit
book	**/ʊ/**	put (2x)			took (2x)		cook[b] (2x)	book[b]		foot	soot
cup	**/ʌ/**	putt	puck	tut	tuck	cup (2x)	cut	but			shut[b]

Note. See Table I2 for the phonetic transcription of AE words. 2x=words repeated twice.
[a]Pseudo words. [b]Words with similar phonological contexts.

However, since it was impossible to find real high frequency words with the seven American English vowels (/i/, /ɪ/, /ɛ/, /æ/, /u/, /ʊ/, and /ʌ/) in all the aforementioned contexts, low frequency words (e.g., *keat*, *kip*, *coot*), pseudo words (e.g., *coop*, *tuke*), and words with similar phonological contexts (e.g., *kept*, *shut*) were

[96] The selection depended on the acoustic quality of the segmented vowels, i.e., vowels with stable formant values.

included so that there were nine words for each vowel (see Appendix L for a list of word frequency). There are very few CVC words with vowel /ʊ/, therefore three words were repeated twice so as to have nine words for this vowel (see Table 11). Although vowel /ʌ/ was integrated as a distractor, we opted to include the same number of words as with the other vowels so that the participants considered it to be part of the group of the target sounds.

In terms of context, the following flanking consonants were voiceless stops so as to avoid duration variability.[97] The preceding consonants included not only voiceless stops but also two voiced stops because these were not expected to influence vowel duration. Moreover, the choice of obstruent consonantal contexts (stops and fricatives) was expected to facilitate segmentation.

Due to the difficulty L2 learners (may) have when reading unfamiliar or nonexistent words, the target words were preceded by a picture depicting a high frequent CVC word that rhymed with the word to be read (see Figure 12). These words were inserted in the carrier sentence "Say (CVC word) now".

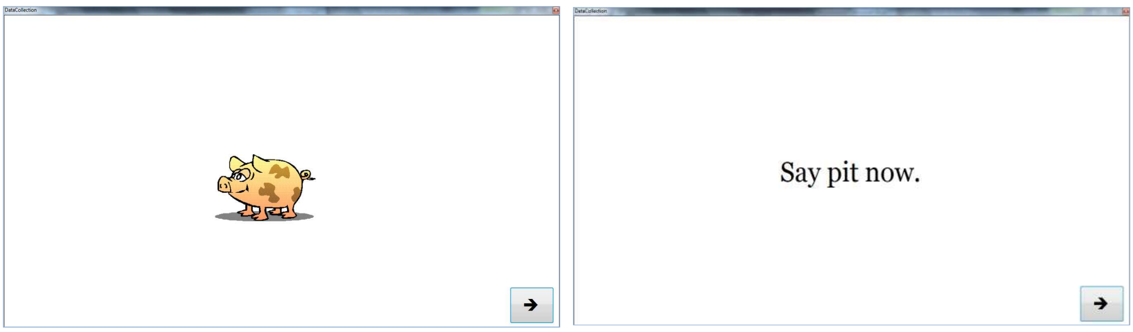


Figure 12. The computer screens of the AE production test. The first screen shows the picture of the key word used to elicit the accurate pronunciation of the target word inserted in the carrier sentence, displayed on the second screen.

The words that were illustrated by pictures were not considered for analysis. Each participant read the 63 sentences three times (Appendix K). However, only one production for each context was considered for analysis (7 vowels x 9 contexts), yielding 63 tokens per participant. The total number of vowels considered for analysis in each test was 2142 (63 tokens x 34 participants). As explained previously, participants took the same production test three times throughout the experiment

[97] Vowel duration in American English is influenced by the nature of the following consonant. Vowels tend to be shorter when followed by voiceless consonants and longer before voiced consonants (Lisker, 1974).

(pretest, posttest and delayed posttest), thus producing 6174 vowels, which were acoustically analyzed.

3.2.2.2 Perception

In this section, a detailed description of the perception identification test and the generalization test used in this study will be provided.

Perceptual performance on the identification of the target American English vowel contrasts was measured three times (*pretest, posttest, and delayed posttest*) in isolated CVC words to study the effect of training, and in monosyllabic words with a more complex syllabic structure to investigate the robustness of learning in generalizing to untrained contexts and talkers (*generalization test*).

3.2.2.2.1 Identification test

The participants' perception of the target American English vowels was tested three times (*pretest*, *posttest*, *delayed posttest*) with a seven alternative forced-choice identification task.[98] The undergraduate students had to identity the AE vowels in naturally spoken words[99] produced by six NSs of American English, so that the participants were always exposed to different voices in each sequence. The AE native talkers consisted of a group of three women and three men, aged between 22 and 50 (mean=28.83 years, SD=11.39 years), from Davenport, Iowa, and Sacramento, California. More background information about these AE talkers is presented in Table 12.

[98] Logan and Pruitt (1995) explain that "in an identification task, a single stimulus is presented, and the listener is required to assign a label to the stimulus" (p. 357), and, therefore, it may also be named "labeling task".

[99] We chose to use natural speech stimuli over synthetic tokens in the categorial discrimination and identification tasks, because (1) they offer a rich diversity of acoustic cues that help forming robust perceptual categories (Pisony & Lively, 1995), and (2) although listener accuracy may differ for the two stimulus types, both share similar patterns of cross-language differences (Beddor & Gottfried, 1995, p. 216). Moreover, studies comparing the results of vowel training programs with natural and synthetic stimuli did not find significant differences in the perceptual performance of groups trained with different types of stimuli (e.g., Nobre-Oliveira, 2007).

Table 12

Biodata of the American English NSs.

Talker	Sex	Age	Place of Birth and Residence	Occupation
T1	M	32	Sacramento, CA	BA student in History
T2	M	18	Sacramento, CA	High school student
T3[a]	M	50	Sacramento, CA	University Professor
T4[a]	F	27	Sacramento, CA	BA student in History
T5[a]	F	24	Davenport, IA	MA student in OccupationalTherapy
T6	F	22	Davenport, IA	MA student in OccupationalTherapy

Note. T=talker; CA=California; IA=Iowa.
[a] Talkers also included in the training tasks.

Data collection of the AE vowels produced by the female NSs from Davenport is described in Rauber, Rato and Silva (2010), and the recording procedures of the NSs from Sacramento are reported in Rauber (2010). The recordings of the talkers used for the stimuli of the perception test and the vowel training tasks are the same as those used in the aforementioned studies, with permission. The naturally spoken words recorded by these NSs were, however, segmented, edited and organized according to the design of each auditory task of this experiment.

The stimuli comprised CVC words with the three AE vowel contrasts /i/-/ɪ/, /ɛ/-/æ/, /u/-/ʊ/, and the distractor vowel /ʌ/, which appeared twice in the following contexts: /pVt/, /tVt/, /tVk/, /kVt/, and /bVt/ (see Table 13). The total number of tokens that participants had to identify was 210 (7 vowels x 5 contexts x 6 talkers). The identification test was preceded by a familiarization test of 28 tokens (7 vowels x 2 contexts x 2 talkers). Overall, in the same testing session, students identified 238 vowels. The test included a three-point Likert scale from 1 (*poor*) to 3 (*good*) so that informants could rate category goodness-of-fit[100] after listening to each token.[101] Stimuli presentation was randomized[102] for each participant, and each of the 210 stimuli was

[100] Goodness ratings indicate how good an exemplar of a given token is, as judged by a listener. For instance, a listener has to identify the vowel in /kæt/, and has to give a rating from (1) "this is a poor exemplar of /æ/" to (3) "this is a good exemplar of /æ/" (Reetz & Jongman, 2009).
[101] The choice of having a three-point Likert scale was a methodological decision based on previous observation of participants' behavior when performing a test and on analysis of perception tests results. With larger scales, participants tend to have difficulty choosing a level, especially when taking long tests. Although measurements may not be as precise as with larger scales, we opted to have only three levels so that it would be easier and quicker to respond. Moreover, we believe that their answers were, nonetheless, accurate.
[102] Randomization organizes the stimuli in arbitrary order.

presented only once, that is, participants could not listen to the same token more than once, and could not change a response once given.

Table 13

Identification Test Stimuli

Vowels	bVt	tVk	tVt	kVt	pVt
/æ/	bat	tack	tat	cat	pat
/ɛ/	bet	tech	Tet	Ket	pet
/ɪ/	bit	tick	tit	kit	pit
/i/	beat	teak	teat	keep	Pete
/ʊ/	book (2x)	took	-	cook	put
/u/	boot	tuke	toot	coot	poop
/ʌ/	but	tuck	tut	cut	putt

Note. The stimuli presented in the ID test included trained and untrained words. Therefore, the words in the gray columns were also included in the training tasks, whereas the other stimuli were only added to the ID test (see Appendix L for wordlist frequency and Table I3 for the phonetic transcription of the stimuli).

It is important to highlight the fact that three of the five contexts, namely /pVk/, /tVt/, and /kVt/, were not included in the perception training tasks, as well as tokens produced by talkers T1, T2, and T6. This decision was based on the aim of comparing identification performance of the target vowels in trained and untrained contexts and produced by familiar and unfamiliar talkers. However, the designation of *untrained* or *unfamiliar* contexts and talkers is debatable because, once participants take the identification pretest, we cannot claim that the stimuli used remain unfamiliar or untrained anymore, when heard in the posttest and delayed posttest. This is also the reason why we decided to add a generalization test to this study with new untrained talkers and contexts, that is, with stimuli not included either in the auditory test or training tasks.

As Logan and Pruitt (1995) acknowledge, the choice of appropriate labels for an identification task may be a problem, because it may influence listeners' performance. Taking this into account and considering the participants' profile as undergraduate students of English Phonetics, a combination of two types of labels was used, namely orthographic representation and seemingly like phonetic transcription. Although there were two constraints, that is, the students' scarce knowledge of the International Phonetic Alphabet (IPA) and the computer program's limitation of not allowing the use of proper phonetic symbols, we chose to use the following labels: *heed* /i/, *hid* /I/, *head* /E/, *had* /ae/, *who'd* /u/, *hood* /U/ and *hud* /^/ (see Figure M1).

The perception test was set up in *TP – S*, version 1.0[103] (Rauber, Rato, Kluge, & Santos, 2011), a software application for designing speech perception experiments.

3.2.2.2.2 Generalization test

Logan and Pruitt (1995) explain that generalization occurs when:

> there is transfer of training to new tasks, to the productions of novel talkers, to new productions from the same talker(s) used in training, to new contexts (i.e., to stimuli in which the contrasting phones occur in phonetic contexts not presented in training), or to stimuli containing novel phonetic categories that share acoustic/phonetic features with the training stimuli. (p. 371)

In short, generalization to new conditions suggests that robust learning occurred. Therefore, we decided to test generalization by means of a similar seven-alternative forced-choice identification task with new stimuli produced by novel talkers. The stimuli used in the generalization test were selected from the edited audio materials of *Focus on Pronunciation* (Lane, 2005) and *Pronounce It Perfectly in English* (Yates, 2005).[104] In the generalization test, the stimuli consisted of 42 new monosyllabic words (see Table 14) produced by five native American speakers, yielding a total of 84 tokens with the seven AE vowels (7 vowels x 6 words x 2 repetitions). The stimuli included not only CVC tokens in new phonetic contexts, but also monosyllables with more complex syllabic structures (i.e., with word-initial consonant clusters such as /pl/, /bl/, /sl/, /sk/, /st/, and /fl/, and word-final clusters as /st/, and /lf/).

[103] The first version of *TP –S* was not an open source application; however, a new updated and free version of the software, *TP 3.1*. (Rauber, Rato, Kluge, & Santos, 2013), is available on http://www.worken.com.br/tp.

[104] In the generalization test, the productions of five talkers (two male speakers and one female speaker from Yates (2005), and one man and one woman from Lane (2005)) were selected. No biodata about the talkers are provided, but all of them speak the American English variety. Since these are professional speakers of standard AE and the regional dialect was not a controlled variable considered for analysis of the generalization test results, we considered that the lack of this information was not relevant.

Table 14

Stimuli used in the Generalization Test

	Monosyllabic Words		
Vowel	**CVC**	**CCV(C)(C)**	**CVCC**
/i/	seed, feel, leak	bleed, ski	feast
/ɪ/	fill, lick, mitt	slipped	fist, wrist
/ɛ/	said, met, wreck	pled, bled, slept	
/æ/	pad, back, sad, rack, mat	plaid	
/u/	cooed, fool, Luke	blue, flu, stewed	
/ʊ/	push, could, full, look	stood	wolf
/ʌ/	cud, bug, luck, rough	blood, flood	

Note. See Table I4 for the phonetic transcription of stimuli, and Appendix L for wordlist frequency.

The generalization test was set up in the software application *TP – S*. It also included a category goodness-of-fit scale from 1 (*poor*) to 3 (*good*) so that participants could rate their categorical perception of the target vowels.

3.2.3 Training Materials

Although the same identification tests (pretest, posttest, delayed posttest, and generalization test) were administered to the trainees and the controls, different training tasks were designed for each group in line with the purposes of the study, as it will be explained further in this section. The administration of the training program was organized as follows (Tables 15 and 16).

Table 15

Summary of the Training Program of the Experimental Group

EG	Session	Vowels	Tasks
1st block	1	high front vowels /i/-/ɪ/	Categorial AX DISC task 2AFC ID task
	2	high back vowels /u/-/ʊ/	Categorial AX DISC task 2AFC ID task
	3	mid – low front vowels /ɛ/-/æ/ low front – mid central vowels /æ/-/ʌ/	Categorial AX DISC task 2AFC ID task
2nd block	4	vowels /i/, /ɪ/, /ɛ/, /æ/, /u/, /ʊ/, /ʌ/	Oddity DISC task (ABX) 7AFC ID task
	5	vowels /i/, /ɪ/, /ɛ/, /æ/, /u/, /ʊ/, /ʌ/	Oddity DISC task (ABX) 7AFC ID task

Note. EG=experimental group; DISC=discrimination; ID=identification; AFC=alternative forced-choice.

Table 16

Summary of the Training Program of the Control Group

CG	Session	Consonants	Tasks
1st block	1	voiceless dental and alveolar fricatives /θ/-/s/	Categorial AX DISC task 2AFC ID task
	2	voiced dental fricative and alveolar stop /ð/-/d/	Categorial AX DISC task 2AFC ID task
		voiceless and voiced dental fricatives /θ/-/ð/	2AFC ID task
2nd block	3	bilabial and alveolar nasal stops /m/-/n/	Categorial AX DISC task 2AFC ID task
	4	alveolar and velar nasal stops /n/-/ŋ/	Categorial AX DISC task 2AFC ID task
	5	nasal stops /m/, /n/, /ŋ/	3AFC ID task

Note. EG=Experimental group; DISC=Discrimination; ID=Identification; AFC=alternative forced-choice.

In the following subsections, we will describe each of the training tasks, which were all designed in the same software application used for the perception tests. First, the materials used in the training program of the experimental group will be detailed, and then the tasks administered to the control group will be presented.

3.2.3.1 Experimental group

Taking into consideration the positive results of previous training studies (e.g., Aliaga-Garcia & Mora, 2007; Aliaga-Garcia, 2010; Iverson & Evans, 2007; Iverson & Pinet, 2008; Lambacher et al., 2005; Lengeris, 2009; Wang, 2008), and the uncontested claim that stimuli variability is necessary to improve L2 speech perception (Logan et al., 1991), we decided to adopt a high variability phonetic training (HVPT) method. This training procedure, which includes stimuli spoken by multiple NSs, directs L2 learners' attention towards relevant phonetic cues by exposing them to acoustic variability. Therefore, the tokens used in the perceptual training tasks were obtained from natural speech produced by 12 native speakers of American English from California and Iowa, aged between 24 and 51 (mean=36.17 years, SD=9.44 years), whose background information is presented in Table 17.

Table 17

Biodata of the American English Talkers

Talker	**Sex**	**Age**	**Place of Birth and Residence**	**Occupation**
T3[a]	M	50	Sacramento, CA	University Professor
T4[a]	F	27	Sacramento, CA	BA student in History
T5[a]	F	24	Davenport, IA	MA student in Occupational Therapy
T7	F	42	Sacramento, CA	Administrative assistant
T8	F	44	Sacramento, CA	Medical assistant
T9	F	51	Sacramento, CA	University Professor
T10	M	26	Sacramento, CA	BA student in International Business
T11	M	36	Sacramento, CA	Social Sciences teacher
T12	M	25	Sacramento, CA	BA student in Computer Science
T13	F	36	Davenport, IA	Administrative assistant
T14	F	34	Davenport, IA	Administrative assistant
T15	F	39	Davenport, IA	Administrative assistant

Note. T=talker; IA=Iowa; CA=California.[a] Talkers also included in the perception test.

The procedures followed to collect data from these native speakers of American English are described in Rauber, et al. (2010), and Rauber (2010). Within this group of 12 AE talkers, the productions of nine were only used in the training tasks, and the tokens of three talkers (viz. T3, T4, and T5) were also included in the identification test. Although the number of talkers was higher in the training auditory exercises, the number of phonological contexts was lower (/bVt/, /tVk/, /sVt/, and /hVd/) than in the identification test. In addition, of the four frameworks included in the training tasks, two contexts (shown in the gray columns of Table 18) were also included in the perception test.

Table 18

Stimuli used in the Training of the Experimental Group

Vowels	**bVt**	**tVk**	**sVt**	**hVd**
/æ/	bat	tack	sat	had
/ɛ/	bet	tech	set	head
/ɪ/	bit	tick	sit	hit[a]
/i/	beat	teak	seat	heed
/ʊ/	book[a]	took	soot	hood
/u/	boot	tuke	suit	hoot[a]
/ʌ/	but	tuck	shut[a]	hut[a]

Note. The stimuli in the gray columns were also included in the perception ID test. See Table I5 for the phonetic transcription of stimuli and Appendix L for wordlist frequency.
[a] Different phonological contexts.

To train participants' ability to perceive vowels categorically we included two types of perception tasks in the training program, namely identification and discrimination tasks, which will be briefly described. Although Wang and Munro (2004) concluded that identification tasks have yielded better results than discrimination training in terms of improvement of perceptual performance, we decided to include both because they develop different perceptual abilities. Identification tasks promote the formation of new perceptual categories that are robust to acoustic variability, that is, encourage learners to group perceptually similar phonetic segments into the same category, whereas discrimination exercises direct on listeners' attention to cues that contrast perceptual categories by encouraging them to listen to between-category differences. As Pisoni and Lively (1995) explain, discrimination training promotes "acquired distinctiveness" and identification training promotes "acquired equivalence" (p. 445). A more detailed description of the perception ID and DISC tasks used in the training sessions of the experimental group will follow.

3.2.3.1.1 Identification tasks

The design of the identification exercises included in the training was similar to the 7AFC identification test, because the focus was on listeners' ability to assign a linguistic label to a set of vowel segments differing in durational and quality acoustic parameters. The main differences regarding the perception test were on the number of available labels in each task, the addition of trial-by-trial feedback, the absence of category goodness-of-fit judgments, and the possibility of listening to each token more than once.

The stimuli used in the categorial two-alternative identification tasks of the first block of the training program were recorded by 12 NSs of American English, and the tokens included in the seven-alternative identification tasks of the second block were spoken by nine talkers (viz. T3, T4, T8, T9, T10, T12, T13, T14, T15) so that participants could train in high variable conditions, that is, with stimuli produced in different contexts by multiple speakers.

In the first block of the training program, the three sessions included a 2AFC identification task with 96 tokens (2 vowels x 4 contexts x 12 repetitions). The third session, however, focused on two vowel contrasts, namely the target contrast /ɛ/-/æ/ and the vowel contrast with the distractor /æ/-/ʌ/. The 2AFC ID task with the distractor vowel had only 64 tokens (2 vowels x 4 contexts x 8 speakers). In the second block of the auditory training, the design of the 7AFC identification tasks was identical to the identification test, but the number of stimuli was smaller. Thus, the total number of tokens of each 7AFC ID task was 168 (7 vowels x 4 contexts x 6 repetitions).

3.2.3.1.2 Discrimination tasks

Discrimination refers to "the act of differentiating two or more stimuli, presented in some predefined format" (Logan & Pruitt, 1995, p. 354). Several variants of discrimination tasks have been used in speech perception research, but Logan and Pruitt (1995) suggest that these can be divided into three basic categories, viz. (1) same-different (AX) tasks, (2) ABX tasks, and (3) category change tasks. In this study, only the two former types of tasks were administered in the training sessions. Therefore, a brief description of each category will be further provided.

In the AX categorial (same-different) discrimination exercises, the participants' task was to indicate whether or not two stimuli in randomized word pairs were exemplars of the same phonetic category. Therefore, listeners had to form some kind of mental representation of the phonetic categories under comparison instead of directly comparing stimuli on the basis of physical identity alone. Stimuli in same pairs were physically different tokens drawn from the same phonetic category, while stimuli in different pairs were drawn from distinct categories. The stimuli used in the AX discrimination tasks were produced by two female native AE talkers (T7 and T8) and two male talkers (T3 and T12). Since there were four tokens with different phonological contexts in the set of the training stimuli, namely *book*, *shut*, *hit*, and *hoot* (see Table 18), these were not included in the discrimination tasks because they did not possibly form minimal pairs. Therefore, the minimal pairs used in the discrimination tasks are presented in Table 19.

Table 19

Minimal Pairs used in the Discrimination Tasks

Vowel Contrast	**Minimal Pairs**
/i/-/ɪ/	beat-bit, teak-tick, seat-sit
/u/-/ʊ/	took-tuke (2x), suit-soot
/ɛ/-/æ/	bet-bat, tech-tack, set-sat
/æ/-/ʌ/	bat-but (2x), tack-tuck

The total number of trials in each of the three AX categorial discrimination tasks was 72 (3 contexts x 4 sequences x 6 repetitions). The AX task administered in the third session, which included the distractor vowel /ʌ/, had only 48 trials (3 contexts x 4 sequences x 4 repetitions). The isolated CVC words had to be paired for the AX task, so we concatenated the word pairs with sequences AX-XA-AA-XX. For example, if the vowel contrast to be trained was /i/-/ɪ/, the organization of the stimuli was: beat-bit (AX); beat-bit (XA); beat-beat (AA); and bit-bit (XX). The interstimulus interval (ISI) for all trials was 1.2 s, which was the same used by Guion et al. (2000).

In the first block of the training program, the three sessions started with an AX discrimination task with 72 stimuli, followed by a 2AFC identification task with 96 tokens, yielding a total of 168 tokens per session. The third session, which focused on two vowel contrasts, yielded a total of 280 trials (see Table 20).

Table 20

Number of Trials in each Training Task of Session 3

Training Session 3	**2AFC ID task** nº trials	**AX DISC task** nº trials
/ɛ/-æ/	96	72
/æ/-ʌ/	64	48
	160	120
total nº	**280**	

In the second block of the training program, two oddity discrimination tasks were used to train vowel perception. This discrimination task was similar to an ABX or AXB categorical task in that sequences of three stimuli for trial, that is, triads produced by three different talkers were presented, in which one stimulus differed categorically from the other two (Flege, Munro, & Fox, 1994; Gottfried, 1984; Nozawa, 2009). The odd stimulus could be presented in any of the three possible positions within the stimuli sequence, and listeners indicated whether it was in the first, second, or third position. In these tasks, participants heard three physically different stimuli in each trial and identified the position of the token that had a categorically different vocalic segment from the other two. Each vowel contrast was tested by different trials, which contained an odd item, and catch trials, which had three physically different tokens of the same vowel produced by three talkers, following the same procedure described by Guion et al. (2000). This encouraged learners to respond only to phonetically relevant differences, not to any auditorily detectable difference. To successfully discriminate the target vowels, participants had to recognize the categorical identity of a set of physically different tokens of the same vowel category while ignoring acoustic differences among instances of the category, which are phonetically irrelevant to their categorical identity.

In the two oddity discrimination tasks, each vowel contrast was trained by means of two catch trials (sequences of three physically different tokens containing the same target vowel), and six change trials (sequences in which there was an odd item, that is, a categorically different vowel among the three stimuli). The odd vowel appeared equally in all three possible positions. For example, to practice the high front vowel contrast /i/-/ɪ/, two catch trials, such as *seat-seat-seat*, and *sit-sit-sit*, were administered. In addition, three change trials in which /i/ was the odd item, and three triads in which /ɪ/ was the odd item, for example, *seat-seat-sit* (ABX), *seat-sit-seat* (AXB), and *sit-seat-seat* (XAB), were also presented (see Tables N1 and N2).

Although the memory demands imposed by an oddity discrimination task make it more difficult for listeners to discriminate sounds compared to other modes of stimulus presentation, as acknowledged by Logan and Pruitt (1995), we decided to include it in the second block of the auditory training of the experimental group, because they develop not only the perceptual discrimination ability, but also phonological short-term memory.

In the two ABX tasks only three contexts were included, namely /sVt/ and /bVt/ in the first task, and /tVk/ in the second task (see Tables N1 and N2). Stimuli used in the oddity DISC tasks were produced by three AE talkers (T3, T4 and T7), totaling 64 trials with a 1.2 s ISI, in each auditory exercise (4 vowel contrasts x 8 sequences (6 change trials + 2 catch trials) x 2 repetitions). The total number of trials presented in the second block of the training was 232. The two sessions consisted of an oddity DISC task with 64 tokens, followed by a 7AFC ID task with 168 stimuli (Table 21).

Table 21

Number of Stimuli per Training Session of the Experimental Group

EG	Session	Vowels	Tasks	Nº Stimuli	Total Nº
1st block	1	/i/-/ɪ/	Categorial AX DISC task	72	168
			2AFC ID task	96	
	2	/u/-/ʊ/	Categorial AX DISC task	72	168
			2AFC ID task	96	
	3	/ɛ/-/æ/	Categorial AX DISC task	72	280
			2AFC ID task	96	
		/æ/-/ʌ/	Categorial AX DISC task	48	
			2AFC ID task	64	
2nd block	4	/i/, /ɪ/, /ɛ/, /æ/, /u/, /ʊ/, /ʌ/	Oddity DISC task (ABX)	64	232
			7AFC ID task	168	
	5	/i/, /ɪ/, /ɛ/, /æ/, /u/, /ʊ/, /ʌ/	Oddity DISC task (ABX)	64	232
			7AFC ID task	168	

The total number of stimuli presented throughout the second block of the training program was higher than in the first block, because the introductory part of these two sessions was shorter, as it will be explained in Section 3.3.

3.2.3.2 Control group

In a pretest-posttest experimental design, the use of control groups is, according to Logan and Pruitt (1995), “relatively uncontroversial” (p. 371). However, to make sure that improvements in performance between the pretest and the posttest are attributable to vowel training and not to the effect of task familiarization, that is, the successive exposure of participants to the stimuli and familiarization with the perception tasks, the control group also experienced auditory training. Therefore, although the length of the training program was the same (viz. five sessions), and the type of perception tasks was similar, the focus of the training of the control group was on two groups of consonantal segments that present perceptual difficulty to Portuguese L2 learners, namely the dental fricatives /θ/ and /ð/ and the nasals /m/, /n/, and /ŋ/. The controls’ training program also consisted of identification and discrimination tasks.

However, in terms of identification tasks the maximum number of alternatives was three. The discrimination practice included only AX categorial tasks, because there were no productions of the same token recorded by three different talkers available. Thus, an oddity DISC task could not be included (see Table 22).

Table 22

Number of Stimuli per Training Session of the Control Group

CG	Session	Consonants	Tasks	Nº stimuli	Total Nº
1st block	1	/θ/-/s/	Categorial AX DISC task	96	
			2AFC ID task	72	168
	2	/ð/-/d/	Categorial AX DISC task	96	240
			2AFC ID task	72	
		/θ/-/ð/	2AFC ID task	72	
2nd block	3	/m/-/n/	Categorial AX DISC task	80	
			2AFC ID task	72	152
	4	/n/-/ŋ/	Categorial AX DISC task	80	
			2AFC ID task	72	152
	5	/m/, /n/, /ŋ/	3AFC ID task	162	162

The stimuli used in the training of the control group were selected from pronunciation teaching materials, thus, the AE dental fricatives were produced by five NSs of American English, and the stimuli used in the training of nasals were produced by two NSs.[105] The training program was divided into two blocks but did not follow the same logic as the training design of the experimental group, based on degree of difficulty. The two blocks of training were selected according to a certain set of consonantal segments. In the first block, the focus was on the dental fricative contrast and, in the second block, on nasals.

3.2.3.2.1 Identification tasks

The alternative forced-choice identification task consisted in labeling a certain phonetic segment from two or more options, as explained earlier. The control group was trained with five 2AFC tasks and one 3AFC task. Immediate feedback was provided

[105] The audio materials used in the consonant training tasks are from two pronunciation books, namely *Pronounce It Perfectly in English* (Yates, 2005) and *Focus on Pronunciation* (Lane, 2005). In the first block of sessions, the productions of three male talkers and two female talkers were used from both resources, and in the second block only a female talker and a male talker from Lane (2005) were selected.

after each trial, and listeners could replay the same stimulus up to three times, before choosing a label. The total number of stimuli in each 2AFC identification exercise was 72 (2 consonants x 18 contexts x 2 repetitions), and 162 in the 3AFC ID task (3 consonants x 18 consonants x 3 repetitions) (see Tables O1 and O2 for a list of the stimuli used in the perception tasks).

3.2.3.2.2 Discrimination tasks

The isolated words used in the discrimination tasks of the first block were paired with an ISI of 1.2 s. These trials of minimal pairs followed the sequence AX, XA, AA, XX, such as *breed-breathe* (AX); *breathe-breed* (XA); *breed-breed* (AA), and *breathe-breathe* (XX), yielding a total of 96 trials (2 consonants x 4 sequences x 12 contexts). The AX tasks of the second block included a total of 80 trials (2 consonants x 4 sequences x 10 contexts) (see Tables O3 and O4).

3.3 Procedures

In this section, the procedures followed to administer the perception and production tests, and the training program will be thoroughly reported. The first part of this section will focus on the testing procedures, namely on the production and perception data collection, and the second part will describe the training program of both the experimental and the control groups.

The experiment was conducted within the *English Phonetics and Phonology* course. Hence, the perception tests and training were administered in the practical classes and tutorials of this course, which included a two-hour theoretical class for the whole group of undergraduate students (including those not participating in the study), a one-hour tutorial, and a one-hour practical class for each group, per week. The production data collection was scheduled individually, according to students' availability in a certain period of time (see the experiment's timeline in Appendix P).

3.3.1 Testing Procedures

The procedures followed to collect perception and production data are described next.

3.3.1.1 Production

This section describes the data collection procedure of the production tests. First, the procedures followed to collect the European Portuguese vowels spoken by the Portuguese students will be reported and then the collection of the American English vowels production by the same students in the three testing moments of the experiment will be presented.

The recordings of the Portuguese L2 participants' productions were done individually in a sound-attenuated booth[106] at the University of Minho with an Edirol R-09HR digital recorder at a 44 Hz sampling rate, with 16-bit accuracy, and a unidirectional Edirol CS-15 microphone (see Figure Q1 for photographs of the booth and the recording equipment). All the recordings were saved as wav sound files and edited after data collection so as to normalize intensity and remove noise. Therefore, a *Praat* script (Appendix R) was run to normalize the audio files, and noise was removed using the *Audacity* 2.0.2 (2012) software.

3.3.1.1.1 European Portuguese vowels

The production of the seven European Portuguese vowels was collected by means of a sentence-reading-aloud task that was set up in a custom-designed computer program using CSharp (see Figure 11), which automatically randomized the presentation of the tokens to avoid ordering effects.

Prior to testing, oral instructions were given in Portuguese by the researcher, and students were familiarized with the test. The seven pictures used to elicit the target vowels were shown on a computer screen to confirm if they were correctly named (Appendix S), and participants read the example sentence.

[106] The room used to collect data was a booth in a conference room used by interpreters. Although it was not completely soundproof, it was an adequate sound-attenuated booth.

Although the dissylabic CVCV words with the target vowels were included in sentences to avoid list-reading intonation, L2 learners were asked to read the tokens with a falling intonation, natural loudness, and normal speech rate, and pause between a set of words. Since it was the researcher who was responsible for clicking on the mouse to move each slide forward, whenever the recording was not satisfactory, participants were asked to reread the sentence. For example, as a result of the orthographic accents <ê, é, ô, ó>, some participants tended to confuse vowel height, thus, producing, for example, the open vowel /ɛ/ instead of the mid-closed /e/. Whenever this happened, they were asked to repeat the sentence. Moreover, when they did not pause voluntarily, the researcher paused the recording and told them to rest and drink some water, which was always at their disposal. Recordings took approximately eight to ten minutes. The collection of the EP vowels was done in the same recording session as the AE vowels, before the perception pretest, but between the two tests participants had to take a break of, at least, five minutes.

3.3.1.1.2 American English vowels

Participants' productions of the seven American English vowels were elicited by means of a sentence-reading-aloud task, which included the naming of a picture before each sentence. The test was assembled in the same customized computer program used for the EP production test.

Before starting the recording sessions, oral instructions were given in Portuguese, and participants were familiarized with the test. The pictures used to elicit the target vowels were shown on a computer screen to confirm if they were correctly identified (Appendix T), and a sample sentence was read. The L2 learners were asked to read the sentences at a normal speech rate, with a falling intonation, and pause between a set of sentences. Whenever the recording was not satisfactory due to, for example, noise, slow or fast speech rate, participants were asked to reread the sentence. Contrarily to what happened in the collection of the EP vowels, if the AE vowels were not accurately produced in terms of vowel quality, participants were not asked to repeat the sentence, provided the picture was correctly labeled, and the word illustrated by the picture rhymed with the word in the carrier sentence. Recordings took approximately between 10-12 minutes.

Production of the AE vowels was collected three times. The first time was in the pretest phase, in the same recording session as the collection of the EP vowels. The second time was immediately after the training sessions were over (posttest), and the third time was two months after the training program was completed (delayed posttest). Except for a few cases in the posttest phase,[107] the recordings always preceded the identification tests to avoid the possible effects of exposure to the audio stimuli on production (Lambacher et al., 2005, p. 237).

3.3.1.2 Perception

The two groups of L2 participants – the trainees and the controls – took the same perception pretest, posttest, delayed posttest and generalization test. The perception tests were set up in *TP-S*, version 1.0, and included a three-point category goodness-of-fit scale. Stimulus presentation was automatically randomized, and tokens were played only once.

3.3.1.2.1 Identification test

The identification pretest was administered two weeks after the second semester began so that in the first weeks the production data were collected. The posttest was done six weeks immediately after the training began and the delayed posttest two months later (see the timeline of the experiment in Appendix P). The same test was administered at the three moments.

The perception tests (pre-, post-, and delayed posttest) were administered in a quiet computer lab[108] at UM. The *TP–S* software was installed in all the computers, so that the test could run on several computers simultaneously. Each participant performed the test individually with NGS MSX6 Pro stereo headphones.

Before the perception test started, the researcher explained the procedures briefly, and handed out a sheet of paper with detailed instructions in English (Appendix U), which participants read before starting the test. These included information about the design of the test, the working of the computer program, and the testing sequence,

[107] There were five students who missed the recording sessions and were not able to record the data before the identification posttest.
[108] The computer lab where participants took the perception tests had 50 computers so that the two groups of participants could complete the tests at the same time. The two weekly training sessions took place in a smaller computer lab (see photo in Appendix Z) and, thus, the experimental group and the control group attended the sessions separately.

that is, the indication of a familiarization trial test before the actual test. The participants were advised to keep the instructions with them throughout the test, so that they could check them whenever in doubt. The researcher was also present and available to help and monitor the students. They were also advised to pause, and rest after a set of tokens. The first version of *TP–S* did not have the option of inserting pauses in a test, therefore, the only predefined pause the test had was between the trial test (28 tokens) and the identification test (210 tokens). This pause gave students the possibility to clarify questions regarding the labels, the stimuli, and the computer program.

The perception test was, thus, preceded by a short practice trial (28 tokens) to familiarize participants with the task, the stimuli and the range of possible labels. Familiarization procedures are important and appropriate when the phonetic contrasts are unfamiliar to listeners and test performance is to be maximized (Beddor & Gottfried, 1995).

In the test, informants listened to a series of stimuli with one of the seven AE vowels, through headphones, and identified them by pressing labeled buttons on a computer screen with a mouse. Immediately after selecting a label, they clicked on one of the three buttons of a scale ranging from (1) *poor* to (3) *good* to rate goodness-of-fit of the selected AE vowel category (see the computer screen of the identification test in Figure M1). When the test was completed, a window with the total scores was displayed (see Figure M2). The reason to show participants the global result of their performance in the tests was mainly motivational. This type of cumulative feedback,[109] which provides information about session-by-session performance, may not affect learning in the same way as immediate feedback, but may maintain students' motivation through the course of the training program (Logan and Pruitt, 1995).

The computer program also measured response time (RT)[110] in seconds. RT started counting immediately after the stimulus was played and stopped with the click of the mouse. The response buttons were blocked while the stimulus was presented, thus, participants could not press them before the audio file was over. This measure was included with the purpose of assessing participants' sensitivity to within-category differences, since poor exemplars of a category take longer to identify than good exemplars (Beddor & Gottfried, 1995; Reetz & Jongman, 2009). By comparing the

[109] Logan and Pruitt define cumulative feedback as the type of feedback that "spans more than one trial" (1995, p. 363).

[110] Response time, or reaction time, is a measure of amount of processing. For example, the more ambiguous a stimulus is, the longer it takes for a listener to process it, and the time to perform a certain reaction, such as pressing a button, will be longer (Reetz & Jongman, 2009, p. 269).

measures of RTs among the three test phases, we can verify whether or not there was a decrease of time spent in the identification of the target vowel contrasts. On average it took participants around 20-30 minutes to complete the test, but they were allowed to stop and rest when necessary.

3.3.1.2.2 Generalization test

First, the researcher explained the instructions orally in English, and then handed them out in separate sheets of paper (see Appendix V). The perception posttest and the generalization test were administered on the same day, but participants had to take a five-minute break between each test. It took one hour, on average, to complete the two tests. Since informants were familiarized with the identification test already, the generalization test did not include practice trials.

In the generalization test, students had to listen to a series of monosyllabic words with one of the seven AE vowels and give their responses with a mouse click by selecting one of seven buttons on the computer screen. After labeling the vowel, they were asked to rate category goodness-of-fit of each AE vowel using a three-point scale, from 1 (*poor*) to 3 (*good*).

3.3.2 Training Procedures

Taking into consideration the fact that students were taking an *English Phonetics and Phonology* course, and that instructional variations have little effect on listeners' performance in identification and discrimination tasks (Beddor & Gottfried, 1995), instructions with a brief articulatory description of the target phonetic segments were given at the beginning of each training session (see Appendix W), and instructions for the auditory tasks were handed out in a separate piece of paper (see Appendix X). Both the experimental and the control groups were given articulatory-visual descriptions of the phonetic segments trained on each specific session. Nevertheless, although the training sessions included instructional information to focus participants' attention on the acoustic properties of vowels, production of the vocalic segments was not encouraged.

The perception training tasks were set up in *TP–S*, and the procedures were similar to the ones described in the previous section. However, there were a few differences in the design of the training tasks, which will be detailed next. The trials in the training tasks were followed by immediate trial-by-trial feedback,[111] and at the end of each session cumulative feedback[112] was also provided to learners. Immediate feedback was given by means of visual information after each response (see Figures Y1 and Y2). If the identification of the target segment was correct, participants could listen to the next trial, but if they identified the vowel incorrectly, a message was displayed, and they had to listen to the stimulus again and select the correct answer.[113] The correct/incorrect feedback was, thus, enhanced by adding the repetition of each stimulus that was incorrectly identified so that listeners would not only realize that they had made an error, but they would also be presented with the stimulus again to hear it associated with its correct label (Logan & Pruitt, 1995, p. 363). Moreover, before selecting a label participants could listen to the same token up to a maximum of three times.

The perception training program consisted of five 45-minute sessions that followed the same organization: (1) articulatory-visual description of the target segments; (2) instructions for each task; (3) discrimination task; and (4) identification task. Although both the experimental and the control groups had similar training sessions on the same day, divided into two blocks, they were organized in a different way. For that reason, we will describe the training program of the experimental group first, and then the program of the control group (see Tables 15 and 16).

The training tasks were administered in *TP-S* running in several computers simultaneously in a quiet computer lab (see Figure Z1). Participants trained individually, and heard the stimuli at a comfortable listening level over headphones. Instructions were provided prior to the administration of exercises (see Appendix X), and participants completed the training practice in the same order: discrimination task followed by an identification exercise. Stimuli presentation was randomized for each participant. The training tasks were administered once per week for a total of five

[111] Feedback is the information provided to participants about their performance. Trial-by-trial feedback is defined as information given to participants about their performance after each trial. This type of feedback, given immediately after each trial, is optimally designed to help informants modify their performance, i.e., to help them adjust their performance to accuracy of responses (Logan & Pruitt, 1995, p. 361).
[112] Cumulative feedback consisted of information about students' overall correct ID and DISC scores measured in percentages.
[113] Iverson and Brown (2007) followed a similar procedure.

weeks. Each training session lasted around 55-60 minutes. The training sessions occurred in the one-hour practical weekly class of the *English Linguistics 1* course.

3.3.2.1 Experimental group

The training tasks of the experimental group were identical to the perception test, except that immediate trial-by-trial feedback was provided, each token could be played up to three times before choosing a response, and no category goodness-of-fit scale was added. After each trial, feedback in the form of a green tick for correct answers or the red-colored sentences "Incorrect answer! Click on the Replay button to listen again and correct your answer" for incorrect answers was shown on the computer screen (Figure Y1).

Training progressed in terms of degree of difficulty, from AX discrimination and 2AFC identification tasks to more complex exercises, namely oddity DISC and 7AFC labeling tasks. In the first block of the training program, the three sessions started with an AX discrimination categorial task followed by a 2AFC identification task, and in the second block sessions started with an oddity DISC task followed by a 7AFC ID task (see Table 15). In the AX discrimination tasks, trainees were told that they would listen to sequences of two CVC words spoken by different native American English talkers and that they had to decide if the vowels of the two words matched or not, that is, if they were the same (different physical tokens of the same vowel) or if they were different (tokens with categorically different vowels). In the 2AFC identification tasks, learners listened to CVC words with one of two vowels of the following vowel pairs /i/-/ɪ/, /æ/-/ɛ/, /u/-/ʊ/, and /æ/-/ʌ/. They were asked to assign a label to the vowel that was spoken by clicking on one of two buttons on the screen.

In the oddity task that initiated the one-hour training sessions of the second block, students were told that the three stimuli in each trial were always spoken by different talkers, so they should ignore differences in speakers' voices as much as possible. They were informed that two situations could occur: (1) one of the three stimuli would be categorically different from the other two, and they would have to point out in which position it was produced (*change trials*); and (2) the trial would consist of three tokens with the same vowel produced by three different talkers (*catch trials*).

Learners were visually presented with four buttons on a computer screen with the numbers 1, 2, 3 and “same”. The “same” button was provided to signal the cases in which no differences in vowel category were detected (see Figure Y1). Participants selected ‘‘1,’’ ‘‘2,’’ or ‘‘3’’ if they judged a stimulus in one of those three sequential positions to be different from the other two stimuli, and they selected ‘‘same’’ if all three examples were considered to be instances of the same vowel. In the 7AFC tasks, the procedures were the same as in the perception tests, with some differences regarding feedback, as already mentioned.

The training sessions consisted of three moments, namely 10-15 minutes for brief explanation of articulatory parameters, 15-20 minutes to complete the discrimination task, and 20-25 minutes for the identification tasks. However, in the second block of the training, only a five-minute summary of the articulatory properties of the seven AE vowels was done, leaving more time to complete the auditory training tasks, which included more stimuli than the tasks in the first two sessions (see Table 21).

3.3.2.2 Control group

The control group was exposed to the same training procedures but the focus was on learning to categorically discriminate and identify consonants. Therefore, the 60-minute training sessions included (1) an articulatory description of the segments to be trained (see Appendix W); (2) brief instructions on how to carry out the exercises (see Appendix X), which were provided in separate sheets of paper; (3) an AX discrimination task; and (4) an identification exercise. Immediate feedback was included in the training tasks, and each trial could be played three times before choosing a response.

The five training sessions were divided into two blocks according to the set of consonants being trained. Four sessions of the training program started with an AX DISC followed by a 2AFC identification task, and the last session consisted of a 3AFC ID task (see Table 16). In the AX discrimination tasks, students were told that they would listen to sequences of minimal pairs produced by two different native American English speakers, and they would have to decide if they were the same (different physical tokens of the same consonant) or if they were different (tokens with categorically different consonants). In the 2AFC identification tasks, participants listened to tokens with one of the two consonants of the following consonant pairs /θ/-/s/, /d/-/ð/,/θ/-/ð/, /m/-/n/, /n/-/ŋ/, and in the 3AFC task they were presented tokens with one of the three nasals /m/-/n/-/ŋ/. Participants were asked to identify the consonant that was spoken by clicking on one of the available buttons on the computer screen.

3.4 Analysis

The analyses carried out to examine perception and production data are summarized next.

3.4.1 Assessment of Production Data

Production data were acoustically analyzed to investigate how European Portuguese and American English vowels were produced by Portuguese EFL learners. Therefore, duration and the first three formants of the vowels were measured, using *Praat* 5.3.39 (Boersma & Weenink, 2013).

After normalizing intensity and removing noise, the recordings of the four production tests (EP vowels' test and AE vowels' pretest, posttest, delayed posttest) were analyzed following similar procedures as those described in Rauber (2010) and Nobre-Oliveira (2007). The waveform and the wideband spectrogram of the production data were visualized in *Praat* 5.3.39, and an annotation text with three tiers was created for each audio file so that the production data could be segmented and labeled. In the first tier, the words with the target vowels were segmented and labeled, in the second tier the phonetic context was transcribed, and in the third tier the target vowels were segmented and transcribed. Since the order of presentation of the stimuli used to collect these data was randomized and the words illustrated by pictures were not considered for analysis, segmenting words in the first tier was necessary to locate the target vowels. To segment the vowels, boundaries were selected at the first and last zero crossings, where the first positive and the last negative peaks with considerable amplitude could be seen, marking the beginning and ending of each target vowel. After segmentation, the vowel was transcribed. The procedures described above are illustrated in Figure 13, which exemplifies the segmentation and labeling of the vowel [i] produced by one of the male participants.

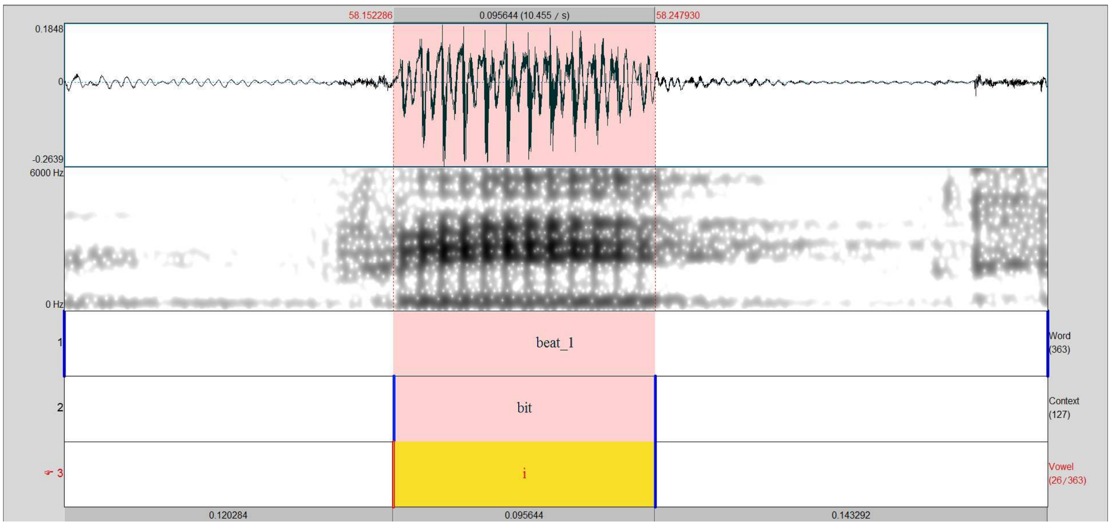


Figure 13. Segmentation of vowel [i].

After segmentation, duration and formants F1, F2 and F3 were measured from the central 40% part of the target vowels. To compare the AE productions of each participant in the pretest, posttest, and delayed posttest, and verify whether there was improvement after training, the aforementioned acoustic parameters were measured, but only duration, F1 and F2 were analyzed for the purpose of this study.

To factor out vocal tract variation in acoustic measurements, normalization[114] was carried out with NORM[115] (Tyler & Thomas, 2010). The method chosen to normalize vowel formant data was the Lobanov method, a vowel-extrinsic formula.[116] This formula was applied to normalize the European Portuguese and the American English vowels produced by L2 participants so that female and male speakers' productions were comparable. Furthermore, in order to have comparable vowel spaces, the production data of Portuguese EFL learners were normalized according to the minimum and maximum formant values[117] of the vowels produced by the American speakers (for a detailed description about the normalization procedure, see Rauber, 2010, pp. 90-91). The *Praat* script run to normalize data is shown in Appendix AA.

[114] Different speakers have different vocal tract sizes and shapes, which in turn cause their formant resonances to differ. Therefore, vowel normalization is crucial to compare the different realizations by different speakers. The four purposes of normalization are: (1) to eliminate variation caused by physiological differences among speakers; (2) to preserve cross-linguistic/sociolinguistic differences; (3) to preserve phonological distinctions among vowels; and (4) to model the cognitive processes that allow human listeners to normalize vowels uttered by different speakers.

[115] NORM is a web-based interface to the vowels R package, which is designed to manipulate, normalize, and plot vowel formant data.

[116] NORM uses the formula: Fn/V/N= (Fn/V/-MEANn)/Sn, where Fn/V/N is the normalized value for Fn/V/ (i.e., for formant *n* of vowel V). MEANn is the mean value for formant *n* for the speaker in question and Sn is the standard deviation for the speaker's formant *n*. /http://ncslaap.lib.ncsu.edu/tools/norm/norm1.php/.

[117] In sum, the reference values are calculated as follows: F1 min=lowest F1 mean-SD; F2 min=lowest F2 mean-1SD; F1 max=highest F1 mean+1SD; F2 max=highest F2 mean+1SD.

After normalizing the formant values, the median F1 and F2 values of each vowel were calculated, as well as the Euclidean distance (ED) between the vowels of three target pairs (/i/-/ɪ/, /ɛ/-/æ/, and /u/-/ʊ/) and two pairs with the distractor vowel (/æ/-/ʌ/, and /ʌ/-/ʊ/). The Euclidean distance[118] is the space in Hertz (Hz) between two vowels of a pair in terms of F1 and F2 values (for details on Euclidean distance, see Bion et al., 2006, p. 1364). Given that we were not interested in the comparison of absolute values between the two groups, but in the space between the vowels of each pair, this procedure allowed the comparison between the Euclidean distance of the vowels produced by the Portuguese and the American participants. The ED calculations were also done with a *Praat* script (see Appendix AB).

After measuring these values, the vowels produced by the EFL learners in the pretest, posttest and delayed posttest were plotted in *Praat* 5.3.39 and compared to each other to verify whether there was improvement in terms of decrease in overlap of the AE vowels. All the measures were then exported to *IBM SPSS* v. 20 for statistical analysis.

3.4.1.1 Duration

To measure duration, the start and end points of each of the 7602 vowel productions (1428 EP vowels + 6174 AE vowels) were segmented manually in the sound wave. As explained previously, each vowel was manually segmented and labeled in the sound wave by using the *Praat* program, version 5.3.39. Both the beginning and the end of the selection were at a zero crossing of the waveform, that is, when the wave sound crosses zero amplitude. The start and end points were considered to be the first and the last periodic pulses on the waveform that had considerable amplitude, and whose periodic shape resembled a vocalic segment. As explained earlier, voiceless obstruent contexts were selected as flanking consonants to facilitate segmentation, and particularly duration measurements, since these consonants allow a more precise identification of the first and last periodic pulses of the vowels.

The duration ratios of the three target vowel pairs were calculated by dividing the mean duration values (ms) of /i/, /u/, and /æ/ by the mean duration of /ɪ/, /ʊ/, /ɛ/, respectively, following the same procedure as Wang (2008). The duration ratio of the

[118] The ED reflects how much the F1 and F2 values of a vowel are distant from the F1 and F2 values of another vowel within the productions of a particular group of participants.

vowel pair /æ/-/ʌ/, with the distractor vowel, was also considered for analysis. According to Wang (2008), the higher the values of the ratios are, the greater the differences between the two vowels of the pair (p. 124). Therefore, the duration ratios were expected to reflect duration differences between the vowels of the four contrasts.

3.4.1.2 The first three formants

The formant frequencies of F1, F2, and F3 of each vowel were measured in the 40% central portion of the target vowels by running a *Praat* script (Appendix AC), and a table with these acoustic measurements was automatically created by another *Praat* script (Appendix AD). To make comparisons between different phases of the experiment, the F1 and F2 values were plotted in Hertz with inverted scales to approximate traditional articulatory vowel charts. To interpret the acoustic characteristics of the vowels under study, the first procedure was to calculate the mean,[119] median,[120] and standard deviation[121] (SD) of the acoustic measurements of the EP and AE vowel productions. The medians and means were plotted in different vowel plots, and the SD was also shown by means of ellipses around the vowel symbols, which represented the mean value (see the scripts used to create vowel plots in Appendices AE and AF). To compare the mean or median results of sets of vowels and verify if they differed statistically or not, tests of significance were used. The extent to which the L2 speakers' production of a vowel differed from the phonetic norm of American English was estimated by computing the difference between the Euclidean distance of the participants' F1 and F2 values and that of the AE speakers' F1 and F2 values.

3.4.2 Assessment of Perception Data

The perception data were analyzed after extracting the results from *TP-S* experiment files. The correct identification percentage scores of the perception pretest, posttest and delayed posttest were compared, and statistical tests were run with *IBM*

[119] The mean is the center of a distribution of scores (Field, 2009).
[120] The median is the middle score of a set of ordered observations, i.e., the average of two scores that fall either side of what would be the middle value (Field, 2009). The median is a more robust measure than the mean because it is less sensitive to extreme scores, especially in highly skewed distributions.
[121] The SD is a measure of variability which indicates how spread the data are from the center, i.e., how much they deviate from the mean (Field, 2009).

SPSS v. 20 to verify whether there were any training effects on the participants' perception of the target vowels. Data were also statistically analyzed to verify if there was generalization of learning. Moreover, the percentage of correct and incorrect identifications was organized in confusion matrices for comparisons. Other variables, such as response times and category goodness-of-fit ratings were also considered for statistical analysis. The data obtained from the questionnaires were related to test results, whenever necessary. The statistical analyses will be reported in the next chapter in more detail, along with the description and discussion of the results.

CHAPTER 4

RESULTS AND DISCUSSION

To count leaves is not less meaningful than to count the stars.
- David Ignatow

In this chapter, a detailed description of the perception and production data analyses is provided, and their results are presented. The first part is dedicated to the effects of auditory training on perceptual performance of Portuguese EFL learners and the second part to its effects on production. Although the three American English vowel contrasts /i/-/ɪ/, /ɛ/-/æ/, and /u/-/ʊ/ were the focus of investigation, the six target vowels and the distractor /ʌ/ were also analyzed separately. Therefore, results of analyses by vowel and by vowel contrast are reported in distinct sections. Furthermore, analyses of individual trainees' perceptual and production performance conclude the two main parts of this chapter. Finally, the effects of perceptual training on vowel perception and production are discussed and then the interface between these two domains is briefly examined.

4.1 Speech Perception

4.1.1 Validation of perceptual testing and training materials

Seven native American English speakers participated in the present study in order to validate the perceptual testing and training materials undertaken by the experimental group and, thus, provide baseline data. Table 23 presents the mean scores of the thirteen perception tasks performed by L1 participants.

Table 23

Mean (%) Scores of Correct Identification and Discrimination in the Perception Tests and Training Tasks performed by Native American English Speakers

Perception tests	**Mean and Standard Deviation**
Identification test	97.14 (*SD*=1.10)
Generalization test	96.59 (*SD*= 3.03)
Training tasks	**Mean and Standard Deviation**
Identification training task: seven-vowel set	97.15 (*SD*= 2.83)
Identification training task: /i/-/ɪ/	99.55 (*SD*= 1.18)
Identification training task: /ɛ/-/æ/	99.55 (*SD*= 1.18)
Identification training task: /u/-/ʊ/	99.11 (*SD*= 2.36)
Identification training task: /æ/-/ ʌ/	98.66 (*SD*= 2.46)
AX discrimination task: /i/-/ɪ/	100.00
AX discrimination task: /ɛ/-/æ/	98.08 (*SD*= 2.46)
AX discrimination task: /æ/-/ ʌ/	99.40 (*SD*= 1.58)
AX discrimination task: /u/-/ʊ/	99.40 (*SD*= 1.57)
Oddity discrimination task 1: seven-vowel set	97.32 (*SD*= 2.16)
Oddity discrimination task 2: seven-vowel set	99.11 (*SD*= 1.52)

The overall perceptual performance of the NSs was above 95% accurate in all tasks, ranging from 96.59% to 100%. Given that the mean scores of correct identification and discrimination were above 95%, and no systematic errors were found, no stimuli were excluded nor changes were made to the tasks after validation, as explained in the previous chapter (cf. Section 3.1.2).

4.1.2 Analysis by Vowel Contrast

4.1.2.1 Exploratory analysis of perception data

In order to examine identification percentage scores (i.e., interval data), exploratory analyses of the perception data obtained from the identification tests (pretest, posttest, delayed posttest, and generalization test) were performed to confirm if they met the assumptions[122] needed to perform parametric tests. The first assumption is that data should have normal distribution. Therefore, to verify whether data was normally distributed two tests were applied, namely the Kolmogorov-Smirnov (K-S) and the Shapiro-Wilk test.[123] In addition, we looked at the values of skewness and kurtosis because they also indicate if the distribution is approximately normal. If these values range between -1 and 1, we can assume that distribution of data is normal, and the further the value is from zero, the more likely it is that data are not normally distributed (Field, 2009, p. 139).

The results of the K-S and Shapiro-Wilk tests showed that the distribution of the vowel pair scores in the four identification tests did not differ significantly from a normal distribution. In the two cases (viz. posttest /u/-/ʊ/ and generalization test /i/-/ɪ/, in the experimental group) in which the K-S test revealed a non-normal distribution, the Shapiro-Wilk test showed no differences from normality.[124] The values of skewness and kurtosis also revealed that the distribution of data was approximately normal.

The second assumption that should be met in order to run parametric tests is homogeneity of variance.[125] Therefore, a Levene's test was run to test homogeneity of variance within both groups of participants. For almost all vowel contrasts there was homogeneity of variances, that is, the variances of vowel contrasts were equal for both the experimental and control groups ($p > .05$), but for the vowel contrast /u/-/ʊ/ in the posttest and for the vowel contrast /i/-/ɪ/ in the generalization test the variances were significantly different in the two groups, $F(1, 32)=10.01$, $p < .01$, and $F(1,32)=8.09$, $p < .05$.

Taking these results into account, parametric tests were applied to assess data, except for intergroup comparisons in the case of vowel contrasts, in which no homogeneity of variances was observed. For these two vowel pairs (/i/-/ɪ/ and /u/-/ʊ/), both non-parametric and parametric tests were applied. A .05 alpha value was used as a significance criterion for all the statistical tests run in this study.

---

[122] The assumptions of parametric tests are: (1) normally distributed data; (2) homogeneity of variance; (3) interval data; (4) independence (Field, 2009, p. 133).

[123] These tests compare the scores in the data sample with a normally distributed set of scores with the same mean and standard deviation. If the test is non-significant (p > .05) it means that the distribution of the data is not significantly different from a normal distribution (i.e., it is normal). If, however, the test is significant (p < .05) then the distribution is significantly different from a normal distribution (i.e., it is non-normal) (Field, 2009, p. 144).

[124] The Shapiro-Wilk test has more power than the K-S test to detect differences from normality, thus this test is sometimes significant when the K-S test is not (Field, 2009, p. 148).

[125] Homogeneity of variance is the assumption that the spread of scores is roughly equal in different groups of cases, or more generally that the spread of scores is roughly equal at different points on the predictor (independent) variable (Field, 2009, p. 152). If none of the four tests (*Based on Mean, Based on Median, Based on Median and with adjusted df, Based on trimmed mean*) has p >.05 there is homogeneity of variances. If the tests have p < .05 there is no homogeneity of variances.

4.1.2.2 Results of the perception pretest

As explained earlier (cf. Section 3.1.1), a preliminary analysis of the pretest results was performed as a screening procedure to verify whether the participants' perceptual performance reached native-like scores and to divide the cohort into two groups by matching vowel accuracy rates between them.

A range of 90-100% total correct percentage identification scores was set as standard and defined as a native-like perceptual accuracy rate. Although three of the participants, viz. E22, C23 and C29 (see Appendix AG), reached this standard of identification accuracy at least in one of the three vowel pairs, none of them met this criterion for all the three target contrasts. Therefore, no participant was withdrawn from the experiment. As briefly explained before, the cohort was divided into an experimental group (n=22) and a control group (n=12), according to their scores in the identification pretest, so that vowel identification accuracy rates would be matched between both groups. As a result, the ranges of identification accuracy scores were 53%-80% (mean 66%) for the experimental group and 52%-86% (mean 69%) for the control group (results are displayed in Table 7). In order to verify whether there were no differences between groups in the pretest, t tests for independent samples were run. The results are displayed in Table 25 and represented in a bar chart (see Figure 16). Moreover, to verify whether vowel duration affected participants' perception, a t test was run to compare the identification rates of segments /i/, /ae/, and /u/ with /ɪ/, /ɛ/ and /ʊ/ of both groups of participants in the pretest. The only significant difference was found in the higher ID score of /æ/ in relation to /ɛ/ (t= 3.72(33), p<.01).

4.1.2.3 Results of the perception pretest and posttest

The effects of perceptual training were assessed through a calculation of differences in the participants' identification scores from pretest to posttest. If the training was effective, the experimental group should reveal a significant increase in the identification scores of the posttest, while the control group should not exhibit such improvement. The mean percentage identification scores in the two perception tests are presented in Table 24 together with the percentage of improvement[126] from the pretest to the posttest, and the results of the *t* test for paired samples.

Table 24

T Test Results of Within-group Comparisons Between the Pretest and Posttest

Group	**Vowel Contrast**	**Pretest** (*n*=34) **Mean (SD)**	**Posttest** (*n*=34) **Mean (SD)**	**% Intragroup Difference**	***t* (df)**
EG	**/ɛ/-/æ/**	66.59 (7.98)	81. 44 (8.67)	14.85	-8.11(21) ***
	/i/-/ɪ/	56.67 (17.41)	83.71 (13.68)	27.04	-6.46(21) ***
	/u/-/ʊ/	63.79 (13.29)	77.42 (12.48)	13.63	-5.07(21) ***
	/æ/-/ʌ/	80.08 (9.00)	85.76 (5.01)	5.68	-3.53(21) **
	/ʊ/-/ʌ/	75.61 (6.35)	83.12 (7.00)	7.51	-4.88(21) ***
CG	**/ɛ/-/æ/**	70.69 (9.60)	71. 80 (8.67)	1.11	-.39(11)
	/i/-/ɪ/	66.53 (15.91)	69.17 (13.68)	2.64	-1.14(11)
	/u/-/ʊ/	62.78 (18.35)	59.58 (12.48)	- 3.1	.68(11)
	/æ/-/ʌ/	80.00 (8.65)	80.56 (9.22)	.56	-.24(11)
	/ʊ/-/ʌ/	71.67 (13.73)	73.47 (12.78)	1.8	-.40(11)

Note. EG=Experimental group; CG=Control group; SD=Standard Deviation; *t* =result of *the t test for paired samples*; (df)=degrees of freedom; levels of significance: * $p < .05$; ** $p < .01$; *** $p < .001$.

The results of the *t* test for paired samples revealed that training had a significant effect on the perceptual ability of experimental participants, because the identification of the three target vowel pairs improved significantly from the pretest to the posttest (see Figure 14), similarly to the labeling of the other two vowel non-target pairs with the distractor vowel. The highest improvement was observed for the high front vowel pair (27.04%), followed by the mid-low front pair (14.85%) and then the high back vowel contrast (13.63%). The participants of the control group did not improve the perception of any of the vowel contrasts (see Figure 15). Their mean scores of accurate identification of the five vowel contrasts after the training with consonants were similar to their scores in the pretest phase. Therefore, no effect of training was observed for the controls.

[126] The percentage of improvement was calculated by subtracting the mean % of the pretest scores from the posttest scores.

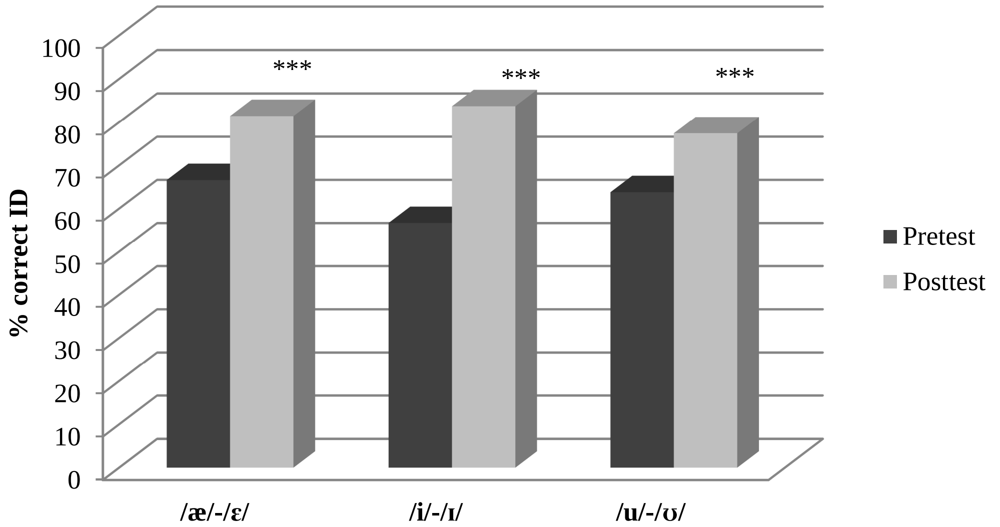


Figure 14. Mean % of correct identification scores of the three vowel pairs in the pretest and posttest for the trained group.

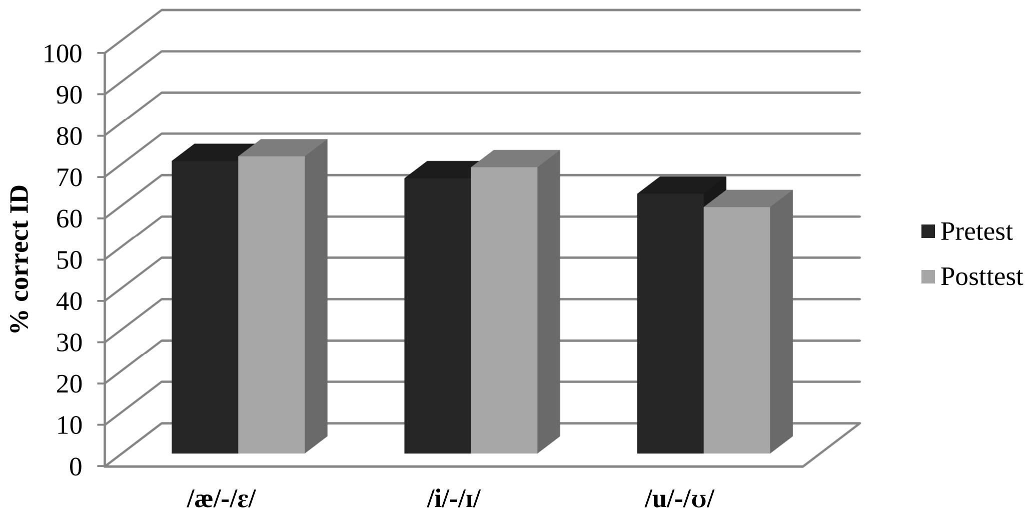


Figure 15. Mean % of correct identification scores of the three vowel pairs in the pretest and posttest for the control group.

In order to verify whether the two groups performed differently in the two test phases, namely before and immediately after training, a comparison was made between the experimental and the control groups' mean scores for each vowel contrast with a *t* test for independent samples (see Table 25). If there were training effects, significant differences in the performances of both groups should be found in the posttest, but not in the pretest.

Table 25

T Test Results of Between-group Comparisons in the Pretest and Posttest

ID test	Vowel Contrast	EG (*n*=22) Mean (SD)	CG (*n*=12) Mean (SD)	% Intergroup difference	*t* (df)
Pretest	**/æ/-/ɛ/**	66.59 (7.98)	70.69 (9.60)	4.1	-1.33(32)
	/i/-/ɪ/	56.67 (17.41)	66.53 (15.91)	9.86	-1.62(32)
	/u/-/ʊ/	63.79 (13.30)	62.78 (18.34)	1.01	.18(32)
	/æ/-/ʌ/	80.08 (9.00)	80.00 (8.65)	.08	.02(32)
	/ʊ/-/ʌ/	75.61 (6.35	71.67 (13.73)	3.94	1.15(32)
Posttest	**/æ/-/ɛ/**	81.44 (8.67)	71.81 (10.45)	9.63	2.88(32) **
	/i/-/ɪ/	83.71 (13.68)	69.17 (16.15)	14.54	2.78(32) **
	/u/-/ʊ/	77.42 (12.48)	59.58 (23.56)	17.84	2.90(32) **
	/æ/-/ʌ/	85.76 (5.01)	80.56 (9.22)	5.2	2.15(32) *
	/ʊ/-/ʌ/	83.12 (7.00)	73.47 (12.78)	9.65	2.86(32) **

Note. EG=Experimental group; CG=Control group; SD=Standard Deviation; *t*=result of *the t test for independent samples*; (df)=degrees of freedom; levels of significance: * $p < .05$; ** $p < .01$; *** $p < .001$.

As expected, there were no significant differences between the perceptual performance of the experimental and the control groups in the pretest (see Figure 16). However, significant differences were found between both groups in the posttest. The target vowel contrasts were significantly better identified by the trained group than by the control group in the posttest (see Figure 17).

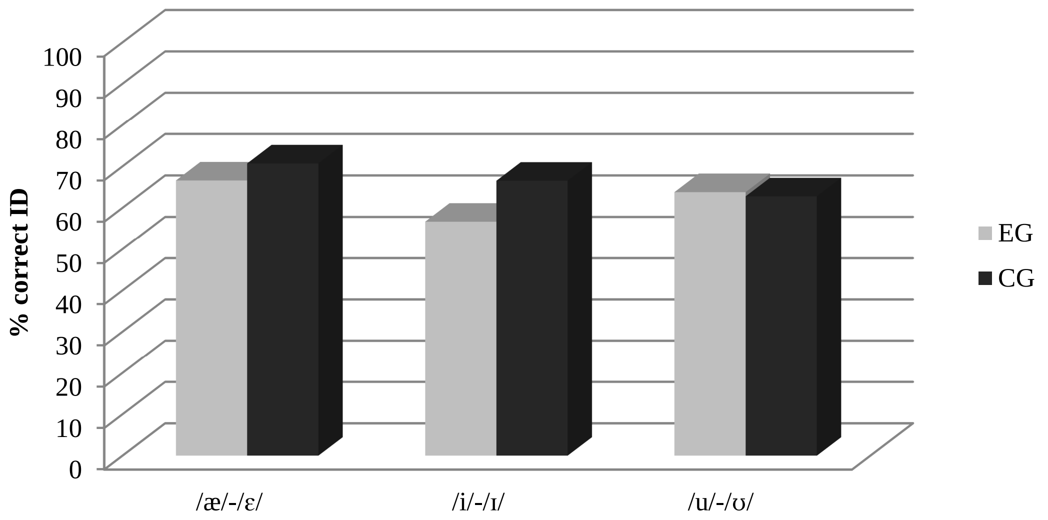


Figure 16. Mean % of correct identification scores of the three vowel contrasts in the pretest for both groups.

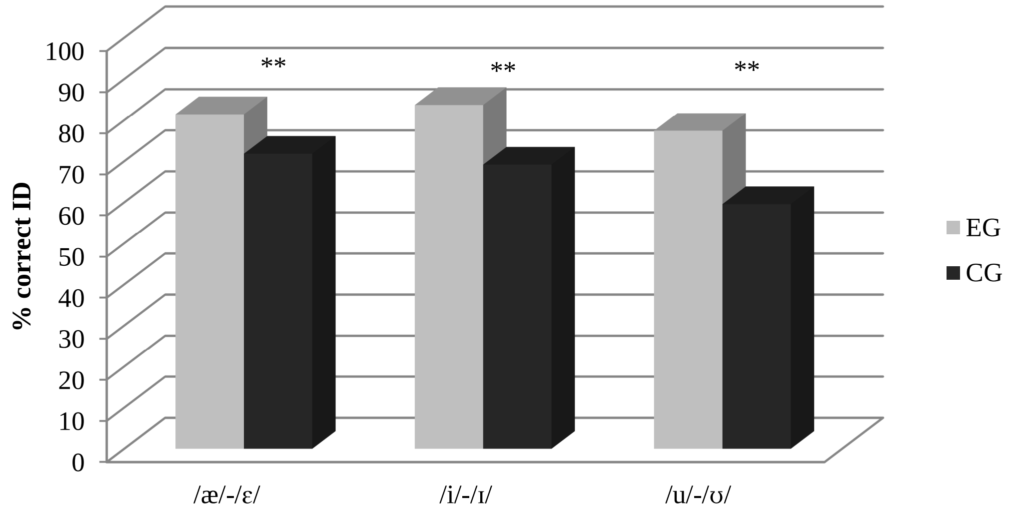


Figure 17. Mean % of correct identification scores of the three vowel contrasts in the posttest for both groups.

A significant effect of training in the identification of the vowel contrasts was found in the trained group but not in the control group. The intergroup identification rate differences were greater for /u/-/ʊ/ (17.84%), followed by /i/-/ɪ/ (14.54%), and then /æ/-/ɛ/ (9.63%). When analyzing the perception scores of the vowel pairs with the distractor vowel, a greater between-group difference was found for /ʊ/-/ʌ/ (9.65%) than for /æ/-/ʌ/ (5.2%).

Since for the vowel contrast /u/-/ʊ/ in the posttest no homogeneity of variances was found between the two groups, $F(1, 32) = 10.01$, $p < .01$, a non-parametric test was also run for this vowel pair. The result of the Mann-Whitney test was identical to the *t* test result. The participants in the trained group outperformed the control group, that is, they had significantly higher perceptual scores than the control group in the identification of /u/-/ʊ/ in the posttest, $U = 69.50$, $p <.05$.

4.1.2.3.1 Analysis of new speaker and new word

Although a generalization test was designed to assess transfer of learning to new talkers and to new syllabically more complex words (not used either in the perception tests or training tasks), the identification test (pretest, posttest, delayed posttest) included both trained and untrained stimuli. There were four types of stimuli, viz. stimuli with trained context and talker (TC-TT), stimuli with untrained context and talker (UC-UT), stimuli with untrained context and trained talker (UC-TT), and stimuli with trained context and untrained talker (TC-UT). Therefore, although no statistical tests were run to attest transfer of learning in the identification posttest, descriptive statistical measures are provided in Table 26 to observe the trained group's performance on new talkers and new words in comparison with their performance on trained words and familiar talkers. Following a procedure similar to Wang's (2008), the trainees' total identification scores of the seven AE vowels in the pre- and posttest were divided into four sets of stimuli.

Table 26

Percentage Correct Scores in the Posttest by Stimulus Type

			UC-UT *n*=968	**TC-UT** *n*=616	**UC-TT** *n*=1760	**TC-TT** *n*=1276	**Total** *N*=4620
EG	**Pretest**	count	610	373	1183	864	3030
		%	63.02	60.55	67.22	67.71	65.60
	Posttest	count	764	495	1465	1070	3794
		%	78.93	80.36	83.24	83.86	82.14
			UC-UT *n*=528	**TC-UT** *n*=336	**UC-TT** *n*=960	**TC-TT** *n*=696	**Total** *N*=2520
CG	**Pretest**	count	353	218	669	488	1728
		%	66.86	64.88	69.69	70.11	68.57
	Posttest	count	360	218	677	498	1753
		%	68.18	64.88	70.52	71.55	69.56

Note. EG=Experimental group; CG=Control group; UC-UT=untrained context-untrained talker; TC-UT=trained context-untrained talker; UC-TT=untrained context, trained talker; TC-TT=trained context-trained talker.

The percentages of correct ID scores per stimulus in the posttest were very consistent among the four stimulus type, ranging between 78.93% in the identification of tokens produced by new talkers in new contexts and 83.86% in the identification of stimuli embedded in familiar contexts and talkers (see Figure 18).

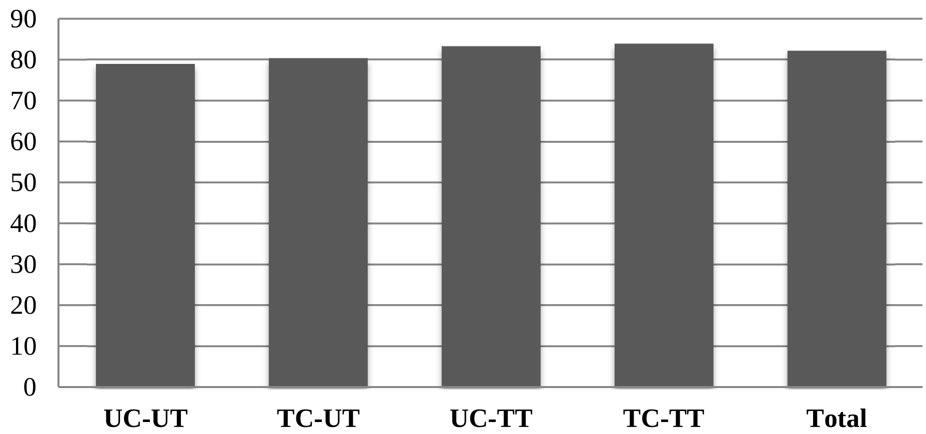


Figure 18. Percentage of correct ID scores in the posttest by stimulus type.

Even though there was a slight difference among the four conditions, with the TC-TT stimulus type having the highest ID scores, and the UC-UT the lowest scores, perceptual learning transferred to new words and new talkers. If generalization did not occur, the ID scores for the UC-UT stimuli would have been substantially lower. Therefore, given that ID scores do not vary much across stimulus-type, we can assume that improvement of trainees' perceptual ability was generalized to untrained tokens.

4.1.2.4 Results of the perception posttest and generalization test

To examine whether there was generalization of perceptual learning to new tokens and new talkers, comparative analyses were undertaken between the posttest and the generalization test's mean % scores of correct identification. *T* tests for paired samples were run to compare identification scores between the two tests (see Table 27). The stimuli of the generalization test consisted of new monosyllabic words produced by novel native American English speakers, whereas the posttest included both trained and untrained stimuli, as aforementioned.

Table 27

T Test Results of Within-group Comparisons Between the Posttest and Generalization Test

Group	Vowel Contrast	Posttest Mean (SD)	Generalization test Mean (SD)	% Intragroup Difference	*t* (df)
EG (*n*=22)	**/æ/-/ɛ/**	81.44 (8.67)	90.28 (6.79)	8.84	- 5.93(21***
	/i/-/ɪ/	83.71(13.68)	89.27 (8.28)	5.56	-3.24(21) **
	/u/-/ʊ/	77.42 (12.48)	67.42 (16.89)	-10	4.00(21) **
	/æ/-/ʌ/	85.76 (5.01)	75.63 (11.81)	-10.13	3.92(21) **
	/ʊ/-/ʌ/	83.12 (7.00)	70.08 (15.12)	-13.04	4.30(21) ***
CG (*n*=12)	**/æ/-/ɛ/**	71.81 (10.45)	72.45 (12.50)	.064	- .25(11)
	/i/-/ɪ/	69.17 (16.15)	68.06 (16.35)	-1.11	.32(11)
	/u/-/ʊ/	59.58 (23.56)	57.17 (20.36)	-2.41	.52(11)
	/æ/-/ʌ/	80.56 (9.22)	66.90 (15.96)	-13.66	4.15(11) **
	/ʊ/-/ʌ/	73.47 (12.78)	58.56 (16.43)	-14.91	4.01(11) **

Note. EG=Experimental group; CG=Control group; SD=Standard Deviation; *t* = result of *the t test for independent samples*; (df)=degrees of freedom; levels of significance: * $p < .05$; ** $p < .01$; *** $p < .001$.

The results of the *t* test demonstrated that the experimental participants had significantly better results in the identification of the two front vowel pairs (/i/-/ɪ/, /æ/-/ɛ/) in the generalization test than in the posttest, whereas the scores for the high back contrast were significantly worse in the generalization test than in the posttest. This seems to indicate that perceptual learning of the vowel contrasts /i/-/ɪ/ and /æ/-/ɛ/ was transferred to new tokens and new talkers but the same did not happen for /u/-/ʊ/ (see Figure 19). For the control group, no perceptual differences were found between the posttest and the generalization test. Their perception of the three vowel contrasts in the posttest was as accurate as in the generalization test (see Figure 20). Moreover, the two groups of EFL learners had significantly lower scores in the identification of the two vowel pairs with the distractor /ʌ/ in the generalization test than in the posttest.

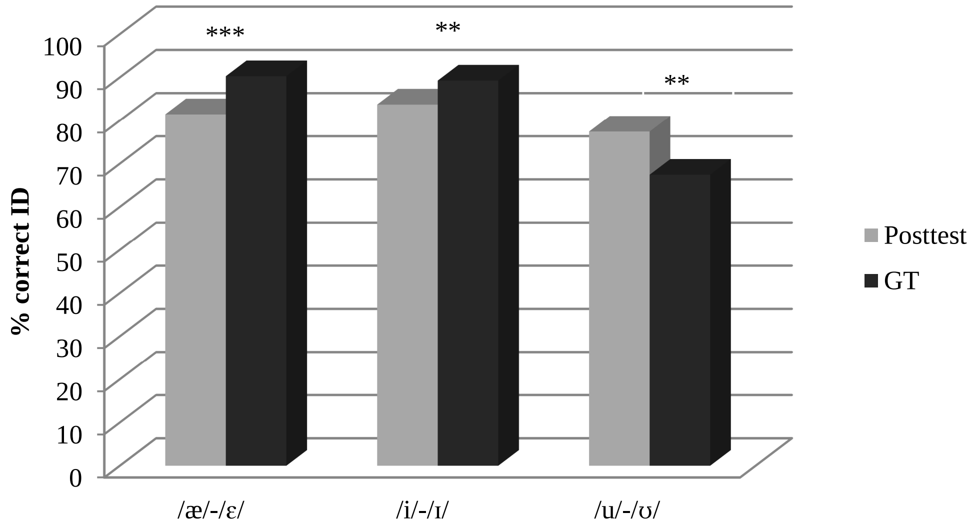


Figure 19. Mean % of correct identification scores of the three vowel pairs in the posttest and generalization test for the trained group.

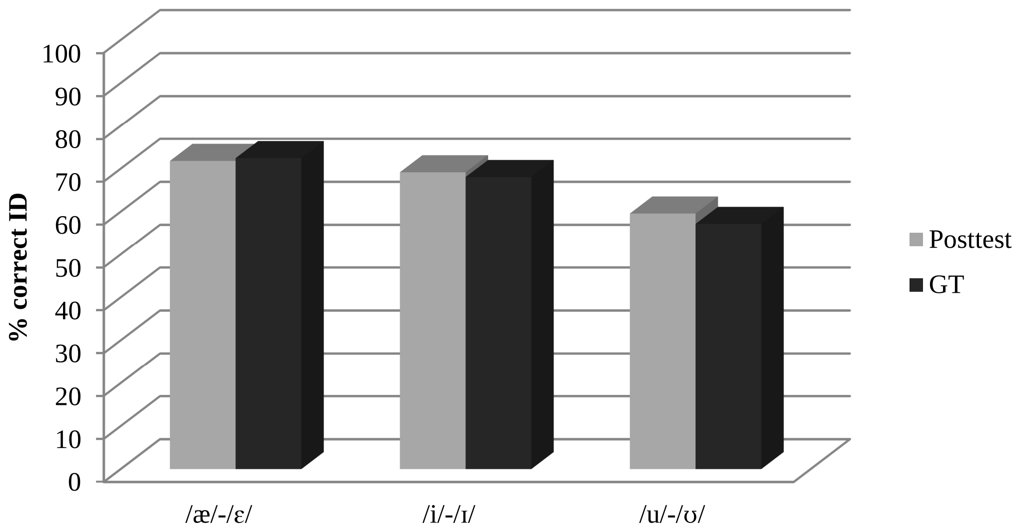


Figure 20. Mean % of correct identification scores of the three vowel pairs in the posttest and generalization test for the control group.

Furthermore, perceptual performance in the identification of the three target vowel pairs was consistent between tests. The two front vowel pairs were more accurately identified than the back vowel contrast in both tests.

Between-group comparisons were performed to verify whether significant different perceptual performances found in the posttest between trainees and controls would also be observed in the generalization test, by running a *t* test for independent samples. The results of the statistical test and the mean identification scores are presented in Table 28.

Table 28

T Test Results of Between-group Comparisons in the Generalization Test

ID test	Vowel Contrast	EG (*n*=22) Mean (SD)	CG (*n*=12) Mean (SD)	*t* (df)
Generalization Test	**/æ/-/ɛ/**	90.28 (6.79)	72.45 (12.50)	5.42(32) ***
	/i/-/ɪ/	89.27 (8.28)	68.06 (16.35)	5.05(32) ***
	/u/-/ʊ/	67.42 (16.89)	57.17 (20.36)	1.57(32)
	/æ/-/ʌ/	75.63 (11.81)	66.90 (15.96)	1.82(32) †
	/ʊ/-/ʌ/	70.08 (15.12)	58.56 (16.43)	2.06(32) *

Note. EG= Experimental group; CG=Control group; SD=Standard Deviation; *t* = result of *the t test for independent samples*; (df)=degrees of freedom; levels of significance: † *p*= <.1.; * *p* < .05; ** *p* < .01; *** *p* < .001;

The two groups behaved differently in the generalization test, with the trained group outperforming the control group in the identification of the two front vowel pairs. Although in the generalization test the trainees had a considerable lower accuracy identification score of the back vowel pair than in the posttest (a difference of -10%, cf. Table 27), it was still higher than the controls' score (Figure 21).

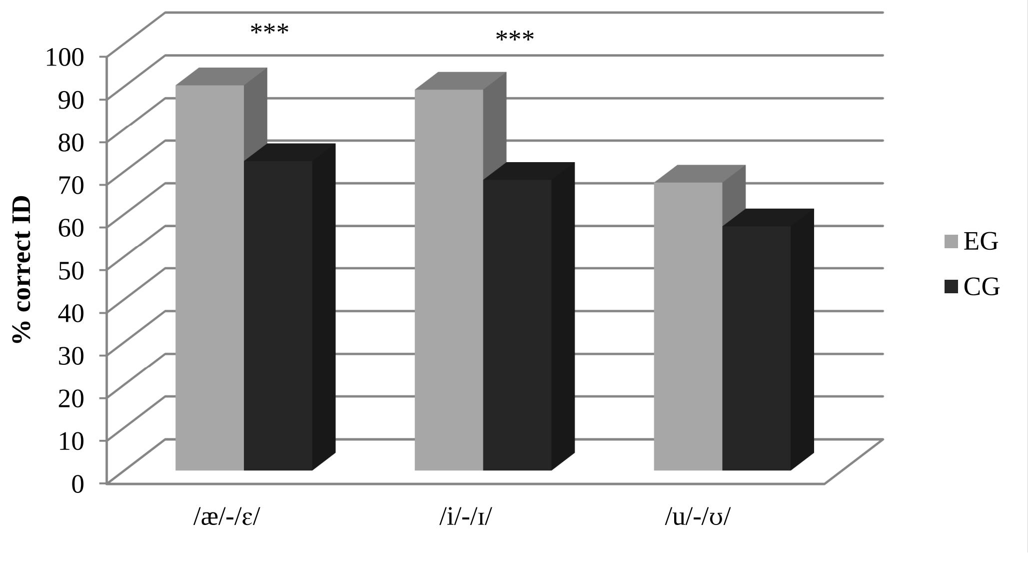


Figure 21. Mean % of correct identification scores of the three vowel contrasts in the generalization test for both groups.

The trainees also had higher scores in the identification of the contrasts with the distractor vowel. The pair /ʊ/-/ʌ/ was significantly more accurately labeled, and /æ/-/ʌ/ was marginally better identified by the trainees than by the controls. Moreover, the results of the trainees in the generalization test are consistent with the identification posttest in that the back vowel contrast was the most difficult to learn.

Given that homogeneity of variances was not found in the identification of /i/-/ɪ/ in the generalization test, a Mann Whitney test was also run for this vowel contrast. The result of the non-parametric test was similar to the *t* test, insofar as a significant difference between the experimental and the control group was found (U=52.50, p <.01).

4.1.2.4.1 Analysis of the perception tests' stimuli

Six native American English speakers (three women and three men) produced the stimuli used in the three identification tests, and other five native AE talkers (three men and two women) produced the tokens included in the generalization test. The vowels of the tokens included in the identification and generalization tests were acoustically analyzed to assess whether both sets of stimuli were comparable. Mean duration, F1 and F2 values are presented in Table 29.

Table 29

Mean duration, F1 and F2 values of AE vowels

	Vowel		**/i/**	**/ɪ/**	**/ɛ/**	**/æ/**	**/ʊ/**	**/u/**	**/ʌ/**
ID_NSs	**Duration (ms)**	**F**	146 (11)	119 (15)	108 (17)	143 (15)	140 (24)	168 (07)	117 (08)
		M	146 (20)	106 (23)	128 (22)	170 (32)	118 (21)	140 (26)	119 (02)
	F1 (Hz)	**F**	267 (117)	537 (71)	805 (41)	1079 (159)	646 (29)	396 (31)	814 (67)
		M	269 (20)	424 (21)	583 (26)	709 (13)	472 (38)	315 (26)	586 (21)
	F2 (Hz)	**F**	2987 (117)	2370 (84)	2097 (60)	1775 (144)	1487 (140)	1740 (207)	1645 (51)
		M	2312 (122)	1882 (100)	1776 (78)	1733 (72)	1116 (248)	1404 (193)	1277 (199)

GT_NSs	Duration (ms)	F	227 (38)	136 (17)	186 (38)	-	143 (07)	214 (36)	199 (43)
		M	186 (50)	124	148 (11)	172 (49)	132 (25)	157 (20)	167
	F1 (Hz)	F	356 (22)	503 (29)	708 (22)	-	591 (39)	329 (3)	835 (28)
		M	263 (23)	478	578 (18)	725 (63)	402 (34)	302 (9)	634
	F2 (Hz)	F	2901 (91)	2236 (135)	2179 (139)	-	1114 2	1589 (78)	1394 (71)
		M	2327 (112)	1697	1641 (9)	1637 (91)	1034 (189)	1474 (85)	1114

Note. ID=identification test; GT=generalization test; NSs=native speakers.

Due to the limited number of productions per vowel in each test, no statistical test could be run to compare productions between groups of talkers. However, the normalized[127] F1 and F2 values of both groups of talkers, plotted in Figure 22, indicate that their productions were similar (see Table 72).

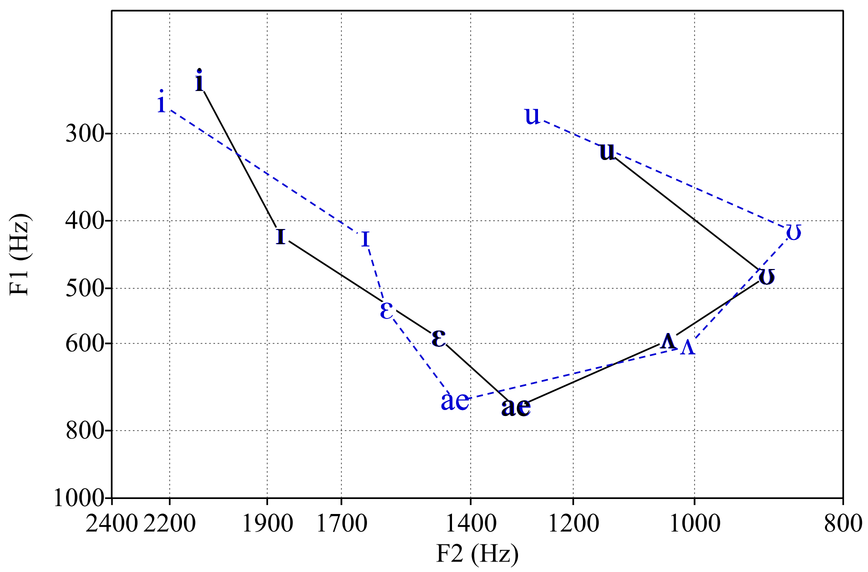


Figure 22. The vowel spaces of the two groups of NSs. Solid line=ID test talkers; dashed line=GT talkers.

Minor differences in terms of duration and formant values may be explained by phonetic context variation. In the identification test, vowels were mainly flanked by

[127] The vowels were normalized using the Lobanov method, available at NORM (http://ncslaap.lib.ncsu.edu/tools/norm/), in order to factor out vocal tract differences between male and female speakers.

voiceless consonants, whereas the phonetic environments in which vowels of the generalization test were embedded were more varied (see Section 3.2.2.2).

We want to emphasize that, although acoustic measurements were not complemented with statistical analyses, the point to be made is that production data between the two groups of AE native talkers were not considerably dissimilar, thus the posttest and generalization test results can be comparable.

4.1.2.5 Results of the perception pretest, posttest and delayed posttest

As mentioned in the previous chapter, the two groups suffered attrition from the posttest to the delayed posttest. There were three dropouts in the experimental group, and one dropout in the control group; thus, statistical analyses to verify effects of perceptual training in the three test phases were performed with the following number of participants: experimental group (*n*=19) and control group (*n*=11). To examine the effect of training in the three test phases (pretest, posttest and delayed posttest), repeated measures analyses of variance were run with a within-subjects design. The results of the ANOVA are listed in Table 30.

Table 30

ANOVA Test Results of the Experimental and Control Groups when Measuring the Effect of Training

Group	Vowel contrast	Pretest Mean (SD)	Posttest Mean (SD)	Delayed Posttest Mean (SD)	*F* (df)[128]
EG (*n*= 19)	**/æ/-/ɛ/**	66.40 (7.10)	81. 32 (8.74)	83.77 (9.14)	35.65(2,36) ***
	/i/-/ɪ/	55.79 (18.50)	85.00 (12.18)	87.63(11.51)	45.18(1.14, 20.57) ***
	/u/-/ʊ/	64.74 (13.38)	78.60 (11.96)	84.12 (10.01)	28.50(1.36, 24.49) ***
	/æ/-/ʌ/	79.83 (8.57)	86.23 (4.19)	88.60 (4.76)	12.69(2,36) ***
	/ʊ/-/ʌ/	75.88 (6.10)	82.72 (7.35)	83.25 (7.19)	13.83(2,32) ***

[128] If Mauchly's test statistic is significant (< .05) one can conclude that there are significant differences between the variances of differences and, therefore, the condition of sphericity is not met. However, if Mauchly's test statistic is non-significant i.e., >.05, then it is reasonable to conclude that the variances of differences are not significantly different. If data violate the sphericity assumption one of the corrections that can be applied is the Greenhouse-Geisser correction (Field, 2009, p. 460). In the cases in which the sphericity assumption was met (Mauchly's test of sphericity was significant), the sphericity results (F ratio) and degrees of freedom were reported, whereas in the cases the assumption was not met, the Greenhouse-Geisser correction was applied, and two degrees of freedom (*df* for the effect of the model and *df* for the residuals of the model) were reported along with the test result.

	/æ/-/ɛ/	71.67 (9.43)	72.58 (10.60)	73.33 (9.13)	.21(2.20)
CG	**/i/-/ɪ/**	68.03 (15.77)	70.61 (16.11)	66.82 (19.13)	.92(2.20)
(*n*= 11)	**/u/-/ʊ/**	62.58 (19.22)	58.33 (24.29)	63.26 (19.05)	.73(2.20)
	/æ/-/ʌ/	80.76 (8.64)	81.21 (9.37)	81.21 (9.25)	.03(1.30,13.03)
	/ʊ/-/ʌ/	71.52 (14.39)	73.03 (13.31)	71.21 (11.67)	.11(1.11, 11.05)

Note. EG=Experimental group; CG=Control group; SD=Standard Deviation; *F*= result of ANOVA; (df) =degrees of freedom; levels of significance: * $p < .05$; ** $p < .01$; *** $p < .001$.

The high variability perceptual training had a significant effect on the identification of the three target vowels for the experimental group, but no effect was found for the control group (see Figures 23 and 24). Significant differences were found in the identification of the three target vowel contrasts in relation to the moment they were tested (pretest, posttest and delayed posttest) for the experimental group, but not for the control group. In addition, a significant effect of training was also found on the perception of the two contrasts with the distractor vowel (/ʊ/-/ʌ/ and /æ/-/ʌ/) for the trainees but not for the controls. The identification of the three vowel contrasts improved significantly after training for the experimental group, and learning remained two months after the high variability phonetic training (HVPT) was over. The control group showed no perceptual improvement in the identification of the three vowel contrasts after the training with consonants.

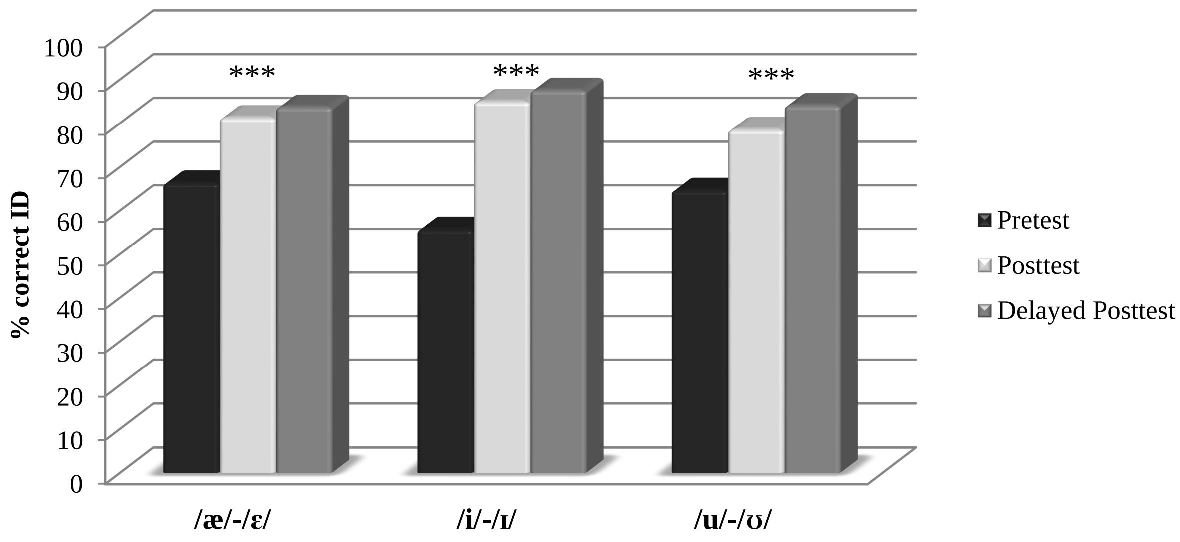


Figure 23. Mean % of correct identification scores of the three vowel contrasts in the pretest, posttest and delayed posttest for the experimental group.

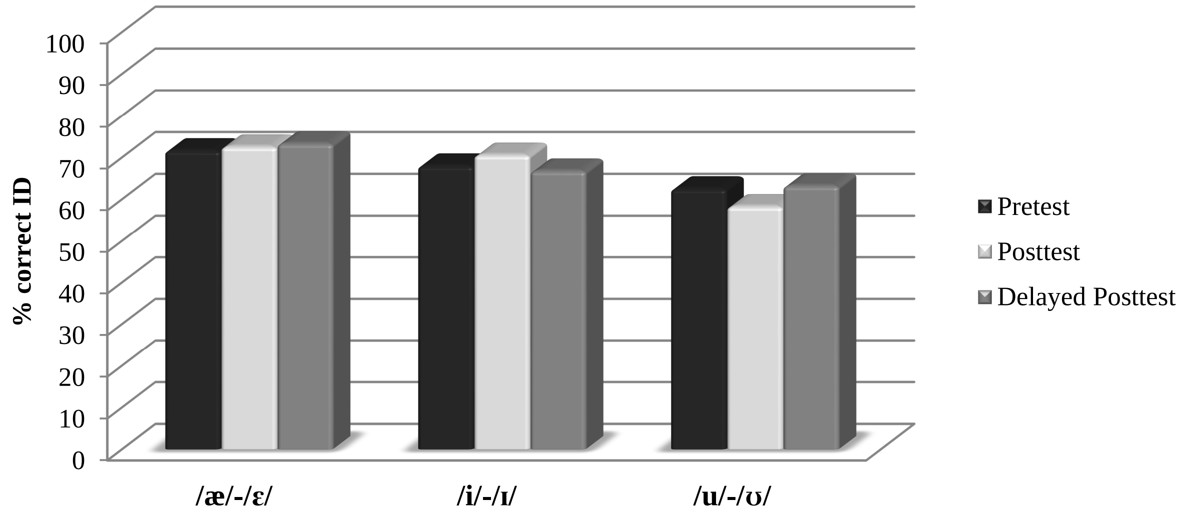


Figure 24. Mean % of correct identification scores of the three vowel contrasts in the pretest, posttest and delayed posttest for the control group.

Post hoc tests that consisted of Bonferroni pairwise comparisons were run for the two groups. For the experimental group, results of the post hoc tests revealed significant differences between pre- and posttests, and between pretest and delayed posttest identification scores. This indicates that the identification of the three target vowel pairs was significantly better immediately after training and two months after the completion of the training program. Moreover, the identification accuracy of the back vowel pair also improved significantly from the posttest to the delayed posttest. No significant differences were found for the control group (see Table 31). In relation to vowel pairs /ʊ/-/ʌ/ and /æ/-/ʌ/, significant differences in ID scores were found both between the pretest and posttest, and between the pretest and delayed posttest for the trainees, whereas these differences were not found for the controls.

Table 31

Pairwise Comparisons of Percentage ID scores

Group	Vowel Contrast	Pretest *vs.* Posttest	Posttest *vs.* Delayed Posttest	Pretest *vs.* Delayed Posttest
	/æ/-/ɛ/	***	ns	***
	/i/-/ɪ/	***	ns	***
Experimental	/u/-/ʊ/	***	**	***
	/æ/-/ʌ/	**	ns	**
	/ʊ/-/ʌ/	**	ns	***

Control	**/æ/-/ɛ/**	ns	ns	ns
	/i/-/ɪ/	ns	ns	ns
	/u/-/ʊ/	ns	ns	ns
	/æ/-/ʌ/	ns	ns	ns
	/ʊ/-/ʌ/	ns	ns	ns

Note. ns=non-significant; values of significance * < .05; ** < .01; *** < .001.

To show the differences in perceptual performance in the three test phases, a calculation of percentage of improvement was made. The results are presented in Table 32.

Table 32

Pairwise Comparisons of Percentage ID Score Differences

Group	**Vowel Contrast**	**% Difference Pretest *vs.* Posttest**	**% Difference Posttest *vs.* Delayed Posttest**	**% Difference Pretest *vs.* Delayed Posttest**
Experimental	**/æ/-/ɛ/**	14.92	2.45	17.37
	/i/-/ɪ/	29.21	2.63	31.84
	/u/-/ʊ/	13.86	5.52	19.38
	/æ/-/ʌ/	6.40	2.37	8.77
	/ʊ/-/ʌ/	6.84	.53	7.37
Control	**/æ/-/ɛ/**	.91	.75	1.66
	/i/-/ɪ/	2.58	-3.79	-1.21
	/u/-/ʊ/	-4.25	4.93	.68
	/æ/-/ʌ/	.45	.0	.45
	/ʊ/-/ʌ/	1.51	-1.82	-.31

Table 32 shows that improvement of both groups in the three testing phases differed significantly. The higher rates of trainees' improvement can be found in the comparison between pretest and delayed posttest, which indicates that this group did not only maintain perceptual learning eight weeks after training was over, but they still improved slightly. The post hoc test results revealed that between posttest and delayed posttest the identification of the back vowel pair improved significantly (p <.01, (5.52%) (cf. Table 31). Although there was some minor variation in terms of percentage of improvement, no significant changes were found in the controls' perceptual ability.

In the previous sections, intergroup analyses were run to verify whether there were differences between trainees and controls, before and immediately after training completion. To examine whether perceptual performance between both groups still differed eight weeks after training was over, a *t* test for independent samples was run. The results are presented in Table 33.

Table 33

T Test Results of Between-group Comparisons in the Delayed Posttest

Test	**Vowel Contrast**	**Experimental Group** (*n*=19) **Mean (SD)**	**Control group** (*n*=11) **Mean (SD)**	***t*(df)**
Delayed posttest	**/æ/-/ɛ/**	83.77 (9.14)	73.33 (9.13)	3.01(28) **
	/i/-/ɪ/	87.63 (11.51)	66.82 (19.13)	3.74(28) ***
	/u/-/ʊ/	84.12 (10.01)	63.26 (19.05)	3.95(28) ***
	/æ/-/ʌ/	88.60 (4.76)	81.21 (9.25)	2.90(28) **
	/ʊ/-/ʌ/	83.25 (7.19)	71.21 (11.67)	3.51(28) **

Note. EG=Experimental group; CG=Control group; SD=Standard Deviation; *t*=result of the *t* test; (df) =degrees of freedom; levels of significance: * $p < .05$; ** $p < .01$; *** $p < .001$.

The results of the *t* test revealed that all vowel contrasts were significantly better identified by trainees than by controls in the delayed posttest. In sum, two months after training was completed, the trained group continued to outperform the control group (see Figure 25).

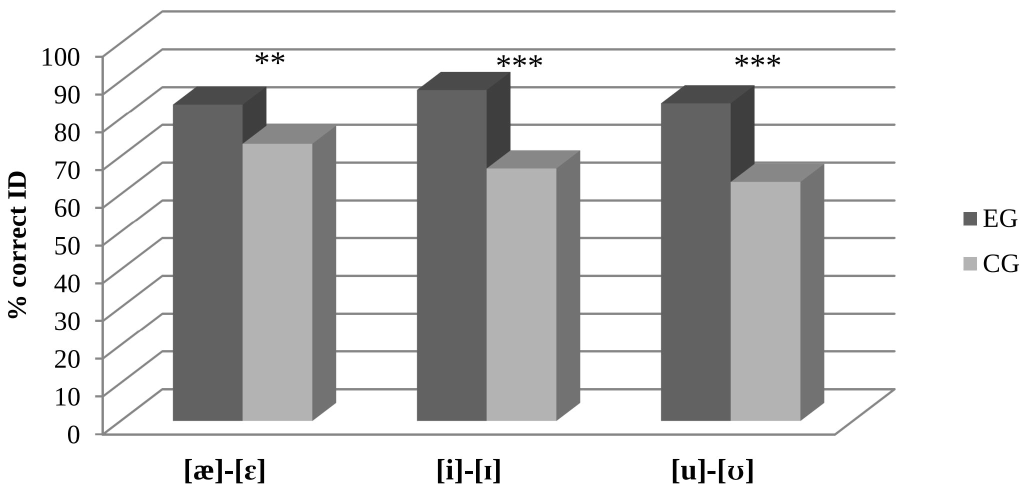


Figure 25. Mean % of correct identification scores of the three vowel pairs in the delayed posttest for the experimental and control groups.

Since for vowel contrast /i/-/ɪ/ no homogeneity of variances was found in the two groups, $F(1,32) = 8.09$, $p < .05$, a non-parametric test was also run for this vowel pair . The result of the Mann-Whitney test indicated that there was a significant difference between the perceptual performance of the experimental and the control groups, U=39.00, p <.01. The experimental group had significantly better perceptual scores in the identification of /i/-/ɪ/ in the delayed posttest than the control group.

In order to verify whether the two groups performed differently in the tests after training, we calculated the percentage of improvement of the control group between the pretest and the posttest, and the posttest and the delayed posttest and interpreted it as task effect. The control group undertook a similar training program focused on English consonants, therefore, we assumed that any improvement observed would reflect task repetition effects, rather than vowel learning effects. The mean percentage of the task effect observed between pretest and posttest was 0.18%, and between the pretest and the delayed posttest it was 1.13%. Therefore, we decided not to run further analysis because the percentage difference was very small.

4.1.2.6 Response time

In order to complement the previous analyses, we compared the measures of response time (RT) among the three test phases to verify whether there was a decrease of time in the identification of the target vowel contrasts. Departing from the assumption that poor exemplars of a vowel category would take longer to identify than good exemplars, we hypothesized that RTs would decrease after training, and shorter RTs would still be observed two months after the training sessions were over. We also hypothesized that RTs would be shorter in the experimental group than in the control group.

Therefore, we first calculated the mean average of RTs (measured in seconds) for each vowel pair per participant, and then we ran normality tests to check the distribution of data and homogeneity of variances. The K-S and the Shapiro-Wilk tests revealed that the following vowel contrasts did not have a normal distribution: /æ/-/ɛ/ in the pre- and posttests, /i/-/ɪ/ in the posttest, and /u/-/ʊ/ in the delayed posttest, in the experimental group; and /æ/-/ɛ/ in the posttest, in the control group. Therefore, following the procedure suggested by Fife-Schaw (2006), non-parametric tests were run together with parametric tests for these vowel pairs, and results were compared. Given that conclusions from both sets of tests were the same in all cases, we opted to report the results of the parametric tests.

To examine whether there was a training effect on RTs, a comparison among the three tests was carried out with a repeated measures ANOVA. The results are displayed in Table 34.

Table 34

ANOVA Results of the Within-group RT Comparisons

Group	**Vowel contrast**	**Pretest RT (s) Mean (SD)**	**Posttest RT (s) Mean (SD)**	**Delayed Posttest RT (s) Mean (SD)**	***F* (df)**
EG (*n*=19)	**/æ/-/ɛ/**	2.59 (1.69)	1.41 (1.08)	.66 (.60)	15.20(1.27, 22.80) ***
	/i/-/ɪ/	1.79 (1.07)	1.52 (1.36)	.83 (.68)	6.63(2, 36) **
	/u/-/ʊ/	2.93 (1.50)	2.17 (1.32)	1.43 (.97)	9.78(1.27,22.89) **
	/æ/-/ʌ/	2.27 (1.52)	1.25 (.82)	.66 (.37)	13.49(1.38,24.83) ***
	/ʊ/-/ʌ/	2.50 (1.59)	1.60 (.96)	1.08 (.63)	13.13(1.30, 23.46) **
CG (*n*=11)	**/æ/-/ɛ/**	2.79 (.72)	2.25 (1.31)	1.48 (1.06)	10.22(2, 20) **
	/i/-/ɪ/	3.12 (1.44)	1.96 (1.15)	1.28 (.71)	.92(2, 20) ***
	/u/-/ʊ/	3.75 (1.24)	3.10 (1.23)	1.79 (1.31)	15.80(2, 20) ***
	/æ/-/ʌ/	3.37 (1.15)	2.77 (1.26)	1.70 (1.13)	20.04(2,20) ***
	/ʊ/-/ʌ/	2.72 (1.00)	2.35 (1.38)	1.37 (.86)	11.00(2,20) **

Note. EG=Experimental group; CG=Control group; SD=Standard Deviation; *F*= result of the ANOVA; (df)=degrees of freedom; levels of significance: * $p < .05$; ** $p < .01$; *** $p < .001$.

Table 34 shows that there was a significant decrease of the mean response times in the identification of all the vowel pairs both for the experimental and the control groups. In other words, there was a significant effect of perceptual training on the RTs of both groups, regardless of being vowel or consonant-centered.

Table 35

Pairwise Comparisons of Bonferroni of Mean Response Times

Group	Vowel Contrast	Pretest *vs.* Posttest	Posttest *vs.* Delayed Posttest	Pretest *vs.* Delayed Posttest
EG	/ɛ/-/æ/	*	**	***
	/i/-/ɪ/	ns	*	**
	/u/-/ʊ/	ns	**	***
	/æ/-/ʌ/	*	*	**
	/ʊ/-/ʌ/	*	**	**
CG	/ɛ/-/æ/	ns	ns	**
	/i/-/ɪ/	*	*	**
	/u/-/ʊ/	ns	*	***
	/æ/-/ʌ/	ns	*	***
	/ʊ/-/ʌ/	ns	*	**

Note. EG=Experimental group; CG=Control group; levels of significance: * $p < .05$; ** $p < .01$; *** $p < .001$.

Bonferroni Pairwise comparisons, presented in Table 35, revealed that the RTs of the experimental group in the identification of the five vowel contrasts decreased significantly, particularly from pretest to delayed posttest, and from posttest to delayed posttest (see Figure 26). From pretest to posttest, the RTs of three vowel pairs reduced, not including /i/-/ɪ/ and /u/-/ʊ/. In the control group, immediately after training, a decrease of RTs was observed for only one vowel pair (/i/-/ɪ/) (see Figure 27). However, between posttest and delayed posttest, and between pretest and delayed posttest, significant changes were observed for all vowel contrasts, except for /ɛ/-/æ/.

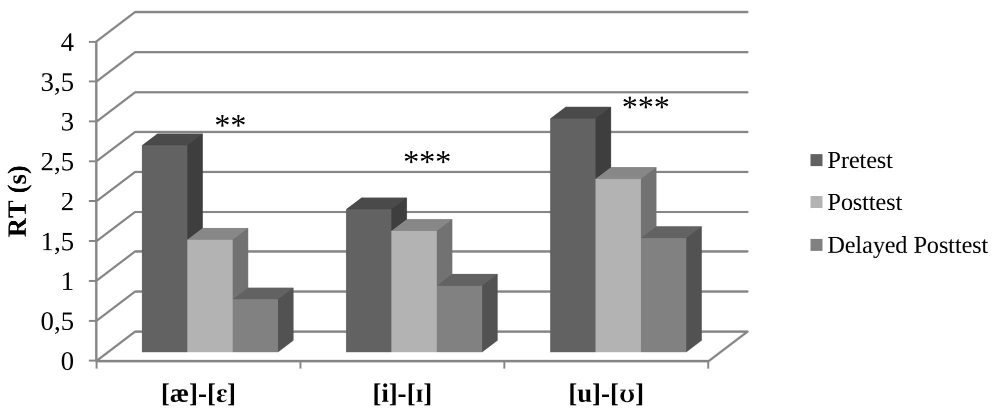


Figure 26. Mean response times of the three vowel pairs for the experimental group.

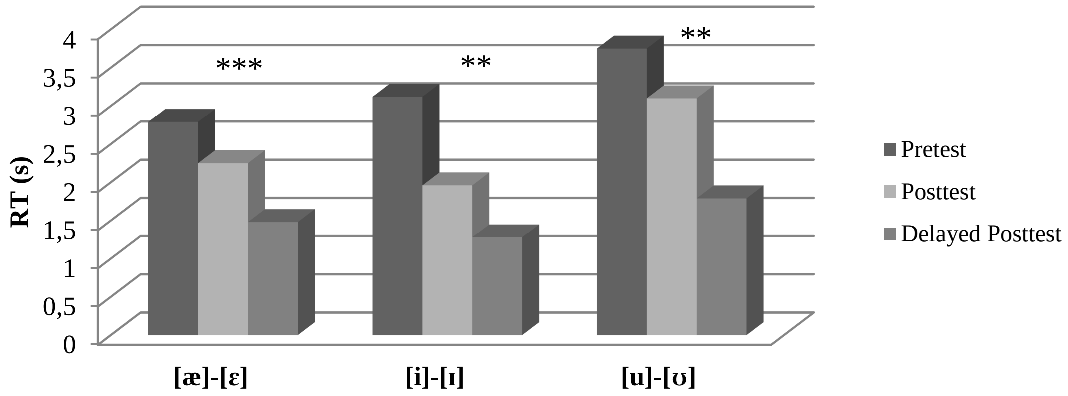


Figure 27. Mean response times of the three vowel pairs for the control group.

If the reduction of RTs in the identification of vowel categories were the result of a better perceptual performance enhanced by auditory training only, that is, if there were no more effects involved, such as task repetition effect, the control group would not show any significant decrease. However, given that the training tasks implied similar behavioral responses from the participants, (viz. hearing a stimulus, identifying it by selecting a button with the computer mouse, and evaluating vowel category goodness-of-fit by clicking on another button) it was not surprising that both groups would decrease RTs. In spite of the significant differences in RTs within each group, our hypotheses that RTs would decrease after training, and shorter RTs would still be observed two months after completion of the training sessions were supported by these findings.

However, to test whether there were differences between groups, we ran *t* tests for independent samples for each vowel contrast in the three tests. The results are presented in Table 36.

Table 36

T Test Results of Between-group Comparisons of RTs

ID test	**Vowel Contrast**	**EG** (*n*=22) **Mean (SD)**	**CG** (*n*=12) **Mean (SD)**	***t* (df)**
Pretest	**/æ/-/ɛ/**	2.48 (1.63)	3.03 (1.08)	-1.04(32)
	/i/-/ɪ/	2.17 (1.52)	3.39 (1.66)	-2.17(32) *
	/u/-/ʊ/	2.82 (1.48)	3.89 (1.28)	-2.11(32) *
	/æ/-/ʌ/	2.20 (1.52)	2.95 (1.23)	-1.46(32)
	/ʊ/-/ʌ/	2.43 (1.58)	3.55 (1.25)	-2.11(32) *
Posttest	**/æ/-/ɛ/**	1.37 (1.01)	2.45 (1.43)	-2.58(32) *
	/i/-/ɪ/	1.52 (1.13)	2.16 (1.30)	-1.51(32)
	/u/-/ʊ/	2.03 (1.29)	3.07 (1.18)	-2.32(32) *
	/æ/-/ʌ/	1.20 (.79)	2.40 (1.33)	-3.35(32) **
	/ʊ/-/ʌ/	1.47 (.96)	2.69 (1.24)	-3.19(32) **
		EG (*n*=19)	**CG** (***n*****=11)**	
Delayed Posttest	**/æ/-/ɛ/**	.66 (.60)	1.48 (1.06)	-2.75(28) **
	/i/-/ɪ/	.88 (.71)	1.28 (.71)	-1.51(28)
	/u/-/ʊ/	1.43 (.97)	1.79 (1.31)	-.85(28)
	/æ/-/ʌ/	.66 (.37)	1.37 (.86)	-3.12(28) **
	/ʊ/-/ʌ/	1.08 (.63)	1.70 (1.13)	-1.93(28)

Note. EG=Experimental group; CG=Control group; SD=Standard Deviation; *t*= result of the *t* test; *(df)* =degrees of freedom; levels of significance: * $p < .05$; ** $p < .01$; *** $p < .001$.

We can observe in Table 36 that the experimental group had overall shorter RTs than the control group in the three ID tests. The *t* test results indicate that, in the pre- and posttests, trainees had significantly shorter RTs than controls, except for vowel contrast /i/-/ɪ/ immediately after training (see Figures 28 and 29). However, these significant differences were not observed for two target vowel pairs in the delayed posttest (see Figure 30). Two months after training, trainees were still significantly quicker than controls in the identification of the low front vowel pair, but not in the identification of the other two vowel pairs, though having slightly shorter RTs.

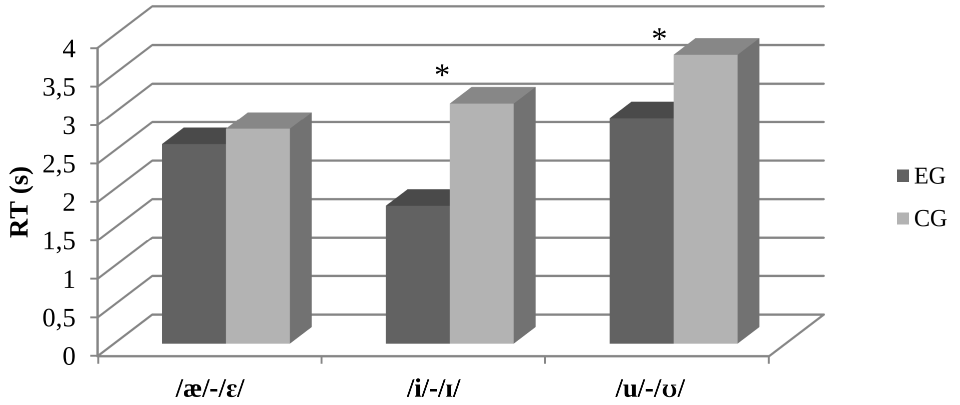


Figure 28. Mean response times of the three vowel pairs in the pretest for both groups.

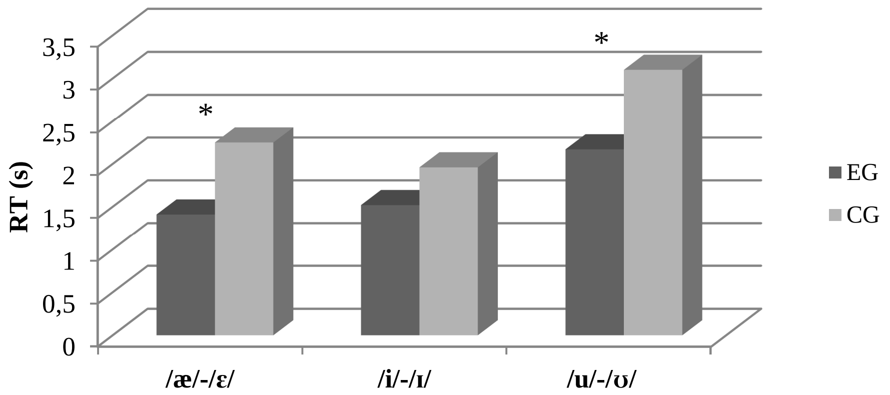


Figure 29. Mean response times of the three vowel pairs in the posttest for both groups.

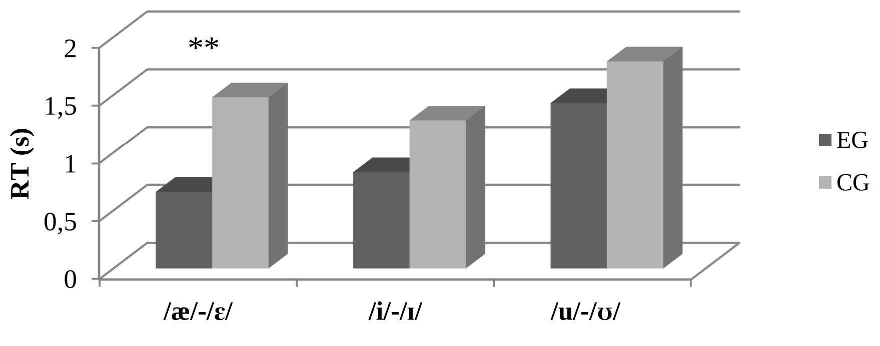


Figure 30. Mean response times of the three vowel pairs in the delayed posttest for both groups.

These findings seem to indicate that the decrease of RTs was more directly related to task effects, that is, to task familiarization, than to vowel training, because there was a gradual decrease from pretest to delayed posttest in both groups. In sum, our hypothesis that RTs would be shorter in the experimental group than in the control group was partially confirmed.

When comparing the RTs with the mean percentage of correct identification, we can conclude that, in both groups, in the three test phases, the longer RTs are related to the identification of the back vowel contrast. The two other target vowel pairs are not so consistently associated with RTs in the perception tests. For example, the experimental group takes less time to identify the mid-low front vowel pair than the high vowel pair, whereas the control group seems to be quicker in labeling the high front pair than the low front pair.

To further investigate whether there was an association between mean percentage of correct identification and mean RTs, a Pearson Correlation Coefficient was applied. Significant negative correlations were found for the two vowel contrasts /æ/-/ɛ/ (r = -.66, p < .001) and /i/-/ɪ/ (r = -.50, p = .002) in the posttest and also in the delayed posttest, /æ/-/ɛ/ (r = -.60, p = .001) and /i/-/ɪ/ (r = -.62, p < .001), respectively. Higher identification scores were associated with shorter RTs. No significant correlations were found for the back vowel contrast in any of the three tests. In the case of the vowel pairs with the distractor vowel (/ʊ/-/ʌ/ and /æ/-/ʌ/) only a marginally significant negative correlation was also found for /æ/-/ʌ/ in the posttest (r = -.317, p= .067).

4.1.3 Analysis by Vowel

To further understand the modification of perceptual mappings of the target American English vowels in the three test phases, separate statistical analyses were carried out for the six target vowels (/i/, /ɪ/, /æ/, /ɛ/, /ʊ/, /u/) and the distractor /ʌ/.

4.1.3.1 Exploratory analysis of perception data

In order to verify whether data were normally distributed, the Kolgomorov-Smirnov (K-S) and the Shapiro-Wilk tests were run. The results revealed that some variables (i.e., vowel identification % scores) differed significantly from a normal distribution. In the experimental group, the scores for vowels /æ/ and /ʊ/ in the posttest, vowels /ɪ/ and /ɛ/ in the delayed posttest, and vowel /æ/ in the generalization test were not normally distributed ($p < .05$). In the case of the control group, only the percentage ID scores for vowel /æ/ had a non-normal distribution in the posttest and generalization test. Since one of the assumptions for parametric tests was violated, we opted to firstly transform data (Martins, 2011, p. 234), and then analyze it again with the transformed values. We used the Log10 transformation to compute the seven variables, and then we ran the K-S and the Shapiro-Wilk tests to verify whether the transformed values had a normal distribution, but they did not. All the transformed values had a non-normal distribution. Moreover, a Levene's test was run to verify homogeneity of variance within both groups of participants. For the majority of vowels there was homogeneity of variances, that is, the variances of vowels were equal for the experimental and control groups ($p > .05$), but for vowels /u/ and /ʊ/ in the posttest (F (1, 32) = 5.49, $p < .05$, and F (1, 32) = 7.09, $p < .05$, respectively), and delayed posttest (F (1, 28) = 10.57, $p < .01$, and F (1, 28) = 5.77, $p < .05$), for vowels /u/ and /i/ in the generalization test (F (1, 32) = 9.03, $p < .01$, and F (1, 32) = 5.16, $p < .05$), and for vowel /æ/ in the delayed posttest (F (1, 28) = 11.09, $p < .01$) the variances were significantly different in the two groups. Thereby, both non-parametric tests and parametric tests were run for these variables with a non-normal distribution, namely the Friedman test and the repeated measures analysis of variance (ANOVA) for intragroup analyses, and the Mann Whitney test and the *t* test for independent samples for intergroup analyses. Both parametric and non-parametric test results revealed to be equivalent. Given that there were no differences in terms of results (that is, the conclusions drawn from both sets of tests were the same in all cases), we decided to report the parametric tests' results, which is, as previously mentioned, a strategy advised by Fife-Schaw (2006). Moreover, parametric tests are more robust and allow the use of multivariate analyses, thus reducing the number of tests performed and, therefore, the probability of Type 1 error (Matos, Santos, Gonçalves, & Martins, 2009, as cited in Martins, 2011, p. 240).

4.1.3.2 Results of the perception pretest and posttest

The same statistical procedures were followed for analysis by vowel. To compare perceptual ability of 22 trainees and 12 controls before and after training, *t* tests for paired samples were carried out. Improvement in the posttest was revealed by calculating the difference between the mean percentage ID scores of each vowel of the pretest and posttest for both groups of L2 participants. The results are displayed in Table 37.

Table 37

T Test Results of Within-group Comparisons for Individual Vowel Identification in the Pretest and Posttest

Group	Vowel	Pretest Mean (SD)	Posttest Mean (SD)	*t* (df)	% Difference
EG (*n*=22)	**/i/**	53.79 (21.56)	85.15(14.79)	-5.04(21) ***	31.36
	/ɪ/	59.55 (24.65)	82.27 (15.55)	-5.61(21) ***	22.72
	/æ/	75.00 (15.35)	81.36 (7.54)	-2.21(21) *	6.36
	/ɛ/	58.18 (18.34)	81.51 (15.21)	-4.39(21) ***	23.33
	/u/	61.52 (23.13)	78.79 (20.36)	-4.00(21) ***	17.27
	/ʊ/	66.06 (11.30)	76.06 (11.20)	-3.54(21) **	10
	/ʌ/	85.15 (12.25)	90.15 (7.09)	-2.08(21) *	5
CG (*n*=12)	**/i/**	70.56 (17.34)	69.72 (18.77)	.16(11)	-0.8
	/ɪ/	62.50 (23.49)	68.61 (23.59)	-2.09(11)	6.11
	/æ/	79.44 (10.62)	75.28 (18.45)	1.06(11)	- 4.16
	/ɛ/	61.94 (17.20)	68.33 (14.87)	- 2.15(11)	6.39
	/u/	62.78 (23.09)	58.06 (28.26)	-1.01(11)	-4.72
	/ʊ/	62.78 (19.22)	61.11 (22.76)	-.24(11)	-1.67
	/ʌ/	80.56 (18.25)	85.83 (11.82)	-1. 57(11)	5.27

Note. EG=Experimental group; CG=Control group; SD=Standard Deviation; *t*= result of the *t* test; (df)=degrees of freedom; levels of significance: * $p < .05$; ** $p < .01$; *** $p < .001$.

The *t* test results by vowel were consistent with the vowel contrast analyses. The trainees' identification abilities improved significantly for all the seven vowels immediately after training (see Figure 31), with an improvement range from 5% to 31.36%, whereas no modification of perceptual performance was found for the participants in the control group (see Figure 32). The higher accurate ID scores were observed for vowels /i/ and /ɪ/, followed by /ɛ/ and /æ/, and then by /u/ and /ʊ/. Furthermore, the highest significant improvement occurred for vowels /i/ and /ɛ/, followed by /ɪ/, /u/, and then /ʊ/ and /æ/. The least significant improvement was observed for the distractor vowel, which had the highest correct ID score in the pre- and posttests (85.15 and 90.15%).

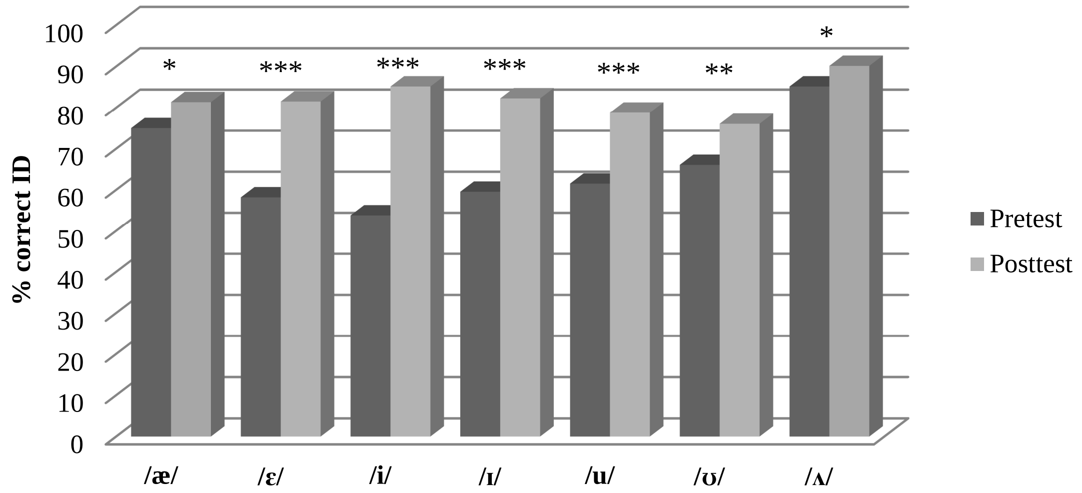


Figure 31. Mean % of correct identification scores of the seven vowels in the pretest and posttest for the experimental group.

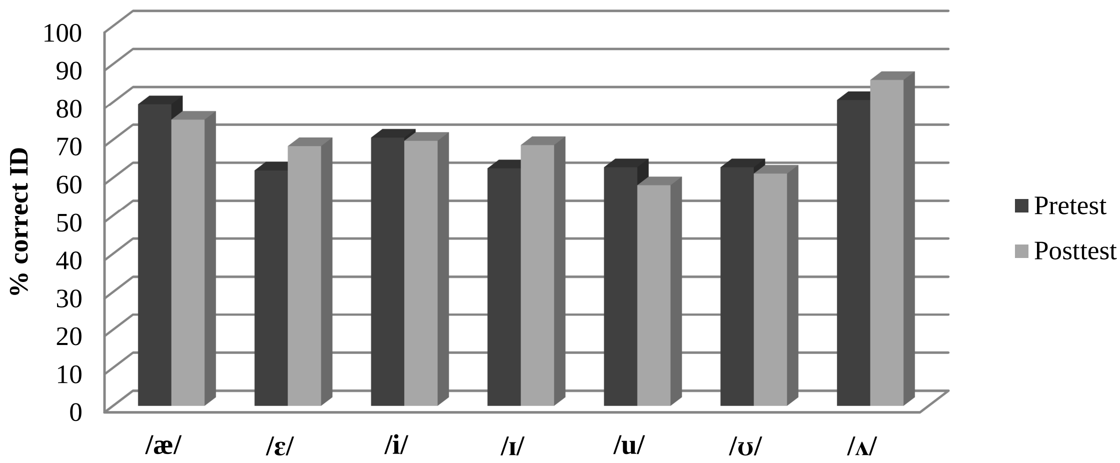


Figure 32. Mean % of correct identification scores of the seven vowels in the pretest and posttest for the control group.

4.1.3.3 Results of the perception posttest and generalization test

A comparison of identification scores by vowel between the posttest and generalization test was carried out by means of a *t* test for paired samples to further examine whether transfer of learning was observable for each of the seven AE vowels. The results are presented in Table 38.

Table 38

T Test Results of Within-group Comparisons for Individual Vowel Identification in the Pretest and Posttest

Group	Vowel	Posttest Mean (SD)	Generalization test Mean (SD)	*t* (df)	% Difference
EG (*n*=22)	**/i/**	85.15 (14.79)	88.38 (10.42)	-1.28(21)	3.23
	/ɪ/	82.27 (15.55)	90.15 (9.54)	-3.43(21) **	7.88
	/æ/	81.36 (7.54)	94.95 (9.38)	-5.47(21) ***	13.59
	/ɛ/	81.51 (15.21)	85.61 (11.96)	-1.44(21)	4.1
	/u/	78.79 (20.36)	78.53 (16.82)	.06(21)	-.26
	/ʊ/	76.06 (11.20)	56.31 (23.08	4.04(21) ***	-19.75
	/ʌ/	90.15 (7.09)	83.84 (13.49)	-2.07(21) *	-6.31
CG (*n*=12)	**/i/**	69.72 (18.77)	71.76 (19.74)	-.45(11)	2.94
	/ɪ/	68.61 (23.59)	64.35 (17.48)	.62(11)	-4.26
	/æ/	75.28 (18.45)	83.80 (21.12)	-2.84(11) *	8.52
	/ɛ/	68.33 (14.87)	61.11 (16.41)	1.49(11)	-7.22
	/u/	58.06 (28.26)	64.35 (29.91)	-1.16(11)	6.29
	/ʊ/	61.11 (22.76)	50.00 (20.10)	1.78(11)	-11.11
	/ʌ/	85.83 (11.82)	67.13 (21.51)	4.46(11) ***	-18.7

Note. EG=Experimental group; CG=Control group; SD=Standard Deviation; *t*= result of the *t* test; (df)=degrees of freedom; levels of significance: * $p < .05$; ** $p < .01$; *** $p < .001$.

The *t* test results additionally revealed that, in the case of the experimental group, the identification of vowels /æ/ and /ɪ/ improved significantly in the generalization test in comparison to the posttest, whereas vowels /ɛ/ and /i/ did not. This finding shows, in more detail, that perceptual assimilation of these L2 vowel segments occurred, and learning transferred to new talkers and new tokens. However, there was a significant decrease (-19.75%) in the identification scores of the high back vowel /ʊ/, which indicates that learning of this segment did not transfer to new tokens, thus suggesting that improvement observed in the posttest was not robust. The identification of the distractor vowel was also significantly poorer in the generalization test than in the posttest in both groups (see Figures 33 and 34).

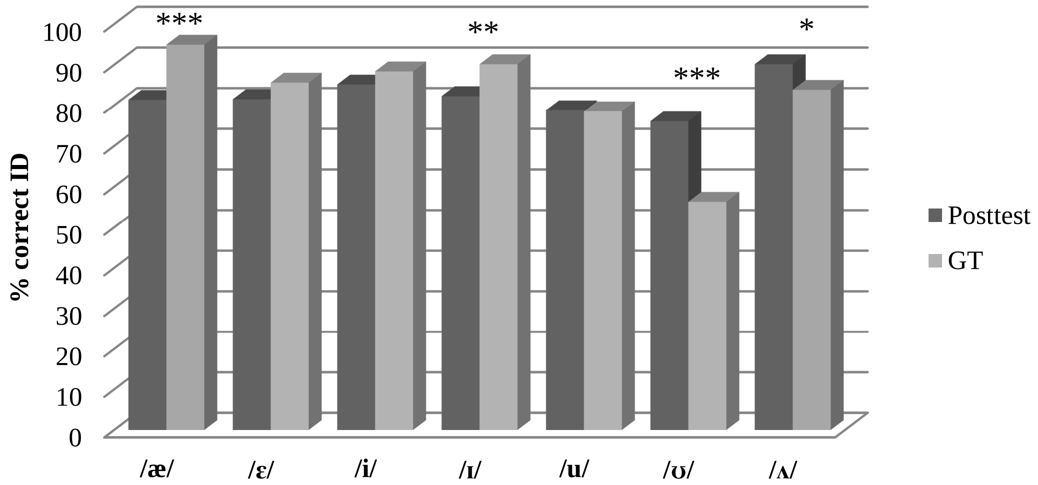


Figure 33. Mean % of correct identification scores of AE vowels in the posttest and generalization test for the experimental group.

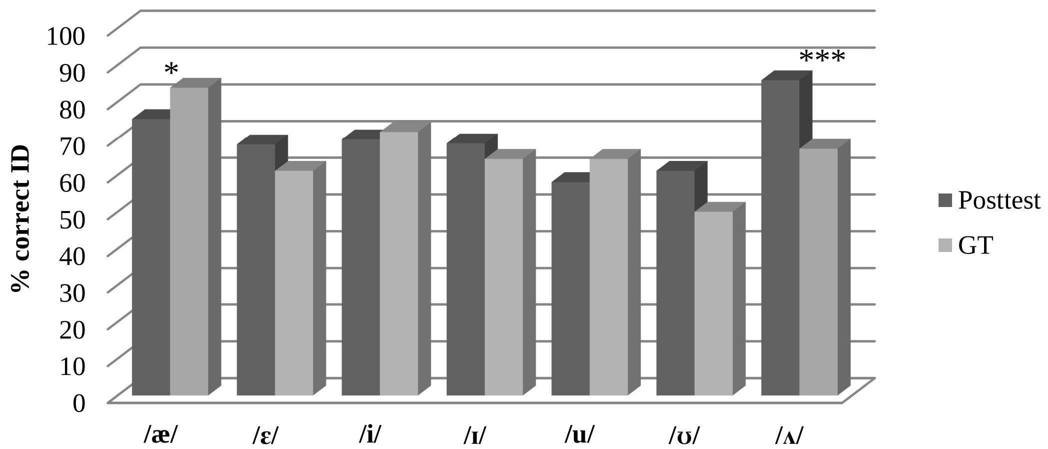


Figure 34. Mean % of correct identification scores of AE vowels in the posttest and generalization test for the control group.

4.1.3.4 Results of the perception pretest, posttest and delayed posttest

Given that there were four dropouts two months after the training was over, the following analyses were carried out with the 19 trainees and 11 controls present in all test phases. To complement the previous analyses, separate repeated measures analyses of variance (ANOVAs) were done to investigate effects of high variability auditory training on L2 learners' identification of AE vowels. The ANOVA results are presented in Table 39.

Table 39

ANOVA Test Results of the Experimental and Control Groups after Participant Dropouts

Group	Vowel	Pretest Mean (SD)	Posttest Mean (SD)	Delayed Posttest Mean (SD)	*F* (df)
EG (*n*=19)	**/æ/**	74.56 (15.99)	82.46 (6.65)	85.96 (4.92)	6.95(1.37, 24.62) **
	/ɛ/	58.24 (19.00)	80.17 (15.53)	81.58 (17.76)	12.48(1.24, 22.32) ***
	/i/	53.86 (23.10)	87.02 (12.27)	86.49 (12.79)	27.23(1.21, 21.72) ***
	/ɪ/	57.72 (25.07)	82.98 (15.23)	88.77 (13.48)	32.65(1.23, 22.18) ***
	/u/	62.81 (23.07)	81.75 (17.40)	85.96 (18.64)	19.98(1.19, 21.36) ***
	/ʊ/	66.67 (12.02)	75.44 (11.98)	75.26 (11.83)	6.32(2, 36) **
	/ʌ/	85.09 (12.24)	90.00 (7.11)	91.23 (8.26)	3.47(2, 36)
	Total	65.56 (7.59)	82.86 (5.80)	85.38 (6.90)	52.85(1.29, 23.25) ***
CG (*n*=11)	**/æ/**	80.60 (10.30)	76.06 (19.14)	80.00 (15.28)	.95(2.20)
	/ɛ/	62.73 (17.81)	69.09 (15.35)	66.67 (16.33)	1.68(2.20)
	/i/	71.21 (18.03)	69.70 (19.69)	66.06 (23.70)	.52(2.20)
	/ɪ/	64.85 (23.11)	71.51 (22.38)	67.58 (23.15)	1.78(2.20)
	/u/	63.03 (24.20)	56.97 (29.38)	59.39 (33.09)	.78(2.20)
	/ʊ/	62.12 (20.02)	59.70 (23.31)	60.00 (23.94)	.08(1.22, 12.15)
	/ʌ/	80.91 (19.10)	86.36 (12.24)	82.42 (12.74)	1.15(2.20)
	Total	69.22 (9.70)	69.91 (9.97)	68.87 (11.07)	.30(1.23, 12.27)

Note. EG= Experimental group; CG=Control group; SD=Standard Deviation; *F*= results of the ANOVA; (df)=degrees of freedom; levels of significance: * $p < .05$; ** $p < .01$; *** $p < .001$.

A significant effect of auditory training was found on the identification of all the seven AE vowels for the participants in the vowel-trained group (experimental group), but not for the consonant-trained participants (control group). Bonferroni pairwise comparisons were carried out for the two groups to further investigate the effects of training (see Table 40). The post hoc test results for the experimental group revealed significant differences between the pretest and posttest for five of the six target vowel segments, but not for the distractor vowel. The low front vowel /æ/ was not significantly better identified after training, but when comparing the pretest and delayed posttest scores, a significant difference ($p < .05$), from 74.56% to 85.96%, was found, which suggests that some perceptual change occurred in between the two test phases. An interesting finding is also the significant improvement of the high front vowel /ɪ/ from the posttest to the delayed posttest, which indicates that perceptual modification still occurred during the eight weeks after conclusion of the training program. Moreover, no significant decrease of identification ability was found for any of the seven vowels. On the contrary, slightly higher accurate scores were observed two months after completion of the training sessions for all the vowels, except /i/ and /ʊ/, which had approximately the same ID rates. These results suggest that robust learning of English vowels was retained.

Table 40

Pairwise Comparisons Between the Pretest and Posttest, the Posttest and Delayed Posttest, and the Pretest and Delayed Posttest

Group	**Vowel**	**Pretest *vs.* Posttest**	**Posttest *vs.* Delayed Posttest**	**Pretest *vs.* Delayed Posttest**
	/æ/	ns	ns	*
	/ɛ/	**	ns	**
	/i/	***	ns	***
	/ɪ/	***	*	***
EG	**/u/**	**	ns	***
	/ʊ/	*	ns	*
	/ʌ/	ns	ns	ns
	/æ/	ns	ns	ns
	/ɛ/	ns	ns	ns
	/i/	ns	ns	ns
	/ɪ/	ns	ns	ns
	/u/	ns	ns	ns
CG	**/ʊ/**	ns	ns	ns
	/ʌ/	ns	ns	ns

Note. EG=Experimental group; CG=Control group; * $p < .05$; ** $p < .01$; *** $p < .001$.

Intergroup analyses were carried out to examine whether the same differences in vowel identification between trainees and controls, which were observed for vowel contrasts, would persist when analyzing each vowel separately. Therefore, *t* tests for independent samples were run, and the results regarding the four identification tests are presented in Table 41.

Table 41

T Test Results of Between-group Comparisons for the Three Tests

ID test	**Vowel**	**EG (*n*=22) Mean (SD)**	**CG (*n*=12) Mean (SD)**	***t* (df)**
Pretest	**/i/**	53.79 (21.56)	70.56 (17.34)	-2.31(32) *
	/ɪ/	59.55 (24.65)	62.50 (23.49)	-.34(32)
	/æ/	75.00 (15.35)	79.44 (10.62)	-.89(32)
	/ɛ/	58.18 (18.34)	61.94 (17.20)	-.58(32)
	/u/	61.52 (23.13)	62.78 (23.09)	-.15(32)
	/ʊ/	66.06 (11.30)	62.78 (19.22)	.63(32)
	/ʌ/	85.15 (12.25)	80.56 (18.25)	.88(32)
Posttest	**/i/**	85.15 (14.79)	69.72 (18.77)	2.64(32) *
	/ɪ/	82.27 (15.55)	68.61 (23.59)	2.04(32) *
	/æ/	81.36 (7.54)	75.28 (18.45)	1.37(32)
	/ɛ/	81.51 (15.21)	68.33 (14.87)	2.43(32) *
	/u/	78.79 (20.36)	58.06 (28.26)	2.47(32) *
	/ʊ/	76.06 (11.20)	61.11 (22.76)	2.58(32) *
	/ʌ/	90.15 (7.09)	85.83 (11.82)	1.34(32)
Gen. Test	**/i/**	88.38 (10.42)	71.76(19.74)	3.23(32) **
	/ɪ/	90.15 (9.54)	64.35 (17.48)	5.60(32***
	/æ/	94.95 (9.38)	83.80 (21.12)	2.14(32) *
	/ɛ/	85.61 (11.96)	61.11(16.41)	5.00(32***
	/u/	78.53(16.82)	64.35 (29.91)	1.78(32)
	/ʊ/	56.31 (23.08	50.00 (20.10)	.80(32)
	/ʌ/	83.84 (13.49)	67.13(21.51)	2.79(32) **
		EG (*n*=19) Mean (SD)	**CG (*n*=11) Mean (SD)**	***t* (df)**
Delayed Posttest	**/i/**	86.49 (12.79)	66.06 (23.70)	3.08(28) **
	/ɪ/	88.77 (13.48)	67.58 (23.15)	3.19(28) **
	/æ/	85.96 (4.92)	80.00 (15.28)	1.58(28)
	/ɛ/	81.58 (17.76)	66.67 (16.33)	2.28(28) *
	/u/	85.96 (18.64)	59.39 (33.09)	2.83(28) **
	/ʊ/	75.26 (11.83)	60.00 (23.94)	2.35(28) *
	/ʌ/	91.23 (8.26)	82.42 (12.74)	2.30(28) *

Note. EG=Experimental group; CG=Control group; SD=Standard Deviation; *t*= result of the *t* test; (df)=degrees of freedom; levels of significance: * *p* < .05; ** *p* < .01; *** *p* < .001.

Before training, no significant differences were found between the experimental and the control groups, except in the identification of the high front tense vowel, which was better identified by the latter group (see Figure 35). After training, significant intergroup differences were observed in the identification of six vowels (/i/, /ɪ/, /ɛ/, /ʌ/, /ʊ/, /u/). Although the trained group identified vowel /æ/ more accurately than the control group did, the difference was not significant (see Figure 36). This pattern was repeated in the delayed posttest, in which the only vowel that was not significantly better identified by the trainees was the front lax vowel /æ/ (see Figure 37). In the generalization test, the identification of the high back vowels did not differ between groups, in particular in the recognition of vowel /ʊ/, for which the percentage difference between groups was approximately 6% (see Figure 38).

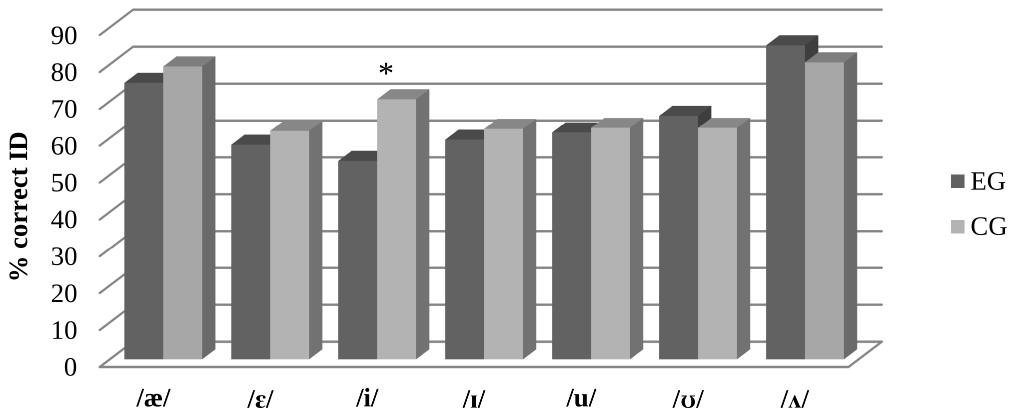


Figure 35. Mean % of correct identification scores of AE vowels in the pretest for the experimental and control groups.

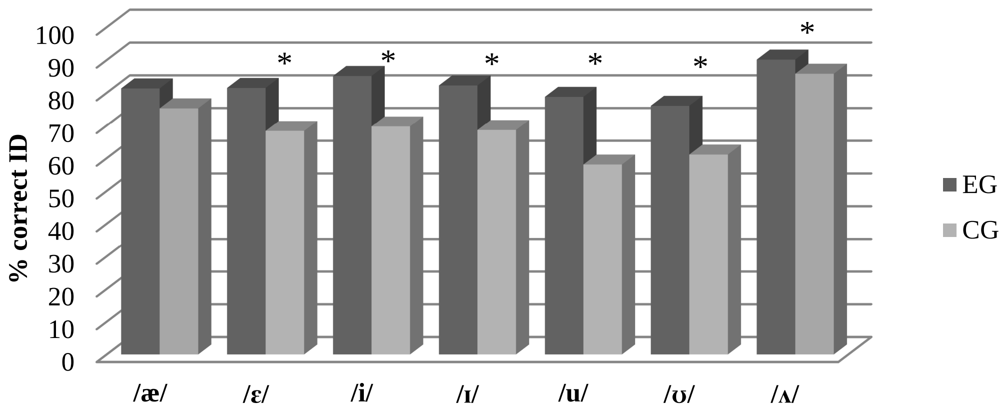


Figure 36. Mean % of correct identification scores of AE vowels in the posttest for the experimental and control groups.

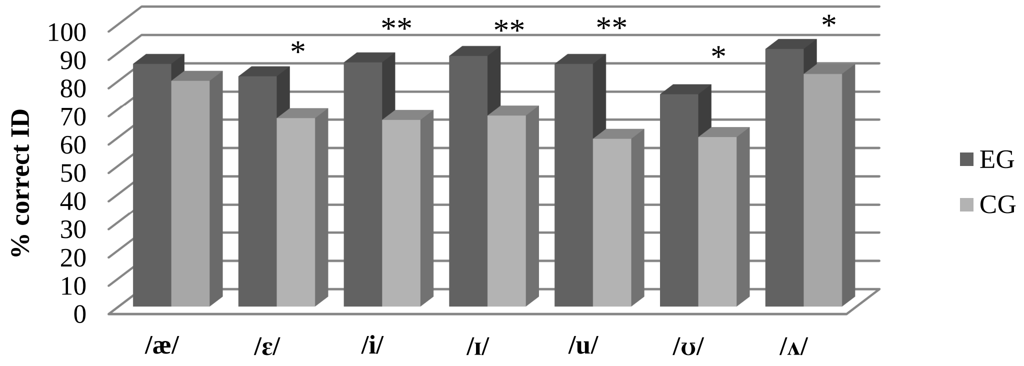


Figure 37. Mean % of correct identification scores of AE vowels in the delayed posttest for the experimental and control groups.

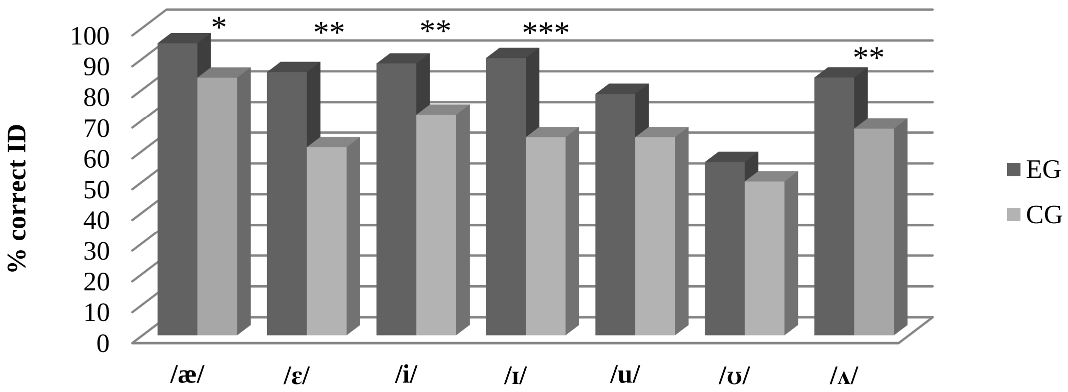


Figure 38. Mean % of correct identification scores of AE vowels in the generalization test for the experimental and control groups.

To investigate whether there was a relationship between ID answer (correct and incorrect) and phonetic context of the stimuli, a Pearson chi-square test was run for the three identification tests (pre-, post- and delayed posttest). A significant correlation was found, χ^2 (9)= 200.57, p<.001. English vowels embedded in the phonetic context /bVk/ were the most accurately identified (86%), followed by context /kVt/ (78%), and /kVk/ (77%). The phonetic contexts in which vowels were less correctly recognized were /pVt/ and /tVk/ with an overall ID score of 59%. Table 42 presents the percentage of correct and incorrect identification scores by phonetic context.

Table 42

Percentage of Correct and Incorrect Identification Scores by Phonetic Context

		% Correct ID	***n***	**% Incorrect ID**	***n***	***N***
Phonetic Context	bVt	69	849	31	375	1224
	kVt	78	637	22	179	816
	kVp	68	138	32	66	204
	pVt	59	722	41	502	1224
	tVt	65	793	35	430	1223
	tVk	59	846	41	582	1428
	bVk	86	351	14	57	408
	kVt	65	133	35	71	204
	pVp	65	132	35	72	204
	kVk	77	157	23	47	204

*Note.*ID=identification; *N* =total number of correct and incorrect ID; *n*=number of correct or incorrect ID.

In addition, a Pearson chi-square test was run to observe if there was a relationship between American English talkers and ID answer (correct and incorrect), and results revealed that there was a significant correlation , χ^2 (9)= 460.87, p<.001.
The talker with the highest rate of correct ID (98%) was T6, a female AE speaker from Iowa (see Table 43). However, the number of tokens produced by this NS was the lowest (only 196 tokens) in the set of stimuli presented in the three identification tests. Two Californian male and female speakers, T3 and T4, respectively, were 67% correctly identified by all participants. The stimuli produced by male speaker T1 were the least accurately identified (69%).

Table 43

Percentage of Correct and Incorrect Identification Scores by Talker

AE Talker	**% Correct ID**	***n***	**% Incorrect ID**	***n***	***N***
T1 CA	69	3397	31	1503	4900
T2 CA	73	858	27	318	1176
T3 CA	76	5058	24	1606	6664
T4 CA	76	5221	24	1639	6860
T5 IA	72	564	28	220	784
T6 IA	98	192	2	4	196

*Note.*ID=identification; *N* =total number of correct and incorrect ID; *n*=number of correct or incorrect ID.

4.1.3.5 Analysis of perceptual confusion matrices

To further analyze L2 learners' vowel identification responses, the mean percentage rates of category recognition, that is, the correct and incorrect identifications were tabulated in confusion matrices for each perception test. Rows in the matrices correspond to the vowels that were heard in the identification tests, and columns correspond to the vowel categories to which listeners mapped the sounds heard.

Table 44

Confusion Matrix of the Pretest

Group	Stimulus heard	Identified vowel /æ/	/ɛ/	/i/	/ɪ/	/u/	/ʊ/	/ʌ/
EG	/æ/	**75.0**	23.5	0.5	0.2	-	0.3	0.6
	/ɛ/	30.6	**58.2**	3.5	3.9	-	1.1	2.1
	/i/	1.5	5.5	**53.8**	37.6	0.5	0.6	0.6
	/ɪ/	0.3	8.6	29.4	**59.5**	0.5	0.8	0.9
	/u/	0.3	0.2	-	0.5	**61.5**	30.3	7.3
	/ʊ/	0.2	-	0.2	-	13.6	**66.1**	20.0
	/ʌ/	1.5	0.9	0.2	0.3	4.9	7.1	**85.2**
CG		/æ/	/ɛ/	/i/	/ɪ/	/u/	/ʊ/	/ʌ/
	/æ/	**79.4**	19.7	-	0.6	-	-	0.3
	/ɛ/	26.9	**61.9**	1.9	6.9	-	0.8	1.4
	/i/	-	4.7	**70.6**	24.4	-	0.3	-
	/ɪ/	3.1	9.2	24.2	**62.5**	-	0.6	0.6
	/u/	-	-	-	-	**62.8**	32.2	5.0
	/ʊ/	-	-	-	-	14.2	**62.8**	13.1
	/ʌ/	3.1	1.1	-	-	8.9	6.4	**80.6**

The confusion matrix of the pretest (Table 44) indicates that vowels were consistently misidentified as their counterparts. For example, /ɪ/ was consistently misidentified as /i/ and vice versa, /ɛ/ was inaccurately mapped onto /æ/, and vice versa, and /ʊ/ was labeled as /u/, and vice versa. The distractor vowel, when incorrectly identified, was either heard as /ʊ/ or /u/, and seldom as /æ/, contrary to what would be expected given their acoustic proximity in the AE vowel space. The high front lax vowel /ɪ/ was also sporadically identified as /ɛ/, and vice versa. In line with previous findings reported by Rato et al. (2013), vowel /æ/ was more accurately identified by both groups than vowel /ɛ/. In sum, the perceptual confusion in the identification of the target vowel categories showed evidence that they had not been fully established in the participants' L2 phonological system.

Table 45

Confusion Matrix of the Posttest

Group	Stimulus heard	Identified vowel /æ/	/ɛ/	/i/	/ɪ/	/u/	/ʊ/	/ʌ/
EG	**/æ/**	**81.4**	18.6	-	-	-	-	-
	/ɛ/	12.6	**81.5**	-	5.8	-	-	0.2
	/i/	-	0.9	**85.2**	13.9	-	-	-
	/ɪ/	0.5	10.6	6.7	**82.3**	-	-	-
	/u/	0.2	-	-	-	**78.8**	14.9	5.3
	/ʊ/	0.3	-	-	-	4.9	**76.1**	18.8
	/ʌ/	0.2	0.2	-	-	1.1	8.5	**90.2**
		/æ/	**/ɛ/**	**/i/**	**/ɪ/**	**/u/**	**/ʊ/**	**/ʌ/**
CG	**/æ/**	**75.3**	23.6	0.3	-	-	-	0.8
	/ɛ/	22.2	**68.3**	1.9	6.9	-	-	0.6
	/i/	-	3.1	**69.7**	26.1	-	0.3	-
	/ɪ/	0.8	8.9	21.1	**68.6**	-	0.3	0.3
	/u/	-	0.3	0.6	-	**58.1**	36.9	4.2
	/ʊ/	-	0.3	-	-	19.4	**61.1**	20.0
	/ʌ/	0.8	0.3	-	-	5.6	7.5	**85.8**

Although there was a decrease of inaccurately identified vowels by the trained group in the posttest (Table 45), the pattern of perceptual confusion was the same, that is, vowels were interchangeably misidentified as their counterparts, except for vowel /ɪ/, which was mapped to /ɛ/ more frequently than to its tense correspondent /i/. The two groups exhibited different perceptual assimilation patterns in relation to this vowel, with the controls confusing it with /i/ more often than with /ɛ/, similarly to what they did in the pretest. After training, vowel /ʌ/ continued to be misidentified as /ʊ/. The back vowel /ʊ/ was also most frequently confused with /ʌ/ than with /u/ by participants in both groups. However, this pattern of misidentifying /ʊ/ as /ʌ/ was still much more salient in the vowel-trainees.

Table 46

Confusion Matrix of the Delayed Posttest

Group	Stimulus heard	Identified vowel						
		/æ/	**/ɛ/**	**/i/**	**/ɪ/**	**/u/**	**/ʊ/**	**/ʌ/**
EG	**/æ/**	**86.0**	13.9	-	-	-	-	0.2
	/ɛ/	11.9	**81.6**	0.3	6.1	0.3	0.3	1.0
	/i/	0.4	-	**86.5**	13.2	-	-	-
	/ɪ/	0.2	6.8	4.2	**88.8**	-	-	-
	/u/	-	-	-	-	**86.0**	11.6	2.5
	/ʊ/	-	-	-	-	6.1	**75.3**	18.6
	/ʌ/	0.5	-	-	-	1.1	7.2	**91.2**
		/æ/	**/ɛ/**	**/i/**	**/ɪ/**	**/u/**	**/ʊ/**	**/ʌ/**
CG	**/æ/**	**80.0**	19.1	0.3	-	-	-	0.6
	/ɛ/	24.2	**66.7**	0.6	8.5	-	-	-
	/i/	0.3	5.5	**66.1**	28.2	-	-	-
	/ɪ/	-	10.0	22.4	**67.6**	-	-	-
	/u/	-	-	-	-	**59.4**	38.8	2.6
	/ʊ/	-	0.3	-	-	13.9	**60.0**	25.8
	/ʌ/	1.8	0.3	-	-	3.6	11.8	**82.4**

Table 46 shows that the experimental group mapped five of the six target vowels onto the expected vowel categories, that is, to their closest counterparts, two months after training was over. Vowel /ɪ/ was the only exception, being frequently misidentified as /ɛ/, rather than as /i/. The same pattern occurred in the posttest. Conversely, controls identified the high lax vowel /ɪ/ more often as its tense correspondent /i/ than as /ɛ/. Vowel /æ/ was also still more accurately identified by both groups than /ɛ/. Vowel /ɛ/ continued to be misidentified not only as /æ/ but also as /ɪ/. Vowel /ʊ/ was still most confused with /ʌ/ than with /u/ by both experimental and control participants.

Table 47

Confusion Matrix of the Generalization Test

Group	Stimulus heard	Identified vowel /æ/	/ɛ/	/i/	/ɪ/	/u/	/ʊ/	/ʌ/
EG	**/æ/**	**95.0**	4.8	-	-	-	-	0.2
	/ɛ/	13.1	**85.6**	-	0.8	-	-	0.5
	/i/	-	-	**88.4**	11.6	-	-	-
	/ɪ/	-	1	8.8	**90.2**	-	-	-
	/u/	-	-	-	-	**78.5**	19.0	2.5
	/ʊ/	0.3	-	-	-	22.2	**56.3**	21.2
	/ʌ/	1.3	-	-	-	4.3	10.6	**83.8**
		/æ/	**/ɛ/**	**/i/**	**/ɪ/**	**/u/**	**/ʊ/**	**/ʌ/**
CG	**/æ/**	**83.8**	16.2	-	-	-	-	-
	/ɛ/	30.0	**60.7**	-	9.3	-	-	-
	/i/	-	0.4	**71.8**	27.8	-	-	-
	/ɪ/	-	1.4	34.3	**64.3**	-	-	-
	/u/	-	-	-	-	**64.4**	30.6	5.0
	/ʊ/	0.4	-	-	-	38.0	**50.0**	11.6
	/ʌ/	4.6	-	-	-	6.4	21.8	**67.2**

In the generalization test, vowel /æ/ was the least misidentified vowel in both groups, whereas its counterpart /ɛ/ was more frequently mapped to /æ/ but also to /ɪ/, in particular by controls. Vowel /ʊ/ was still the most inaccurately labeled vowel, being confused with /u/ or /ʌ/. The distractor vowel was most often misidentified as /ʊ/ but also as /u/ or as /æ/. The high front vowels were interchangeably confused with each other.

4.1.3.6 Analysis of category goodness-of-fit

Participants rated category goodness-of-fit of the seven American English vowels with a three-level Likert scale ranging from 1 (*poor*) to 3 (*good*) in the identification tests (pretest, posttest, delayed posttest and generalization test), but in order to compare the results among the same set of stimuli, only the results of the pretest, posttest and delayed posttest are reported. The means of the goodness-of-fit ratings were calculated, and comparisons were carried out for the three identification tests (see Tables 48 and 49). Our prediction was that higher ratings of goodness-of-fit would be observed after the high-variability phonetic training for the participants in the experimental group, but not for the controls. To compare the ratings between the pretest and posttest, we ran a *t* test for paired samples for the two groups, whose results are presented in Table 48.

Table 48

T Test Results of Within-group Comparisons of Category Goodness-of-fit

Group	**Vowel**	**Pretest Mean (SD)**	**Posttest Mean (SD)**	***t* (df)**
EG (*n*=22)	**/æ/**	2.22 (.49)	2.37 (.39)	-1.42(21)
	/ɛ/	2.02 (.46)	2.25 (.47)	-2.16(21)*
	/i/	2.21 (.47)	2.37 (.43)	-1.38(21)
	/ɪ/	1.96 (.42)	2.29 (.46)	-3.03(21) **
	/u/	2.25 (.48)	2.26 (.34)	-.09(21)
	/ʊ/	2.29 (.52)	2.4 (.31)	-1.37(21)
	/ʌ/	2.23 (.48)	2.54 (.30)	-3.17(21) **
CG (*n*=12)	**/æ/**	2.52 (.41)	2.43 (.35)	.58(11)
	/ɛ/	2.37 (.42)	2.33 (.45)	.28(11)
	/i/	2.43 (.41)	2.33(.39)	.80(11)
	/ɪ/	2.19 (.47)	2.22 (.42)	-.25(11)
	/u/	2.45 (.34)	2.39 (.33)	.49(11)
	/ʊ/	2.43 (.43	2.42 (.30)	.06(11)
	/ʌ/	2.45 (.50)	2.48 (.29)	-.24(11)

Note. EG=Experimental group; CG=Control group; SD=Standard Deviation; *t*= result of the *t* test; (df)=degrees of freedom; levels of significance: * $p < .05$; ** $p < .01$; *** $p < .001$.

After the auditory training, EFL learners in the experimental group rated the American English vowels /ɛ/, /ɪ/, and /ʌ/ significantly higher than before training. No significant changes regarding category goodness-of fit were found either for the other four vowels or for the control group. To closer examine goodness-of-fit ratings for the two attrited groups, repeated measures ANOVAs were run, and results are shown in Table 49.

Table 49

ANOVA Results of Within-group Comparisons of Category Goodness-of-fit

Group	Vowel	Pretest Mean (SD)	Posttest Mean (SD)	Delayed Posttest Mean (SD)	*F* (df)
EG (*n*=19)	**/æ/**	2.22 (.41)	2.34 (.40)	2.58 (.41)	6.48(2,36) **
	/ɛ/	2.01 (.41)	2.19 (.48)	2.44 (.48)	7.69(2,36) **
	/i/	2.22 (.44)	2.36 (.46)	2.55 (.41)	4.53(2,36) *
	/ɪ/	1.96 (.37)	2.26 (.48)	2.54 (.44)	14.91(2,36) ***
	/u/	2.26 (.45)	2.24 (.35)	2.49 (.43)	2.98(2,36) †
	/ʊ/	2.30 (.46)	2.41 (.31)	2.56 (.29)	3.64(1.54, 27.68)
	/ʌ/	2.24 (.42)	2.52 (.31)	2.65 (.32)	10.55(2,36) ***
CG (*n*=11)	**/æ/**	2.49 (.42)	2.42 (.37)	2.32 (.54)	.40(2,20)
	/ɛ/	2.33 (.41)	2.31 (.47)	2.12 (.51)	.72(2,20)
	/i/	2.42 (.43)	2.34 (.40)	2.34 (.54)	.14(2,20)
	/ɪ/	2.16 (.48)	2.23 (.43)	2.12 (.57)	.17(2,20)
	/u/	2.42 (.34)	2.34 (.31)	2.34 (.54)	.12(2,20)
	/ʊ/	2.40 (.45)	2.38 (.28)	2.27 (.52)	.23(2,20)
	/ʌ/	2.41 (.51)	2.45 (.29)	2.39 (.50)	.06(2,20)

Note. EG=Experimental group; CG=Control group; SD=Standard Deviation; *F*= result of the ANOVA; (df)=degrees of freedom; levels of significance: † $p <.1$; * $p < .05$; ** $p < .01$; *** $p < .001$.

A significant effect of perceptual training was found on goodness-of-fit ratings of five AE vowels for the trainees (/ɛ/, /æ/, /i/, /ɪ/, /ʌ/), but not for the controls (Table 49). A marginally significant effect was also found on the rating of vowel /u/ by the experimental group. Bonferroni pairwise comparisons were run for the two groups to further investigate the effects of training (see Table 50).

Table 50

Pairwise Within-group Comparisons of Category Goodness-of-fit

Group	**Vowel Contrast**	**Pretest *vs.* Posttest**	**Posttest *vs.* Delayed Posttest**	**Pretest *vs.* Delayed Posttest**
EG	**/æ/**	ns	ns	**
	/ɛ/	ns	ns	**
	/i/	ns	ns	*
	/ɪ/	†	*	***
	/u/	ns	*	ns
	/ʊ/	ns	ns	ns
	/ʌ/	*	ns	**
CG	**/æ/**	ns	ns	ns
	/ɛ/	ns	ns	ns
	/i/	ns	ns	ns
	/ɪ/	ns	ns	ns
	/u/	ns	ns	ns
	/ʊ/	ns	ns	ns
	/ʌ/	ns	ns	ns

Note. EG=Experimental group; CG=Control group; levels of significance: † p <.1; * p < .05; ** p < .01; *** p < .001.

The post hoc test results for the experimental group revealed significant differences mainly between the pretest and delayed posttest. Except for the high back vowels /u/ and /ʊ/, the goodness-of-fit ratings improved significantly from the beginning of the training sessions to the delayed posttest phase. Significant higher ratings were also found between the post- and delayed posttests for two vowels, /ɪ/ and /u/. A marginally significant (p <.01) increase was also observed for the /ɪ/ category goodness-of-fit ratings between the pretest and posttest (see Figure 39). As predicted, the control group did not show any significant changes regarding evaluation of vowel category goodness-of-fit after training (see Figure 40).

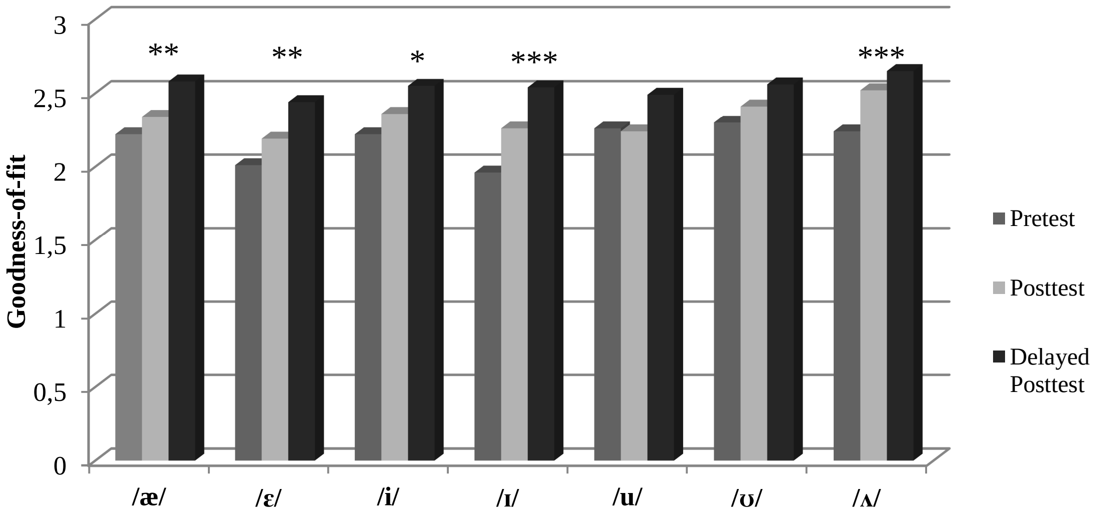


Figure 39. Mean ratings of AE vowel category goodness-of-fit for the experimental group.

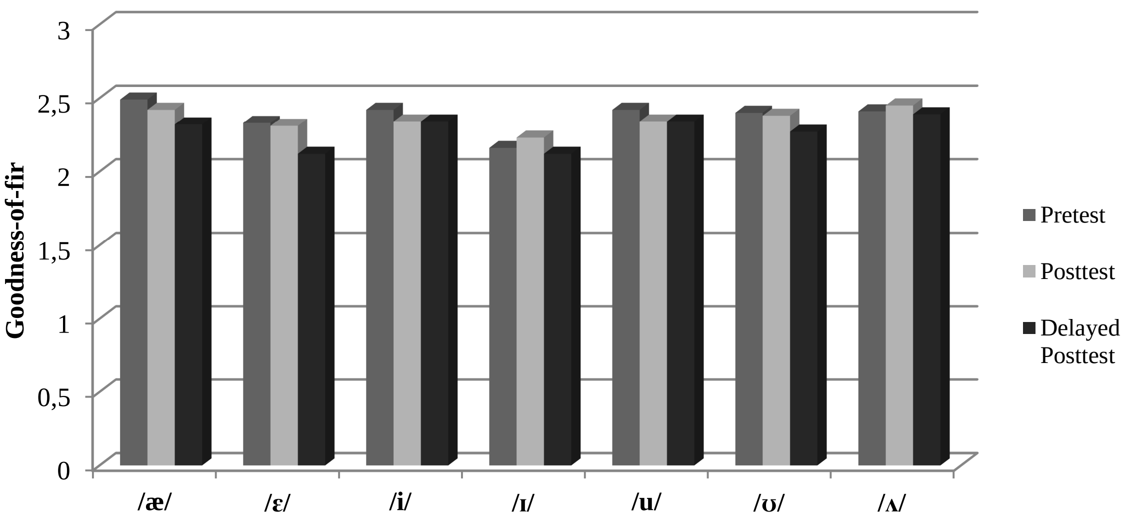


Figure 40. Mean ratings of AE vowel category goodness-of-fit for the control group.

Comparisons between the experimental and the control groups were carried out to find whether there were differences in terms of category goodness-of-fit ratings. Table 51 exhibits the results of the *t* tests for independent samples that were performed.

Table 51

T Test Results of Between-group Comparisons of Category Goodness-of-fit

ID test	Vowel Contrast	EG (*n*=22) Mean (SD)	CG (*n*=12) Mean (SD)	*t* (df)
Pretest	**/i/**	2.22 (.49)	2.52 (.41)	-1.37(32)
	/ɪ/	2.02 (.46)	2.37 (.42)	-1.42(32)
	/æ/	2.21 (.47)	2.43 (.41)	-1.79(32)
	/ɛ/	1.96 (.42)	2.19 (.47)	-2.20(32) *
	/u/	2.25 (.48)	2.45 (.34)	-1.31(32)
	/ʊ/	2.29 (.52)	2.43 (.43	-.79(32)
	/ʌ/	2.23 (.48)	2.45 (.50)	-1.24(32)
Posttest	**/i/**	2.37 (.39)	2.43 (.35)	.28(32)
	/ɪ/	2.25 (.47)	2.33 (.45)	.44(32)
	/æ/	2.37 (.43)	2.33 (.39)	-.49(32)
	/ɛ/	2.29 (.46)	2.22 (.42)	-.45(32)
	/u/	2.26 (.34)	2.39 (.33)	-1.07(32)
	/ʊ/	2.4 (.31)	2.42 (.30)	.21(32)
	/ʌ/	2.54 (.30)	2.48 (.29)	.60(32)
ID test	**Vowel Contrast**	**EG (*n*=19) Mean (SD)**	**CG (*n*=11) Mean (SD)**	***t* (df)**
Delayed Posttest	**/i/**	2.58 (.41)	2.32 (.54)	1.22(28)
	/ɪ/	2.44 (.48)	2.12 (.51)	2.30(28) *
	/æ/	2.55 (.41)	2.34 (.54)	1.45(28)
	/ɛ/	2.54 (.44)	2.12 (.57)	1.70(28)
	/u/	2.49 (.43)	2.34 (.54)	.86(28)
	/ʊ/	2.56 (.29)	2.27(.52)	1.94(28) †
	/ʌ/	2.65 (.32)	2.39 (.50)	1.75(28)

Note. EG=Experimental group; CG=Control group; *t*=results of the *t* test; levels of significance: † p <.1; * $p < .05$; ** $p < .01$; *** $p < .001$.

The results of *t* tests revealed that, before training, the only significant intergroup difference was in the goodness-of-fit rating of vowel /ɛ/, which was rated higher by the controls (2.19) than by the trainees (1.96). After training, no significant differences were found for either groups. However, two months after the training was over, the participants of the experimental group rated vowel /ɛ/ significantly higher, and vowel /ʊ/ marginally higher than the controls. Although there were no more significant differences, we can observe that ratings of the vowel-trained group were higher for all vowels than for the consonant-trained group.

To complement the previous analyses and to investigate whether there was a relation between mean percentage correct identification and category goodness-of-fit ratings, a Pearson Correlation Coefficient was applied for each vowel. Significant positive correlations were found for the following vowels, presented in Table 52.

Table 52

Significant Correlations between Percent Correct ID and Goodness-of-fit

Vowel	**Perception test**	**Pearson Coefficient**	**Significance Level**
/i/	pretest	*r* = .370	*p* < .05
/æ/	pretest	*r* = .371	*p* < .05
/ʊ/	pretest	*r* = .365	*p* < .05
/u/	pretest	*r* = .355	*p* < .05
/i/	delayed posttest	*r* = .548	*p* < .001
/ɪ/	delayed posttest	*r* = .450	*p* < .05
/ɛ/	delayed posttest	*r* = .455	*p* < .05

Note. *r*=results of Pearson's correlation test.

In addition, marginally significant positive correlations were found for vowels listed in Table 53.

Table 53

Marginally Significant Correlations between Percent Correct ID and Goodness-of-fit

Vowel	**Perception test**	**Pearson Coefficient**	**Significance Level**
/ʌ/	pretest	*r* = .316	*p* < .01
/ɪ/	posttest	*r* = .300	*p* < .01
/ɛ/	posttest	*r* = .309	*p* < .01
/ʌ/	delayed posttest	*r* = .349	*p* < .01

Note. *r*=results of Pearson's correlation test.

In sum, higher identification scores were associated with higher goodness-of-fit ratings for the aforementioned vowels in the three perception tests.

4.1.4 Analysis of individual trainees' perceptual performance

The mean percentage scores of correct identification of the high front /i/-/ɪ/, high back /u/-/ʊ/ and mid-low front /ɛ/-/æ/ vowel contrasts at pretest, posttest, delayed posttest and generalization test are presented in Tables AH1-3. The tables include columns with the difference in percentage identification scores for each vowel pair between pretest and posttest, and between posttest and both delayed posttest and generalization test. These percentage differences were calculated by subtracting the results of the second perception tests from the corresponding first tests. The analyses were carried out only for participants in the vowel-trained group so as to divide them into four different subgroups corresponding to their overall perceptual performance in the training program. We followed, to a great extent, the division into four subgroups of EFL learners proposed by Munro (2013), viz. (1) native-like learners; (2) fast learners; (3) slow learners; and (4) puzzling learners, being the latter designation our suggestion to refer to performances that were not consistent across perception tests.

At pretest, except for participant E22 who performed at a native-like level (i.e. with ≥ 90% ID scores) in the identification of two vowel contrasts (/i/-/ɪ/ and /u/-/ʊ/), no other L2 learner reached scores above 90%. However, at posttest, eight learners (E1, E2, E3, E10, E11, E13, E16, and E22) obtained ≥ 90% scores in the identification of /i/-/ɪ/ and four improved in the recognition of /æ/-/ɛ/ (E2, E14, E16, E21) and /u/-/ʊ/ (E1, E9, E10, E14) to a native-like level. Within this group, participant E22, despite having 95% correct ID score in the identification of the high front vowels, still improved to 98% after training, although there was almost no space for improvement. However, from this group of learners only two were included in the native-like group, namely E1 and E14, because they were consistently good at identifying the three vowel contrasts after training. A general analysis shows that the two high vowel contrasts had higher ID scores, which reached 98%, than the /æ/-/ɛ/ pair, whose highest correct ID score was 93%.

Conversely, there were also some trainees who, though showing some gains, had ID scores ranging from 52% to 72% in the perception of the vowel targets. At posttest, in the labeling of /i/-/ɪ/, four students (E12, E15, E19 and E20) had relatively low scores (52%-69%); in the identification of /æ/-/ɛ/ the number increased to six (E3, E8, E11, E15, E17 and E18), with scores ranging from 68% to 72%; and in the case of the back vowel contrast there were seven learners (E2, E7, E8, E15, E18, E19 and E21) who had ID scores between 52% and 72%. However, not all of these students should be included in the slow learners' group, because their perceptual performance was not consistently poor, that is, although the gain in the identification of one or two of the target vowel contrasts was poor (for example, E18 improved only .01% from pre- to posttest in the ID of the /æ/-/ɛ/) or even inexistent (for instance, E21 had lower scores in the post- than in the pretest in the ID of /u/-/ʊ/ with a difference of -6.66%), their overall performance was not equally poor across vowel contrasts. Some of these EFL learners were, thus, best fit in the puzzling group. When comparing the posttest scores, we concluded that the high front vowel pair had less low scores, followed by the mid-low front pair and then the high back vowel contrast. Generally speaking, the majority of EFL learners were included in the fast learners' group because their improvement rates were constantly high (for instance, E2, E5, E6, E8, E9, E10, E12, E13 and E16). A small group of four (E3, E7, E18, E9) could be included in the slow learners' group, because in comparison to the fast learners' group, their improvement rates were not as high, varying from 3% to 15%. The perceptual ability of seven participants (E4, E11, E15, E 17, E20, E21, and E22) in the identification of the vowel contrasts was rather puzzling, because their gains varied from high to low scores or even reflected some loss from pretest to posttest. In particular, participant E11 had improvement rates from 10% and 13% to 77% in the case of the high front vowels (specifically, from 23% to 100% correct ID).

When looking at the accurate ID scores in the delayed posttest, we concluded that several participants reached a native-like perceptual performance (i.e., ≥90% scores). In the identification of /i/-/ɪ/ by the 19 participants, 11 (E1, E2, E3, E4, E7, E9, E10, E11, E13, E14 and E22) obtained scores above 91% (92%-100%); in the labeling of /æ/-/ɛ/, eight learners (E1, E2, E7, E9, E11, E13, E20 and E21) had ID scores between 90% and 95%; and in the perception of /u/-/ʊ/, seven participants had ID scores between 92% and 98%. Despite the native-like results in the recognition of one or two vowel pairs, only four participants reached an overall native-like performance at the end of this study, namely E1, E9, E11, and E13 who had ≥90% ID scores in the perception of the English target vowels. Participant E14 can be also included in this group, despite having a lower ID score in the identification of the mid-low vowel pair in the delayed posttest (87%). In terms of low scores, only one participant (E12) had a correct ID score below 70% (i.e., 63%) for the high front pair, and two participants (E2, and E8) had ID scores of 68% and 69%, respectively, for the high back vowel contrast.

In comparison with the posttest results, in which there were only three cases of decline in perceptual ability (one in the ID of the high front vowel pair, from 57% to 52%, and two in the ID of the back vowel contrast, from 75% to 68%, and from 90% to 80%), in the delayed posttest there were several cases in which there was a decrease in identification accuracy of the target vowel contrasts. Three learners had lower ID scores in the identification of the high back vowel pair (from -3% to -5%), five in the ID of the high front vowel contrast (from -2% to -7%) and seven in the identification of the mid-low front vowels (from -2 to -10%). These results were expected because there were no training sessions between the posttest and delayed posttest. However, high rates of improvement were surprisingly found for the three vowel pairs by some participants. For example, five trainees improved between 7% to 12% in the identification of the mid-low front vowels, eight students had a gain varying from 7% to 15% in the labeling of the high front vowel pair, and the most unexpected finding was the high number of students who still improved in the identification of the high back vowel contrast two months after the training was over. Improvement rates of 11 learners varied from 6% to 16%, which indicates that learning occurred for the vowel-trained group even after the training was over, but not for the consonant-trained group.

When comparing the generalization test (GT) with the posttest results, almost all students transferred learning of the two front vowel contrasts to new words and new talkers. Only participants E12 and E16 did not generalize perceptual performance to identification of the high front vowels, though they still had high ID scores (88% to 86%, and 92% to 83%) in the generalization test. In the identification of the mid-low vowel pair only six of the 22 students (E1, E2, E11, E13, E16, and E22) had lower ID scores in the generalization test. Conversely, in line with the statistical test results described in the previous sections, the majority of EFL learners did not generalize learning of the back vowel pair to new stimuli. Only three (E3, E11 and E16) learners revealed higher (3%-16%) ID scores in the generalization test than in the posttest. The other trainees' ID scores differences ranged from -2% (E2) to -39% (E6) in relation to the posttest. In the GT, native-like performance was observed in 13 participants' perception of four front vowels, and on two trainees' recognition of the high back vowel contrast. In the case of the mid-low front vowel pair, ID scores varied from 78% to 100%; for the high front vowels scores, the range was from 69% to 100%; and the range of scores in the ID of the back vowel contrast varied greatly from below chance level (<50%) to 94%. Three learners (E6, E15 and E18) frequently misidentified the back vowels, having only 39% of correct scores, and five learners (E2, E8, E19, E20 and E21) had scores at chance level (50%-58%). These results indicate that robust learning of the back vowels did not occur for this group of learners, because they were not able to identify the vowels in new monosyllabic words. Due to this fact, that is, to the poorer perceptual identification of the back vowel pair in the generalization test, none of the learners of the posttest native-like group achieved ≥90% scores in the identification of all the vowel contrasts. For instance, participants E2, E5, E9, E13, E14, E17 and E22 had native-like ID scores in all the target vowels, except in the back vowels. Nevertheless, there were also two learners (E3 and E11) who had ID scores above 91% in two vowel pairs, except in the mid-low front vowel contrast.

As expected, trainees' perceptual ability to map the heard vowels onto the correct AE phonetic categories varied greatly across vowel contrasts and identification tests. Overall, after training, several learners reached native-like (e.g., E1, E9, E11, E13 and E14) or near native-like (e.g., E2, E3, E7, E19, E16 and E22) perceptual performance, and these trainees were included either in the native-like or in the fast learners' group, because they had steady high ID scores in the tests after auditory training. Furthermore, there were seven learners (E4, E5, E8, E12, E17, E18, and E19) whose improvement was not as great and quick as the fast learners', especially due to the difficulties in identifying the back vowels; thus, they were included in the slow learners' group. Moreover, the perceptual behavior of four students was puzzling when comparing their performance across tests. For example, participant E6 had an unexplained decrease from 78% to 39% in the identification of the back vowels in the generalization test. However, difficulty with the back vowel contrast in comparison with the other vowels (which had scores between 78% and 94%) could not be confirmed in the delayed posttest, because this participant dropped out. The same applies to participant E15, who had poor results in the identification of the two high vowel contrasts, but not in the recognition of the mid-low pair in the posttest and generalization tests. The same pattern was observed for participants E20 and E21 who had much higher (i.e., native-like) scores in the perception of /æ/-/ɛ/ than of /i/-/ɪ/ or /u/-/ʊ/.

Although, as previously mentioned, the group of participants of this study was rather homogeneous in terms of EFL background (see Appendix G, for a summary of data), the native-like performance of two students at the end of the training study might be associated with their greater EFL experience. Participant E1, for instance, started learning English in a formal classroom context at the age of six, and at the time of the study he had learned English for 12 years (see Table F1). In comparison with the mean AOL of 10 years and LFI of 8 years, this learner had more experience with English than the other participants. Participant E13 also differed from the other trainees because he studied English at a private foreign language school for eight years (see Table F2). Nonetheless, there were other students that reached a native-like or near native-like performance, but who did not have more EFL experience than their colleagues, at least as far as we could attest from the background questionnaire data.

4.2 Speech Production

4.2.1 Acoustic analysis of European Portuguese vowels

The production of seven European Portuguese oral vowels (/i/, /e/, /ɛ/, /a/, /ɔ/, /o/, /u/) was recorded by the 34 participants of this study. However, the productions of only 32 participants are reported due to the fact that two L2 learners did not meet the birth place and residence in the Minho region criteria. In total, 1428 vowels were segmented but only 1344 vowels (7 vowels x 6 contexts x 32 participants) were analyzed acoustically. The mean and median acoustic measurements of the vowels are presented in Table 54 organized by gender. The vowel segments were produced by 16 female (50%) and 16 male (50%) informants, whose ages ranged from 18 to 42 (mean=23.25, SD=6.9). This group of undergraduate students was quite homogeneous in terms of birth place and residence. Apart from the two participants who were excluded from this analysis,[129] the 32 students were born and lived in the Minho region at the time of this study. Within this group of people from the west northern region, 26 (81.3%) belonged to the district of Braga, four (12.5 %) to the district of Porto, and two (6.3%) to the district of Viana do Castelo. The mean and median duration (measured in seconds), F1 and F2 values (measured in Hertz) are presented in Table 54.

[129] The two participants who were not included in the analysis were born and lived in other regions of Portugal than the Minho region (viz. Bragança and Madeira) until moving to Braga to study at the university.

Table 54

Averages of EP Vowel Duration, F1 and F2 values

Vowel EP		/i/ Mean (SD)	/i/ Mdn	/e/ Mean (SD)	/e/ Mdn	/ɛ/ Mean (SD)	/ɛ/ Mdn	/a/ Mean (SD)	/a/ Mdn	/ɔ/ Mean (SD)	/ɔ/ Mdn	/o/ Mean (SD)	/o/ Mdn	/u/ Mean (SD)	/u/ Mdn
Duration (ms)	F	85 (16)	82	105 (20)	100	115 (17)	114	119 (17)	119	119 (18)	119	105 (17)	107	89 (15)	87
	M	73 (10)	73	91 (13)	92	100 (12)	101	101 (13)	98	100 (13)	101	95 (12)	98	77 (15)	78
F1 (Hz)	F	399 (50)	398	481 (36)	487	658 (57)	652	887 (72)	859	695 (59)	671	509 (58)	512	438 (44)	443
	M	328 (30)	327	433 (30)	427	560 (41)	554	728 (81)	694	580 (44)	571	453 (29)	449	362 (27)	359
F2 (Hz)	F	2539 (116)	2526	2103 (119)	2106	2059 (120)	2101	1532 (108)	1525	1101 (91)	1087	1069 (75)	1076	922 (81)	916
	M	2092 (153)	2134	1762 (126)	1746	1690 (111)	1696	1339 (81)	1324	1008 (53)	1009	1009 (94)	1012	908 (58)	906

Note. F=female speakers; M=male speakers; ms=milliseconds; Hz=Hertz. The standard deviations are between parentheses, and every cell represents 16 speakers.

The mean F1 and F2 values of the seven European Portuguese vowels were plotted, and the standard deviation is shown through ellipses around the vowel symbol, which represents the mean value. Figures 41 and 42 illustrate the vowel spaces of 16 Portuguese women and 16 men.

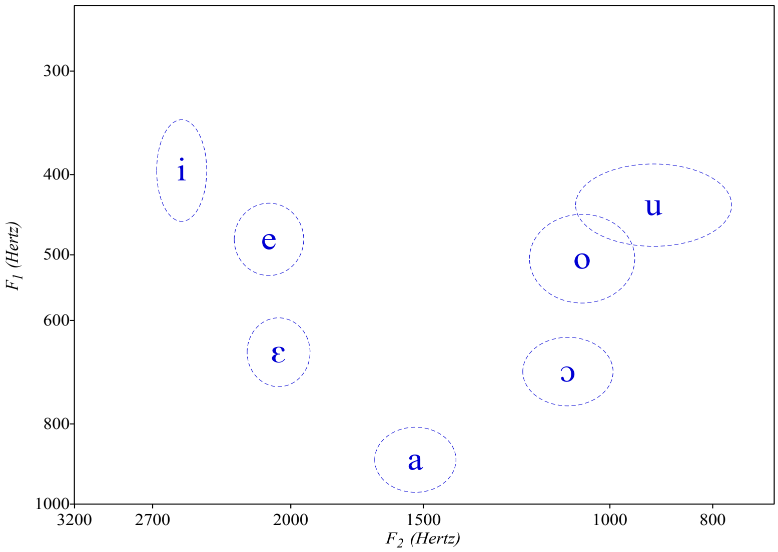


Figure 41. EP vowel space of 16 women.

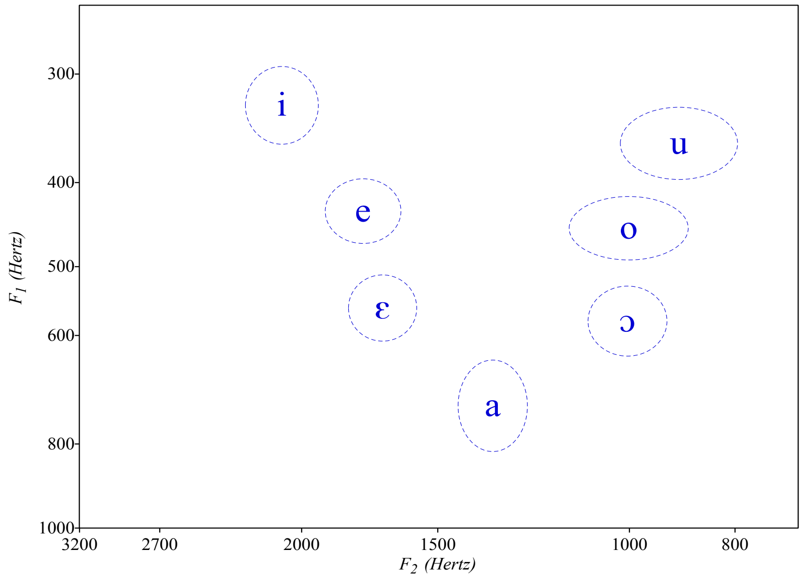


Figure 42. EP vowel space of 16 men.

The F1 and F2 formant values were normalized using the Lobanov method to remove variation caused by vocal tract size differences between male and female speakers. Therefore, the normalized F1 and F2 values were also plotted (see Figure 43).

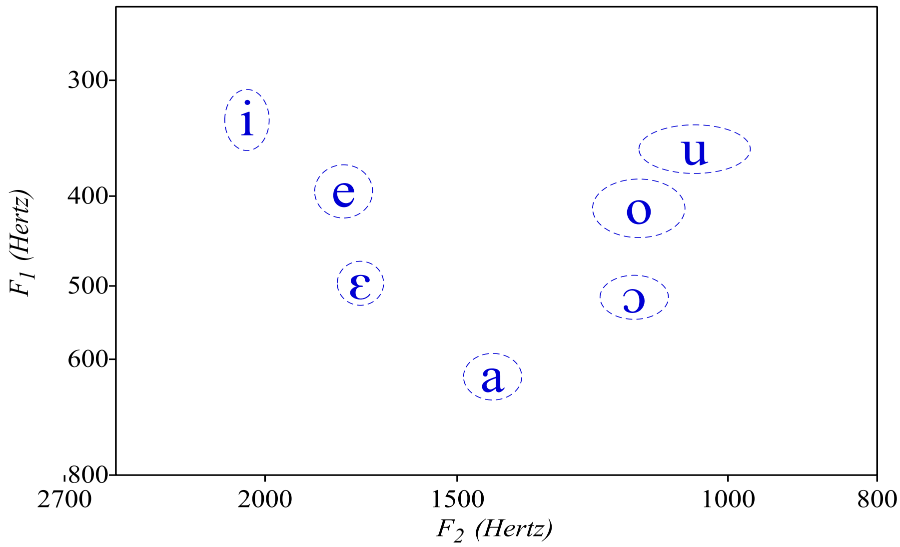


Figure 43. EP vowel space of 32 Portuguese speakers.

As can be observed, the Portuguese vowel space is symmetrical, each front vowel has a back vowel counterpart, which is slightly lower and vowel /a/ is in low central position. This configuration corroborates previous measurements of European

Portuguese vowels produced by speakers from Lisbon (Delgado-Martins, 1973; Escudero et al., 2009).

4.2.2 Acoustic analysis of American English vowels

Production data of seven American English vowels (/i/, /ɪ/, /æ/, /ɛ/, /ʌ/, /ʊ/, /u/) were collected three times, viz. before the auditory training (pretest), immediately after it was over (posttest), and eight weeks later (delayed posttest). Acoustic parameters of vowel duration, F1 and F2 values were measured for each participant and organized into four groups, as follows: (1) ten women in the experimental group; (2) twelve men in the experimental group; (3) eight women in the control group; (4) and four men in the control group. Given that there were four participants who dropped out in the delayed posttest, the number of participants per groups altered. Thereby, the group of trainees in the delayedd posttest phase included eight female and eleven male speakers, while the control group consisted of seven women and four men. Furthermore, the whole set of data was divided into three moments, namely the pretest, posttest and delayed posttest. The mean, median and standard deviation values of vowel duration, F1 and F2 are shown in Tables 55-57, following the organization previously explained.

Table 55

Averages of AE Vowel Duration, F1 and F2 values at Pretest

	Vowel		/i/		/ɪ/		/ɛ/		/æ/		/u/		/ʊ/		/ʌ/	
			Mean (SD)	**Mdn**	**Mean (SD)**	**Mdn**	**Mean (SD)**	**Mdn**	**Mean (SD)**	**Mdn**	**Mean (SD)**	**Mdn**	**Mean (SD)**	**Mdn**	**Mean (SD)**	**Mdn**
EG	**Dur (ms)**	**F**	109 (30)	95	107 (28)	92	130 (25)	116	139 (26)	124	115 (31)	99	108 (30)	98	115 (29)	99
		M	99 (15)	99	94 (13)	96	118 (16)	114	121 (16)	119	103 (15)	108	98 (16)	97	099 (16)	97
	F1 (Hz)	**F**	415 (57)	426	446 (51)	438	720 (55)	714	793 (117)	734	445 (45)	422	478 (52)	461	696 (52)	688
		M	343 (36)	332	367 (43)	367	596 (58)	575	612 (65)	581	364 (30)	350	387 (38)	388	564 (67)	545
	F2 (Hz)	**F**	2582 (148)	2560	2427 (189)	2445	2093 (105)	2061	2014 (165)	2002	1289 (261)	1165	1221 (221)	1151	1632 (100)	1625
		M	2205 (193)	2187	2115 (187)	2050	1784 (148)	1723	1768 (143)	1731	1195 (149)	1140	1123 (137)	1049	1496 (82)	1478
	Vowel		**/i/**		**/ɪ/**		**/ɛ/**		**/æ/**		**/u/**		**/ʊ/**		**/ʌ/**	
			Mean (SD)	**Mdn**	**Mean (SD)**	**Mdn**	**Mean (SD)**	**Mdn**	**Mean (SD)**	**Mdn**	**Mean (SD)**	**Mdn**	**Mean (SD)**	**Mdn**	**Mean (SD)**	**Mdn**
CG	**Dur (ms)**	**F**	87 (14)	87	83 (15)	90	110 (19)	110	113 (018)	114	89 (21)	91	82 (14)	85	91 (19)	88
		M	91 (11)	84	88 (11)	80	114 (12)	109	115 (13)	112	97 (01)	94	93 (11)	92	97 (12)	92
	F1 (Hz)	**F**	419 (28)	407	477 (76)	461	690 (66)	681	738 (120)	709	453 (39)	444	477 (54)	454	690 (68)	670
		M	322 (34)	298	329 (27)	313	560 (30)	572	569 (36)	575	37 (61)	357	362 (55)	361	546 (40)	534

Vowel		/i/ Mean (SD)	Mdn	/ɪ/ Mean (SD)	Mdn	/ɛ/ Mean (SD)	Mdn	/æ/ Mean (SD)	Mdn	/u/ Mean (SD)	Mdn	/ʊ/ Mean (SD)	Mdn	/ʌ/ Mean (SD)	Mdn
F2 (Hz)	F	2643 (108)	2596	2440 (107)	2424	2097 (110)	2044	2021 (115)	1996	1336 (148)	1256	1333 (113)	1326	1734 (114)	1699
	M	2130 (169)	2049	2108 (167)	2046	1780 (135)	1716	1795 (127)	1719	1033 (79)	1043	996 (79)	993	1503 (11)	1500

Note. F=female speakers; M=male speakers; ms=milliseconds; Hz=Hertz. The standard deviations are between parentheses.

Table 56

Averages of AE Vowel Duration, F1 and F2 values at Posttest

	Vowel	Sex	/i/ Mean (SD)	Mdn	/ɪ/ Mean (SD)	Mdn	/ɛ/ Mean (SD)	Mdn	/æ/ Mean (SD)	Mdn	/u/ Mean (SD)	Mdn	/ʊ/ Mean (SD)	Mdn	/ʌ/ Mean (SD)	Mdn
EG	Dur (ms)	F	104 (26)	100	108 (26)	101	128 (21)	116	131 (20)	121	110 (25)	106	107 (24)	102	107 (25)	100
		M	99 (18)	100	101 (017)	103	115 (17)	110	124 (20)	122	105 (19)	101	101 (18)	98	103 (17)	103
	F1 (Hz)	F	417 (49)	406	504 (29)	510	742 (58)	732	824 (83)	787	439 (39)	428	549 (57)	536	707 (62)	714
		M	352 (42)	344	447 (38)	438	604 (48)	601	656 (88)	615	385 (33)	372	446 (44)	438	580 (56)	580
	F2 (Hz)	F	2600 (186)	2575	2264 (123)	2262	2070 (102)	2049	1966 (167)	1979	1303 (212)	1336	1337 (125)	1346	1633 (124)	1621
		M	2188 (226)	2157	1908 (144)	1929	1760 (126)	1751	1724 (128)	1675	1209 (129)	1179	1213 (130)	1191	1472 (89)	1449
	Vowel	**Sex**	**/i/ Mean (SD)**	**Mdn**	**/ɪ/ Mean (SD)**	**Mdn**	**/ɛ/ Mean (SD)**	**Mdn**	**/æ/ Mean (SD)**	**Mdn**	**/u/ Mean (SD)**	**Mdn**	**/ʊ/ Mean (SD)**	**Mdn**	**/ʌ/ Mean (SD)**	**Mdn**
CG	Dur (ms)	F	81 (22)	82	078 (20)	77	100 (25)	95	105 (26)	105	83 (021)	90	79 (19)	83	84 (212)	86
		M	79 (01)	74	75 (01)	69	105 (12)	99	105 (01)	99	85 (01)	80	83 (01)	82	89 (11)	86
	F1 (Hz)	F	423 (45)	408	476 (84)	448	705 (63)	698	770 (138)	706	435 (56)	416	475 (78)	444	662 (98)	669
		M	332 (48)	295	344 (52)	310	559 (38)	575	562 (50)	572	355 (57)	323	367 (53)	335	540 (45)	545
	F2 (Hz)	F	2610 (136)	2596	2433 (91)	2393	2112 (119)	2108	1993 (131)	1937	1373 (182)	1315	1322 (162)	1310	1715 (122)	1691
		M	2092 (164)	1974	2063 (162)	1942	1762 (111)	1685	1751 (93)	1701	1081 (49)	1087	1031 (62)	1055	1498 (24)	1483

Note. F= female speakers; M=male speakers; ms=milliseconds; Hz= Hertz. The standard deviations are between parentheses.

Table 57

Averages of AE Vowel Duration, F1 and F2 values at Delayed Posttest

	Vowel	Sex	/i/ Mean (SD)	Mdn	/ɪ/ Mean (SD)	Mdn	/ɛ/ Mean (SD)	Mdn	/æ/ Mean (SD)	Mdn	/u/ Mean (SD)	Mdn	/ʊ/ Mean (SD)	Mdn	/ʌ/ Mean (SD)	Mdn
EG	Dur (ms)	F	105 (32)	95	103 (26)	103	121 (25)	113	127 (24)	124	101 (28)	98	106 (26)	114	107 (27)	103
		M	94 (13)	87	97 (13)	93	109 (01)	106	120 (15)	116	98 (19	96	96 (15)	94	96 (13)	92
	F1 (Hz)	F	405 (39)	405	506 (30)	509	739 (53)	727	834 (94)	817	441 (25)	441	566 (59)	549	713 (88)	697
		M	354 (38)	359	461 (40)	475	600 (53)	595	646 (72)	640	383 (27)	392	450 (49)	452	571 (54)	583
	F2 (Hz)	F	2616 (187)	2563	2231 (138)	2233	2072 (139)	2066	1959 (118)	1986	1299 (217)	1292	1314 (91)	1293	1612 (185)	1622
		M	2168 (206)	2119	1901 (117)	1912	1767 (123)	1726	1722 (118)	1694	1199 (120)	1168	1218 (144)	1213	1493 (92)	1491

	Vowel	Sex	/i/		/ɪ/		/ɛ/		/æ/		/u/		/ʊ/		/ʌ/	
			Mean (SD)	**Mdn**	**Mean (SD)**	**Mdn**	**Mean (SD)**	**Mdn**	**Mean (SD)**	**Mdn**	**Mean (SD)**	**Mdn**	**Mean (SD)**	**Mdn**	**Mean (SD)**	**Mdn**
CG	**Dur (ms)**	**F**	74 (14)	74	76 (19)	74	96 (14)	100	99 (20)	104	79 (19)	83	75 (21)	75	80 (17)	82
		M	83 (16)	74	80 (18)	71	110 (24)	97	108 (24)	95	91 (24)	75	86 (20)	75	92 (20)	081
	F1 (Hz)	**F**	412 (34)	407	511 (85)	500	707 (58)	711	800 (137)	795	458 (73)	433	514 (83)	521	672 (94)	644
		M	342 (41)	318	355 (33)	333	553 (30)	550	558 (40)	556	375 (55)	358	391 (49)	370	539 (24)	533
	F2 (Hz)	**F**	2651 (110)	2623	2423 (132)	2380	2076 (99)	2040	1962 (104)	1914	1345 (151)	1413	1326 (117)	1284	1708 (134)	1718
		M	2091 (167)	1984	2056 (160)	1948	1735 (119)	1680	1744 (117)	1674	1053 (74)	1083	1033 (98)	1057	1476 (13)	1467

Note. F=female speakers; M=male speakers; ms=milliseconds; Hz=Hertz. The standard deviations are between parentheses.

Following the same procedure as for EP vowels, the productions of the English vowels were first normalized to eliminate gender differences (see normalized vowel F1 and F2 values in Tables AI1-3). In addition, to have comparable vowel spaces, the productions of L2 learners were normalized according to the minimum and maximum formant values of the vowels produced by native American English speakers. The normalized F1 and F2 mean values were plotted as well as standard deviations, which were illustrated by ellipses around the vowel symbol, so as to show the vowel spaces of the L2 participants in the experimental and control groups in the three test phases (see Appendix AJ).

When examining the vowel spaces for each group per production test, we observed that in the pretest the two high back vowels (/ʊ/-/u/) overlapped in both groups, but in particular in the control group. The same pattern of overlap was visible for the mid-low front contrast (/ɛ/-/æ/), since both vowels were produced with similar F1 and F2 values. The high front vowels (/i/-/ɪ/) also overlapped, but there was some distance between them in the vowel space. Vowel /i/ was higher and more fronted than /ɪ/. In the posttest, the trainees exhibited a different vowel space from the pretest. There was no overlap between the two high front vowels due to the increase of the F1 value of vowel /ɪ/ from 319 Hz in the pretest to 383 Hz in the posttest, that is, the vowel lowered. Although slightly overlapped, the high back vowels were more distant from each other, which was caused by an F1 rise of vowel /ʊ/ from 345 Hz to 399 Hz in the posttest. Moreover, there was an upward movement of vowel /ɛ/ which was caused by a decrease of F1 from 594 Hz to 573 Hz (see Tables AI1-3). The same changes in F1 values were not found in the control group. The delayed posttest vowel space of the trainees was similar to that of the posttest. The high front vowels were quite distant from each other. Additionally, the back vowels were also partially overlapped, but some distance separated them. Still regarding the results of the delayed posttest, the vowel space of the controls was similar to that of the posttest; however, some movements occurred between these two test phases. Both back vowels moved downwards. The first formant of vowel /ʊ/ increased from 345 Hz in the posttest to 377 Hz in the delayed posttest, and the F1 of vowel /u/ rised from 320 Hz to 345 Hz. The high front lax vowel /ɪ/ moved downwards with an F1 increase from 333 Hz to 355 Hz (see Tables AI1-3).

In addition, medians of F1 and F2 values were plotted for the seven AE vowels in each production test (see Figures 44-46). Each vowel plot includes the F1 and F2 medians of vowels produced by both the trainees and the controls.

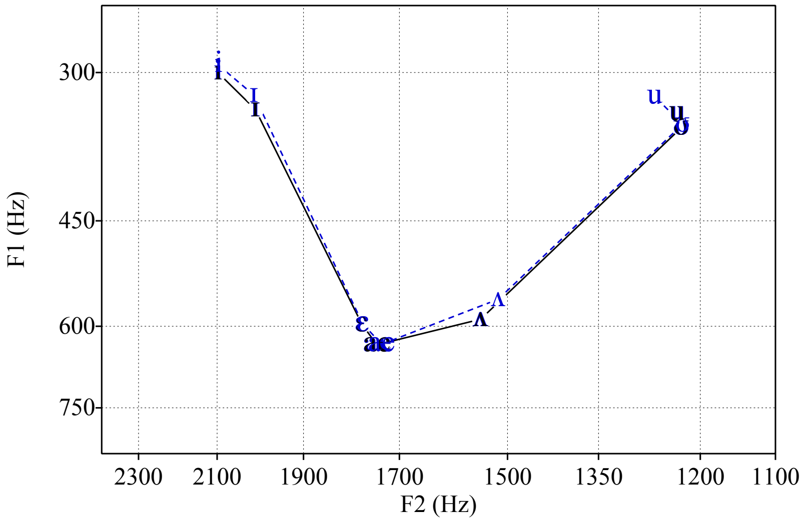


Figure 44. Vowel space of the experimental group (dashed line) and the control group (solid line) at pretest.

Figure 44 shows that, at pretest, the participants of both groups had equivalent vowel spaces before training, that is, the AE vowels were equally distributed in terms of height and frontness/backness. There was overlap of vowels /i/, /ɛ/, /æ/, and /ʊ/, partial overlap of vowel /ɪ/, and no overlap of vowels /u/ and /ʌ/, although they were very close.

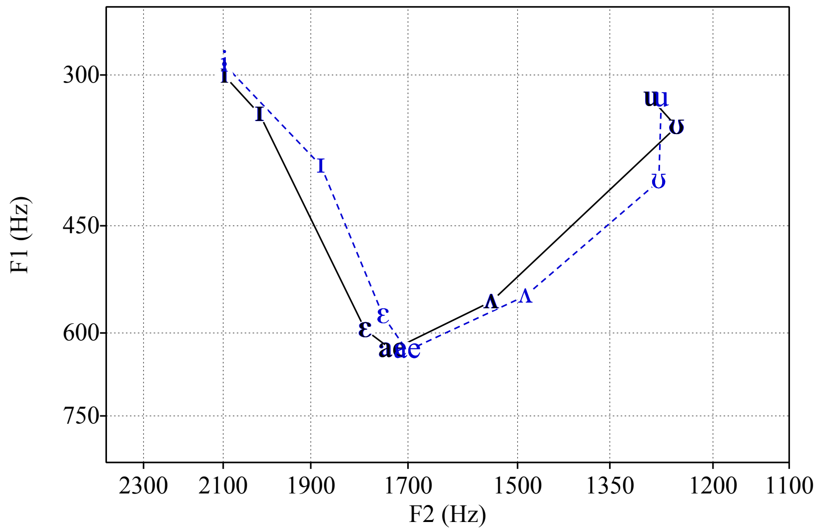


Figure 45. Vowel space of the experimental group (dashed line) and the control group (solid line) at posttest.

Intergroup differences in terms of L2 learners' vowel space were shown immediately after perceptual training was completed. The main differences were the previously described downward movement of /ɪ/ and /ʊ/ exhibited by the trainees. This movement contributed to increase the acoustic distance between these two lax vowels and their tense counterparts /i/ and /u/. These changes were not observed in the controls' productions, although vowel /ɪ/ lowered, to some extent, given the increase of the median F1 value from 307 Hz to 323 Hz (see Tables AI1-2). Speakers in the experimental group pronounced vowel /ɛ/ with a lower median F1 value than controls, which caused the rising of this vowel in the vowel space, consequently increasing the distance between /ɛ/ and /æ/.

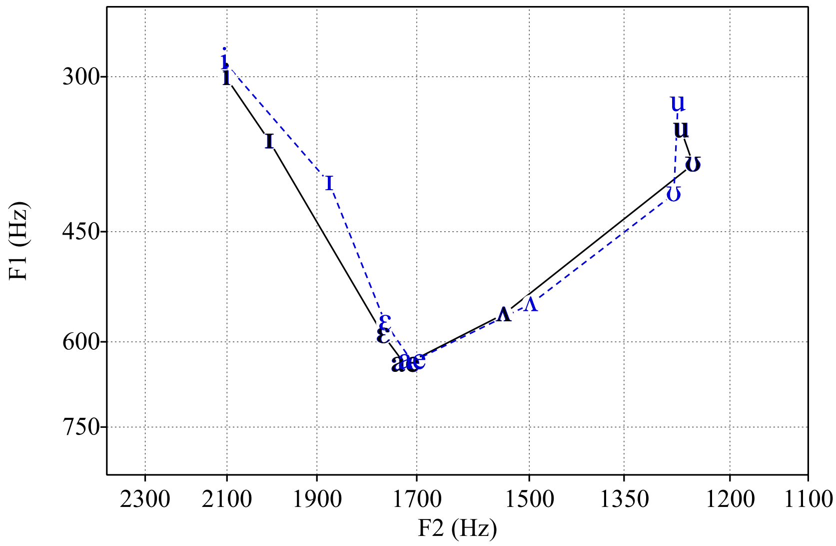


Figure 46. Vowel space of the experimental group (dashed line) and the control group (solid line) at delayed posttest.

Eight weeks later, at delayed posttest, similar between-group differences were found regarding median F1 values of vowels /i/, /ɪ/, /ɛ/, /u/, and /ʊ/. Vowels /ɪ/ and /ʊ/ were pronounced with higher F1 values by trainees than by controls, which contributed to the enhancement of the acoustic distance between their tense counterparts. Controls produced vowels /i/, /u/ and /ɛ/ with higher F1 values than the experimental participants. Therefore, these vowels, produced by controls, were lower than the same vowels pronounced by trainees. The distance between the distractor vowel /ʌ/ and /æ/ was also greater in the experimental group than in the control group, being /ʌ/ produced slightly higher and more posterior than /æ/.

To further examine within-group differences in the three production tests, the median F1 and F2 values of the American English vowels were plotted for each group in the three production tests (see Figures 47 and 48).

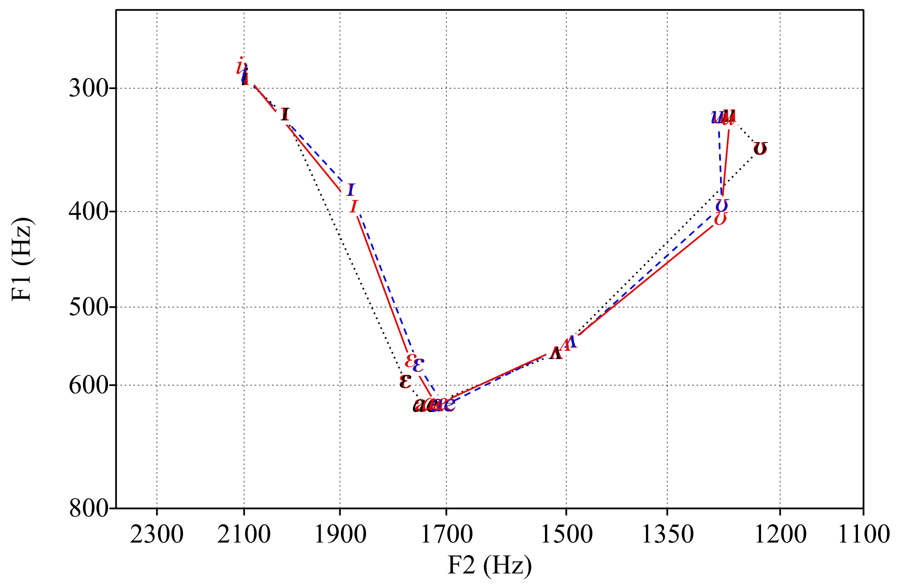


Figure 47. Vowel space of the experimental group at pretest (dotted line), posttest (dashed line) and delayed posttest (solid line).

When comparing the three production tests, we can observe that some movement occurred for three vowels, namely /ɪ/, /ɛ/, and /ʊ/. The high vowels /ɪ/ and /ʊ/ increased their F1 values, which caused their lowering in the vowel space, thus enlarging their distance in relation to their correspondent tense vowels /i/ and /u/. Vowel /ɛ/ also moved to a higher position within the vowel space with the decreasing of the first formant value. Although the most apparent changes were observed from pretest to posttest, some movement was still observed from posttest to delayed posttest for the same vowels.

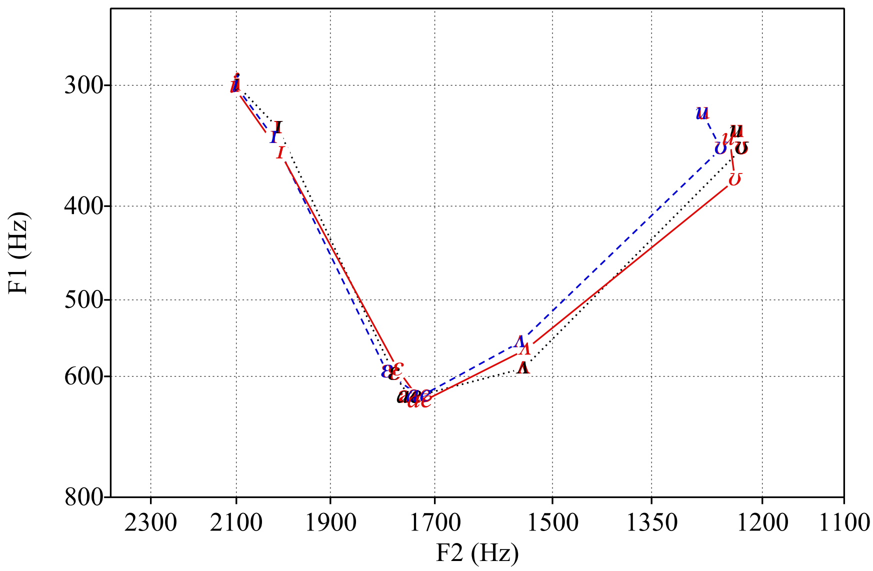


Figure 48. Vowel space of the control group at pretest (dotted line), posttest (dashed line) and delayed posttest (solid line).

As regards the control group, although there was an increase in the acoustic distance between /i/ and /ɪ/, and /u/ and /ʊ/, due to a lowering of the lax vowels, in particular in the delayed posttest, the intragroup differences were not as noticeable as for the experimental group (see Tables AI1-3).

Given that the focus of production analyses was not the comparison of absolute values between and within the two groups of EFL learners, no statistical tests per vowel were run. Rather, statistical tests were carried out for Euclidean distances (measured in Hertz) between vowels of five contrasts (/i/-/ɪ/, /æ/-/ɛ/, /æ/-/ʌ/, /u/-/ʊ/, and /ʌ/- /ʊ/).

4.2.3 Exploratory analysis of production data

In order to analyze the distribution of production data, K-S and Shapiro-Wilk tests were run for both Euclidean distances and duration ratios, and Levene's tests were carried out to verify homogeneity of variance within both groups of participants.

The results of the K-S and Shapiro-Wilk tests revealed that the Euclidean distance of a few vowel pairs were not normally distributed (viz. /i/-/ɪ/, and /æ/-/ɛ/ in the pretest of the experimental group, and /i/-/ɪ/ and /æ/-/ɛ/ in the pre-, posttest and delayed posttest of the control group); thus, both non-parametric and parametric tests were run for these variables. In terms of homogeneity of variance, significantly different variances were found in the experimental and control groups for the ED of the back vowel pair in the posttest and delayed posttest, $F(1, 32)=6.15$, $p < .05$, and $F(1, 28)=7.55$, $p < .05$, respectively. Hence, a non-parametric test to compare production between both groups was applied together with a parametric test.

The results of the Levene's tests for duration ratios showed that variances were equal for both groups. However, the distribution of duration ratios of the vowel pair /æ/-/ɛ/ was not normal in the posttest and delayed posttest in the case of the experimental group, and the same vowel pair was not normally distributed in the controls' pretest and posttest, along with the high front pair in the posttest. Given that conclusions from both types of tests were the same, we decided to report the parametric tests' results, as advised by Fife-Schaw (2006).

4.2.4 Results of the production pretest and posttest

4.2.4.1 Euclidean distance

To compare the Euclidean distances between the vowels of five contrasts (/i/-/ɪ/, /æ/-/ɛ/, /æ/-/ʌ/, /u/-/ʊ/, and /ʌ/-/ʊ/) produced before and after perceptual training sessions and assess the effect of training, paired-samples *t* tests were run. Differences in terms of Euclidean distances were calculated by subtracting the posttest from the pretest EDs. Results are presented in Table 58.

Table 58

T Test Results of Within-group Comparisons of ED

Group	Vowel Contrast	Pretest Mean (SD)	Posttest Mean (SD)	Difference (Hz)	*t* (df)
EG (*n*=22)	**/æ/-/ɛ/**	54.71 (72.57)	86.20 (74.18)	31.49	-3.18(21) **
	/i/-/ɪ/	98.58 (109.22)	239.96 (96.67)	141.38	-5.97(21) ***
	/u/-/ʊ/	75.18 (51.50)	127.94 (88.35)	49.76	-2.58(21) *
	/æ/-/ʌ/	249.15 (58.25	237.79 (63.21)	-11.36	.80(21)
	/ʌ/-/ʊ/	354.34 (69.66)	330.54 (66.78)	-23.8	2.92(21) **
CG (*n*=12)	**/æ/-/ɛ/**	59.38 (66.20)	76.70 (87.18)	17.32	-1.38(11)
	/i/-/ɪ/	100.84 (124.17)	90.49 (98.17)	-10.35	.80(11)
	/u/-/ʊ/	41.52 (27.79)	62.90 (36.29)	21.38	-2.33(11) *
	/æ/-/ʌ/	207.49 (63.26)	209.47 (84.39)	2.25	-.12(11)
	/ʌ/-/ʊ/	393.08 (92.69)	353.76 (94.44)	-39.32	2.39(11) *

Note. EG=Experimental group; CG=Control group; SD=Standard Deviation; *t*= result of the *t* test; (df)=degrees of freedom; levels of significance: * $p < .05$; ** $p < .01$; *** $p < .001$.

Euclidean distances between the vowels of the three target pairs increased significantly from pretest to posttest for the participants in the experimental group. The ED of the high front vowel pair was the one to increase the most, followed by the high back vowel contrast and then the mid-low front pair. The ED between the distractor vowel /ʌ/ and /ʊ/ decreased significantly. Differently, the control group produced the high back contrast with a significantly larger ED, but the ED of the vowel pair /ʌ/-/ʊ/ also decreased significantly in the posttest.

4.2.4.2 Duration

Duration ratios of the target vowel pairs were calculated by dividing the mean duration values (ms) of /i/, /u/ and /æ/ by the mean duration of /ɪ/, /ʊ/, and /ɛ/, respectively, as in Wang (2008). In addition, the duration ratio of /æ/-/ʌ/ was also calculated. As Wang (2008) explains, “the higher the values of the ratios are, the greater the differences between the two vowels of the pair are” (p. 124). Therefore, the duration ratios reflect vowel duration differences between the two vowels of each pair. Table 59 presents the results of the paired-samples *t* tests that were carried out to verify whether there were differences regarding duration ratios between pretest and posttest.

Table 59

T Test Results of Within-group Comparisons of Duration Ratios

Group	Vowel Contrast	Pretest Mean (SD)	Posttest Mean (SD)	*t* (df)
EG (*n*=22)	**/æ/-/ɛ/**	1.05 (.04)	1.06 (.08)	-.62(21)
	/i/-/ɪ/	1.04 (.07)	.98 (.11)	2.99(21) **
	/u/-/ʊ/	1.06 (.07)	1.04 (.09)	1.50(21)
	/æ/-/ʌ/	1.23 (.10)	1.23 (.14)	.003(21)
CG (*n*=12)	**/æ/-/ɛ/**	1.03 (.04)	1.04 (.08)	-.53(11)
	/i/-/ɪ/	1.05 (.05)	1.04 (.093)	.41(11)
	/u/-/ʊ/	1.07 (.08)	1.04 (.07)	.75(11)
	/æ/-/ʌ/	1.24 (.10)	1.23 (.13)	.17(11)

Note. EG= Experimental group; CG = Control group; SD = Standard Deviation; *t*= result of the *t* test; (df) = degrees of freedom; levels of significance: * $p < .05$; ** $p < .01$; *** $p < .001$.

Except for vowel contrast /i/-/ɪ/, for which the duration ratio decreased significantly from pretest to posttest for the trainees' group, no significant differences were found for the other vowel contrasts in both groups. To understand these results better, we ran *t* tests for each vowel segment to compare vowel length between the pre- and posttests, and we also compared duration between vowels of the target contrasts at each test phase. Contrary to expectations, vowel /ɪ/ became significantly longer after training (t=*-2.16*(21), p <.05), and vowel /i/ was shorter than /ɪ/ at posttest, though the difference was not significant (t=*-1.27*(21), p>.05).

4.2.5 Results of the production pretest, posttest and delayed posttest

4.2.5.1 Euclidean distance

Further analyses were carried out to investigate whether there was an effect of auditory training on the Euclidean distances between the vowels of the five vowel contrasts in the three production tests. Due to a participant droupout rate of 11.76% in the delayed posttest, analyses were carried out with 30 participants. The results of the repeated measures analysis of variance (ANOVA) are displayed in Table 60.

Table 60

ANOVA Results of Within-group Comparisons of EDs

Group	Vowel contrast	Pretest Mean (SD)	Posttest Mean (SD)	Delayed Posttest Mean (SD)	*F* (df)
EG (*n*=19)	**/æ/-/ɛ/**	57.33 (77.82)	77.49 (72.79)	90.81 (64.27)	1.48(1.18, 21.15)
	/i/-/ɪ/	108.34 (114.70)	245.49 (97.67)	260.32 (89.31)	13.8(2,36) ***
	/u/-/ʊ/	76.90 (52.88)	142.11 (86.66)	150.40 (79.71)	4.94(2,36) *
	/æ/-/ʌ/	245.61 (60.04)	242.39 (65.55)	237.60 (78.61)	125(2,36)
	/ʌ/-/ʊ/	357.77 (71.92)	333.52 (69.68)	347.32 (86.79)	723(1.21, 21.77)
CG (*n*=11)	**/æ/-/ɛ/**	61.21 (69.11)	68.55(86.50)	75.02 (76.70)	369(1.29, 12.87)
	/i/-/ɪ/	83.69 (114.37)	79.94 (95.56)	114.42 (110.80)	2.80(2,20)
	/u/-/ʊ/	43.14 (28.55)	59.31 (35.76)	62.55 (34.96)	2.79(2,20)
	/æ/-/ʌ/	215.23 (60.10)	219.80 (80.16)	206.10 (69.23)	162 (2,20)
	/ʌ/-/ʊ/	391.84 (97.11)	353.92 (99.05)	355.04 (102.67)	2.26 (2,20)

Note. EG=Experimental group; CG=Control group; SD=Standard Deviation; *F*= result of the ANOVA; (df)=degrees of freedom; levels of significance: * $p < .05$; ** $p < .01$; *** $p < .001$.

ANOVA results showed that auditory training had a significant effect on production performance of two vowel pairs by trainees. The EDs of the two high vowel contrasts (/i/-/ɪ/, and /u/-/ʊ/) increased significantly. However, no substantial effect was found in the controls' Euclidean distance of any AE vowel pair. A more detailed analysis was performed with Bonferroni post hoc tests (see Table 61).

Table 61

Bonferroni Pairwise Comparisons of EDs

ED	Vowel Contrast	Pretest vs. Posttest	Pretest vs. Delayed Posttest	Posttest vs. Delayed Posttest
EG	**/æ/-/ɛ/**	†	ns	ns
	/i/-/ɪ/	***	**	ns
	/u/-/ʊ/	*	*	ns
	/æ/-/ʌ/	ns	ns	ns
	/ʌ/-/ʊ/	ns	ns	ns
CG	**/æ/-/ɛ/**	ns	ns	ns
	/i/-/ɪ/	ns	ns	ns
	/u/-/ʊ/	ns	ns	ns
	/æ/-/ʌ/	ns	ns	ns
	/ʌ/-/ʊ/	ns	ns	ns

Note. EG=Experimental group; CG=Control group; levels of significance: † <.1; * $p < .05$; ** $p < .01$; *** $p < .001$.

As reported earlier, the trainees produced significantly greater EDs for vowel pairs /i/-/ɪ/ and /u/-/ʊ/ in the posttest and delayed posttest than in the pretest (see Figure 49). The ED of /æ/-/ɛ/ was also considerably larger immediately after perceptual training. As regards controls, no significant differences in EDs were found for any pair (see Figure 50). Table 62 presents the results of ED differences, calculated by subtracting the EDs of the posttest from the pretest, of the delayed posttest from the pretest, and of the delayed posttest from the posttest.

Table 62

Pairwise Comparisons of ED Differences

ED	Vowel Contrast	Pretest vs. Posttest	Pretest vs. Delayed Posttest	Posttest vs. Delayed Posttest
EG	**/æ/-/ɛ/**	20.16	33.48	13.32
	/i/-/ɪ/	137.15	151.98	14.83
	/u/-/ʊ/	65.21	73.5	8.29
	/æ/-/ʌ/	-3.22	-8.01	-4.79
	/ʌ/-/ʊ/	-24.25	-10.45	13.8
CG	**/æ/-/ɛ/**	7.34	6.47	13.81
	/i/-/ɪ/	-3.75	34.48	30.73
	/u/-/ʊ/	16.17	3.24	19.41
	/æ/-/ʌ/	4.57	-13.7	-9.13
	/ʌ/-/ʊ/	-37.92	1.12	-36.8

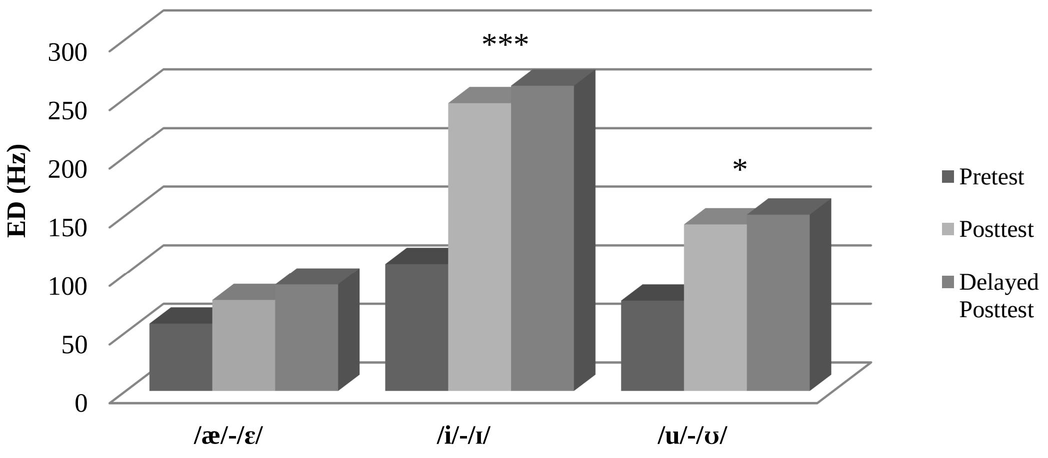


Figure 49. EDs of the target vowel pairs for the experimental group.

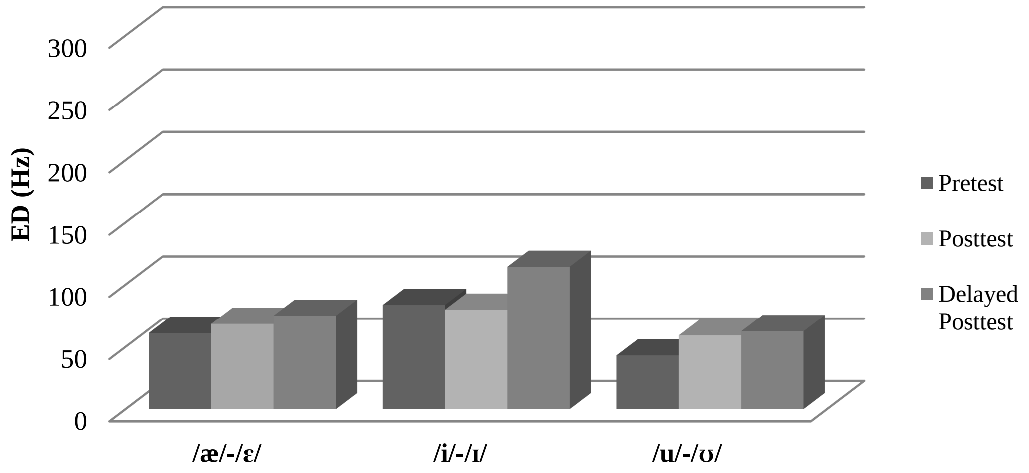


Figure 50. EDs of the target vowel pairs for the control group.

To examine whether there were intergroup differences in terms of production performance, *t* tests for independent samples were carried out for each test phase.

Table 63

T Test Results of Between-group Comparisons of Vowel EDs

ED	Vowel Contrast	EG (*n*=22) Mean (SD)	CG (*n*=12) Mean (SD)	*t* (df)
Pretest	**/æ/-/ɛ/**	54.71 (72.57)	59.38 (66.20)	-.19(32)
	/i/-/ɪ/	98.58 (109.22)	100.84 (124.17)	-.06(32)
	/u/-/ʊ/	75.18 (51.50)	41.52 (27.79)	2.09(32) *
	/æ/-/ʌ/	249.15 (58.25	207.49 (63.26)	1.93(32)
	/ʌ/-/ʊ/	354.34 (69.66)	393.08 (92.69)	-1.38(32)
Posttest	**/æ/-/ɛ/**	86.20 (74.18)	76.70 (87.18)	.34(32)
	/i/-/ɪ/	239.96 (96.67)	90.49 (98.17)	4.29(32) ***
	/u/-/ʊ/	127.94(88.35)	62.90 (36.29)	2.42(32) *
	/æ/-/ʌ/	237.79 (63.21)	209.47 (84.39)	1.11(32)
	/ʌ/-/ʊ/	330.54 (66.78)	353.76 (94.44)	-.84(32)
		EG (*n*=19) Mean (SD)	**CG (*n*=11) Mean (SD)**	***t* (df)**
Delayed Posttest	**/æ/-/ɛ/**	90.81 (64.27)	75.02 (76.70)	.60(28)
	/i/-/ɪ/	260.32 (89.31)	114.42 (110.80)	3.95(28) ***
	/u/-/ʊ/	150.40 (79.71)	62.55 (34.96)	3.45(28) **
	/æ/-/ʌ/	237.60 (78.61)	206.10 (69.23)	1.10(28)
	/ʌ/-/ʊ/	347.32 (86.79)	355.04 (102.67)	-.22(28)

Note. EG=Experimental group; CG=Control group; SD=Standard Deviation; *t*= result of the *t* test; (df)=degrees of freedom; levels of significance: * *p* < .05; ** *p* < .01; *** *p* < .001.

The results of the *t* tests showed that, before training, the only significant intergroup difference was in the production of the back vowel contrast, which trainees pronounced with a significantly greater ED than the controls. Given that only the perception pretest results were used to assign the 34 participants to two homogeneous groups in terms of perceptual ability, differences regarding pronunciation were not taken into account at pretest.

The experimental group produced the high back vowel contrast with a significantly greater ED than the control group in the three production tests (see Figures 51-53). After training, in the posttest and delayed posttest, the ED of the high front vowel contrast was also greater when produced by the trainees. Of the three target vowel contrasts, only the ED of /æ/-/ɛ/ did not differ significantly between groups. However, when compared with native speakers' productions, the ED of this contrast was also the smallest in relation to the other two vowel pairs.

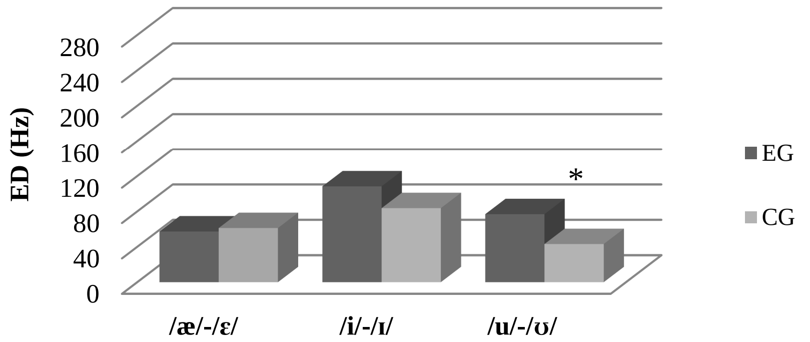


Figure 51. EDs of the three vowel pairs in the production pretest for both groups.

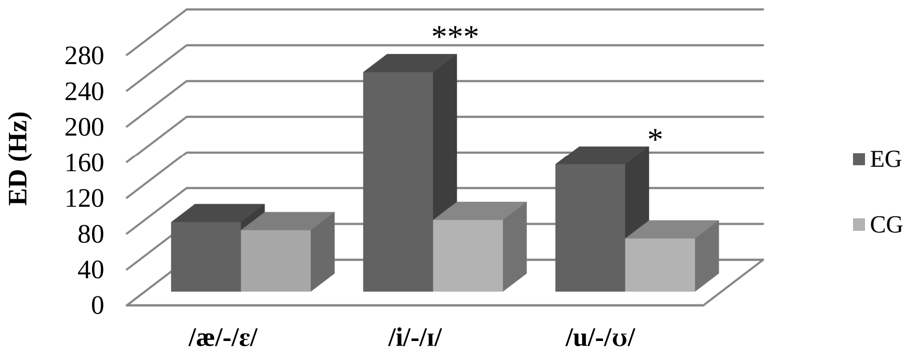


Figure 52. EDs of the three vowel pairs in the production posttest for both groups.

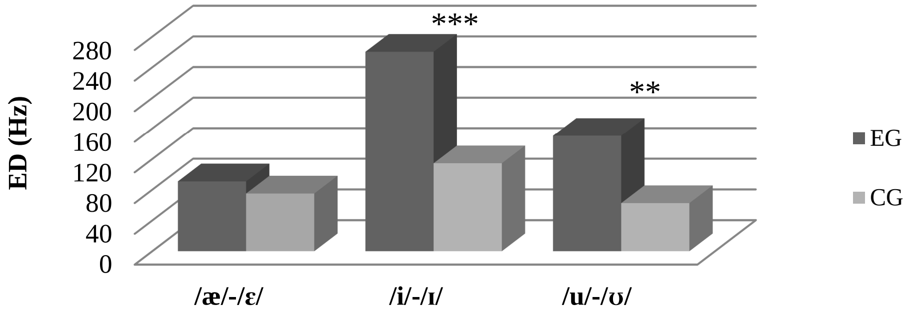


Figure 53. EDs of the three vowel pairs in the production delayed posttest for both groups.

To understand pronunciation accuracy of EFL learners in the production of AE vowels better, we compared EDs among three groups of speakers, namely the two groups of L2 speakers and a group of native American English speakers. The acoustic measurements (duration, F1 and F2 values) of a group of nine AE NSs reported in Rauber (2010) were used, with permission from the author. These values were used as reference to carry out statistical tests. One-way ANOVAs were run to examine intergroup differences in terms of EDs. Descriptive statistical information is provided for the group of AE NSs together with the ANOVA results for each production test in Table 64.

Table 64

One-way ANOVA Results of Intergroup Comparisons of Vowel EDs

	/i/-/ɪ/	**/æ/-/ɛ/**	**/u/-/ʊ/**	**/æ/-/ʌ/**	**/ʌ/-/ʊ/**
Mean and *SD* of ED NS (*n*=9)	386.36 (67.45)	109.19 (32.32)	219.74 (124.64)	162.65 (69.51)	298.11 (69.96)
			Pretest		
***F* (df)**	25.68(2,40) ***	2.40(2,40)	19.42(2,40) ***	6.52(2,40) **	3.94(2,40) *
			Posttest		
***F* (df)**	26.92(2,40) ***	.54(2,40)	8.36 (2,40) **	3.63 (2,40) *	1.38 (2,40)
			Delayed Posttest		
***F* (df)**	22.09(2,36) ***	.735(2,36)	8.98(2,36) **	3.16(2,36) †	1.23(2,36)

Note. NNs=native speakers; SD=Standard Deviation; *F*= result of the ANOVA; (df)=degrees of freedom; levels of significance: † <.1; * *p* < .05; ** *p* < .01; *** *p* < .001.

At pretest, significant differences were found for four of the five contrasts. The only vowel contrast that did not differ significantly was /æ/-/ɛ/. In fact, this pair did not differ significantly across groups in the three tests. This indicates that the ED of this vowel contrast as produced by L2 participants did not differ significantly from the AE speakers. This was expected, to some extent, because the native ED between the two vowels of this pair is not as large as in the other vowel contrasts analyzed in this study. Given that the number of participants differed across groups, we chose to run Gabriel's post hoc test to compare groups' EDs in the three testing moments (Field, 2009, p. 374). Results are presented in Table 65.

Table 65

Gabriel's Pairwise Test Results for Vowel EDs

	Vowel Contrast	**Native vs. Experimental**	**Native vs. Control**	**Experimental vs. Control**
Pretest	**/æ/-/ɛ/**	ns	ns	ns
	/i/-/ɪ/	***	***	ns
	/u/-/ʊ/	***	***	ns
	/æ/-/ʌ/	**	ns	ns
	/ʌ/-/ʊ/	ns	*	ns
Posttest	**/æ/-/ɛ/**	ns	ns	ns
	/i/-/ɪ/	**	***	***
	/u/-/ʊ/	*	**	ns
	/æ/-/ʌ/	*	ns	ns
	/ʌ/-/ʊ/	ns	ns	ns
Delayed Posttest	**/æ/-/ɛ/**	ns	ns	ns
	/i/-/ɪ/	**	**	***
	/u/-/ʊ/	ns	**	*
	/æ/-/ʌ/	*	ns	ns
	/ʌ/-/ʊ/	ns	ns	ns

Note. Levels of significance: * $p < .05$; ** $p < .01$; *** $p < .001$.

Taking the three target vowel contrasts into account, we observed that the ED of the high front pair was significantly greater when produced by the experimentals than by the controls in the two tests after training. In the delayed posttest, the ED of the high back contrast was also produced significantly larger by the participants in the experimental group than in the control group. As far as the natives' target vowel productions are concerned, their EDs were significantly larger than those of the participants in both groups, except for the pair /æ/-/ɛ/ in the three production tests. In the delayed posttest, no significant differences were found between the natives and the vowel-trainees in the production of the back and the mid-low front vowel pairs, which indicates that their productions of these vowel contrasts reached a native-like level. In sum, at the end of the study, L2 participants in the experimental group reached a native-like production performance for two of the three vowel contrasts. Moreover, though the ED of /i/-/ɪ/ still differed significantly from the L1 speakers', it improved significantly after training.

4.2.5.2 Duration

Duration ratios were further analyzed to verify the extent to which perception training affected vowel length and to compare the production of the target vowel contrasts between the two groups in the three tests. An increase of ratios should be expected if the length difference between the two vowels of each pair was greater after auditory training. The repeated measures ANOVA results are shown in Table 66.

Table 66

ANOVA Test Results of Duration Ratios

Group	Vowel Contrast	Pretest Mean (SD)	Posttest Mean (SD)	Delayed Posttest Mean (SD)	*F* (df)
EG (*n*=19)	/æ/-/ɛ/	1.04 (.05)	1.06 (.08)	1.08 (.11)	2.94(2,36) †
	/i/-/ɪ/	1.04 (.06)	.98 (.11)	.98 (.09)	5.82(2,36) **
	/u/-/ʊ/	1.05 (.05)	1.03 (.09)	.99 (.10)	3.51(2,36) *
	/æ/-/ʌ/	1.22 (.10)	1.24 (.15)	1.23 (.12)	.27(2,36)
CG (*n*=11)	/æ/-/ɛ/	1.03 (.04)	1.04 (.08)	1.01 (.08)	.76(2,20)
	/i/-/ɪ/	1.05 (.05)	1.03 (.08)	1.01 (.08)	1.93(2,20)
	/u/-/ʊ/	1.06 (.08)	1.03 (.06)	1.06 (.06)	.60(2,20)
	/æ/-/ʌ/	1.24 (.11)	1.23 (.13)	1.22 (.12)	.25(2,20)

Note. EG=Experimental group; CG=Control group; SD=Standard Deviation; *F*= result of the ANOVA; (df)=degrees of freedom; levels of significance: † <.1; * $p < .05$; ** $p < .01$; *** $p < .001$.

The ANOVA results revealed that perceptual training did not positively affect vowel duration ratios. In fact, the only significant effects found were negative, that is, the duration ratios of both high vowel contrasts decreased after training, which indicates that the durational difference between the vowels of these two pairs diminished. The only marginal difference observed was for the duration ratio of /æ/-/ɛ/ that was slightly higher in the posttest and delayed posttest. However, the result of the Bonferroni post hoc procedure did not reveal any significance for that vowel contrast (see Table 67). A significant decrease of duration ratios for the high front vowel pair was found between the pretest and both the posttest and delayed posttest. There was also a marginal significant decrease of the high back vowel duration ratio between pretest and delayed posttest.

Table 67

Bonferroni Pairwise Comparisons of Duration Ratios

Duration	**Vowel Contrast**	**Pretest vs. Posttest**	**Pretest vs. Delayed Posttest**	**Posttest vs. Delayed Posttest**
EG	**/æ/-/ɛ/**	ns	ns	ns
	/i/-/ɪ/	*	*	ns
	/u/-/ʊ/	ns	†	ns
	/æ/-/ʌ/	ns	ns	ns
CG	**/æ/-/ɛ/**	ns	ns	ns
	/i/-/ɪ/	ns	ns	ns
	/u/-/ʊ/	ns	ns	ns
	/æ/-/ʌ/	ns	ns	ns

Note. Levels of significance: † <.1; * $p < .05$; ** $p < .01$; *** $p < .001$.

The production performance of both trainees and controls was compared in the three test moments by means of *t* tests for independent samples (see Table 68).

Table 68

T Test Results of Between-subjects Comparisons of Duration Ratios

Test	**Vowel Contrast**	**EG (*n*=22) Mean (SD)**	**CG (*n*=12) Mean (SD)**	***t* (df)**
Pretest	**/æ/-/ɛ/**	1.05 (.04)	1.03 (.04)	1.28(32)
	/i/-/ɪ/	1.04 (.07)	1.05 (.05)	-.48(32)
	/u/-/ʊ/	1.06 (.07)	1.07 (.08)	-.051(32)
	/æ/-/ʌ/	1.23 (.10)	1.24 (.10)	-.086(32)
Posttest	**/æ/-/ɛ/**	1.06 (.08)	1.04 (.08)	.62(32)
	/i/-/ɪ/	.98 (.11)	1.04 (.093)	-1.65(32)
	/u/-/ʊ/	1.04 (.09)	1.04 (.07)	-.22(32)
	/æ/-/ʌ/	1.23 (.14)	1.23 (.13)	.02(32)
	Vowel Contrast	**EG (*n*=19) Mean (SD)**	**CG (*n*=11) Mean (SD)**	***t* (df)**
Delayed Posttest	**/æ/-/ɛ/**	1.08 (.11)	1.01 (.08)	2.02(28) †
	/i/-/ɪ/	.98 (.09)	1.01 (.08)	-.79(28)
	/u/-/ʊ/	.99 (.10)	1.06 (.06)	-2.02(28) †
	/æ/-/ʌ/	1.23 (.12)	1.22 (.12)	.23 (28)

Note. EG=Experimental group; CG=Control group; SD=Standard Deviation; *t*= result of the *t* test; (df)=degrees of freedom; levels of significance: † <.1; * $p < .05$; ** $p < .01$; *** $p < .001$.

No significant differences were found between the two groups of L2 participants. However, marginally significant differences were found for the duration ratios of vowel pairs /æ/-/ɛ/ and /u/-/ʊ/. The ratios were significantly higher for the former pair and significantly lower for the latter pair in the experimental group than in the control group.

Having as reference the American English NSs vowel duration values provided by Rauber (2010), one-way ANOVAs were carried out to compare duration ratios among the three groups of speakers. The first row of Table 69 shows the NSs' ratios and the other three rows present present the ANOVAs results.

Table 69

One-way ANOVA Results of Intergroup Comparisons of Duration Ratios

	/i/-/ɪ/	/æ/-/ɛ/	/u/-/ʊ/	/æ/-/ʌ/
Mean and *SD* of Dur NSs (*n*=9)	2.44 (.37)	1.50 (.27)	1.16 (.17)	1.32 (.13)
		Pretest		
***F* (df)**	222.00(2,40) ***	46.01(2,40) ***	2.98(2,40) †	2.42(2,40)
		Posttest		
***F* (df)**	200.88(2,40) ***	35.10(2,40) ***	4.38(2,40) *	1.60(2,40)
		Delayed Posttest		
***F* (df)**	29.13(2,36) ***	195.55(2,36) ***	6.48(2,36) **	2.04(2,36)

Note. NS=native speaker; SD=Standard Deviation; *F*= result of the ANOVA; (df)=degrees of freedom; levels of significance: † <.1; * *p* < .05; ** *p* < .01; *** *p* < .001.

The data from the one-way ANOVAs showed that the duration ratios among the three groups differed significantly in all vowel contrasts, except for the pair with the distractor vowel /ʌ/. Furthermore, at pretest, vowel duration differences were only marginally significant in the case of the high back pair. The explanation for these findings is associated with the high duration ratios of the NSs group.

Table 70

Gabriel's Pairwise Test Results for Vowel Duration Ratios

	Vowel Contrast	Native vs. Experimental	Native vs. Control	Experimental vs. Control
Pretest	/æ/-/ɛ/	***	***	ns
	/i/-/ɪ/	***	***	ns
	/u/-/ʊ/	ns	ns	ns
	/æ/-/ʌ/	ns	ns	ns
Posttest	/æ/-/ɛ/	***	***	ns
	/i/-/ɪ/	***	***	ns
	/u/-/ʊ/	*	ns	ns
	/æ/-/ʌ/	ns	ns	ns
Delayed Posttest	/æ/-/ɛ/	***	***	ns
	/i/-/ɪ/	***	***	ns
	/u/-/ʊ/	**	ns	ns
	/æ/-/ʌ/	ns	ns	ns

Note. Levels of significance: † <.1; * *p* < .05; ** *p* < .01; *** *p* < .001.

Gabriel's post hoc test revealed that there were no differences between trainees and controls (see Table 70), that is, their ratios were very similar. However, significant differences between the AE NSs and both groups were found for two of the vowel contrasts (/ɛ/-/æ/ and /i/-/ɪ/) in the three production tests. Duration ratios of the segmental contrast /u/-/ʊ/ did not differ between controls and natives in any of the three tests, whereas they did significantly between the vowel-trainees and the L1 speakers after training, but not before. This seems to indicate that training had an effect on the production ratios of trainees but in the unexpected direction, that is, instead of having increased, ratios decreased. This might be explained by the fact that their attention was redirected to vowel spectral differences. At the onset of training, L2 learners' seemed to rely more on duration than at its offset.

4.2.6 Analysis of individual trainees' production performance

When comparing Euclidean distances for each vowel contrast per participant, one of the first observations that has to be made is that there were very few cases in which the ED was greater in L2 learners' vowel productions than the EDs measured in AE NSs productions (see Table AK1). In the pretest, there was only one participant who produced vowels /ɛ/-/æ/ with an exaggerated ED of 339 Hz. In the posttest, the same participant (E8) continued to pronounce the same vowels with a greater ED than NSs, but surprisingly eight weeks after training the ED reduced substantially to 81 Hz. Another similar case was observed in the productions of participant E6, who also produced the vowels of the mid-low front contrast with a higher ED than the AE native speakers. Nonetheless, since this participant dropped out from the training program at the delayed posttest stage, no further examination could be carried out. After training, the ED of the back vowel pair, produced by participant E18, was also higher (383 Hz) than that measured for native speakers. In the delayed posttest, two cases of higher EDs were detected for vowel contrast //ɛ/-/æ/ in the production data of learners E3 and E11, and one case regarding the high front vowels in the production of participant E13. The back vowel pair was also produced with a higher ED by informants E3, E4, E11 and E14 than by AE native speakers. These cases of larger EDs can be interpreted as attempts to reach native-like production, given that by making the vowels of the target pairs more distant from each other in the vowel space, there is less chance of them being overlapped. Moreover, within this group of students, who produced the vowel contrasts with a greater ED than NSs, there were some that reached a native-like or near native-like perceptual performance, viz. participants E11, E13, and E14.

When comparing between the pretest and posttest, we can observe that overall ED distances increased, except in some cases regarding the back vowel contrast. Participants E2, E9, E14, and E16, who exhibited a near native-like perceptual ability in the perception tests, and participant E15 produced the back vowel pair after training with a smaller ED than before training. Two months after training, learners E2 and E9 produced this pair with an even smaller Euclidean distance. Conversely, learner E14, as previously mentioned, produced the vowel contrast with an exaggerated ED. Given that the other two participants (E15 and E16) dropped out, there were no data to analyze. Two EFL learners (E7 and E22), despite having produced the three vowel contrasts with greater EDs immediately after training, produced them with smaller EDs eight weeks later. In particular, in the puzzling case of E22, the production of the two front contrasts was less accurate than before training, since the EDs were even smaller, despite the improvement at the production posttest. Apart from these cases, the EDs of the other vowel contrasts increased after auditory training.

Given that production measures of EDs were greater than percentage ID scores and varied extensively among L2 speakers, no division of learners into different subgroups was made in terms of production performance. Furthermore, since no significant effect of training was found on vowel duration ratios, no individual analyses were carried out for this production measure but individual results can be seen in Table AK2.

4.3 Analysis of Follow-up Questionnaires

Immediately after the training program was completed, the experimental group answered a follow-up questionnaire to evaluate the training sessions. The overall information is summarized in Appendix AL, and individual answers are shown in Appendix AM.

The results of this questionnaire were consistent with the overall results of trainees' perceptual performance after training. The vowel contrast that EFL learners found easier to distinguish throughout the training was /i/-/ɪ/ (54.5%). However, it is important to mention that the vowel pair with the distractor vowel /æ/-/ʌ/ was not included as an option. This might have been the easiest contrast to learn, since the distractor vowel was, in the students' opinion, the easiest to identify during training (45.5%). According to trainees, the most difficult vowel contrast to learn was /u/-/ʊ/ (54.5%) and the most difficult vowel to identify was /ʊ/ (54.5%). In terms of auditory exercises, discrimination tasks (63.3%) were assessed as being easier than closed-set identification tasks. Overall, learners agreed that the duration of the training sessions was adequate (81.8%), and length of the training program was sufficient (81.8%) to learn the seven English vowels. Eleven students (50%) reported not feeling tired at the end of the training sessions, but nine (40.9%) indicated some degree of tiredness, though most of them informed that they did not feel much tired (27.3%). In terms of motivation, 15 students (68.2 %) stated that they were motivated, and five (22.7%) that they were highly motivated throughout the training. Finally, 16 students (72.7%) reported being concentrated, three (13.6%) highly concentrated and only one (4.5%) acknowledged not being concentrated.

At the end of the questionnaire, students were asked to give their opinion about two aspects: the software that was used in the experiment for perception training and testing, and the training sessions. *TP-S* software was assessed as being very user-friendly and practical. Two participants mentioned that the option of repeating the same stimuli (available in the training) should have also been available in the identification tests. However, this suggestion was not related to the software, but to test design. Some students also emphasized that they would have benefited if they could do the training exercises at home.

4.4 General Discussion

4.4.1 Cross-language vowel perception

The research objectives of this study pertaining to L2 vowel perception included examining whether: (1) perceptual training would have a positive effect on the identification of English vowels by Portuguese EFL learners; (2) potential perceptual learning would be generalized to new words and new speakers; (3) training effects would remain two months after training sessions were over; (4) learners would categorize the non-native vowels /ɪ/, /æ/, and /ʊ/ into the preexisting L1 segments /i/, /ɛ/, and /u/, or if they would create new phonemic categories for the L2 sounds.

The characteristics of the perceptual training adopted in this study are summarized next. The auditory training included naturally spoken stimuli (viz. the target vowels, embedded in CVC words, within four phonetic contexts) produced by twelve native American English speakers. Due to the high number of talkers, we described it as a high-variability phonetic training. Furthermore, it was a computer-based training insofar as a software application was developed to design perception tasks and tests to enhance learners' motivation and interest in training, as well as to provide visual corrective feedback both on a trial-by-trial (i.e., immediate) and a session-by-session (i.e., cumulative) basis. Moreover, in comparison with other vowel training studies (e.g., Nobre-Oliveira, 2007; Wang, 2008; Wang & Munro, 2004) it was not self-paced nor individualized, that is, a predefined number of training sessions was set to control for potential training effects. The training program was divided into two blocks. The first block consisted of three 45-minute training sessions, which included a discrimination task and a closed-set identification task, focused on each vowel contrast, and the second block included two other 45-minute sessions that were dedicated to the seven-vowel set. Although the whole program included five sessions, only three were devoted to each vowel pair; thus, the perception training was, to a great extent, short-termed in comparison with other vowel training experiments with a similar number of target vowels (e.g., Aliaga-Garcia & Mora, 2007; Lambacher et al., 2005; Nishi & Kewley-Port, 2007). Furthermore, the phonetic training was carried out in a formal EFL classroom environment, in particular within the *English Phonetics and Phonology* course, with Portuguese undergraduate students with an intermediate (B1) English proficiency level. This group of participants was pre- and post-tested (immediately after training was over and eight weeks later) on the identification of seven AE vowels and divided into two groups. Twenty-two students attended vowel-centered training sessions (experimental group), and twelve undertook consonant-centered training (control group). To assure that both groups were comparatively homogeneous in terms of English experience and had similar pre-training perceptual performance, two screening measures were taken into consideration, namely the background questionnaire data and the percentage of accurate identification scores in the pretest.

The EFL learners' perceptual performance in the four testing moments is briefly discussed next. Overall, pre-training perceptual performance of both groups did not differ in terms of L2 vowel identification. Experimental participants had a total vowel ID score of 66% and the controls 69%. Between-group comparisons in the identification of the vowel contrasts also revealed that they did not differ significantly. However, when percentage ID scores were analyzed for each vowel, a difference was found in the identification of /i/, with the trainees having a significantly lower ID score (53.79%) than the controls (70.56%). The low ID rate of /i/, which was frequently misidentified as /ɪ/, explains the difficulty with the high front vowel contrast, which had the lowest ID score of the three target pairs. Given that there was more space to perceptual learning, it was in fact the vowel contrast that improved the most after training (27.04%), followed by /ɛ/-/æ/ (14.85%), and then the back vowel contrast /u/-/ʊ/ (13.63%). Although the vowel pairs with the distractor vowel had high ID scores at pre-training, there was also significant gain after auditory training. Contrariwise, the control group did not show improvement (the calculated overall pre- and posttest difference was 0.18%). Since the controls undertook a similar consonant-centered training program with the same number of sessions, which consisted of discrimination and identification tasks, it might be argued that their perceptual performance would improve due to task effects. However, no significant perceptual improvement was found for any of the three vowel contrasts or for any of the seven vowels, which seems to indicate that task repetition solely did not promote modification of perceptual performance.

Nevertheless, significant training effects were observed in terms of response times after training for both groups. As expected, the repetition of performing identical perception tasks during the training reduced the time of reacting to aural stimuli. Moreover, familiarization with the computer software and with the auditory exercises made participants respond gradually quicker from pretest to delayed posttest. Nonetheless, vowel-trainees had significantly shorter RTs than consonant-trainees in the identification of /ɛ/-/æ/, /u/-/ʊ/, /æ/-/ʌ/ and /ʌ/-/ʊ/ at posttest. Two months later, significant intergroup differences were still found in the identification of /ɛ/-/æ/ and /æ/-/ʌ/. This seems to suggest that learners in the experimental group gradually improved their perceptual assessment of AE vowel categories from poor to good exemplars of L2 categories, while the controls took more time to identify exemplars. Although the measure of response time provided additional corroborative evidence that perceptual learning occurred, its use might be debatable, since it primarily reflected behavioral adaptation and familiarization to tasks. Nevertheless, RTs can complement analyses, in particular if target sounds are embedded in longer tokens, such as sentences.

The ratings of category goodness-of-fit, which provided further information about perceptual performance, were consistent with ID scores. Higher ID scores were correlated with higher goodness-of-fit ratings for five vowels at pretest, two vowels at posttest and four vowels at delayed posttest (see Tables 52-53). Vowel training significantly affected the evaluation of AE vowel exemplars by participants in the experimental group, but did not influence controls' ratings. In sum, vowel-trainees' perceptual improvement in the identification of the target vowels was accompanied by an increase of confidence in vowel categorization that was shown by a significant increase of goodness ratings and a reduction of response times. However, to have more statistically robust results regarding goodness-of-fit ratings, the range of the scale should have been greater than three points. A larger scale (of seven or nine points) would have allowed the calculation of fit indexes, which combine both the identification and the goodness-of-fit data into a single metric (Guion, Flege, Akahane-Yamada, & Pruitt, 2000). The fit indexes would have provided a means to raise the ID scores that were considered good tokens of the vowel category and to lower the scores of identifications that were selected because they had no good competitors (*ibid.*, p. 2723). However, the choice of having a small scale was to facilitate participants' task when undertaking the seven-alternative forced-choice identification tests with 210 stimuli. We hypothesized that choosing a button from seven options, that is, identifying a vowel from seven possible options, was already demanding, thus if, for example, a seven-point scale was added to rate goodness-of-fit it would have been even more wearying in terms of attention and short-term memory.

A significant effect of high variability perceptual training was found in the identification of the five vowel contrasts in relation to the moment they were tested for the experimental group, but not for the control group. Improvement in perceptual ability to recognize the AE vowel categories was observed not only immediately after training, but also eight weeks later. In particular, the identification scores of the back vowel contrast /u/-/ʊ/ increased significantly (from posttest to delayed posttest) indicating that learning still occurred after training was over. Consequently, this target contrast reached slightly higher scores (84.12%) at delayed posttest than /ɛ/-/æ/ (83.77%). One of the reasons that might explain the improved learning after training was completed is the integration of some phonetic transcription practice of English speech samples in classes, which kept the acquired perceptual awareness of vowel distinctions activated. The controls, who maintained their posttest ID scores with no significant loss, might have also benefited from hearing extra spoken input, though it was not vowel-focused. Given that perceptual awareness of vowel contrasts was not explicitly raised from consonant-centered training, their ability to perceptually process vowel segments did not improve, but did not decline either. Moreover, the identification scores of the non-target AE vowel contrasts were maintained by the trainees, that is, the ID score differences between posttest and delayed posttest were non-significant, although slightly higher, which suggests that long-term effects were observed. This finding supports previous research that reported long-term retention effects one month (e.g. Nobre-Oliveira, 2007) and three months (e.g., Nishi-Kewley, 2007; Wang, 2008; Wang & Munro, 2004) after conclusion of vowel training. A longer time gap between posttest and delayed posttest was desirable, but due to academic calendar constraints, namely the end of the semester and the probability of having a high level of attrition if delayed posttest was administered at the beginning of the following academic year (i.e., five months later), the delayed posttest could not be administered later than two months after training sessions were over.

Immediately after auditory training, two identification tests were administered, namely the posttest and the generalization test. The results showed that the identification of two of the AE vowel contrasts, embedded in new words with the syllabic structures CVC and (C)CV(C)(C) and produced by novel talkers, were significantly more accurate in the generalization test than in the posttest, being the back vowel pair the only exception because it had significantly lower ID scores in the generalization test. In short, generalization of accurate perception of the front vowels occurred, but not of the back vowels. This is explained by the significantly lower ID score of /ʊ/, but not of /u/ in relation to the posttest. The back lax vowel had a difference of -20% in the generalization test (from 76% to 56%), reaching chance level. The confusion matrix also showed that /ʊ/ was either misidentified as /u/ (22.2%) or as /ʌ/ (21.2%), which is an evidence that this vowel category was not still robustly assimilated as an L2 sound. The goodness ratings for this vowel did not differ significantly after training either, which seems to suggest that EFL learners were not more certain about the categorization of this vowel than before training. Given that the perception tests consisted of closed-set identification tasks, in which there was no response option such as “none (of these options)” or “I don’t know”, participants had to select one of the seven English vowel labels, even if they did not know which vowel most perceptually resembled /ʊ/. Therefore, it might have been helpful to understand cross-language perceptual assimilation difficulties with the AE vowels better if there had been an option as the aforementioned. The vowel pairs with the distractor /ʌ/ (viz. /æ/-/ʌ/ and

/ʌ/-/ʊ/) also had significantly lower ID scores in the generalization test, which might be explained by the poorer identification of /ʌ/ (with a decrease of -6.31%) and /ʊ/ (with a decrease of -19.75%) in relation to posttest.

In sum, both groups of EFL learners had an accented perception of the target vowel contrasts at pretest, but after training the degree of accentedness in the experimental group decreased significantly reaching a near-native-like perceptual performance (ranging from 84% to 88% accurate ID scores). Positive training effects were still observed two months later, with the retention of learning of the high vowel pairs and significant improvement of the back vowel contrast. Generalization of learning occurred for the two front vowel pairs, but not for the back vowels, due to the perceptual difficulty in identifying vowel /ʊ/. Similar findings in terms of perceptual performance improvement were also reported, for instance, by Aliaga-Garcia and Mora (2009), Lambacher (2005), Lengeris (2009), Nobre-Oliveira (2007), Pereira and Hazan (2013), and Wang (2008).

Moreover, the analysis of individual differences, described previously, revealed that some participants were faster in learning how to distinguish the AE vowels, thus achieving native-like and near-native-like perceptual performances, whereas a few had more difficulty and took more time to accurately perceive the non-native vowels after three 45-minute training sessions. This finding suggests that after each session scores could have been analyzed so that individual learners' difficulties with the AE target contrasts could have been identified, and extra tasks could have been assigned to these students. However, this was not the aim of the experiment, because the effects of individualized self-paced training are more difficult to account for when comparing group performances.

The pretest results revealed that, before training, the least difficult pair to be perceived by both groups of participants was /ɛ/-/æ/, and /u/-/ʊ/ was the most difficult contrast for controls, while /i/-/ɪ/ was the most difficult for trainees. In a study carried out with a group of learners with the same EFL background (Rato et al., 2013), ten male participants exhibited the same gradual degree of perceptual difficulty in the perception of the two front vowel contrasts, being the mid-low contrast more easily identified than the high front vowel pair.

Due to the similarity between the Brazilian Portuguese and the European Portuguese vowel systems, the predictions about which of the three vowel contrasts would be most difficult or easiest to perceive before training were based on Nobre-

Oliveira (2007) and Rauber’s (2010) studies that investigated the perception and production of American English vowels by Brazilian EFL learners. Hence, if Portuguese learners followed the same pattern, the least difficult pair to perceive would be /i/-/ɪ/, followed by /u/-/ʊ/, and then /ɛ/-/æ/. As far as perception is concerned, the first two contrasts are more dissimilar due to their F1-F2 acoustic distance than /ɛ/-/æ/. Thereby, acoustic dissimilarity was expected to reflect perceptual dissimilarity. However, as aforementioned, this was not the pattern observed in this group of EFL learners. Moreover, due to the spectral proximity of vowels in the mid-low vowel pair, it was expected to be the most challenging to identify. Hence, a closer examination of vowel perceptual assimilation patterns helps to understand the pre-training performance of both groups of learners. First, vowel /æ/ was the most accurately identified among the six target vowels, and in comparison to the preexisting L1 /ɛ/ it had significantly higher ID scores in both the experimental (t =-2.65(21), p<.05) and control groups (t =-2.86(11), p<.05), as also observed by Rato et al. (2013). One of the possible reasons for this identification pattern was advanced by researchers:

> According to Escudero et al. (2009), the F1 value of /ɛ/ produced by (male) Portuguese monolinguals from Lisbon is 455 Hz, considerably lower than the English /ɛ/ (538 Hz). Thus, the male L2 speakers might have considered the English /ɛ/ so low that they often identified it as /æ/. (*ibid*.)

Although the acoustic F1 values of vowel /ɛ/ produced by this group of Portuguese speakers from the Minho region differed from those reported by Escudero et al. (2009), the male Portuguese speakers still produced /ɛ/ with a lower F1 (554 Hz) than AE men (582 Hz) and the same in relation to women, who produced a higher Portuguese /ɛ/ (652 Hz) than the AE /ɛ/ (704 Hz). In addition, another fact might explain the high percentage of correct identification of AE /æ/, namely its F1 value (820 Hz for AE women, and 671 Hz for AE men) being acoustically (i.e., physically) close to the F1 value of EP /a/ (859 Hz for women and 694 Hz for men). Therefore, it seems that English /æ/ was perceptually differentiated from /ɛ/ and correctly assimilated as /æ/ because it is spectrally close to Portuguese /a/.

Even though no cross-language perceptual assimilation task, in which learners identify L2 sounds in terms of L1 categories, was administered to assess the degree of (dis)similarity between the American English and European Portuguese vowel systems, we hypothesized that, similarly to Brazilian EFL learners (Nobre-Oliveira, 2007;

Rauber, 2010), and according to the PAM-L2 (Best & Tyler, 2007), the assimilation pattern of English /ɛ/-/æ/ by Portuguese learners would be a case of single category assimilation (SG), because both sounds would be equivalent to the same L1 phonological category (/ɛ/), that is, /ɛ/ and /æ/ were expected to be heard as homophones. Therefore, learning to perceive a difference between single-category assimilated phones would depend on whether they were perceived as good or poor exemplars of the Portuguese segment /ɛ/. However, this pattern was not observed. Rather, according to the results of the confusion matrix at pretest, a case of category-goodness assimilation (CG) seems more adequate to describe the perceptual mapping of these two L2 sounds, which were heard as instances of the same native phoneme but with different category goodness ratings. It seems that English /æ/ and /ɛ/ were perceived not as Portuguese /ɛ/, but as /a/, thus participants chose the L2 sound that most resembled /a/, namely /æ/. Moreover, one of these two L2 sounds was perceived as being more deviant (/ɛ/) than the other (/æ/). Therefore, the prediction was that a new phonemic category was likely to be established for the deviant /ɛ/, while the L2 phone that was perceived as a better exemplar was likely to be perceived as equivalent to the L1 category /a/ and no new category was expected to be formed. Given that no cross-linguistic perceptual assimilation task was administered, the previous observation is only a hypothesis, based on Flege's analysis of L2 Spanish patterns of misidentification (1992, as cited in Flege, Bohn & Jang, 1997, p. 442), which concluded that realizations of English /æ/ were misidentified as the Spanish vowel /a/ by Spanish monolinguals. Flege (1995) also reported discrimination failure of vowel contrasts /æ/-/ɛ/, /ɛ/-/ɪ/, /u-/ʊ/, and /ʌ/-/ɑ/ by Portuguese speakers (p. 249). However, the researcher explains that Portuguese perceivers may have managed to discriminate /æ/-/ɑ/ by virtue of identifying English /æ/ in terms of Portuguese /ɛ/, and English /ɑ/ in terms of Portuguese /a/. On the other hand, Major (1987) observed that the /ɛ/ of the seven-vowel Portuguese inventory was often realized as /æ/.

According to Piske (2008, as cited in Moyer, 2013, p. 162), "FL learners have a writing system in mind, and inevitably refer to L1 sound-symbol correspondences when learning to hear and speak L2". Therefore, orthography might have also played a role in the perception of /æ/-/ɛ/, because the labels of the identification test included not only the representation of the sound, but also a key word, namely "had /ae/" and "head /E/". Nevertheless, this does not fully explain the higher rate of correct identification of /æ/ in comparison to /ɛ/, because the grapheme-sound relation of both is identical, and though sound /ɛ/ was represented by the spelling <ea>, next to it had the symbol /E/, which is a transparent representation of the sound.

The SML (Flege, 1995) hypothesizes that similar vowels pose more difficulties for L2 listeners than new vowels due to the mechanism of perceptual equivalence classification, whereas identical vowels are perceived as L1 categories, and therefore no difficulty is predicted for their discrimination. In sum, from the set of the English target vowels, only two (viz. /i/ and /u/) can be considered identical to Portuguese categories, while AE /ɪ/, /ɛ/, /æ/, and /ʊ/ can be considered similar to L1 categories, namely to the EP vowels /i/, /ɛ/, /a/ and /u/; thus, they were the most challenging vowels to learn.

According to Iverson and Evans (2007a, 2007b) and Escudero and Williams (2001), small native vowel inventories, such as Portuguese, result in L2 single-category assimilation cases. However, the analysis of the confusion matrices suggests that category-goodness assimilation patterns were observed instead. The pre-training perception of the two high front vowel pairs seemed to be cases of category-goodness assimilation, whose predicted level of distinction between vowels is intermediate, thus English vowels /i/ and /ɪ/ were mapped onto L1 /i/, and vowels /u/ and ʊ/ onto L1 /u/. The discrimination between the vowels of each target pairs was moderate, at pretest, and new L2 categories were predicted to be created for the deviant sounds /ɪ/ and /ʊ/ after training, which would be observed in the increase of recognition accuracy. The post-training results suggest that perceptual patterns altered to a great extent, and new categories were established for L2 vowels, providing evidence of two category-assimilation patterns at posttest.

After training, a modification of perceptual patterns was observed for the trained group, but not for the controls. The least difficult contrast to perceive was /i-ɪ/, followed by /ɛ-æ/, and then /u-ʊ/, which indicates that trainees' perceptual mappings changed immediately after training. The control group maintained the same patterns throughout the different phases of the experiment. The modification of perceptual mappings was observed in the significant improvement in the correct ID scores of all the target vowels. The degree of perceptual improvement (from pre- to posttest) was the following, from the highest to the lowest: (1) high front tense vowel /i/ (31.36%); (2) mid front vowel /ɛ/ (23.33%); (3) high front lax vowel /ɪ/ (22.72%); (4) high back tense vowel /u/ (17.27%); high back lax vowel /ʊ/ (10%), and low front vowel /æ/ (6.36%).

The degrees of perceptual difficulty, ranging from least difficult to most difficult, in the identification of the three vowel contrasts before and after auditory training by the two groups of L2 learners are shown in Table 71.

Table 71

Degrees of Perceptual Difficulty in the Identification of the Target Vowel Contrasts

	PERCEPTION	least difficult		most difficult
EG	**pretest**	/ɛ-æ/	/u-ʊ/	/i-ɪ/
	posttest	/i-ɪ/	/ɛ-æ/	/u-ʊ/
	delayed posttest	/i-ɪ/	/u-ʊ/*	/ɛ-æ/*
CG	**pretest**	/ɛ-æ/	/i-ɪ/	/u-ʊ/
	posttest	/ɛ-æ/	/i-ɪ/	/u-ʊ/
	delayed posttest	/ɛ-æ/	/i-ɪ/	/u-ʊ/

Note. The vowel pairs marked with an asterisk had identical ID scores at the delayed posttest.

Two months after the training was over, there was still a significant improvement in the identification of the back vowel contrast by the trainees, who perceived /u-ʊ/ (84.12%) almost as accurately as /ɛ-æ/ (83.77%); thus, performing similarly to the Brazilian EFL participants in both Nobre-Oliveira (2007) and Rauber's (2010) studies. In conclusion, the perceived dissimilarity between the vowels of each pair increased after short-term auditory training from moderate levels of perception to good near-native-like patterns.

Other misidentification patterns that were not predicted are discussed next. Vowel /ɪ/ was misidentified not only as /i/ but also, though much less frequently, as /ɛ/, which was a pattern reported by Flege (1994, as cited in Flege 1995, p. 294) for Portuguese speakers, Nobre-Oliveira (2007) for Brazilian speakers, and Flege and MacKay (2004) for native speakers of Italian. English /ʊ/ was not only confused with /u/ but mostly with /ʌ/. One of the possible explanations might be advanced if we examine the normalized acoustic F1 and F2 values of vowels that were used as stimuli in the identification tests (see Table 72) and look at the corresponding vowels spaces of native American English talkers (Figure 22). The F2 values of vowels /ʌ/ and /ʊ/ were closer than those of /u/ and /ʊ/ in the stimuli of both the generalization and identification tests, which may have led L2 listeners to misperceive /ʊ/ as either /u/ and /ʌ/ sounds. Furthermore, both /ʊ/ and /ʌ/ might have been mapped onto the Portuguese L1 category /ɐ/, which is located in a close F1-F2 position in the vowel space to English L2 segments /ʌ/ and /ʊ/. One of the American English speakers, who validated the perception tests and training tasks, commented that “many of the stimuli were midway between /ʌ/ and /ʊ/”, which indicates that even a NS had some difficulty in identifying and distinguishing between /ʊ/ and /ʌ/, and that the vowel tokens of this vowel pair were perceptually close. Another less probable factor that might have influenced EFL listeners’ perception might be the lack of orthographic transparency of English, especially in the representation of the back vowels. It seems that when L2 listeners heard a high back vowel, in particular /ʊ/, but sometimes also /u/, they visualized the Portuguese orthographic representation for segment /u/ in stressed position,[130] namely <u>, and chose the label in the ID tests that had this spelling rather than <o> or <oo> (the ID tests’ labels were “*who’d* /u/”, “*hood* /U/” and “*hud* /^/”). However, this is not a satisfactory explanation, because if this had been the case, vowel /u/ would have also been more often misidentified as vowel /ʌ/ than as /ʊ/.

Table 72

Normalized F1 and F2 Values of AE Vowels Included in the ID Tests

		/i/	**/ɪ/**	**/ɛ/**	**/æ/**	**/ʊ/**	**/u/**	**/ʌ/**
ID	**F1 (Hz)**	253.85	420.65	589.09	741.41	480.12	318.39	593.35
	F2 (Hz)	2104.25	1862.42	1468.31	1307.96	896.84	1140.42	1040.20
GT	**F1 (Hz)**	272.20	424.15	537.81	726.91	414.56	282.93	605.29
	F2 (Hz)	2227.19	1639.51	1588.17	1433.08	861.19	1275.57	1009.92

Note. ID=identification test; GT=generalization test.

Another relevant finding regarding L2 vowel perception is related to the performance of the control group in the identification tests (posttest, delayed posttest and generalization test) after training, because, if perceptual improvement was solely

[130] Vowel /u/ is orthographically represented by <u> in stressed position (e.g., <sul>, <ˈu.va>) and <o> in unstressed position <ˈpa.to> or in function words, such as the definite article <o>.

caused by task effect, that is, if it resulted from the repetition of performing perceptual tasks, significant improvement should have been observed in controls' identification scores after training with consonants. However, no perceptual improvement was found for any of the three vowel contrasts, which seems to indicate that learning was not promoted by task repetition. The claim that phonetic training has to be centered on the target segments in order to be effective was supported by the finding that a vowel-centered training was beneficial for learning the three vowel contrasts, whereas the consonant-centered training was not. Although the natural stimuli used in the training program of the controls included the target vowels (e.g., <bass-bath>, <miss-myth>), their attention was directed to consonant contrasts. However, this should be further investigated, for instance, by matching the stimuli presented to the two different groups (e.g., <bass-bath> and <Bess-Beth> to simultaneously train the vowel contrast /ɛ/-/æ/ and the consonant contrast /s/-/θ/) and comparing their perceptual performance on both sets of segments after training. See Carlet (in progress) for an ongoing study exploring this issue further.

4.4.2 Cross-language vowel production

One of the main research objectives of this study was to investigate whether perceptual training would have an effect on vowel production. To evaluate its effects, pronunciation of the target American English vowels was tested three times by means of a reading-sentence task, before training, immediately after it, and two months later. The production performance of both groups was compared in the three testing moments to verify whether there were intergroup differences in production accuracy of the target vowel pairs.

In addition, the production of European Portuguese vowels was also elicited once, at the beginning of the experiment. The main aim of collecting the production of European Portuguese vowels was to provide information about the distribution of EP vowels in the vowel space of this specific group of participants from the Minho region to understand their cross-language perceptual assimilation patterns and production difficulties better, since acoustic measurements hitherto documented are based on oral vowels produced by informants from Lisbon (e.g., Delgado-Martins, 1973, 1975; Escudero et al., 2009), which might not provide the most adequate referential EP acoustic values for analysis of this specific oral corpus. Therefore, this set of data is not extensively discussed, but rather used to establish acoustic comparisons between L1 and L2 vowel inventories. Moreover, the acoustic values used for American English vowels were the ones reported by Rauber (2010) and they should be understood only as an example of a possible reference and not as the desirable or attainable model of AE vowel production. Hence, F1 and F2 values are sparsely mentioned to discuss L2 learners' productions.

As aforementioned, the two groups of EFL learners were pre- and post-tested for production ability to investigate whether perceptual training affected participants' pronunciation of the target vowels, in particular the articulatory distinction between /i/ and /ɪ/, /ɛ/ and /æ/, and /u/ and /ʊ/. To assess long-term effects, the same production test was replicated two months after completion of the auditory training. Conversely to perception assessment, no generalization test was designed to assess carryover effects of potential production improvement to new contexts and new words. To assess production performance two measures were used, namely the Euclidean distance between the vowels of each target contrast and duration ratios.

At pretest, there was overlap in the production of vowels /æ/ and /ɛ/ by both groups of participants, and partial overlap of the high vowel contrasts (see Figures AJ1 - 2). No between-group differences were found in the production of the two high vowel pairs, but the ED of /u/-/ʊ/ was significantly greater (t=2.09(32), p<.05) when produced by trainees than by controls; thus, overlap was mostly visible for the latter group than for the former. In terms of duration ratios, no differences were found between groups. When comparing EFL learners' EDs and duration ratios to native AE reference values, we observed that the EDs of two of the target vowel contrasts produced by Portuguese speakers were significantly smaller than those of AE speakers. The ED between /æ/ and /ɛ/, though not significantly different, was also much smaller when produced by both groups of L2 learners (54.71 Hz and 59.38 Hz) than by AE NSs (109.19 Hz). Duration ratios were also significantly lower for vowel contrasts /æ/-/ɛ/ and /i/-/ɪ/ in comparison to NSs, which indicated that a native-like durational distinction between vowels in each pair was not made. Given that production performance was not a criterion to assign participants to groups, the pretest measures of ED and duration were not controlled between participants, which explains the significant intergroup difference in the articulation of /u/-/ʊ/ at pretest. However, if we look at Figure 44, we can see that the pre-training vowel spaces of both groups were equivalent and vowels were equally distributed in terms of height and frontness/backness. Moreover, both acoustic values of ED and duration indicate that almost no spectral or durational distinction was made between vowels of each target contrast by this group of EFL learners. In sum, L2 vowels /ɪ/, /æ/ and /ʊ/ were produced as /i/, /ɛ/ and /u/, respectively, and segments /æ/ and /ɛ/ were produced with median F1 and F2 values closer to AE /ɛ/ than to Portuguese /ɛ/ (see Appendix AN).

Immediately after training, the EDs of the target vowel contrasts (/æ/-/ɛ/, /i/-/ɪ/, and /u/-/ʊ/) increased significantly; thus, vowels /i/-/ɪ/ were no longer overlapped, /u/-/ʊ/ were only slightly overlapped, and /æ/-/ɛ/ partially overlapped as produced by the experimental participants (see Figure AJ3). However, it seems that the more aware participants were of the spectral difference between vowels, the less they were about the durational distinction. The controls also produced the /u/-/ʊ/ contrast with a higher ED in relation to pretest, but increase was not as high (21.38 Hz) as that observed for the trainees (49.76 Hz). The vowels of the pair /i/-/ɪ/ were closer at posttest (90.49 Hz) than at pretest (100.84 Hz), but the difference was not significant (see Figure AJ4). Both groups differed significantly in the production of /i/-/ɪ/ (t=4.29(32), p<.001) and /u/-/ʊ/ (t=2.42(32), p<.05). Eight weeks later, they continued to differ significantly in the pronunciation of the same vowel pairs. The EDs of /i/-/ɪ/ (t=3.95(28), p<.001) and /u/-/ʊ/ (t=3.45(28), p<.01) were significantly greater in the productions by the vowel-trainees than by the consonant-trainees. Despite not being a significant difference, the ED of /æ/-/ɛ/ was somewhat greater in the trainees' productions. At the end of the study, the experimental participants achieved a near-native like production in terms of vowel quality, given that the only vowel contrast ED that still differed significantly from NSs was the high front pair (see Table 73). However, the distance between the vowels of each pair was much higher in the vowel-trainees than in the consonant trainees (260 Hz and 114 Hz, respectively).

Table 73

Degrees of EDs of the Target Vowel Contrasts

	PERCEPTION	**Greater ED**		**Smaller ED**
EG	**pretest**	/i-ɪ/	/u-ʊ/	/ɛ-æ/
	posttest	/i-ɪ/	/u-ʊ/	/ɛ-æ/
	delayed posttest	/i-ɪ/	/u-ʊ/	/ɛ-æ/
CG	**pretest**	/i-ɪ/	/ɛ-æ/	/u-ʊ/
	posttest	/i-ɪ/	/ɛ-æ/	/u-ʊ/
	delayed posttest	/i-ɪ/	/ɛ-æ/	/u-ʊ/
NS		/i-ɪ/	/u-ʊ/	/ɛ-æ/

Note. EG=Experimental group; CG=control group; NS=native American English speakers

Table 73 shows that in terms of spectral distinction between vowels of the three target contrasts, the degree of Euclidean distance, from greater to smaller, produced by experimental participants followed the same direction as that of NSs, whereas the control group's did not. The back vowel contrast was pronounced with the smallest ED by the controls.

In terms of duration ratios and in comparison to NSs, EFL learners did not distinguish vowels by length. Overall, the duration ratios of vowel pairs decreased immediately after training, and in the particular case of the high vowel pair /i/-/ɪ/ duration ratio was significantly lower (t=2.99(21), p<.01) in the productions of the vowel-trainees. In fact, it seemed that training had the reverse effect. By redirecting learners' attention to spectral differences, durational distinctions became less evident. At the onset of training, L2 learners' seemed to rely more on duration than at its offset. At pretest, the durational difference between vowels of the pairs /æ/-/ɛ/, /i/-/ɪ/, /u/-/ʊ/, and /æ/-/ʌ/ was significant,[131] but immediately after training and two months later the durational distinction between the vowels of the high vowel contrasts was not significant, which indicates that EFL learners produced vowels with similar durational values.[132] To some extent, this corroborates Bohn's (1995) desensitization hypothesis, because at pretest it seemed that spectral differences were insufficient to distinguish vowel contrasts, that is, L2 learners were not previously sensitized to rely on quality;

[131] *T* tests were run to compare durational values at the three test phases. The results at pretest, for the experimental group were the following: /æ/-/ɛ/ (t=4.87(21), p<.001), /i/-/ɪ/ (t=2.40(21), p<.05), /u/-/ʊ/ (t=2.70(21), p<.001), and /æ/-/ʌ/ (t=16.02(21), p<.001).

[132] The results at posttest were: /æ/-/ɛ/ (t=3.14(21), p<.01), /i/-/ɪ/ (t=-1.27(21), p>.05), /u/-/ʊ/ (t=1.83(21), p>.05), and /æ/-/ʌ/ (t=10.72(21), p>.05). At delayed posttest, the results were: /æ/-/ɛ/ (t=3.55(18), p<.01), /i/-/ɪ/ (t=-.63(18), p>.05), /u/-/ʊ/ (t=-34(18), p>.05), and /æ/-/ʌ/ (t=10.26(18), p<.001).

thus, duration differences were used to differentiate the four English contrasts, whereas after training spectral differences seemed to override vowel length. As training progressed, awareness of spectral vowel dissimilarities contributed to the decrease of durational differences. Traditionally, in the English foreign language classroom, the two high vowel contrasts are taught as being differentiated by length, rather than by quality, which might explain the significant durational differences between vowels of the target pairs at the beginning of the experiment. The brief articulatory descriptions that preceded each training session focused exclusively on spectral differences between vowels and no reference was made to duration with the purpose of directing learners' attention to vowel quality only. When a few participants inquired about the long-short vowel distinction, the researcher clarified that the main phonological feature that differentiates vowel contrasts in English is quality, not duration, and no aural examples were provided. This finding seems to indicate that training had an effect on the duration ratios of trainees but in the inverse direction, that is, instead of having increased, ratios decreased.

In conclusion, perceptual training had a significant effect on pronunciation accuracy of the target vowels, given that learning was transferred to production, in particular in terms of vowel quality. The trainees achieved a native-like pronunciation performance in two of the three target vowel contrasts and a near-native-like ability in distinguishing the high front vowels. No such production transfer was found in the study by Nobre-Oliveira (2007), and only modest gains were reported by Aliaga-Garcia and Mora (2009). Both studies assessed production accuracy by means of acoustic measurements only. Wang (2008) followed two evaluation procedures, by combining acoustic measurements of vowel duration with results of intelligibility tests administered to native English listeners, and observed no effect of training on production. Similarly, Pereira and Hazan (2013), who assessed vowel intelligibility by NESs and measured Euclidean distance and duration, found not much spectral change, that is, no significant changes in terms of ED and duration were observed. However, Lengeris (2008, 2009) reported that post-training tokens of English vowels articulated by Greek trainees were more accurately identified by native English listeners than pre-training tokens, which suggests that intensive laboratory training is effective in production improvement.

4.4.3 Interface between cross-language vowel perception and production

One of the questions that this study aimed to discuss was the interface between English vowel perception and production. Thereby, perceptive and productive abilities were assessed through distinct evaluation procedures, viz. calculation of percentage of correct vowel identification scores and acoustic measurements of ED (Hz) and duration ratios (ms). Given the different types of evaluation criteria and the non-linear relation between them, no correlation tests could be run to assess whether perceptual and production performances were related. Although higher ID scores were expected to be related with higher EDs, this did not seem to occur for the vowel contrast /ɛ-æ/, but no statistical test was run to state that more robustly. To have comparable measures, scores of intelligibility tests submitted to NEs should be used instead. However, this procedure was not included.

Nonetheless, a few observations can be made regarding the interface between L2 vowel perception and production. The pre-training perceptive and productive performances of the 34 EFL learners were greatly interrelated because perceptual difficulties in identifying the target vowel contrasts were related to overlapping segments in the F1-F2 vowel space. An accented pronunciation, in which almost no articulatory distinction was made between vowels of the target pairs, was reflected in an accented perception, that is, in the difficulty to recognize English vowels. At pretest, the most difficult to perceptually categorize were the high vowel contrasts, and the least difficult was the mid-low front pair. These results were consistent with acoustic measurements, given that at the onset of training the ED of the mid-low front pair did not differ significantly from that of native speakers of American English, whereas the ED of the high contrasts did.

The post-training performance in perceptually categorizing vowels showed a significant improvement that was transferred to production. Significantly higher scores in the identification tests were related to significantly higher EDs, in particular in the case of the more near-native-like EDs between the two high front vowels. The ED of the /æ/-/ɛ/ contrast, despite increasing significantly from pre- to posttest, still did not differ from that of NESs. The lowest ED increase of the back-vowel contrast indicates that it was the pair that least improved in terms of production accuracy, which was again supported by perception ID scores. At posttest, it was the contrast with the lowest results. Moreover, at the generalization test, which was at the same time as the posttest,

/ʊ/-/u/ was the only contrast that did not generalize to new tokens and new talkers. This might be explained by the EFL learners' difficulties in identifying the high back lax vowel embedded in tokens with consonant clusters pronounced by novel native American English speakers. Aliaga-Garcia (2013) reported a similar finding, that is, less improvement for back vowels in production, in a HVPT training study with Catalan/Spanish EFL learners.

Long-term training effects were found in both perceptual and production abilities of experimental participants. The identification scores were maintained eight weeks after training was over, and the perception of the high back vowel contrast improved significantly. The EDs of the three target vowels did not differ significantly from posttest to delayed posttest, but the values of the EDs at delayed posttest were somewhat higher, which indicates that there was no loss of production ability in distinguishing vowels of the target contrasts. Once again, the same tendency was found in perception test results, that is, slightly higher ID scores two months after training was over. No significant effect of training was found either in terms of ID scores or ED measures for the control group, which relates, to some extent, their perceptual difficulties to production inaccuracy. The pair that was better produced by controls, that is, that had a similar ED (though smaller than the trainees') to NSs was the mid-low front vowel contrast, whose vowels were also the least difficult to perceptually discriminate.

The positive effects of training on perception and, in particular, on production might have also been enhanced by the specific context in which this perceptual training experiment was carried out, namely in the *English Phonetics and Phonology* course. Although this was a theoretical-based course, after perceptual training was concluded, some phonetic transcription practice of different varieties of English speech samples was included in classes. However, no articulatory training complemented the course, and production of English sounds was not encouraged throughout the course. The amount of English input was also limited to audio files. Nonetheless, though this specific classroom context may have facilitated learning because perceptual awareness was complemented with articulatory descriptions of speech sounds, it seems not to have affected participants' overall performance significantly because neither significant improvement nor robust learning was observed for the controls. Similar non-significant effects of pronunciation instruction were found by Rauber, Rato and Silva (2010) in a study about the learning of English vowels by Mandarin speakers. The reported non-significant improvement in the production and perception of English vowels after an eight-month upper-intermediate EFL course that included pronunciation instruction suggests that systematic and segment-focused training is needed to promote substantial improvement and phonological learning.

CONCLUSION

The pedagogical objectives of the present study were to raise Portuguese EFL learners' awareness of L1-L2 vowel (dis)similarities and to improve their ability to perceive and produce English vowel sounds. Specifically, it aimed at investigating the effectiveness of perceptual training in promoting phonological learning. The improved post-training perceptual and articulatory performance of participants in the experimental group provides evidence that the main objectives were achieved. The main findings of the study are summarized next by relating them to the research questions.

The first hypothesis predicting that, with a limited number of sessions, high variability perceptual training would have a positive effect on the perception of English vowel contrasts /i/-/ɪ/, /ɛ/-/æ/, and /u/-/ʊ/ by native European Portuguese speakers was confirmed. After training, the degree of accented perception of Portuguese learners decreased significantly, reaching a near-native-like perceptual performance. Positive training effects were still observed eight weeks later, with the retention of learning of the front vowel pairs and significant improvement of the back vowel contrast, which confirmed the second hypothesis. Not only were long-term positive effects observed, but there was also evidence of delayed effects of perceptual learning. The third hypothesis was partially confirmed, because generalization of learning occurred for the two front vowel pairs, but not for the back vowel contrast. Perceptual training had a significant effect on pronunciation accuracy of the target vowels, since learning was transferred to production, specifically in terms of vowel quality. The trainees achieved a native-like Euclidean distance between the vowels of the high back and the mid-low front contrasts and a near-native-like acoustic distance between the vowel segments of the high front pair. This finding supported the prediction of the fourth hypothesis because, even without any specific articulatory training, an improvement in production accuracy was observed, highlighting the interface between vowel perception and production. The results of the identification pretest revealed that, before training, the least difficult pair to perceive was /ɛ/-/æ/, and the most difficult was /i/-/ɪ/. After training, the least difficult contrast to perceive was /i/-/ɪ/, followed by /ɛ/-/æ/, and then /u/-/ʊ/. Two months after training was over, the easiest pair to perceive was /i/-/ɪ/, and then both /u/-/ʊ/ and /ɛ/-/æ/ were similarly challenging. The fifth hypothesis, which predicted that the least difficult pair to perceive and produce would be /i/-/ɪ/, followed by /u/-/ʊ/, and then /ɛ/-/æ/, was only partially confirmed after training.

Concerning the role of acoustic cues, the sixth hypothesis anticipated that learners would rely on both vowel duration and quality to produce the English vowels. Acoustic measurements of duration ratios and Euclidean distances revealed that, although at pretest the target English contrasts were differentiated in terms of duration, after training spectral differences were preferred instead. Training seemed to have redirected learners' attention to spectral cues and, consequently, duration was given less importance. The focus on spectral quality was so much that duration differences, which were observed at pretest, were no longer present at tests taken after training.

Traditionally, second language/foreign language (L2/FL) research has been conducted in two distinct contexts, the classroom and the laboratory. Laboratory-based research has the advantage of allowing the control of experimental variables, assigning participants to experimental groups, and using control groups, all of which are difficult to implement in classroom-based research. In particular, it is ethically and logistically difficult to have a control group that does not receive or is excluded from potential benefits caused by the experiment (Mackey & Gass, 2005). According to Loewen & Philp (2012), L2/FL classroom research consists of four types: observational studies, (non)interventionist quasi-experimental studies, and action research. The present method was a combined approach to classroom interventionist research in that, on the one hand, it consisted of an experimental study carried out with a high level of control in a computer lab with specific software for perception tasks; on the other hand, it was conducted in a classroom within a regular undergraduate class.

Combining the roles of researcher and teacher had disadvantages and some advantages. On the one hand, although there was no penalty for declining to participate, the fact that the teacher was the one requesting the students' participation may have influenced their willingness to undertake the training program. From the group of 67 students enrolled in the *English Phonetics and Phonology* course, 49 volunteered to participate, though only the elicited data of 34 participants were analyzed, as explained previously. To avoid participant attrition, an extra mark was given for completing the training program; however, at the end of the semester, four students dropped out and did not take the delayed perception and production posttests. Another limitation found when assigning learners to two groups was timetable restrictions, that is, although some students had been assigned to a certain group after matching their perception test results, they had to be allocated to a different group due to timetable constraints. Other difficulties found were situations such as: (1) when a participant missed a training session, the researcher had to meet with that student before the following session to guarantee that the same instructions were given and the two training tasks were completed; (2) a one-week gap (for Easter holidays) interrupted the training program; (3) after training was over, the researcher could not immediately provide all the perception tasks to students to avoid interference in the delayed posttest, thus, only after the end of the semester did they have access to all the training materials; and (4) other uncontrollable variables, such as stress, anxiety and tiredness, caused by tests in other courses and by other classes, might have affected participants' performance during training. In addition, the fact that training was neither individualized nor self-paced did not meet the teacher's pedagogical concern in promoting effective and learner-adapted EFL learning, or the principle of learner centeredness. The ambiguity of playing both roles that implied, to some extent, distinct concerns and objectives, was sometimes difficult to manage. For example, withholding information about the specific objectives of the study, before training, was challenging to the researcher, and controlling the amount of input, avoiding the production of the target segments or providing extra articulatory descriptions was contrary to the teacher's natural tendency to reinforce and repeat information. The fact that both groups had classes together and participants knew each other contributed to some disclosure of information that the researcher would have preferred to withhold. For example, in the first weeks, when recordings were taking place, learners already knew they would have to utter sentences in English, and during the experiment, they realized that the training sessions were targeting different English segments.

Nonetheless, given the fact that it was a longitudinal experiment with three testing moments and five training sessions, it was an advantage being both the teacher and researcher because it avoided participant attrition in the two main phases of the experiment. Logistically, booking computer labs with adequate equipment and booths to record participants' productions during four weeks was facilitated by the fact that these spaces were used for classes. More importantly, the non-nativeness of the teacher, who knew the target language as a foreign language, might have promoted a better understanding of the learners' process of phonological learning. Jenkins (2009) explains this advantageous situation as follows: "the non-native speaker has been through the same process of learning the same language, often through the same L1 filter, and she knows what it is like to have made the foreign language, in some sense, her own, to have appropriated it for particular purposes" (p. 120). Finally, the experiment was completely integrated and in line with the objectives of the course.

In sum, in spite of some drawbacks, the dichotomy research vs. teaching proved to be beneficial to both dimensions, because the experiment was tightly controlled (e.g., participants' homogeneity regarding EFL background, experience, and proficiency level; between-group simultaneity of experimental phases; same testing and training context), and training program promoted learning in a short period of time, and provided further understanding of how to implement effective ways of improving FL learners' phonological skills.

As expected, considerable individual differences between EFL learners were found not only in pre-training performance but also in how each trainee responded to training. Further examination of learning progression throughout the five training sessions should be carried out to understand inter-participant variability better. Hence, we acknowledge that an individual analysis of each training session performance would be required to detect optimal (and non-optimal) learning stages, and individual learning differences. Consequently, training could have been adapted and individualized so as promote effectiveness of perceptual training to all students (e.g., Nobre-Oliveira, 2007; Pierce, 2013; Wang, 2008). Therefore, further analysis of identification and discrimination scores at each session will be carried out in future research to understand the pre- and post-training differences described in the previous chapter better.

Regarding the tool used in the experiments, the perception tests and training tasks were designed with the first version of *TP-S*, a software application developed for speech perception experiments. For that reason, it had some limitations at the time of data collection. For example, it did not include the possibility of changing a response once given by clicking on an "oops" button.

The research question regarding English vowel perceptual assimilation patterns by European Portuguese speakers was based on previous research with Brazilian Portuguese speakers (e.g., Major, 1987; Nobre-Oliveira, 2007; Rauber, 2010) and with speakers of similar small-vowel-set languages such as Spanish/Catalan (e.g., Aliaga-Garcia, 2010; Cebrian, 2006; Flege, Bohn, & Jang, 1997; Flege, Munro, & Fox, 1994; Flege, 1994, as cited in Flege, 1995), and Italian (e.g., Flege, MacKay, & Meador, 1999; Flege & MacKay, 2004). However, a perceptual assimilation task (PAT)[133] to assess the degree of cross-linguistic similarity between the vowel systems of European

[133] In a PAT, listeners identify L2 sounds in terms of L1 categories and provide goodness of fit ratings (Cebrian et al., 2010).

Portuguese and American English would have provided auditory measures of perceptual similarity between native (L1) and non-native (L2) vowels, and would have helped both to predict learners' difficulties in the perception and production of English vowels better, and to understand the participants' perceptual performance further, that is, to which L1 vowels were L2 vowels assimilated (Guion et al., 2000).

The addition of a "none" option in the identification tests could also have helped to understand the cases in which the English vowels were not categorized as any of the non-native sounds. However, it could also have been problematic, because it might have interfered with listeners' attentional effort to decide among a set of sounds. In other words, if a "none" option were included, every time participants had difficulty perceiving a vowel sound they would probably choose the easiest non-effortless option. For example, one of the native AE speakers, who validated the experiment's materials, emphasized that "sometimes there was no correct choice for what the vowel sounded like to me, because there were no /e/, /o/ or /y/ choices". Regardless of this comment, her overall score was above 95% correct.

The decision to analyze the results of the identification tests and the category goodness ratings separately was due to the small size of the goodness-of-fit scale (viz. a three-point Likert scale), and to the findings of previous research that led to the following conclusion: "Differences in the goodness ratings add only minor additional information, whereas the consistency with which listeners categorize multiple tokens of L2 categories appear to be a rather reliable indicator of L1/L2 similarity" (Strange, 2007, p. 54). Nonetheless, the combination of both measures (ID correct scores and goodness ratings) into a single variable, that is, into a fit index (see, for example, Cebrian, Mora, & Aliaga-Garcia, 2010; Guion et al., 2000) would have provided a more robust measure of participants' perceptual performance in the three testing moments. Moreover, a larger goodness-of-fit scale would have helped to understand better the perceptual assimilation patterns described previously, as Strange (2007) suggests.

To assess generalization of perceptual learning further, the identification of the target vowels could have been tested at sentence level. For example, an ambiguous sentence such as "The sheep is over there" could be presented aurally so that the vowel within the content word could be identified, or, as Pereira and Hazan (2013) suggested, true/false or acceptable/non-acceptable judgments of sentences such as "The ship is eating the grass" could be used to assess generalization of phonological learning. A generalization test to assess production transfer of learning to new phonetic contexts and to new monosyllabic (with consonant clusters), disyllabic or polysyllabic words could have been designed to examine, for example, if perceivers' difficulties with the high back vowel in the generalization test would be also observed in the pronunciation of untrained words with novel phonetic contexts. In addition, spontaneous or semi-spontaneous speech samples could have been elicited to test whether controlled laboratory-based production performance would generalize to more natural speech events.

To avoid interference of orthography, especially in the case of the back vowel contrast, production data could have been collected with a delayed repetition task, in which the participant hears a token with the target segment produced by a native speaker, and repeats the same token after a few seconds (see, for example, Lacabex & Lecumberri, 2010). For example, the listener hears "The word is ship", and then has to say "Now I say ship". This task should be, however, accompanied by the visualization of a picture corresponding to the token to assure that the targeted pronunciation is elicited. If only aural tokens are presented, the productive performance of L2/FL learners will depend exclusively on their perceptive skills, and information about which ability comes first, that is, whether production precedes perception or vice-versa, is lost. Piske (2013) suggests that instead of listening to a word with the target segment, L2/FL perceivers can simultaneously hear the definition of a word and see the corresponding picture. This method seems to be very effective because it avoids the interference of orthography and does not implicitly test perception as the delayed repetition task. However, its effectiveness will depend on learners' vocabulary size; thus, might not be adequate to test beginner learners' pronunciation skills. Souza and Mora (2013) suggest that to test pronunciation skills in different speech conditions the delayed repetition technique should be contextualized within spoken interactions by means of question answering.

Directly linked with learners' vocabulary size is the frequency of occurrence of English sounds. On the one hand, a vocabulary size test could have been administered to test whether participants with larger L2/FL vocabulary sizes would perceive non-native vowel sounds more accurately (e.g., Fullana, 2013). On the other hand, as Fullana (*ibid.)* suggests, further analysis of frequency of occurrence of the target sounds in an English frequency wordlist can provide additional information whether there is a correlation between vowel frequency and L2/FL learners' segmental perception. However, we agree with Hardman (2013) regarding the problematic use of L1 word frequency tools to assess L2 and FL frequency of words because L1 word familiarity differs from L2 familiarization with non-native words. Therefore, although frequency indexes were provided for the stimuli used in the tests, no analyses were carried out to verify whether there was a correlation between pronunciation accuracy and word frequency. Nonetheless, listeners' word familiarity could have been tested by means of a questionnaire, in which participants would have to say how familiar they were with words used in the study, and additionally translate the word, as suggested by Fullana (2013). However, the moment at which the familiarization test is administered in a longitudinal study must be carefully selected because it may influence participants' performance. If it is administered before the experiment, too much information may be disclosed, even if the test includes a balanced number of distractors, and if it is at the end, it is no longer a reliable measure because, at that moment, participants will be already familiarized with the words used in the experiment.

To simultaneously avoid the influence of word frequency, that is, of top-down processing and the (positive or negative) influence of orthography, that is, of sound-symbol mappings that are latent even in the absence of written input (Moyer, 2013),[134] the perceptual training tokens could have consisted of naturally produced nonce words. However, the findings of a nonce-word-based training study with ESL students revealed that though there were some gains on the ability to discriminate tense-lax vowel contrasts in nonce words, they were not generalized to real word contrasts (Pierce, 2013). The researcher explains the absence of transfer to real words by suggesting that "it is possible that a previously phonologized word is so strongly linked to an orthographic representation that an additional level of perceptual training is required". In the present study, a very few nonce and low frequency words were included to guarantee a balanced number of vowels occurring in the same phonetic contexts, but no analysis was carried out to investigate the relation between processing of language meaning and processing of acoustic information.

To complement acoustic measurements, and to further assess whether native English speakers would categorize vowels produced by L2 participants as L1 categories, an identification test followed by intelligibility judgments on a trial-by-trial basis could be presented to a group of NESs to evaluate vowel tokens produced by EFL learners at

[134] Moyer (2013) adds that the learner "sees" the word mentally while speaking or hearing it. However, this can be an advantage to L2 learning, as Basseti (2009, as cited in Moyer, 2013) explains: "orthographic input provides a visual and permanent analysis of the auditory input, which may complement a defective perception and thus enable learners to produce phonemes they have difficulty perceiving (p. 163).

the three testing moments. The percentage of accurately identified segments could then be used as an indicator of intelligibility and be correlated with perception tests scores. To examine whether acoustic differences in vowel production by the EFL learners would be judged as non-native, near- or native-like pronunciation, accent ratings could have been provided by American English listeners.

To investigate cue weighting in the categorization of English vowels further, that is, which acoustic cues (duration and/or spectral quality) EFL learners relied on to perceive the target vowels, an identification test with synthetic vowel stimuli could have been administered (e.g., Cebrian, 2006; Rauber, 2010). The results of the production pre- and posttests suggest that, at the onset of training, EFL learners relied more on duration, which is a nonfunctional phonological feature in European Portuguese, to distinguish target vowels than after training. This finding confirmed, to some extent, the desensitization hypothesis (Bohn, 1995). However, due to the type of naturally spoken stimulus included in the perception tests, no precise claims can be made in terms of perceptual cue weighting. Findings suggest that EFL listeners did not rely on durational differences between vowels of the target pairs to perceptually categorize English vowels from the onset of the experiment, because, generally, the longer vowels were not significantly more accurately identified than the shorter vowels. Nevertheless, the apparently non-use of durational cues could be only corroborated with a controlled test design in terms of duration manipulation. Though it would have been pertinent to pre- and posttest cue weighting with synthetic stimuli, the use of synthetic stimulus materials in the perception training tasks was not considered because we wanted to focus EFL learners' attention on vowel quality, rather than on duration. Moreover, although a combination of synthetic and natural stimuli seems to result in vowel perceptual learning (e.g., Wang, 2008; Wang & Munro, 2004), when comparing the training gains of participants trained with natural stimuli with those trained with synthetic stimuli, no significant difference might be found (e.g., Nobre-Oliveira, 2007). Moreover, to illustrate how EFL listeners perceived vowel formant frequencies, the vowel tokens produced by the native American English talkers could have been represented, for instance, with a Bark scale, which is said to reflect human pitch perception (Reetz & Jongman, 2009).

Finally, further analysis could have examined the effect of talker and phonetic context on the perception of the target vowels. Given that both variables were controlled, to a great extent, in the experiment design, that is, in terms of phonetic context vowels were mostly flanked by voiceless stop contexts, and talkers varied in terms of regional accent only, not in terms of English variety (e.g., Wong, 2012), in this study, only the correlations between each of these variables and correct/incorrect identification were reported. In the specific case of the identification test talkers, the baseline group of NESs did not identify regional accents consistently. For example, one of the native American English listeners affirmed: "the English used in the testing was standard American English. I thought if there had been someone with a southern American accent I would have had problems". This seems to indicate that both Southern Californian and Midwestern Iowan American English talkers were perceived as speaking standard English; conversely, another NES said: "I noticed there were multiple dialects represented by the talkers". Nevertheless, stimulus variables, viz. phonetic context and English regional accent, were thoroughly controlled, and although correlation tests were run, their effect on perception should be further examined to so as to understand EFL learners' perceptual difficulties better.

To investigate whether there was a "familiar talker advantage", that is, if trained listeners perceived more vowels correctly for familiar than for unfamiliar talkers, further analysis could have been carried out. Levi, Winters, and Pisoni (2011) found that the familiar talker advantage is not a result of voice familiarity in general, but rather it is tied to the context, which the talkers have become familiar with. Future research should look at this aspect in more detail.

Despite the limitations of the present study, we believe that detailed descriptions of the phonetic realizations of American English phonological categories and analyses of perceptual categorization patterns by Portuguese EFL learners were appropriate to assess perceptual training effects and establish a relationship between vowel production and perception. Furthermore, a gradual progression in terms of perceptual tasks is advised from low to high variability training tasks (e.g., Sebastián-Gallés, 2005). Including the whole set of English vowels in the training program seems to be more effective than only focusing on a given vowel pair (Pereira & Hazan, 2013). Therefore, further investigation should test this hypothesis.

Future research could be also carried out with EFL learners with various levels of experience with the L2/FL to further understand how L1-L2 phonetic (i.e., perceptual and production) similarity patterns change as a result of learners' training. Although there are several studies that have investigated the effect of age of onset of learning (AOL) on the perception and production of non-native sounds in naturalistic (i.e., immersion) contexts, longitudinal examination of training effects, in particular, long-term effects in formal EFL instruction settings with children can provide more information regarding non-native phonological learning in FL contexts. In Portugal, the teaching of English at primary school is an extracurricular (that is, not mandatory) activity; thus, not all children begin the mandatory English instruction at fifth grade with the same proficiency level and amount of English experience. A study focusing on this age group would provide some understanding whether pronunciation training in a formal instructional setting at this level of education would be beneficial in the long term, though, for instance, Iverson, Pinet, and Evans (2012) reported that inexperienced and experienced learners receive similar benefits from training. Furthermore, the delayed effect of perceptual learning that was observed for the back vowel contrast provides promising evidence that phonological attainment might be promoted by an intensive short-term perceptual training followed by classroom L2 input.

Data from one of the volunteers that participated in the training program, but was excluded from analysis due to a four-year immersion in an English-speaking country, revealed that, to some extent, as expected, perceptual learning might be more challenging after a certain age (the participant was 54 years old at the time of the study) due to a possibly gradual loss of hearing capability. The vowel-centered training seemed not to affect the informant’s perceptual and productive performance given that no gains were found, despite the immersion experience. More than investigating whether age above 40 years old is detrimental to perceptual learning, an examination of which training techniques or approaches would promote phonological learning effectively when there is some hearing loss (either in the case of older people or partially hearing impaired people) could contribute, to some extent, to the development of adequate pronunciation teaching materials.

The findings of this study provide further evidence that short-term high-variability perceptual training carried out in an EFL classroom is effective in modifying categorization and pronunciation of non-native sounds, thus contributing to knowledge about phonological learning of English sounds by Portuguese learners. In terms of teaching materials it provided data that might contribute to an appropriateness of pronunciation activities regarding vowel segments and directed at Portuguese EFL learners.

Moreover, the data collected, in particular, in terms of the acoustic-phonetic realizations of English vowels might offer useful information for the design of products and tools that require automatic speech recognition (ASR) technologies, specifically multimedia devices addressed at Portuguese clients. Thorough and detailed information regarding English mispronunciation patterns can help adapting ASR system conventions to a better recognition of Portuguese accented speech. Given that English is the main language of popular culture, the pronunciation of words and expressions (e.g., titles of songs and movies, and names of music bands and musicians) is nowadays elicited by many tools, such as iPhones/smartphones, smart TVs, car audio equipment, and the "machines" have to be able to recognize not only native accents but also non-native speech.

In sum, the present study has provided evidence that short-term perceptual training is effective in promoting phonological learning of non-native vowels in the foreign language classroom, improving both perception and production of the target sounds.

REFERENCES

Albini, A. B. (2012). *Os efeitos de um treinamento de percepção na aquisição do padrão acentual pré-proparoxítono da língua inglesa por estudantes brasileiros* (Unpublished doctoral dissertation). Universidade Federal do Paraná, Curitiba, Brazil.

Aliaga-Garcia, C. (2010). Measuring perceptual cue weighting after training: A comparison of auditory vs. articulatory training methods. In K. Dziubalska-Kołaczyk, M. Wrembel, & M. Kul (Eds.), *New Sounds 2010: Proceedings of the Sixth International Symposium on the Acquisition of Second Language Speech* (pp. 2-7).

Aliaga-Garcia, C. (2013). *Effects of identification and articulatory training methods on L2 vowel production*. Paper presented at New Sounds 2013, 17-19 May, Concordia University, Montreal, Canada.

Aliaga-Garcia, C., & Mora, J. C. (2007). Assessing the effects of phonetic training on L2 sound perception and production. In A. Rauber, M. Watkins, & B. Baptista (Eds.), *New Sounds 2007: Proceedings of the Fifth International Symposium on the Acquisition of Second Language Speech* (pp. 25-28).

Aliaga-Garcia, C., & Mora, J. (2009). Assessing the effects of phonetic training on L2 sound perception and production. In M. Watkins, A. Rauber, & B. Baptista (Eds.), *Recent Research in Second Language Phonetics/Phonology: Perception and Production* (pp. 2-31). Newcastle upon Tyne: Cambridge Scholars Publishing.

Aliaga-Garcia, C., Mora, J., & Cerviño-Povedano, E. (2010). Phonological short-term memory and L2 speech learning in adulthood. In K. Dziubalska-Kołaczyk, M. Wrembel, & M. Kul (Eds.), *New Sounds 2010: Proceedings of the Sixth International Symposium on the Acquisition of Second Language Speech* (pp. 12-18).

Baker, W., Trofimovich, P., Flege, J., Mack, M., & Halter, R. (2008). Child-adult differences in second-language phonological learning: The role of cross-language similarity. *Language and Speech*, *51*(4), 317-342. doi: 10.1177/0023830908099068

Barroso, H. (1999). *Forma e Substância da Expressão da Língua Portuguesa*. Coimbra: Almedina.

Becker, M. (2013). *Inteligibilidade da língua inglesa sob o paradigma de lingua franca: percepção de discursos de falantes de diferentes L1s por brasileiros* (Unpublished doctoral dissertation). Federal University of Paraná, Curitiba, Brazil.

Beddor, P., & Gottfried, T. (1995). Methodological Issues in Cross-Language Speech Perception Research with Adults. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross Language Research* (pp. 207-232). Timonium, MD: York Press.

Best, C. (1995). A Direct Realist View of Cross-Language Speech Perception. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross Language Research* (pp. 171-204). Timonium, MD: York Press.

Best, C., & Tyler, M. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In O.-S. Bohn, & M. Munro (Eds.), *Language Experience in Second Language Speech Learning – In honor of James Emil Flege* (pp. 13-34). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Bion, R., Escudero, P., Rauber, A., & Baptista, B. (2006). Category formation and the role of spectral quality in the perception and production of English front vowels. *Interspeech 2006 – ICSLP*, 17-21.

Birdsong, D. (2004). Second Language Acquisition and Ultimate Attainment. In A. Davies, & C. Elder (Eds.), The Handbook of Applied Linguistics (pp. 82-105). London: Blackwell.

Boersma, P. (2007). Some listener-oriented accounts of *h*-aspiré in French. *Lingua, 117*, 1989-2054. doi:10.1016/j.lingua.2006.11.004

Boersma, P., & Hamann, S. (Eds.) (2009). *Phonology in Perception*. Berlin: Mouton de Gruyter.

Bohn, O.-S. (1995). Cross-language Speech Perception in Adults: First Language Transfer Doesn't Tell It All. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross Language Research* (pp. 275-300). Timonium, MD: York Press.

Bohn, O.-S. (2000). Linguistic relativity in speech perception: An overview of the influence of language experience on the perception of speech sounds from infancy to adulthood. In S. Niemeier, & R. Dirven (Eds.), *Evidence for Linguistic Relativity* (pp. 1-28). Amsterdam, Philadelphia: John Benjamins.

Bohn, O.-S. & Caudery, T. (2007). *The sounds of English - An activity-based course in English Phonetics and Phonology*. Denmark: University of Aarhus.

Bohn, O.-S. & Flege, J. (1990). Interlingual identification and the role of foreign language experience in L2 vowel perception. *Applied Psycholinguistics, 11*, 303-328. doi: 10.1017/S0142716400008912

Bohn, O.-S. & Flege, J. (1992). The production of new and similar vowels by adult German learners of English. *Studies in Second Language Acquisition, 14*, 131-158. doi: 10.1017/S0272263100010792

Bohn, O.-S., & Steinlen, A. (2003). Consonantal context affects cross-language perception of vowels. In M. Solé, D. Recasens, & J. Romero (Eds.), *Proceedings of the 15th International Congress of Phonetic Sciences* (pp. 2289-2292).

Bradlow, A. (2008). Training non-native language sound patterns: Lessons from training Japanese adults on the Englisn /r/ /l/ contrast. In J. Hansen Edwards, & M. Zampini (Eds.), *Phonology and Second Language Acquisition* (pp. 287-308). Amsterdam/Philadelphia: John Benjamins Publishing Company.

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*(4), 2299-2310. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3507383/

Bradlow, A., Yamada, R., Pisoni, D., & Tohkura, Y. (1999). Training Japanese Listeners to identify English /r/ and /l/: Long-term retention of learning in perception and production. *Perception & Psychophysics*, *61*(5), 977-985. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3472521/

Bresnahan, M., Ohashi, R., Nebashi, R., Liu, W. Y., & Shearman, S. M. (2002). Attitudinal and affective response toward accented English. *Language and Communication*, *22*, 171-185. doi: 10.1016/S0271-5309(01)00025-8

Callan, D. Tajima, K., Callan. A., Kubo, R., Masaki, S., & Yamada, R. (2003). Learning-induced neural plasticity associated with improved identification performance after training of a difficult second-language phonetic contrast. *NeuroImage*, *19*, 113-124. doi: 016/S1053-8119(03)00020-X

Carlet, A. (ongoing). Second language perception and production of English consonants and vowels by Catalan speakers: The effect of implicit learning in a cross-training study. Unpublished manuscript. Universitat Autonoma de Barcelona, Spain.

Cebrian, J. (2006). Experience and the Use of Duration in the Categorization of L2 Vowels. *Journal of Phonetics*, *34*, 372-387.

Cebrian, J. (2009). Interaction of L2 phonotactics and L1 syllable structure in L2 vowel production. In M. Watkins, A. Rauber, & B. Baptista (Eds.), *Recent Research in Second Language Phonetics/Phonology: Perception and Production* (pp. 234-243). Newcastle upon Tyne: Cambridge Scholars Publishing.

Cebrian, J., Mora, J. C., & Aliaga-Garcia, C. (2010). Assessing crosslinguistic similarity by means of rated discrimination and perceptual assimilation tasks. In K. Dziubalska-Kołaczyk, M. Wrembel, & M. Kul (Eds.), *New Sounds 2010: Proceedings of the Sixth International Symposium on the Acquisition of Second Language Speech* (pp. 77-82).

Celce-Murcia, M., Brinton, D., & Goodwin, J. (1996). *Teaching pronunciation*. Cambridge: Cambridge University Press.

Chládková, K., & Escudero, P. (2012). Comparing vowel perception and production in Spanish and Portuguese: European versus Latin American dialects. *Journal of the Acoustical Society of America*, *131*(2), 119-125. doi: 10. 1121/1.3674991

Clopper, C., Pisoni, D. & Jong, K. (2005). Acoustic characteristics of the vowel systems of six regional varieties of American English. *Journal of the Acoustical Society of America*, *118*(3), 1661-1676. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3432912/

Collins, B., & Mees, I. (2008). *Practical phonetics and phonology: A resource book for students* (2nd ed.). London and New York: Routledge.

Cruz-Ferreira, M. (1999). *Handbook of the International Phonetic Association*. Cambridge: Cambridge University Press.

Crystal, D. (2003). *English as a Global Language* (2nd ed.). Cambridge: Cambridge University Press.

Darcy, I., Mora, J., & Daidone, D. (2013). *Phonological attention control, inhibition, and second language speech learning.* Paper presented at New Sounds 2013, 17-19 May, Concordia University, Montreal, Canada.

De Bot, K. (1992). A bilingual production model: Levelt's speaking model adapted. *Applied Linguistics*, *13*, 1-24.

Delgado-Martins, M. R. (1973). Análise acústica das vogais tónicas do português. In M. R. Delgado-Martins (2002), *Fonética do Português. Trinta Anos de Investigação* (pp. 41-52). Lisboa: Editorial Caminho.

Delgado-Martins, M. R. (1975). Vogais e Consoantes do Português: estatística de ocorrências, duração e intensidade. In M. R. Delgado-Martins (2002), *Fonética do Português. Trinta Anos de Investigação* (pp. 53-63). Lisboa: Editorial Caminho.

Delgado-Martins, M. R. (1998). *Ouvir Falar – Introdução à Fonética do Português*. Lisboa: Editorial Caminho.

Diehl, R., Lotto, A., & Holt, L. (2004). Speech Perception. *Annual Review of Psychology, 55*, 149-179. doi: 10.1146/annurev.psych.55.090902.142028

Dörnyei, Z., & Csizér, K. (2012). How to Design and Analyze Surveys in Second Language Acquisition Research. In A. Mackey. & S. Gass (Eds.), *Research Methods in Second Language Acquisition: A Practical Guide* (pp. 74-94). West Sussex: Wiley-Blackwell.

Emiliano, A. (2009). *Fonética do Português Europeu: Descrição e Transcrição*. Lisboa: Guimarães Editores.

Escudero, P. (2005). *Linguistic Perception and Second Language Acquisition. Explaining the attainment of optimal phonological categorization* (Doctoral dissertation). Retrieved from http://www.fon.hum.uva.nl/theses/

Escudero, P. (2009). Linguistic Perception of "similar" L2 sounds. In P. Boersma, & H. Silke (Eds.) *Phonology in Perception* (pp. 151-190). Berlin: Mouton de Gruyter.

Escudero, P., Boersma, P., Rauber, A., Bion, R. (2009). A cross-dialect acoustic description of vowels: Brazilian and European Portuguese. *Journal of the Acoustical Society of America, 126*(3), 1379-1393. doi: 10.1121/1.3180321

Escudero, P., & Chládková, K. (2010). Spanish listeners' perception of American and Southern British English vowels. *Journal of the Acoustical Society of America, 128*(5), EL254-259. doi: 10.1121/1.3488794

Escudero, P., & Williams, D. (2011). Perceptual assimilation of Dutch vowels by Peruvian Spanish listeners. *Journal of the Acoustical Society of America, 129*(1), EL1-7. doi: 10.1121/1.3525042

Fant, G. (1960). *Acoustic Theory of Speech Production*. The Hague, Netherlands: Mouton.

Field, A. (2009). *Discovering Statistics using SPSS* (3rd ed.). London: Sage.

Fife-Schaw, C. (2006). Levels of Measurement. In G. M. Breakwell, S. Hammond, C. Fife-Schaw, & J. A. Smith (Eds.), *Research Methods in Psychology* (3rd ed.) (pp. 50-63). London: Sage.

Flege, J. (1984). The effect of linguistic experience on Arabs' perception of the English /s/ vs. /z/ contrast. *Folia Linguistica, 18*, 117-138. doi: 10.1515/flin.1984.18.1-2.117

Flege, J. (1987). The production of "new" and "similar" phones in a foreign language: evidence for the effect of equivalence classification. *Journal of Phonetics*, *15*, 47-56. Retrieved from http://www.jimflege.com/files/Flege_new_similar_JP_1987.pdf

Flege, J. (1988). Effects of speaking rate on tongue position and velocity of movement in vowel production. *Journal of the Acoustical Society of America, 84*(3), 901-916. Retrieved from http://jimflege.com/files/Flege_speaking_rate_JASA_1988.pdf

Flege, J. (1989). Chinese subjects' perception of the word-final English /t/-/d/ contrast: Performance before and after training. *Journal of the Acoustical Society of America, 58*(5), 1684-1697. Retrieved from http://jimflege.com/files/Flege_Chinese_subjects_JASA_1989.pdf

Flege, J. (1991). Age of learning affects the authenticity of voice-onset time (VOT) in stop consonants produced in a second language. *Journal of the Acoustical Society of America, 89*(1), 395-411. Retrieved from http://jimflege.com/files/Flege_authenticity_VOT_JASA_1991.pdf

Flege, J. (1993). Production and perception of a novel, second-language phonetic contrast. *Journal of the Acoustical Society of America, 93*(3), 1589-1608. Retrieved from http://www.jimflege.com/files/Flege_novel_contrast_JASA_1993.pdf

Flege, J. (1995). Second Language Speech Learning: Theory, Findings and Problems. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross Language Research* (pp. 233-277). Timonium, MD: York Press.

Flege, J., Bohn, O.-S. & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics*, *25*, 437-470. Retrieved from http://jimflege.com/files/Flege_Bohn_effect_experience_JP_1997.pdf

Flege, J., & Hillenbrand, J. (1986). Differential use of temporal cues to the /s/-/z/ contrast by native and non-native speakers of English. *Journal of the Acoustical Society of America*, *79*(2), 508-517. doi: 10.1121/1.393538

Flege, J., MacKay, I., & Meador, D. (1999). Italian speakers' perception and production of English vowels. *Journal of the Acoustical Society of America*, *106*(5), 2973-2987. Retrieved from http://www.utdallas.edu/~assmann/hcs6367/flege_makay_meador99.pdf

Flege, J., & MacKay, I. (2004) Perceiving vowels in a second language. *Studies in Second Language Acquisition*, *26*, 1-34. doi: 10.1017/50272263104261010

Flege, J., Munro, M., & Fox, A. (1994). Auditory and categorical effects on cross-language vowel perception. *Journal of the Acoustical Society of America, 95*(6), 3623-3641. doi: 10.1121/1.409931

Flege, J., Munro, M., & MacKay, I. (1995a). Factors affecting strength of perceived foreign accent in a second language. *Journal of the Acoustical Society of*

America, *97*(5), 3125-3134. Retrieved from http://jimflege.com/files/Flege_Munro_factors_affecting_JASA_1995.pdf

Flege, J., Munro, M., & MacKay, I. (1995b). Effects of age of second-language learning on the production of English consonants. *Speech Communication*, *16*(1), 1-26. doi: 10.1016/0167-6393(94)00044-B

Flege, J., Schirru, C., & Mackay, I. (2003). Interaction between the native and second language phonetic subsystems. *Speech Communication*, *40*, 467-491. Retrieved from http://jimflege.com/files/Flege_Schirru_interaction_between_SC_2003.pdf

Fowler, C. A. (1986). An event approach study of speech perception from a direct-realist perspective. *Journal of Phonetics*, *14*, 3-28. Retrieved from http://www.haskins.yale.edu/Reprints/HL0543.pdf

Fowler, C. A. (2003). Speech production and perception. In A. Healy, & R. Proctor (Eds.), *Handbook of psychology*, *vol. 4*: *Experimental psychology* (pp. 237-266). New York: John Wiley & Sons.

Fowler, C. A., & Magnuson, J. S. (2012). Speech Perception. In M. Spivey, K. McRae, & M. Joanisse (Eds.), *The Cambridge Handbook of Psycholinguistics* (pp. 3-20). doi: 10.1017/CBO9781139029377

Fullana, N., & Mackay, I. (2013). *L2 learners' segmental perception in an instructed setting: An investigation of the effects of vocabulary size and frequency of occurrence of English /æ/, /ɒ/, and /ʌ/*. Paper presented at New Sounds 2013, 17-19 May, Concordia University, Montreal, Canada.

Fullana, N., & Mora, J. (2009). Production and perception of voicing contrasts in English word-final obstruents: Assessing the effects of experience and starting age. In M. Watkins, A. Rauber, & B. Baptista (Eds.), *Recent Research in Second Language Phonetics/Phonology: Perception and Production* (pp. 97-117). Newcastle upon Tyne: Cambridge Scholars Publishing.

Giegerich, H. (1992). *English Phonology – An introduction*. Cambridge: Cambridge University Press.

Giles, H., Williams, A., Mackie, D. M., & Rosselli, F. (1995). Reactions to Anglo- and Hispanic-American Accented Speakers: Affect, Identity, Persuasion, and the English-only Controversy. *Language and Communication*, *15*(2), 107-120. doi: 10.1016/0271-5309(94)00019-9

Gonçalves, A. (2007). *Que língua falamos, quando falamos inglês? Para o estudo do inglês como língua de comunicação internacional* (Unpublished master's thesis). Universidade de Lisboa, Lisboa.

Gottfried T. L. (1984). Effects of consonant context on the perception of French vowels. *Journal of Phonetics*, *12*, 91-114. Retrieved from http://www2.lawrence.edu/fast/gottfrit/Gottfried1984.pdf

Gottfried, T. L. (2007). Music and language learning: Effect of musical training on learning L2 speech contrasts. In O.-S. Bohn, & M. J. Munro (Eds.), *Language experience in second language speech learning: In honor of James Emil Flege* (pp. 221-237). Amsterdam: John Benjamins.

Gottfried, T., Miller, J., & Payton, P. (1990). Effect of speaking rate on the perception of vowels. *Phonetica*, *47*, 155-172.

Green, D. W. (1986). Control, activation and resource: A framework and a model for the control of speech in bilinguals. *Brain and Language*, *27*(2), 210-223. doi: 10.1016/0093-934X(86)90016-7

Guion, S. G., Flege, J., Akahane-Yamada, R., & Pruitt, J. (2000). An investigation of current models of second language speech perception: The case of Japanese adults' perception of English consonants. *Journal of the Acoustical Society of America*, *107*(5), 2711-2724. Retrieved from http://www.bcs.rochester.edu/courses/crsinf/561/ARCHIVES/S06/Guion-Flege_JASA.pdf

Hagiwara, R. (1997). Dialect variation and formant frequency: The American English vowels revisited. *Journal of the Acoustical Society of America*, *102*(1), 655-658. doi: 10.1121/1.419712

Hardison, D. (2003). Acquisition of second-language speech: Effects of visual cues, context, and talker variability. *Applied Psycholinguistics*, *24*, 495-522. doi:10.1017.S0142716403000250

Hardison, D. (2004). Generalization of computer-assisted prosody training: Quantitative and qualitative findings. *Language Learning and Technology*, *8*(1), 34-52. Retrieved from http://llt.msu.edu/vol8num1/hardison/default.html

Hardman, J. (2013). *Accentedness and intelligibility of Mandarin-accented English for Chinese, Koreans and Americans*. Paper presented at New Sounds 2013, 17-19 May, Concordia University, Montreal, Canada.

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

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. doi: 10.1016/j.specom.2005.04.007

Hewings, M. (2004). *Pronunciation practice activities: A resource book for teaching English pronunciation.* Cambridge: Cambridge University Press.

Hillenbrand, J., Getty, L., Clark, M., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America*, *97*(5), 3099-3111. Retrieved from http://homepages.wmich.edu/~hillenbr/Papers/HillenbrandGettyClarkWheeler.pdf

Hirata, Y., & Kelly, S. (2010). Effects of lips and hands on auditory learning of second-language speech sounds. *Journal of Speech, Language and Hearing Research*, *53*(2), 298-310. doi: 10.1044/1092-4388

Horslund, C., & Bohn, O.-S. (2013). *The effect of native language inventory size and structure on L2 speech learning: Production and perception of English phones by L1 Danish and L1 Finnish speakers*. Paper presented at New Sounds 2013, 17-19 May, Concordia University, Montreal, Canada.

Iverson, P., & Evans, B. G. (2007a). Auditory training of English vowels for first-language speakers of Spanish and German. *Proceedings of the 16th International Congress of Phonetic Sciences, Germany*, 1625-1628.

Iverson, P., & Evans, B. G. (2007b). Learning English vowels with different first language vowels systems: Perception of formant targets, formant movement, and duration. *Journal of the Acoustical Society of America, 122,* 2842-2854. doi: 10.1121/1.2783198

Iverson, P., & Evans, B. G. (2009). Learning English vowels with different first-language vowel systems II: Auditory training for native Spanish and German Speakers. *Journal of the Acoustical Society of America*, *126*(2), 866-877. doi: 10.1121/1.3148196

Iverson, P., Hazan, V., & Bannister, K. (2005). Phonetic training with acoustic cue manipulations: A comparison of methods for teaching English /r/-/l/ to Japanese adults. *Journal of the Acoustical Society of* America, *118*, 3267-3278. doi: 10.1121/1.2062307

Iverson, P., & Pinet, M. (2008). Training English vowels for French speakers with varying English experience. *Journal of the Acoustical Society of America*, *123*(5), 3734-3734. doi: 10.1121/1.2935238

Iverson, P., Pinet, M., & Evans, B. (2012). Auditory training for experienced and inexperienced second-language learners: Native French speakers learning English vowels. *Applied Psycholinguistics*, *33*(1), 145-160. doi: 10.1017/S0142716411000300

Jamieson, D., & Morosan, D. (1986). Training non-native speech contrasts in adults: Acquisition of the English /ð/-/θ/contrast by francophones. *Perception and Psychophysics*, 40(4), 205-215. Retrieved from http://link.springer.com/article/10.3758%2FBF03211500

Jenkins, J. (2007). *English as a Lingua Franca: Attitude and Identity*. Oxford: Oxford University Press.

Jenkins, J. (2009). *World Englishes* (2nd ed.). London and New York: Routledge.

Jenkins, J., & Yeni-Komshian, G. (1995). Cross-language speech perception: Perspective and promise. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross Language Research* (pp. 463-479). Timonium, MD: York Press.

Johnson, K. (2003). *Acoustic and Auditory Phonetics* (2nd ed.). Malden: Blackwell Publishing.

Kachru, B. (1992). *The Other Tongue: English Across Cultures*. Illinois: University of Illinois Press.

Kerschhofer-Puhalo, N. (2010). Preferences in L2-vowel perception. In K. Dziubalska-Kołaczyk, M. Wrembel, & M. Kul (Eds.), *New Sounds 2010: Proceedings of the Sixth International Symposium on the Acquisition of Second Language Speech* (pp. 245-250).

Kessinger, R., & Blumstein, S. (1998). Effects of speaking rate on voice onset time and vowel production: Some implications for perception studies. *Journal of Phonetics*, *26*(2), 117-128. doi: 10.1006/jpho.1997.0069

Kim, J.-E. (2003). *Native language and non-linguistic influences on the production of English vowels by speakers of Korean: An acoustic study* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database.

Kim, Y., & Hazan, V. (2010). Individual variability in perceptual learning of L2 Speech sounds and its cognitive correlates. In K. Dziubalska-Kołaczyk, M. Wrembel, & M. Kul (Eds.), *New Sounds 2010: Proceedings of the Sixth International Symposium on the Acquisition of Second Language Speech* (pp. 251-256).

Kuhl, P., & Iverson, P. (1995). Linguistic Experience and the "Perceptual Magnet Effect". In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross Language Research* (pp. 121-154). Timonium, MD: York Press.

Lacabex, E. G., & Lecumberri, M. L. (2012). *Over/generalization effects in the production of English weak and full vowels in unstressed syllables after vowel reduction training: evidence for L2 sound learning in progress*. Paper presented at EUROSLA 22 – The 22nd Annual Conference of the European Second Language Association, 5-8 September, Poznan, Poland.

Lacabex, E. G., Lecumberri, M., & Cooke, M. (2008a). Identification of the contrast full-vowel schwa: training effects and generalization to a new perceptual context. *Ilha do Desterro, 55*, 173-196. doi: 10.5007/2175-8026.2008n55p173

Lacabex, E. G., Lecumberri, M., & Cooke, M. (2008b). Perception of English Vowel Reduction by Trained Spanish Speakers. In A. Rauber, M. Watkins, & B. Baptista (Eds.), *New Sounds 2007: Proceedings of the Fifth International Symposium on the Acquisition of Second Language Speech* (pp. 293-299).

Lacabex, E. G., Lecumberri, M., & Cooke, M. (2009). Training and Generalization Effects of English Vowel Reduction for Spanish Listeners. In M. Watkins, A. Rauber, & B. Baptista (Eds.), *Recent Research in Second Language Phonetics/Phonology: Perception and Production* (pp. 32-34). Newcastle upon Tyne: Cambridge Scholars Publishing.

Lacabex, E. G., & Lecumberri, M. (2010). Investigating training effects in the production of English weak forms by Spanish learners. In K. Dziubalska-Kołaczyk, M. Wrembel, & M. Kul (Eds.), *New Sounds 2010: Proceedings of the Sixth International Symposium on the Acquisition of Second Language Speech* (pp. 137-143).

Ladefoged, P. (1993). *A Course in Phonetics* (3rd ed.). Fort Worth: Harcourt Brace College Publishers.

Ladefoged, P. (1999) American English. *Handbook of the International Phonetic Association: A Guide to the Use of the International Phonetic Alphabet*, pp. 41-44.

Ladefoged, P. (2003). *Phonetic Data Analysis – An Introduction to Fieldwork and Instrumental Techniques*. Malden: Blackwell Publishing.

Ladefoged, P. (2005). *Vowels and Consonants: An Introduction to the Sounds of Languages* (2nd ed.). Malden & Oxford: Blackwell Publishing.

Lambacher, S. G., Martens, W. L., Kakehi, K., Marasinghe, C. A., & Molholt, G. (2005). The effects of identification training on the identification and production of American English vowels by native speakers of Japanese. *Applied Psycholinguistics, 26*, 227-247. doi: 10.1017/S0142716405050150

Lane, L. (2005). *Focus on Pronunciation 3*. New York: Longman.

Leather, J. (1990). Perceptual and productive learning of Chinese lexical tone by Dutch and English speakers. In A. James, & J. Leather (Eds.), *New Sounds 90:*

Proceedings of the 1990 Amsterdam symposium on the acquisition of second language speech (pp. 72-97).

Lee, S., & Cho, M. (2010). The precedence of perception over production in the acquisition of English stress by EFL learners: An OT account. In K. Dziubalska-Kołaczyk, M. Wrembel, & M. Kul (Eds.), *New Sounds 2010: Proceedings of the Sixth International Symposium on the Acquisition of Second Language Speech* (pp. 287-292).

Lengeris, A. (2008). The effectiveness of auditory phonetic training on Greek native speakers' perception and production of Southern British English vowels. *Proceedings of the 2nd ISCA Workshop on Experimental Linguistics* (pp. 133-136).

Lengeris, A. (2009). *Individual differences in second-language vowel learning* (Doctoral dissertation). Department of Speech, Hearing and Phonetic Sciences, Division of Psychology and Language Sciences, University College London. Retrieved from http://eprints.ucl.ac.uk/

Lengeris, A., & Hazan, V. (2007). Cross-language perceptual assimilation and discrimination of Southern British English vowels by native speakers of Greek and Japanese. *Proceedings of the 16th International Congress of Phonetic Sciences*, *Germany*, 1641-1644.

Lengeris, A., & Hazan, V. (2010). The effect of native vowel processing ability and frequency discrimination acuity on the phonetic training of English vowels for native speakers of Greek. *Journal of the Acoustical Society of America*, *128*(6), 3757-3768. doi: 10.1121/1.3506351

Lenneberg, E. (1967). *Biological foundations of language*. New York: Willey.

Lehiste, I. (1970). *Suprasegmentals*. Cambridge & London: MIT Press.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge & London: MIT Press.

Levelt, W. J., Roelofs, A., & Meyer, A. (1999). A theory of lexical access in speech production. *Behavioral and brain sciences*, *22*, 1-75. doi: 10.1017/S0140525X99001776

Levi, S. V., Winters, S. J., & Pisoni, D. B. (2011). Effects of cross-language voice training on speech perception: Whose familiar voices are more intelligible?. *Journal of the Acoustical Society of America*, *130*(6), 4053-4062. doi: 10.1121/1.3651816

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revisited. *Cognition*, *21*, 1-36. Retrieved from http://www.linguistica.unifi.it/upload/sub/articoli/Liberman-MattinglyMotor-theory.pdf

Lieberman, P., & Blumstein, S. (1988). *Speech physiology, speech perception, and acoustic phonetics*. Cambridge & New York: Cambridge University Press.

Lisker, L. (1974). On "explaining" vowel duration variation. Glossa, *8*, 233-246. Retrieved from http://www.haskins.yale.edu/Reprints/HL0159.pdf

Lively, S. E., Logan, J. S., & Pisoni, D. B. (1993). Training Japanese listeners to identify English /r/ and /l/: II. The role of phonetic environment and talker variability in learning new perceptual categories. *Journal of the Acoustical Society of America*, *94,* 1242–1255. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3509365/

Lively, S. E., Pisoni, D. B., Yamada, R. A., Tohkura, Y., & Yamada, T. (1994). Training Japanese listeners to identify English /r/and /l/: III. Long-term retention of new phonetic categories. *Journal of the Acoustical Society of America, 96,* 2076-2087. Retrieved from http://linguistics.berkeley.edu/~kjohnson/ling290e/lively_et_al_1994.pdf

Loewen, S., & Philp, J. (2012). Instructed second language acquisition. In A. Mackey, & S. Gass (Eds.), *Research Methods in Second Language Acquisition: A Practical Guide* (pp. 53-73). West Sussex: Wiley-Blackwell.

Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English /r/ and /l/: a first report. *Journal of the Acoustical Society of America, 89*(2), 874–886. doi: 10.1121/1.1894649

Logan, J., & Pruitt, J. (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. 351-378). Timonium, MD: York Press.

MacDonald, J., & McGurk, H. (1978). Visual influences on speech perception processes. *Perception and Psychophysics*, *24*(3), 253-257. Retrieved from http://link.springer.com/article/10.3758%2FBF03206096#page-1

MacKay, I., Flege, J. E., & Satomi, I. (2006). Evaluating the effects of chronological age and sentence duration on degree of perceived foreign accent. *Applied Psycholinguistics*, *27*(2), 157-183. doi: 10.1017/S0142716406060231

MacKay, I., & Fullana, N. (2009). Starting age and exposure effects on EFL learners' sound production in a formal learning context. In M. Watkins, A. Rauber, & B. Baptista (Eds.), *Recent Research in Second Language Phonetics/Phonology: Perception and Production* (pp. 43-61). Newcastle upon Tyne: Cambridge Scholars Publishing.

Mackey, A., & Gass, S. (2005). *Second Language Research: Methodology and Design*. New York and London: Routledge.

Major, R. (1987). Phonological Similarity, Markedness, and Rate of L2 Acquisition. *Studies in Second Language Acquisition*, *9*(1), 63-82. doi: 10.1017/S0272263100006513

Martins, C. (2011). *Manual de Análise de Dados Quantitativos Com Recurso ao IBM SPSS: Saber decidir, fazer, interpretar e redigir*. Braga: Edições Psiquilíbrios.

Martins, P., Carbone, I., Pinto., Silva, A. & Teixeira, A. (2008). European Portuguese MRI based speech production studies. *Speech Communication, 50*, 925-952. doi: 10.1016/j.specom.2008.05.019

Massaro, D. W. (1987). *Speech perception by ear and eye: A paradigm for psychological inquiry*. Hillsdale, NJ: Erlbaum.

Mateus, M. H. M. (1990). *Fonética, Fonologia e Morfologia do Português*. Lisboa: Universidade Aberta.

Mateus, M. H. M., Falé, I., & Freitas, M. J. (2005). *Fonética e fonologia do português*. Lisboa: Universidade Aberta.

McClaskey, C., Pisoni, D., & Carrell, T. (1983). Transfer of training of a new linguistic contrast in voicing. *Perception and Psychophysics*, *34*(4), 323-330. Retrieved from http://link.springer.com/article/10.3758%2FBF03203044

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices: A new illusion. *Nature*, *264*, 746-748.

Monette, D., Sullivan, T., & DeJong, R. (2011). *Applied Social Research: A tool for human services* (8th ed.). Belmont, CA: Brooks/Cole, Cengage Learning.

Morley, J. (1991). The pronunciation component in teaching English to speakers of other languages. *TESOL*, *25*(3), 481-521. Retrieved from http://deepblue.lib.umich.edu/bitstream/handle/2027.42/90128/3586981.pdf?sequence=1

Morrison, G., & Escudero, P. (2007). A cross-dialect comparison of Peninsula- and Peruvian-Spanish vowels. *Proceedings of the International Congress of Phonetic Sciences, Germany,* 1505-1508.

Mota, M. (2010). Speech production from a psycholinguistic perspective: A review of L1 and L2 models. In A. Rauber, M. Watkins, R. Silveira, & R. Koerich (Eds.), *The Acquisition of Second Language Speech: Studies in Honor of Professor Barbara O. Baptista* (pp. 37-53). Florianópolis: Editora Insular.

Moyer, A. (2013). *Foreign Accent: The Phenomenon of Non-Native Speech*. Cambridge: Cambridge University Press.

Munro, M. J. (1993). Productions of English vowels by native speakers of Arabic: Acoustic measurements and accentedness ratings. *Language and Speech*, *36*(1), 39-66. Retrieved from http://las.sagepub.com/content/36/1/39.full.pdf

Munro, M. J. (2013). *The phonetic learning process and second language teaching: A synthesis of new findings*. Paper presented at New Sounds 2013, 17-19 May, Concordia University, Montreal, Canada.

Munro, M. J., & Bohn, O.-S. (2007). The study of second language speech learning: A brief overview. In O.-S. Bohn, & M. Munro (Eds.), *Language Experience in Second Language Speech Learning – In honor of James Emil Flege* (pp. 3-11). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Munro, M. J., & Derwing, T. M. (1995a). Foreign Accent, Comprehensibility, and Intelligibility in the Speech of Second Language Learners. *Language Learning*, *45*(1), 73-97. doi: 10.1111/j.1467-1770.1995.tb00963.x

Munro, M. J., & Derwing, T. M. (1995b). Processing Time, Accent, and Comprehensibility in the Perception of Native and Foreign-Accented Speech. *Language and Speech*, *38*(3), 298-306. doi: 10.1177/002383099503800305

Nishi, K., & Kewley-Port, D. (2007). Training Japanese listeners to Perceive American English vowels: Influence of training sets. *Journal of Speech, Language and Hearing Research*, *50*, 1496-1509. Retrieved from http://www.cs.indiana.edu/~port/teach/641/Nishi_DKP_Traing.Jap.EngVs.JSLHR2007.pdf

Nishi, K., & Kewley-Port, D. (2008). Non-native speech perception training using vowel subsets: Effects of contrast choice and order of training. *Journal of Speech, Language and Hearing Research*, *51*, 1480-1493. doi: 10.1044/1092-4388(2008/07-0109)

Nobre-Oliveira, D. (2007). *The Effect of Perceptual Training on the Learning of English Vowels by Brazilian Portuguese Speakers* (Doctoral dissertation). Retrieved from http://repositorio.ufsc.br/

Nobre-Oliveira, D. (2008). Effects of Perceptual Training on the Learning of English Vowels in Non-native Settings. In A. Rauber, M. Watkins, & B. Baptista (Eds.), *New Sounds 2007: Proceedings of the Fifth International Symposium on the Acquisition of Second Language Speech* (pp. 382-389).

Nobre-Oliveira, D. (2010). Studies on L2 perceptual training: An overview. In A. Rauber, M. Watkins, R. Silveira, & R. Koerich (Eds.), *The Acquisition of Second Language Speech: Studies in Honor of Professor Barbara O. Baptista* (pp. 121-143). Florianópolis: Editora Insular.

Nozawa, T. (2009). Effects of “catch” trials on the discrimination of American English vowels by native Japanese speakers. *Journal of the Acoustical Society of America*, *125*(4), 2752-2752.

Okuno, T. (2013). *Acquisition of L2 vowel duration in Japanese by native English speakers* (Doctoral dissertation). Available from ProQuest Dissertations and Theses.

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. *Proceedings of the AVSP 2001 – Auditory-Visual Speech Processing, Denmark,* 149-154.

Penfield, W. (1965). Conditioning the uncommitted cortex for language learning. *Brain*, *88*, 787-98.

Pereira, Y., & Hazan, V. (2013). *Impact of different training modes on the perception and production of English vowels by L2 learners*. Paper presented at New Sounds 2013, 17-19 May, Concordia University, Montreal, Canada.

Peterson, G., & Barney, H. (1952). Control Methods Used in a Study of vowels. *Journal of the Acoustical Society of America*, *24*(2), 175-184. doi: 10.1121/1.1906875

Pierce, L. (2013). *Nonce word exemplar training: Cutting through the noise of meaningful speech*. Paper presented at New Sounds 2013, 17-19 May, Concordia University, Montreal, Canada.

Piske, T. (2007). Implications of James E. Flege’s research for the foreign language classroom. In O.-S. Bohn, & M. Munro (Eds.), *Language Experience in Second Language Speech Learning – In honor of James Emil Flege* (pp. 301-14). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Piske, T. (2013). *Individual differences in L2 speech perception and production*. Paper presented at New Sounds 2013, 17-19 May, Concordia University, Montreal, Canada.

Piske, T., MacKay, I., & Flege, J. (2001). Factors affecting degree of foreign accent in an L2: a review. *Journal of Phonetics*, *29*, 191-215. Retrieved from http://jimflege.com/files/Piske_MacKay_factors_affecting_JP_2001.pdf

Pisoni, D., Aslin, R., Perey, A., & Hennessy, B. (1982). Some effects of Laboratory Training on Identification and Discrimination of Voicing Contrasts in Stop Consonants. *Journal of Experimental Psychology: Human Perception and Performance*, *8*(2), 297-314. doi: 10.1037/0096-1523.8.2.297

Pisoni, D., & Lively, S. (1995). Variability and invariance in speech perception: A new look at some old problems in perceptual learning. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross Language Research* (pp. 433-459). Timonium, MD: York Press.

Polka, L., & Bohn, O.-S. (1996). A cross-language comparison of vowel perception in English-learning and German-learning infants. *Journal of the Acoustical Society of America*, *100*(1), 577-592.

Polka, L., & Bohn, O.-S. (2003). Asymmetries in vowel perception. *Speech Communication*, *41*, 221-231. doi: 10.1016/S0167-6393(02)00105-X

Polka, L., & Bohn, O.-S. (2011). Natural Referent Vowel (NRV) framework: An emerging view of early phonetic development. *Journal of Phonetics*, *39*, 467-478. doi:10.1016/j.wocn.2010.08.007

Polka, L., & Werker, J. F. (1994). Developmental changes in perception of non-native vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance*, *20*(2), 421-435. doi: 10.1037/0096-1523.20.2.421

Poulisse, N., & Bongaerts, T. (1994). *First language use in second language production*. Amsterdam: John Benjamins.

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*(3), 1684–1696. doi: 10.1121/1.2161427

Rato, A., Rauber, A., Soares, L., & Lucas, L. (2013). *Challenges in the perception and production of English front vowels by native speakers of European Portuguese*. Manuscript submitted for publication.

Rauber, A. S. (2010). *Acoustic characteristics of Brazilian English vowels: perception and production results*. Saarbrücken: Lambert Academic Publishing.

Rauber, A. S., Escudero, P., Bion, R., & Baptista, B. (2005). The interrelation between the Perception and Production of English Vowels by Native Speakers of Brazilian Portuguese. *Proceedings of Interspeech 2005* (pp. 2913-2916).

Rauber, A. S., Rato, A., Silva, A. (2010). Percepção e produção de vogais anteriores do inglês por falantes nativos de mandarim. *Diacrítica*, *24*(1), 5-23. Retrieved from http://ceh.ilch.uminho.pt/publicacoes/Diacritica24-1.pdf

Rauber, A. S., Rato, A., Kluge, D. C., & Santos, G. R. (2013). TP 3.1 Software: A tool for designing audio, visual, and audiovisual perceptual training tasks and perception tests. *Proceedings of INTERSPEECH 2013, France*, 2095-98.

Reetz, H., & Jongman, A. (2009). *Phonetics: Transcription, Production, Acoustics and Perception*. West Sussex, UK: Wiley-Blackwell.

Reis, M., & Nobre-Oliveira, D. (2008). Effects of Perceptual Training on the Identification and Production of the English Voiceless Plosives Aspiration by Brazilian EFL Learners. In A. Rauber, M. Watkins, & B. Baptista (Eds.), *New*

Sounds 2007: Proceedings of the Fifth International Symposium on the Acquisition of Second Language Speech (pp. 398-407).

Rochet, B. & Chen, F. (1992). Acquisition of the French VOT contrasts by adult speakers of Mandarin Chinese. *ICSLP (International Conference on Spoken Language Processing) Proceedings*. Ed. J.J Ohala et al. Edmonton: University of Alberta. 273-76.

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.

Safronova, E. (2013). *Listeners' cognitive skills in the perceptual judgment of foreign accent in L2 speech*. Paper presented at New Sounds 2013, 17-19 May, Concordia University, Montreal, Canada.

Sebastián-Gallés, N. (2005). Cross-language speech perception. In D. B. Pisoni, & R. E. Remez (Eds.), *The Handbook of Speech Perception* (pp. 546-566). Malden: Blackwell Publishing.

Sereno, J., & Wang, Y. (2007). Behavioral and cortical effects of learning a second language. The acquisition of tone. In O.-S. Bohn, & M. Munro (Eds.), *Language Experience in Second Language Speech Learning – In honor of James Emil Flege* (pp. 241-257). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Sheldon, A., & Strange, W. (1982). The acquisition of /r/ and /l/ by Japanese learners of English: Evidence that speech production can precede speech perception. *Applied Psycholinguistics, 3*(3), 243-261. doi: 10.1017/S0142716400001417

Silveira, R. (2009). Investigating the role of orthography in the acquisition of L2 pronunciation: a case study. In M. Watkins, A. Rauber, & B. Baptista (Eds.), *Recent Research in Second Language Phonetics/Phonology: Perception and Production* (pp. 270-290). Newcastle upon Tyne: Cambridge Scholars Publishing.

Souza, H., & Carlet, A. (2013). *Vowel inventory size and the use of temporal cues in non-native vowel perception by Catalan and Danish EFL learners*. Paper presented at New Sounds 2013, 17-19 May, Concordia University, Montreal, Canada.

Souza, H., & Mora, J. (2013). *Effects of speech rate on the intelligibility of non-native vowels.* Paper presented at New Sounds 2013, 17-19 May, Concordia University, Montreal, Canada.

Stevens, K. (1998). *Acoustic Phonetics*. Massachusetts: MIT Press.

Stevens, K.N. and Blumstein, S.E. (1978). Invariant cues for place of articulation in stop consonants. Journal of the Acoustical Society of America, 64, 1358-1368.

Strange, W. (1995). Cross-language studies of speech perception: a historical review. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross Language Research* (pp. 3-45). Timonium, MD: York Press.

Strange, W. (2007). Cross-language phonetic similarity of vowels: Theoretical and methodological issues. In O.-S. Bohn, & M. Munro (Eds.), *Language Experience in Second Language Speech Learning – In honor of James Emil*

Flege (pp. 35-55). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Strange, W. (2011). Automatic selective perception (ASP) of first and second language speech: A working model. *Journal of Phonetics, 39*(4), 456-466. doi: 10.1016/j.wocn.2010.09.001

Strange, W., Bohn, O.-S., Nishi, K., & Trent, S. A. (2005). Contextual variation in the acoustic and perceptual similarity of North German and American English vowels. *Journal of the Acoustical Society of America, 118*, 1751-1762.

Strange, W., & Dittman, S. (1984). Effects of discrimination training on the perception of /r-l/ by Japanese adults learning English. *Perception and Psychophysics, 36*(2), 131-145. Retrieved from http://link.springer.com/article/10.3758%2FBF03202673

Strange, W., Yamada, R., Fitzgerald, B., & Kubo, R. (1998). Perceptual assimilation of American English vowels by Japanese listeners. *Journal of Phonetics*, *26*(4), 311-344. doi: 10.1006/jpho.1998.0078

Sueyoshi, A., & Hardison, D. (2005). The role of gestures and facial cues in second language listening comprehension. *Language Learning*, *55*(4), 661-699. doi: 10.1111/j.0023-8333.2005.00320.x

Tees, R., & Werker, J. (1984). Perceptual Flexibility: Maintenance or Recovery of the Ability to Discriminate Non-Native Speech Sounds. *Canadian Journal of Psychology*, *38*(4), 579-590. doi: 10.1037/h0080868

Trapp, L. & Bohn, O.-S. (2000). Training Danish Listeners to identify word-final /s/ and /z/: Generalization of training and its effect on production accuracy. In A. James & J. Leather (Eds.) *Proceedings of the Fourth International Symposium on the Acquisition of Second-Language Speech – New Sounds 2000* (pp. 343-350).

Traunmüller, H. & Eriksson, A. (1995) The frequency range of the voice fundamental in the speech of male and female adults. Unpublished manuscript. Retrieved from http://www2.ling.su.se/staff/hartmut/f0_m%26f.pdf

Treiman, R., Kessler, B., & Bick, S. (2003). Influence of consonantal context on pronunciation of vowels: A comparison of human readers and computational models. *Cognition, 88,* 49-78. doi:10.1016/S0010-0277(03)00003-9

Treiman, R., Kessler, B., Devin, J., Bick, S., & Davis, M. (2006). Influence of consonantal context on the reading of vowels: Evidence from children. *Journal of Experimental Child Psychology, 93,* 1-24. doi:10.1016/j.jecp.2005.06.008

Tsukada, K., Birdsong, D., Bialystok, E., Mack, M., Sung, H. & Flege, F. (2003). The Perception and Production of English /ɛ/ and /æ/ by Korean Children and Adults Living in North America. In M. Solé, D. Recasens & J. Romero (Eds.), *Proceedings of the 15th International Congress of Phonetics Sciences* (pp. 1589-1592).

Van Son, R., & Pols, L. (1995). How transitions and local context affect segment identification. *Proceedings of the Institute of Phonetic Sciences of the University of Amsterdam, Netherlands, 19,* 51-69.

Wang, X. (2008). *Perceptual Training for Learning English Vowels – Perception, Production, and Long-Term Retention*. Saarbrücken: VDM Verlag Dr. Müller.

Wang, Y., Jongman, A., & Sereno, J. (2003). Acoustic and perceptual evaluation of Mandarin tone productions before and after training. *Journal of the Acoustical Society of America*, *113*(2), 1033-1043. Retrieved from http://neuroling.ku.edu/~kuppl/documents/JASAtrainprodfinal.pdf

Wang, Y., Jongman, A., & Sereno, J. (2006). Second language acquisition and processing of Mandarin tone. In E. Bates, L. Tan, & O. Tzeng (Eds.), *Handbook of Chinese Psycholinguistics* (pp. 250-257). Cambridge: Cambridge University Press.

Wang, X., & Munro, M. J. (2004). Computer-based training for learning English vowel contrasts. *System*, *32*(4), 539-552. doi:10.1016/j.system.2004.09.011

Wang, Y., Spence, M., Jongman, A., & Sereno, J. (1999). Training American listeners to perceive Mandarin tones. *Journal of the Acoustical Society of America, 106*(6), 3649-3658. doi: 10.1121/1.428217

Wayland, R., & Guion, S. (2004). Training English and Chinese Listeners to Perceive Thai Tones: A Preliminary Report. *Language Learning*, *54*(4), 681-712. doi: 10.1111/j.1467-9922.2004.00283.x

Wayland, R., & Li, B. (2008). Effects of two training procedures in cross-language perception of tones. *Journal of Phonetics*, *36*(2), 250-267. Retrieved from http://www.sciencedirect.com/science/article/pii/S0095447007000277

Werker, J. (1989). Becoming a Native Listener. *American Scientist*, *7*, 54-59. Retrieved from http://www.utdallas.edu/~assmann/hcs6367/werker89.pdf

Werker, J., & Tees, R. (1984). Phonemic and phonetic factors in adult cross-language speech-perception. *Journal of the Acoustical Society of America*, *75*(6), 1866-1878. doi: 10.1121/1.390988

Werker, J., & Tees, R. (1999). Influences on infant speech processing: Toward a new synthesis. *Annual Review of Psychology, 50,* 509-535. doi: 10.1146/annurev.psych.50.1.509

Wilson, S., Saygin, A., Sereno, M., & Iaconobi, M. (2004). Listening to speech activates motor areas involved in speech production. *Nature Neuroscience*, *7*, 701–702. Retrieved from http://stephenw.bol.ucla.edu/papers/mirsp.pdf

Wong, J. (2012). Training the Perception and Production of English /e/ and /æ/ of Cantonese ESL Learners: A Comparison of Low vs. High Variability Phonetic Training. *Proceedings of the 14th Australasian International Conference on Speech Science and Technology* (pp. 37-40).

Wong, J. (2013). *Does proficiency matter? Effects of High Variability Phonetic Training on the Perception and Production of English Vowels by Cantonese ESL learners with high and low proficiency levels*. Paper presented at 3rd International Conference on English Pronunciation: Issues & Practices (EPIP3), 8-10 May, University of Murcia, Spain.

Yamada, R. (1995). Age and Acquisition of Second Language Speech Sounds Perception of American English /r/ and /l/ by Native Speakers of Japanese. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross Language Research* (pp. 305-320). Timonium, MD: York Press.

Yamada, R., Tohkura, Y., Bradlow, A. & Pisoni, D. (1996). Does Training in Speech Perception Modify Speech Production?. *Proceedings of the Fourth International Conference on Spoken Language Processing, 2* (pp. 606-609).

Yates, J. (2005). *Pronounce it perfectly in English* (2nd ed.). New York: Barron's Educational Series.

Yeni-Komshian, G. (1998). Speech Perception. In J. B. Gleason, & N. B. Ratner (Eds.), *Psycholinguistics* (pp. 107-156). Fort Worth, TX: Harcourt Brace.

Ylinen S., Uther, M., Latvala, A., Vepsäläinen, S., Iverson, P., Akahane-Yamada R., & Näätänen, R. (2010). Training the brain to weight speech cues differently: a study of Finnish second-language users of English. *Journal of Cognitive Neuroscience*, *22*(6), 1319-1332. doi: 10.1162/jocn.2009.21272

Software:

Audacity (Version 2.0.2) [Computer program]. Retrieved from http://audacity.sourceforge.net/

Boersma, P., & Weenink, D. (2013). Praat: doing phonetics by computer (Version 5.3.39) [Software]. Retrieved from http://www.praat.org/

Davis, C. (2005). N-Watch [Computer program]. Retrieved from http://www.pc.rhul.ac.uk/staff/c.davis/Utilities/

IBM Corp. (2011). IBM SPSS Statistics (Version 20) [Software]. New York: IBM Corp.

Tyler, K., & Thomas, E. R. (2010). Vowels: Vowel Manipulation, Normalization, and Plotting in R. R package, (Version 1.1.) [Software]. Retrieved from http://ncslaap.lib.ncsu.edu/tools/norm/

Rauber, A., Rato, A., Kluge, D., & Santos, G. (2011). TP-S (Version 1.0). [Software]. Brazil: Worken.

APPENDICES

Appendix A

CEF Descriptors of Phonological Control per Proficiency Level

Proficient user	C2	No descriptor available
	C1	Can vary intonation and place stress correctly in order to express finer shades of meaning
Independent user	B2	Has a clear, natural pronunciation and intonation
	B1	Pronunciation is clearly intelligible even if a foreign accent is sometimes evident and occasional mispronunciations occur
Basic user	A2	Pronunciation is generally clear enough to be understood despite a noticeable foreign accent, but conversational partners will need to ask for repetition from time to time
	A1	Pronunciation of a very limited repertoire of learnt words and phrases can be understood with some effort by native speakers and used to dealing with speakers of his/her language group

Source: Table "Phonological Control" (2001, as quoted in Moyer, 2013, p.165).

Appendix B

Acoustic Measurements of American English Vowels

Table 1

Acoustic Measurements of American English Vowels

vowel			i	ɪ	e	ɛ	æ	ɑ	ɔ	o	ʊ	u	ʌ	ɝ
PB (1952) GA	**F0 (Hz)**	**F**	235	232	-	223	210	212	216	-	232	231	221	218
		M	136	135	-	130	127	124	129	-	137	141	130	133
	F1 (Hz)	**F**	310	430	-	610	860	850	590	-	470	370	760	500
		M	270	390	-	530	660	730	570	-	440	300	850	490
	F2 (Hz)	**F**	2790	2480	-	2330	2050	1220	920	-	1160	950	1400	1640
		M	2290	1990	-	1840	1720	1090	840	-	1020	870	1190	1350
	F3 (Hz)	**F**	3310	3070	-	2990	2850	2810	2710	-	2680	2670	2780	1960
		M	3010	2550	-	2480	2410	2440	2410	-	2240	2240	2390	1690
HGCW (1995) Northern Midwest	**F0 (Hz)**	**F**	227	224	219	214	215	215	210	217	230	235	218	217
		M	138	135	129	127	123	123	121	129	133	143	133	130
	F1 (Hz)	**F**	437	483	536	731	699	936	781	555	519	459	753	523
		M	342	427	476	580	588	768	652	497	469	378	623	474
	F2 (Hz)	**F**	2761	2365	2530	2058	2349	1551	1136	1035	1225	1105	1426	1588
		M	2322	2034	2089	1799	1952	1333	997	910	1122	997	1200	1379
	F3 (Hz)	**F**	3372	3053	3047	2979	2972	2815	2824	2828	2827	2735	2933	1929
		M	3000	2684	2691	2605	2601	2522	2538	2459	2434	2343	2550	1710
	Duration (ms)	**F**	306	237	320	254	332	323	353	326	249	303	226	321
		M	243	192	267	189	278	267	283	265	192	237	188	263
H (1997) Southern California	**F1 (Hz)**	**F**	362	467	440	808	1017	997	-	516	486	395	847	477
		M	291	418	403	529	685	710	-	437	441	323	574	429
	F2 (Hz)	**F**	2897	2400	2655	2163	1810	1390	-	1391		1700	1753	1558
		M	2338	1807	2059	1670	1601	1221	-	1188	1366	1417	1415	1362
	F3 (Hz)	**F**	3495	3187	3252	3065	2826	2743	-	2904	2926	2866	2989	1995
		M	2920	2589	2690	2528	2524	2405	-	2430	2446	2399	2496	1679
R (2010) California (West)	**F0 (Hz)**	**F**	195	192	208	182	178	176	182	200	166	203	177	-
		M	135	131	128	125	118	117	119	135	127	134	123	-
	F1 (Hz)	**F**	308	501	450	704	820	749	705	519	540	335	718	-
		M	276	423	398	582	671	642	628	439	454	306	605	-
	F2 (Hz)	**F**	2766	2121	2386	1910	1808	1293	1239	1492	1554	1782	1695	-
		M	2331	1884	2056	1729	1669	1106	1083	1254	1371	1556	1406	-
	F3 (Hz)	**F**	3310	2975	3024	2839	2668	2654	2659	2735	2750	2730	2747	-
		M	2918	2593	2667	2562	2431	2439	2468	2349	2375	2269	2493	-
	Duration	**F**	130	103	49*	116	167	152	168	48*	114	127	110	-

	(ms)	**M**	140	118	52*	134	179	169	174	49*	128	135	131	-
			i	**ɪ**	**e**	**ɛ**	**æ**	**ɑ**	**ɔ**	**o**	**ʊ**	**u**	**ʌ**	**ɚ**
RRS (2010) Midwest (Iowa)	**F1 (Hz)**	**F**	393	565	-	816	713	-	-	-	-	-	-	-
	F2 (Hz)	**F**	2744	2228	-	1968	1998	-	-	-	-	-	-	-
	Duration (ms)	**F**	134	82	-	104	154	-	-	-	-	-	-	-

Note. PB=Peterson and Barney; HGCW =Hillenbrand, Getty, Clark and Wheeler; H=Hagiwara; R=Rauber; RRS=Rauber, Rato and Silva.
* The asterisks next to the duration values of /e/ and /o/ indicate that, although these vowels were produced as diphthongal realizations /eɪ/ and /oʊ/, only the duration value of the first element is provided.

Table 2

Acoustic Measurements of American English Vowels by Regional Dialect (Clopper et al., 2005)

Vowel			**i**	**ɪ**	**e**	**ɛ**	**æ**	**ɑ**	**ɔ**	**o**	**ʊ**	**u**	**ʌ**
New England	**F1 (Hz)**	**F**	370	557	490	819	941	933	861	581	640	411	808
		M	299	431	413	570	666	659	612	452	450	346	591
	F2 (Hz)	**F**	2852	2321	2566	2053	1986	1380	1310	1259	1554	1275	1580
		M	2245	1950	2103	1807	1742	1121	1064	1074	1305	1134	1337
Mid-Atlantic	**F1 (Hz)**	**F**	364	529	438	819	1022	1043	910	529	629	407	782
		M	280	423	387	540	663	693	630	448	460	331	560
	F2 (Hz)	**F**	2885	2416	2757	2162	2085	1548	1414	1259	1549	1530	1561
		M	2292	1909	2138	1763	1669	1220	1175	1034	1215	1093	1209
North	**F1 (Hz)**	**F**	331	497	469	815	789	947	835	547	550	405	729
		M	292	451	432	607	639	798	638	480	475	334	605
	F2 (Hz)	**F**	2842	2238	2515	1933	2132	1468	1270	1142	1365	1288	1394
		M	2364	2013	2209	1820	1960	1339	1074	1001	1180	1060	1254
Midland	**F1 (Hz)**	**F**	321	464	485	698	874	775	755	531	507	406	713
		M	299	437	426	547	695	689	599	469	473	338	579
	F2 (Hz)	**F**	2821	2258	2459	2076	1950	1244	1229	1310	1472	1457	1571
		M	2288	1948	2091	1797	1773	1155	1049	1141	1281	1246	1329
South	**F1 (Hz)**	**F**	379	567	581	776	972	939	838	628	625	405	776
		M	276	397	422	496	655	676	574	452	399	320	546
	F2 (Hz)	**F**	2980	2324	2544	2131	2013	1398	1301	1497	1641	1586	1764
		M	2181	1882	1980	1800	1779	1108	953	1177	1238	1425	1247
West	**F1 (Hz)**	**F**	338	468	477	791	984	937	859	572	598	430	742
		M	262	421	409	555	696	678	637	440	432	307	579
	F2 (Hz)	**F**	2994	2356	2708	2108	1966	1292	1316	1325	1564	1466	1617
		M	2251	1926	2095	1747	1681	1075	1063	1047	1221	1237	1307

Appendix C

MRI Images of EP Vowels

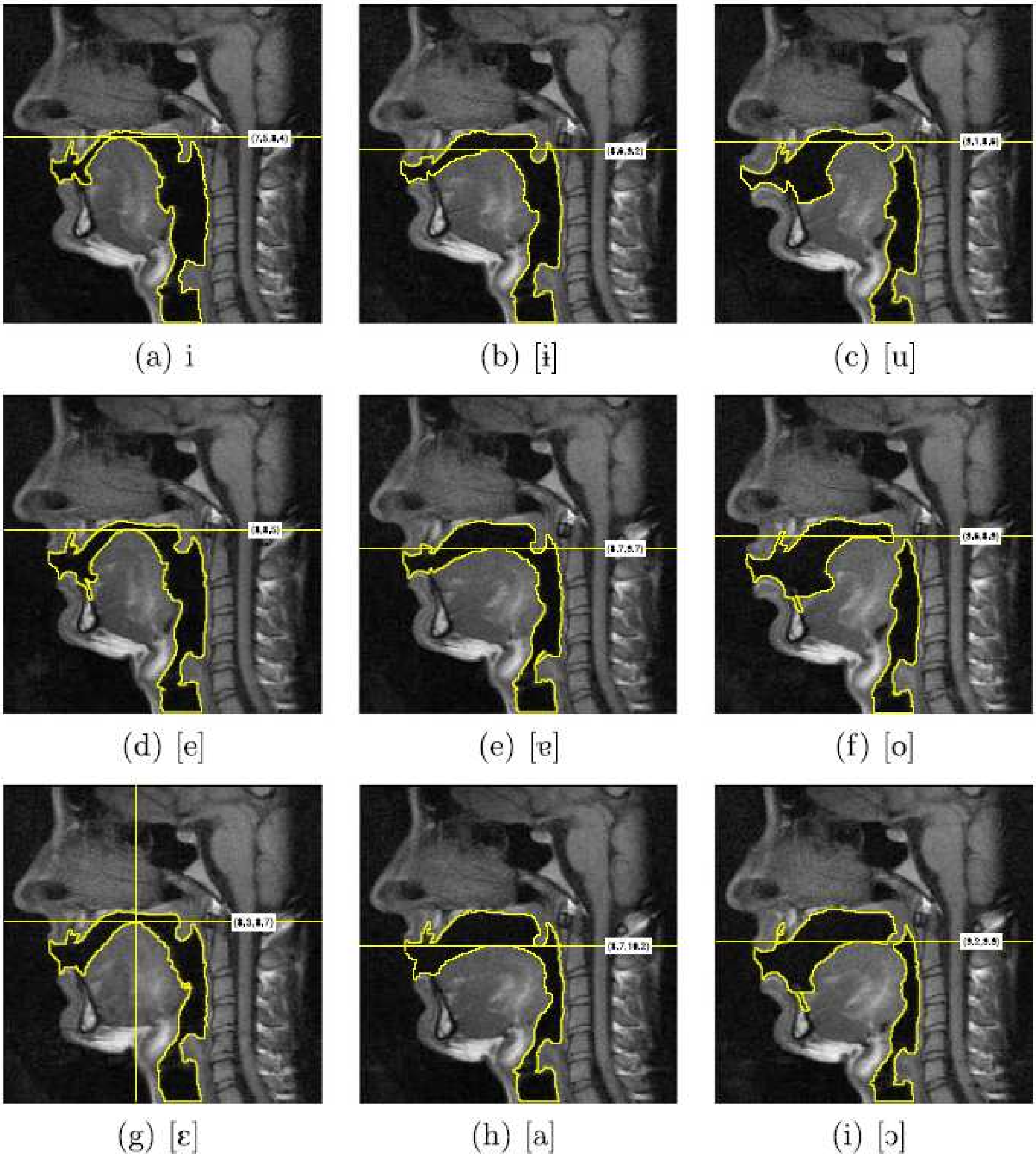


Figure 1. Midsagittal images of the EP oral vowels (Martins et al., 2008, p. 934).

Appendix D

Consent Form

Universidade do Minho
Centro de Estudos Humanísticos

AUTORIZAÇÃO

CONCORDO em participar da investigação sobre a aquisição do inglês por falantes nativos de português europeu e AUTORIZO a investigadora Anabela Rato a gravar frases lidas por mim para posterior análise fonética. Também estou ciente de que farei um teste de perceção da fala e que os dados daí resultantes serão analisados apenas para fins académicos, garantindo o meu anonimato.

E por ser expressão da verdade, firmo a presente autorização que vai datada e assinada para que surta os devidos efeitos legais.

Braga, Portugal, ____________________de 2012.

Nome do participante: ____

Assinatura do participante: ____

Appendix E

Background Questionnaire

QUESTIONÁRIO

Nome: ____

Curso que frequenta: ______________________________ Ano:________________

Telf.: ________________ E-mail: ____

1. Idade: _______________ 2. Naturalidade: _______________

3. Residência: ______________ 3.1. Há quantos anos reside nesta localidade?_______

4. Estuda inglês atualmente? Sim/Não

4.1. Se respondeu sim:

4.1.1. Qual é o seu nível de proficiência, segundo o QECRL[135]? __________________

4.1.2. Onde estuda inglês (universidade pública, instituto de línguas)? ____________

4.1.3. Quantas horas de aulas de inglês tem semanalmente? ___________________

4.1.4. Quantas horas por semana dedica ao estudo da língua inglesa? _______________

5. Estudou inglês durante o seu percurso escolar? Sim /Não

5.1. Se respondeu "Sim", durante quantos anos estudou inglês? __________________

5.2. Com que idade começou a estudar inglês? _________________________________

5.3. As aulas de inglês focavam a comunicação escrita ("writing") e oral ("speaking") e a compreensão escrita ("reading") e oral ("listening")?

____

6. Estudou inglês sem ser na escola (por exemplo, num instituto de línguas)? Sim/Não

6.1. Se respondeu "sim", durante quanto tempo? ______________________________

7. Interrompeu o estudo de inglês durante algum tempo? Sim/Não

7. 1. Se respondeu "sim", durante quanto tempo? ______________________________

8. Já viveu (mais de 1 mês) em algum país de língua inglesa? Sim/Não

8.1. Se respondeu "sim", em que país? ___________________

8.2. Durante quanto tempo? ____________________________

8.3. Quantos anos tinha na altura?________________________

8.4. Frequentou alguma escola naquele país? Sim/Não

9. Qual a variante de inglês que utiliza na oralidade (inglês britânico, americano, sul africano, australiano, canadiano, etc) ____________________________________

10. Conversa em inglês com outros falantes portugueses? Sim/Não

10.1. Com que frequência (semanalmente)? _________________________

[135] Quadro Europeu Comum de Referência para as Línguas (níveis A1 e A2: utilizador elementar; B1 e B2: utilizador independente; C1 e C2: utilizador proficiente).

11. Conversa com frequência em inglês com falantes nativos? Sim/Não

11.1. Com que frequência (semanalmente)? ___________________________

12. Conversa com frequência em inglês com falantes não nativos? Sim/Não

11.2. Com que frequência (semanalmente)? ___________________________

13. Assiste a filmes, séries, programas em inglês sem dobragem? Sim/Não

13.1. Com que frequência?

5/10 min/dia ___ 20/40min/dia ___ 1h/dia ___ 2h/dia____ +_____

14. Ouve música inglesa? Sim/Não

14.1 Com que frequência?

5/10 min/dia ___ 20/40min/dia ___ 1h/dia ___ 2h/dia____ +_____

15. Joga vídeo jogos em inglês? Sim/Não

15.1. Com que frequência?

5/10 min/dia ___ 20/40min/dia ___ 1h/dia ___ 2h/dia____ +_____

16. Lê em inglês (artigos, revistas, livros)? Sim/Não

16.1. Que género de literatura? _______________

16. 2. Com que frequência?

5/10 min/dia ___ 20/40min/dia ___ 1h/dia ___ 2h/dia____ +_____

17. Estuda, estudou ou tem contacto com outra língua estrangeira? Sim/Não

17. 1 Em que contexto? (escola, família…) ____________________________________

17.2. Que língua? ____

17.3. Qual é o seu nível de proficiência, segundo o QECRL? ______________________

18. Há alguém da sua família direta que fale outras línguas para além do Português? Sim/Não

18. 1. Se "sim", que línguas? ____

19. Acrescente alguma informação que julgue pertinente e que não tenha sido contemplada neste questionário.

____

____

Braga, ______________________________________

Appendix F

Individual Data of the Background Questionnaire

Table 1

Participants' Background Information

Part.	Sex	Age	Place of Birth	Residence	LOR	LP	AOL	LFI	Exp. Abroad	Int. EL	ENG var.
EXPERIMENTAL GROUP											
1	M	18	Guimarães	Guimarães	18	B1	6	12	n/a	N	AE
2	M	20	Ponte de Lima	Ponte de Lima	20	B1	10	7	n/a	3	n/a
3	M	42	Barcelos	Barcelos	42	B1	10	7	n/a	26	AE
4	F	19	Guimarães	Guimarães	19	B1	10	7	n/a	2	AE
5	F	18	Braga	Braga	18	B1	10	7	n/a	1	AE
6	M	22	Cab. Basto	Cab. Basto	22	B1	10	8	n/a	4	AE
7	F	20	Braga	Braga	20	B2	10	9	n/a	N	AE
8	M	32	Famalicão	Famalicão	32	B1	10	5	n/a	15	AE
9	F	18	Marco Canav.	Marco Canav.	18	B1	10	7	n/a	1	AE
10	M	20	Guimarães	Guimarães	20	B1	9	11	n/a	N	AE
11	M	20	Guimarães	Braga	0,5	B1	10	8	n/a	1	AE
12	M	30	Braga	Braga	30	B1	10	7	3, UK	12	AE
13	M	19	Barcelos	Barcelos	19	B1	8	8	n/a	1	AE
14	F	18	Famalicão	Famalicão	18	B1	10	7	n/a	1	AE
15	F	18	Santo Tirso	Braga	0,5	B1	10	8	n/a	N	Hybrid
16	F	21	Mirandela	Porto	4	B1	9	12	n/a	N	AE
17	F	26	Braga	Braga	26	B1	11	8	n/a	4	AE
18	M	36	Ermesinde	Ermesinde	36	B1	10	8	n/a	16	AE
19	M	18	Vila Verde	Vila Verde	18	B1	10	7	n/a	N	AE
20	F	19	Guimarães	Guimarães	19	B1	10	7	n/a	11	AE
21	M	18	Guimarães	Guimarães	18	B1	9	8	n/a	N	AE
22	F	19	Arcos Valdevez	Braga	0,5	B1	10	7	n/a	N	AE
CONTROL GROUP											
23	F	18	Braga	Braga	18	B1	8	10	n/a	N	AE
24	F	19	Santo Tirso	Santo Tirso	19	B1	9	8	n/a	N	Hybrid
25	F	26	Rio de Janeiro	Braga	24	C1	9	12	n/a	5	BE
26	M	31	Barcelos	Barcelos	31	B1	10	8	n/a	13	BE
27	F	19	Madeira	Braga	0,5	B1	10	7	n/a	2	BE
28	F	31	Esposende	Esposende	31	B2	11	7	n/a	13	AE
29	F	40	Braga	Braga	40	C1	10	8	n/a	5	Hybrid
30	F	18	Madeira	Guimarães	10	B1	10	7	n/a	1	AE
31	M	23	Famalicão	Famalicão	23	B1	10	11	n/a	N	AE
32	M	28	Braga	Braga	28	B1	11	7	n/a	10	AE
33	M	20	Barcelos	Braga	1,5	B1	10	9	n/a	1	AE
34	F	19	Guimarães	Guimarães	19	B1	10	9	n/a	N	AE

Note. Part=Participant; LOR= Length of residence (years); LP =Language proficiency (A1, A2, B1, B2, C1, C2 levels, CEFR)*;* AOL=Age of learning, i.e., age at which learning of English began*;* LFI=Length of formal instruction, i.e., number of years of L2 learning; Exp. Abroad = Experience abroad in English-speaking countries (months/country)*;* Int. EL=Interruption of English learning (years)*;* ENG var.=English variety spoken; Cab.=Cabeceiras; Canav.=Canaveses.

Table 2

Participants' Language Experience

Part.	6	6.1	10	10.1	11	11.1	12	12.1	13	13.1	14	14.1	15	15.1	16	16.1
EXPERIMENTAL GROUP																
1	N	n/a	Y	2x week	N	n/a	N	n/a	Y	20-40 min/day	Y	20-40 min/day	Y	5-10 min/day	Y	5-10 min/day
2	N	n/a	Y	1x week	N	n/a	Y	2x week	Y	1 h/day	Y	20-40 min/day	Y	20-40 min/day	Y	2 h/day
3	N	n/a	N	n/a	N	n/a	N	n/a	Y	20-40 min/day	Y	20-40 min/day	N	n/a	Y	5-10 min/day
4	N	n/a	Y	2x week	N	n/a	N	n/a	Y	1 h/day	Y	2 h/day	N	n/a	Y	20-40 min/day
5	N	n/a	N	n/a	N	n/a	N	n/a	Y	5-10 min/day	Y	1 h/day	N	n/a	N	n/a
6	N	n/a	N	n/a	Y	1 x week	Y	1x week	Y	2 h/day	Y	2 h/day	Y	5-10 min/day	Y	20-40 min/day
7	N	n/a	Y	daily	N	n/a	Y	daily	Y	1 h/day	Y	20-40 min/day	Y	1 h/day	Y	1 h/day
8	N	n/a	Y	daily	Y	daily	N	n/a	Y	20-40 min/day	Y	5-10 min/day	Y	5-10 min/day	N	n/a
9	N	n/a	Y	1x week	N	n/a	Y	1x week	Y	20-40 min/day	Y	20-40 min/day	N	n/a	N	n/a
10	N	n/a	Y	1x week	N	n/a	N	n/a	Y	20-40 min/day	Y	20-40 min/day	N	n/a	Y	+ 2 h/day
11	N	n/a	N	n/a	N	n/a	N	n/a	Y	+ 2 h/day	Y	2 h/day	N	n/a	Y	20-40 min/day
12	N	n/a	Y	1x week	N	n/a	N	n/a	Y	+ 2 h/day	Y	+ 2 h/day	Y	20-40 min/day	Y	5-10 min/day
13	Y	8	N	n/a	N	n/a	N	n/a	Y	20-40 min/day	Y	1 h/day	Y	20-40 min/day	N	n/a
14	N	n/a	N	n/a	N	n/a	N	n/a	Y	1 h/day	Y	5-10 min/day	Y	5-10 min/day	Y	5-10 min/day
15	Y	3	Y	1x week	Y	1 x week	Y	1x week	Y	1 h/day	Y	+ 2 h/day	N	n/a	Y	20-40 min/day
16	Y	5	Y	daily	N	n/a	N	n/a	Y	1 h/day	Y	2 h/day	Y	1 h/day	Y	1 h/day
17	Y	2	Y	1x week	N	n/a	Y	1x week	Y	5-10 min/day	Y	1 h/day	N	n/a	Y	20-40 min/day
18	N	n/a	N	n/a	N	n/a	Y	3x week	Y	20-40 min/day	Y	2 h/day	Y	1 h/day	N	n/a
19	N	n/a	N	n/a	N	n/a	N	n/a	Y	2 h/day	Y	20-40 min/day	Y	20-40 min/day	N	n/a
20	N	n/a	Y	2x week	N	n/a	N	n/a	Y	20-40 min/day	Y	1 h/day	N	n/a	N	n/a
21	N	n/a	N	n/a	N	n/a	N	n/a	Y	2 h/day	N	n/a	Y	1 h/day	N	n/a
22	N	n/a	Y	3x week	N	n/a	Y	1x week	Y	1 h/day	Y	2 h/day	N	n/a	Y	1 h/day
CONTROL GROUP																
23	N	n/a	Y	2x week	N	n/a	N	n/a	Y	2 h/day	Y	+ 2 h/day	N	n/a	Y	+ 2 h/day
24	N	n/a	N	n/a	N	n/a	N	n/a	Y	+ 2 h/day	Y	2 h/day	N	n/a	N	n/a
25	Y	8	Y	1x week	N	n/a	Y	1x week	Y	20-40 min/day	Y	1 h/day	N	n/a	Y	5-10 min/day
26	N	n/a	N	n/a	N	n/a	N	n/a	Y	1 h/day	Y	1 h/day	N	n/a	Y	20-40 min/day
27	N	n/a	Y	2x week	Y	1 x week	Y	1x week	Y	2 h/day	Y	5-10 min/day	Y	5-10 min/day	Y	20-40 min/day
28	N	n/a	Y	2x week	N	n/a	Y	1x week	Y	5-10 min/day	Y	2 h/day	N	n/a	Y	20-40 min/day
29	N	n/a	N	n/a	N	n/a	N	n/a	Y	5-10 min/day	Y	5-10 min/day	N	n/a	Y	5-10 min/day
30	N	n/a	Y	2x week	N	n/a	Y	2x week	Y	2 h/day	Y	+ 2 h/day	Y	20-40 min/day	Y	1 h/day
31	N	n/a	Y	2x week	N	n/a	Y	daily	Y	2 h/day	Y	2 h/day	Y	20-40 min/day	Y	20-40 min/day
32	N	n/a	N	n/a	N	n/a	Y	1x week	Y	2 h/day	Y	1 h/day	Y	20-40 min/day	Y	5-10 min/day
33	N	n/a	Y	2x week	Y	1 x week	Y	1x week	Y	20-40 min/day	Y	2 h/day	Y	20-40 min/day	Y	5-10 min/day
34	N	n/a	Y	1x week	N	n/a	Y	1x week	Y	2 h/day	Y	2 h/day	Y	1 h/day	Y	5-10 min/day

Notes.
Part = Participant
n/a=not applicable
6. Did you study English at a language private school? Y (yes) / N (no)
6.1 How long did you study English there? (years)
10. Do you speak English with Portuguese speakers? Y (yes) / N (no)
10. 1.How frequently (per week)? (daily; 3 times; twice; once; not applicable)
11. Do you speak English with native English speakers? Y (yes) / N (no)
11. 1. How frequently (per week)? (daily; 3 times; twice; once; not applicable)
12. Do you speak English with other non-native speakers? Y (yes) / N (no)
12. 1 - How frequently (per week)? (daily; 3 times; twice; once; not applicable)
13. Do you watch English TV programs or movies? Y (yes) / N (no)
13.1. How frequently (per day)? (5-10 min; 20-40 min; 1h; 2h; + 2h)
14. Do you listen to English music? Y (yes) / N (no)
14.1. How frequently (per day)? (5-10 min; 20-40 min; 1h; 2h; + 2h)
15. Do you play English video games? Y (yes) / N (no)
15.1. How frequently (per day)? (5-10 min; 20-40 min; 1h; 2h; + 2h)
16. Do you read English literature? Y (yes) / N (no)
16. 2. How frequently (per day)? (5-10 min; 20-40 min; 1h; 2h; + 2h)

Appendix G

Summary of the Statistical Descriptive Data of the Background Questionnaire

	Mean (years)	*SD*	Range (years)
Age	23.03	6.76	18 - 42
AOL	9.71	0.94	6 - 11
LFI	8.18	1.70	5 - 12
	Mean (hours)	***SD***	**Range (hours)**
ENG lessons per week	3.88	0.69	0-4
ENG study per week	2.75	2.28	0-8
		Values (%)	
	F	**M**	***Mo***
Sex	52.9	47.1	2
		Values (%)	
	Yes	**No**	***Mo***
EL in other formal contexts	14.7	85.3	2
EL interruption	64.7	35.3	1
Passive ENG Use			
Speak English with EP speakers	61.8	38.2	1
Speak with native English speakers	14.7	85.3	2
Speak English with non-native speakers	47.1	52.9	2
Active ENG Use			
Watch TV	100	0	1
Listen to music	97.1	2.9	1
Play video games	52.9	47.1	1
Read	73.5	26.5	1
	Mdn	**Cumulative Percent (%)**	**Label**
English Proficiency	1	88.2	*B1*
Passive ENG Use			
Speak English with EP speakers	4	61.8	*once a week*
Speak with native English speakers	5	100	*n/a*
Speak English with non-native speakers	5	100	*n/a*
Active ENG Use			
Watch TV	3	64.7	*1h/day*
Listen to music	3	52.9	*1h/day*
Play video games	3	52.9	*1h/day*
Read	2	52.9	*20-40m/day*
ENG variety	1	76.5	*AE*

Note. ENG=English; EL=English learning; AOL=age of learning; LFI=length of formal instruction; *SD*=standard deviation; F=female participants; M=male participants; *Mo* =Mode; *Mdn* =Median; n/a=not applicable.

Appendix H

Follow-up Questionnaire

Questionário

Nome: ____

1. Qual foi o par de vogais mais **fácil** de distinguir durante o treino?
   a) [i] [ɪ] *heed-hid*   b) [æ] [ɛ] *had-head*   c) [u] [ʊ] *who'd-hood*

2. Qual foi o par de vogais mais **difícil** de distinguir durante o treino?
   a) [i] [ɪ] *heed-hid*   b) [æ] [ɛ] *had-head*   c) [u] [ʊ] *who'd-hood*

3. Qual foi a vogal mais **fácil** de identificar durante o treino?
   a) [i] *heed*   b) [ɪ] *hid*   c) [æ] *had*   d) [ɛ] *head*   e) [u] *who'd*
   f) [ʊ] *hood*   g) [ʌ] *hud*

4. Qual foi a vogal mais **difícil** de identificar durante o treino?
   a) [i] *heed*   b) [ɪ] *hid*   c) [æ] *had*   d) [ɛ] *head*   e)[u] *who'd*
   f) [ʊ] *hood*   g) [ʌ] *hud*

5. Qual foi o tipo de exercícios mais **fácil** de fazer durante o treino:
   a) exercícios de discriminação   b) exercícios de identificação

6. Considera que a **duração** das sessões de treino foi, em média:
   a) a) longa   b) adequada   c) curta

7. Considera que, **para aprender a distinguir** as sete vogais acima referidas, as 5 sessões de treino foram:
   a) insuficientes 2) suficientes 3) mais do que suficientes

8. No fim das sessões de treino, sentia-se **cansada(o)?**
   a) Sim   b) Não

8.1. Se assinalou, sim, indique o grau de **cansaço**:
   a) Muito   b) Razoável   c) Pouco

9. Numa escala de 1 a 4, indique o seu nível de **motivação** ao longo do programa de treino:
   1 - Desmotivado   2 - Pouco motivado   3 - Motivado   4 - Muito motivado

10. Numa escala de 1 a 4, indique o seu nível de **concentração** ao longo do programa de treino:
   1 - Desconcentrado   2 - Pouco concentrado 3 – Concentrado   4 - Muito concentrado

11. Acrescente **comentários** que julgue pertinentes sobre: 1) o *software* (*TP* - *S*) que utilizou para realizar os exercícios de treino e de teste e 2) sobre as sessões de treino.

Appendix I

Phonetic Transcriptions

Table 1

Phonetic Transcription of the EP Words Read by the L2 Participants

Picture	Vowel	pVpV	pVtV	pVkV	tVpV	tVtV	tVkV
bico (beak)	**[i]**	[ˈpipu]	[ˈpitu]	[ˈpiku]	[ˈtipu]	[ˈtitu]	[ˈtiku]
dedo (finger)	**[e]**	[ˈpepu]	[ˈpetu]	[ˈpeku]	[ˈtepu]	[ˈtetu]	[ˈteku]
lego	**[ɛ]**	[ˈpɛpu]	[ˈpɛtu]	[ˈpɛku]	[ˈtɛpu]	[ˈtɛtu]	[ˈtɛku]
pato (duck)	**[a]**	[ˈpapu]	[ˈpatu]	[ˈpaku]	[ˈtapu]	[ˈtatu]	[ˈtaku]
copo (glass)	**[ɔ]**	[ˈpɔpu]	[ˈpɔtu]	[ˈpɔku]	[ˈtɔpu]	[ˈtɔtu]	[ˈtɔku]
côco (coconut)	**[o]**	[ˈpopu]	[ˈpotu]	[ˈpoku]	[ˈtopu]	[ˈtotu]	[ˈtoku]
cubo (cube)	**[u]**	[ˈpupu]	[ˈputu]	[ˈpuku]	[ˈtupu]	[ˈtutu]	[ˈtuku]

Table 2

Phonetic Transcription of the AE Words Read by the L2 Participants

picture	vowel	pVt	pVk	tVt	tVk	kVp	kVt	bvt	dVk	fVt	sVt
feet	**[i]**	[pit]	[pik]	[tit]	[tik]	[kip]	[kit]	[bit]	-	[fit]	[sit]
pig	**[ɪ]**	[pɪt]	[pɪk]	[tɪt]	[tɪk]	[kɪp]	[kɪt]	[bɪt]	-	[fɪt]	[sɪt]
bed	**[ɛ]**	[pɛt]	[pɛk]	[tɛt]	[tɛk]	[kɛpt]	[kɛt]	[bɛt]	[dɛk]	-	[sɛt]
cat	**[æ]**	[pæt]	[pæk]	[tæt]	[tæk]	[kæp]	[kæt]	[bæt]	-	[fæt]	[sæt]
boot	**[u]**	[pup]	-	[tut]	[tuk]	[kup]	[kut]	[but]	[duk]	-	[sut]
book	**[ʊ]**	[pʊt]	-	-	[tʊk]	-	[kʊt]	[bʊt]	-	[fut]	[sʊt]
cup	**[ʌ]**	[pʌt]	[pʌk]	[tʌt]	[tʌk]	[kʌp]	[kʌt]	[bʌt]	-	-	[ʃʌt]

Table 3

Phonetic Transcription of the Identification Test Stimuli

Vowels	bVt	tVk	tVt	kVt	pVt
[æ]	[bæt]	[tæk]	[tæt]	[kæt]	[pæt]
[ɛ]	[bɛt]	[tɛk]	[tɛt]	[kɛt]	[pɛt]
[ɪ]	[bɪt]	[tɪk]	[tɪt]	[kɪt]	[pɪt]
[i]	[bit]	[tik]	[tit]	[kip]	[pit]
[ʊ]	[bʊk]	[tʊk]	-	[kʊk]	[pʊt]
[u]	[but]	[tuk]	[tut]	[kut]	[pup]
[ʌ]	[bʌt]	[tʌk]	[tʌk]	[kʌt]	[pʌt]

Table 4

Phonetic Transcription of the Generalization Test Stimuli

	Monosyllabic Words		
Vowel	CVC	CCV(C)(C)	CVCC
[i]	[fil],[lik],[sid]	[blid],[ski]	[fist]
[ɪ]	[fɪl],[lɪk],[mɪt]	[slɪpt]	[fɪst],[rɪst]
[ɛ]	[sɛd],[mɛt],[rɛk]	[plɛd],[blɛd],[slɛpt]	-
[æ]	[pæd], [bæk], [sæd], [ræk], [mæt]	[plæd]	-
[u]	[kud],[ful],[luk]	[blu],[flu],[stud]	-
[ʊ]	[pʊʃ],[kʊd],[fʊl],[lʊk]	[stʊd]	[wʊlf]
[ʌ]	[kʌd],[bʌg],[lʌk],[rʌf]	[blʌd],[flʌd]	-

Table 5

Phonetic Transcription of the Stimuli used in the Training of the Experimental Group

Vowels	**bVt**	**tVk**	**sVt**	**hVd**
[æ]	[bæt]	[tæk]	[sæt]	[hæd]
[ɛ]	[bɛt]	[tɛk]	[sɛt]	[hɛd]
[ɪ]	[bɪt]	[tɪk]	[sɪt]	[hɪt]
[i]	[bit]	[tik]	[sit]	[hid]
[ʊ]	[bʊk]	[tʊk]	[sʊt]	[hʊd]
[u]	[but]	[tuk]	[sut]	[hut]
[ʌ]	[bʌt]	[tʌk]	[ʃʌt]	[hʌt]

Appendix J

EP Sentences Read by the L2 Participants

[i]
Digo pipo novamente.
Digo pito novamente.
Digo pico novamente.
Digo tipo novamente.
Digo tito novamente.
Digo tico novamente.

[e]
Digo pêpo novamente.
Digo pêto novamente.
Digo pêco novamente.
Digo têpo novamente.
Digo têto novamente.
Digo têco novamente.

[ɛ]
Digo pépo novamente.
Digo péto novamente.
Digo péco novamente.
Digo tépo novamente.
Digo této novamente.
Digo téco novamente.

[a]
Digo papo novamente.
Digo pato novamente.
Digo paco novamente.
Digo tapo novamente.
Digo tato novamente.
Digo taco novamente.

[ɔ]
Digo pópo novamente.
Digo póto novamente.
Digo póco novamente.
Digo tópo novamente.
Digo tóto novamente.
Digo tóco novamente.

[o]
Digo pôpo novamente.
Digo pôto novamente.
Digo pôco novamente.
Digo tôpo novamente.
Digo tôto novamente.
Digo tôco novamente.

[u]
Digo pupo novamente.
Digo puto novamente.
Digo puco novamente.
Digo tupo novamente.
Digo tuto novamente.
Digo tuco novamente.

Appendix K

AE Sentences Read by the L2 Participants

[i]

Say Pete now.
Say peak now.
Say teat now.
Say teak now.
Say keat now.
Say keep now.
Say beat now.
Say feet now.
Say seat now.

[ɪ]

Say pit now.
Say pick now.
Say tit now.
Say tick now.
Say kit now.
Say kip now.
Say bit now.
Say fit now.
Say sit now.

[ɛ]

Say pet now.
Say peck now.
Say tet now.
Say tech now.
Say ket now.
Say kept now.
Say bet now.
Say set now.
Say deck now.

[æ]

Say pat now.
Say pack now.
Say tat now.
Say tack now.
Say cat now.
Say cap now.
Say bat now.
Say fat now.
Say sat now.

[u]

Say toot now.
Say tuke now.
Say coot now.
Say coop now.
Say boot now.
Say duke now.
Say suit now.
Say boot now.
Say poop now.

[ʊ]

Say put now.
Say took now.
Say cook now.
Say book now.
Say foot now.
Say soot now.
Say book now.
Say foot now.
Say soot now.

[ʌ]

Say putt now.
Say puck now.
Say tut now.
Say tuck now.
Say cut now.
Say cup now.
Say but now.
Say shut now.
Say cup now.

Appendix L

CELEX Word Frequency (*N-Watch*)

ITEM	CELEX	CELEX_W	CELEX_S
back	1233.02	1251.45	997.69
bat	10.56	10.9	6.15
beat	53.24	55.36	26.15
bit	240.67	199.88	761.54
bled	3.97	4.22	0.77
bleed	3.69	3.92	0.77
blood	141.68	150.18	33.08
blue	128.72	135.66	40
book	275.47	255.48	530.77
boot	9.72	10.24	3.08
bug	3.24	3.19	3.85
but	5412.79	5306.08	6775.38
cap	30.34	32.11	7.69
cat	41.28	43.86	8.46
cooed	0.78	0.84	0
cook	49.5	50.66	34.62
coop[b]	0	0	0
coot	0.45	0.48	0
could	1880.5	1930.36	1243.85
cud	1.01	1.08	0
cup	60.84	63.07	32.31
cut	177.88	183.86	101.54
deck	19.16	20.6	0.77
duke	38.21	39.46	22.31
fat	86.03	91.75	13.08
feast	12.18	13.01	1.54
feel	371.45	361.93	493.08
feet	229.27	240.66	83.85
fill	41.56	41.75	39.23
fist	17.77	18.92	3.08
fit	69.94	71.69	47.69
flood	15.87	16.69	5.38
flu	4.36	4.4	3.85
fool	36.09	38.01	11.54
foot	116.59	119.04	85.38
full	274.58	281.87	181.54
had	6255.03	6572.83	2196.92
head	456.48	482.47	124.62
heed	2.79	3.01	0

hit	91.34	94.4	52.31
hood	7.43	7.59	5.38
hoot	1.06	1.14	0
hut	22.57	24.1	3.08
keat	0	0	0
keep	349.89	354.52	290.77
kept	209.5	219.16	86.15
ket[a]	0	0	0
kip	0.45	0.48	0
kit	8.49	8.67	6.15
leak	7.82	8.07	4.62
lick	2.91	3.13	0
look	591.4	565	928.46
luck	45.47	46.57	31.54
luke	0	0	0
mat	7.54	8.01	1.54
met	148.16	152.17	96.92
mitt	0	0	0
pack	26.82	27.71	15.38
pad	11.79	12.47	3.08
pat	19.05	19.4	14.62
peak	23.07	22.77	26.92
peak	23.07	22.77	26.92
peck	3.69	3.31	8.46
pet	13.63	14.4	3.85
pete[a]	0	0	0
pick	67.65	66.08	87.69
pit	13.24	13.61	8.46
plaid	2.74	2.95	0
pled	0.34	0.36	0
poop	0.39	0.42	0
puck	1.01	1.08	0
push	43.91	45.06	29.23
put	687.26	667.59	938.46
putt	3.35	3.61	0
rack	8.32	8.8	2.31
rough	43.18	44.94	20.77
sad	46.2	45.6	53.85
said	2774.02	2874.34	1493.08
sat	228.04	243.25	33.85
seat	77.99	81.69	30.77
seed	27.82	28.98	13.08
set	373.91	381.99	270.77
shut	64.97	67.65	30.77
sit	119.94	119.94	120
ski	6.26	5	22.31

slept	33.69	35.6	9.23
slipped	29.83	31.69	6.15
soot	1.9	2.05	0
stewed	2.35	2.53	0
stood	212.74	227.65	22.31
suit	61.62	64.28	27.69
tack	2.68	2.59	3.85
tat	1.06	1.08	0.77
teak	1.12	1.2	0
teak	1.12	1.2	0
teat	4.58	4.94	0
teat	4.58	4.94	0
tech	3.91	4.1	1.54
tet[a]	0	0	0
tick	3.52	3.43	4.62
tit	1.01	1.08	0
took	475.25	494.46	230
toot	0.45	0.48	0
tuck	3.74	3.55	6.15
tuke[b]	0	0	0
tut	0.84	0.36	6.92
wolf	6.7	6.93	3.85
wreck	8.49	8.98	2.31
wrist	20	21.57	0

Note. CELEX: = Total CELEX Word frequency: nº of occurrences in COBUILD/ECT corpus divided by 17.9; CELEX_W = CELEX Written Word Frequency; nº of occurrences in COBUILD written corpus divided by 16.6.; CELEX_S CELEX = Spoken Word Frequency; nº of occurrences in COBUILD spoken corpus divided by 1.3.; CELEX frequency (per million).
[a] Proper nouns. [b] Pseudo words.

Appendix M

Computer Screens of the Identification Tests

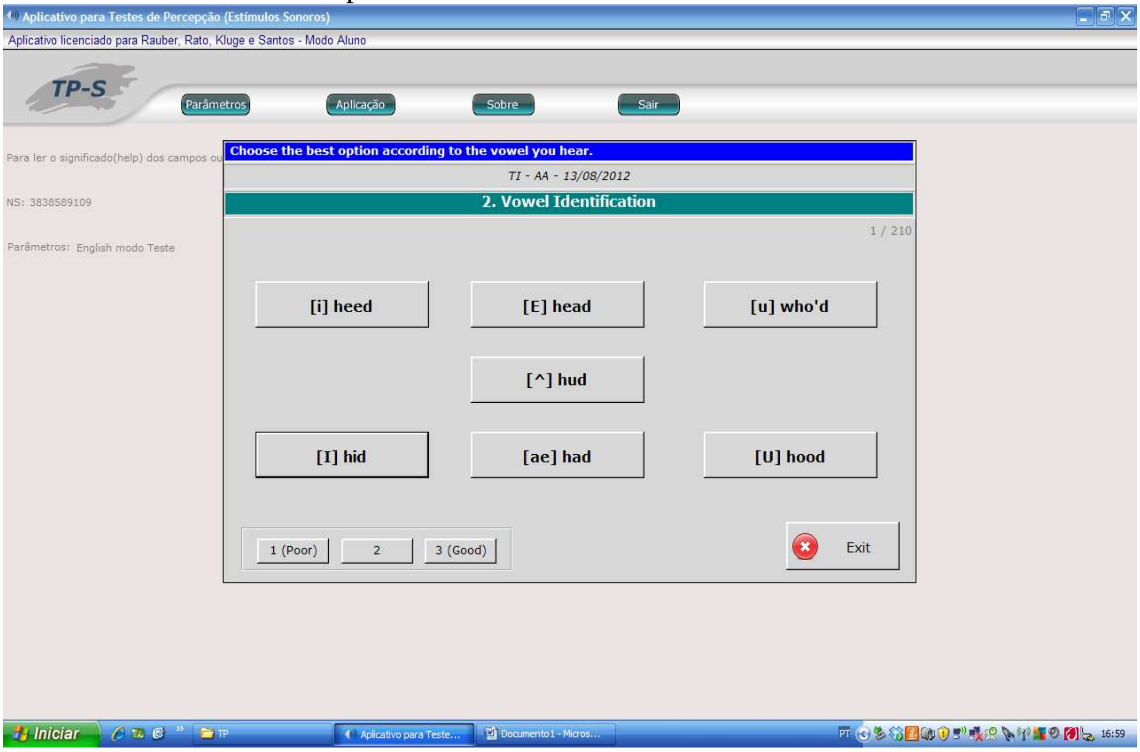


Figure 1. Computer screen of the 7AFC identification test (pretest, posttest, delayed posttest).

This screen includes (1) a short instruction (blue tab); (2) the indication of the type of experiment being run (TI – Identification), followed by the name of the informant and the date (grey tab); (3) the designation of the test (green tab); (4) the number of the stimulus the participant is listening to, followed by the total number of the experiment's stimuli (top right corner); (5) the seven response buttons; (6) a three-point Likert scale to rate category goodness-of-fit; (7) the exit button.

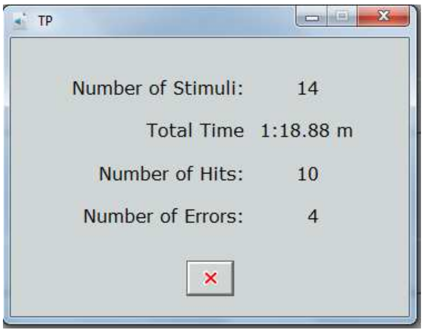


Figure 2. Final message displayed after completion of the identification test.

When the test was completed, a window with information about the total number of stimuli presented, total time spent and total number of correct and incorrect answers was displayed.

Appendix N

Stimuli Organization of the Oddity Discriminations Tasks

Table 1

Oddity Discrimination Task 1

seatM6	**sitF1**	**sitF5**	
sitF1	seatF5	**sitM6**	**Change trials**
sitF5	**sitF1**	seatM6	
seatF1	seatM6	seatF5	Catch trials
sitM6	sitF1	sitF5	
sitF1	**seatF5**	**seatM6**	**Change trials**
seatM6	sitF5	**seatF1**	
seatF5	**seatF1**	sitM6	
satF1	**set M6**	**setF5**	**Change trials**
setM6	satF5	**setF1**	
setF5	**setF1**	satM6	
satF1	satM6	satF5	Catch trials
setM6	setF1	setF5	
setF1	**satF5**	**satM6**	**Change trials**
satM6	setF5	**satF1**	
satF5	**satF1**	setM6	
sootF1	**suit M6**	**suitF5**	**Change trials**
suitM6	sootF5	**suitF1**	
suitF5	**suitF1**	sootM6	
sootF1	sootM6	sootF5	Catch trials
suitM6	suitF1	suitF5	
suitF1	**sootF5**	**sootM6**	**Change trials**
sootM6	suitF5	**sootF1**	
sootF5	**sootF1**	suitM6	
batF1	**but M6**	**butF5**	**Change trials**
butM6	batF5	**butF1**	
butF5	**butF1**	batM6	
batF1	batM6	batF5	Catch trials
butM6	butF1	butF5	
butF1	**batF5**	**batM6**	**Change trials**
batM6	butF5	**batF1**	
batF5	**batF1**	butM6	

Table 2

Oddity Discrimination Task 2

teakM6	**tickF1**	**tickF5**	
tickF1	teakF5	**tickM6**	**Change trials**
tickF5	**tickF1**	teakM6	
teakF1	teakM6	teakF5	Catch trials
tickM6	tickF1	tickF5	
tickF1	**teakF5**	**teakM6**	**Change trials**
teakM6	tickF5	**teakF1**	
teakF5	**teakF1**	tickM6	
tackF1	**tech M6**	**techF5**	**Change trials**
techM6	**tackF5**	**techF1**	
techF5	**techF1**	tackM6	
tackF1	tackM6	tackF5	Catch trials
techM6	techF1	techF5	
techF1	**tackF5**	**tackM6**	**Change trials**
tackM6	techF5	**tackF1**	
tackF5	**tackF1**	techM6	
tookF1	**tuke M6**	**tukeF5**	**Change trials**
tukeM6	tookF5	**tukeF1**	
tukeF5	**tukeF1**	tookM6	
tookF1	tookM6	tookF5	Catch trials
tukeM6	tukeF1	tukeF5	
tukeF1	**tookF5**	**tookM6**	**Change trials**
tookM6	tukeF5	**tookF1**	
tookF5	**tookF1**	tukeM6	
tackF1	**tuck M6**	**tuckF5**	**Change trials**
tuckM6	tackF5	**tuckF1**	
tuckF5	**tuckF1**	tackM6	
tackF1	tackM6	tackF5	Catch trials
tuckM6	tuckF1	tuckF5	
tuckF1	**tackF5**	**tackM6**	**Change trials**
tackM6	tuckF5	**tackF1**	
tackF5	**tackF1**	tuckM6	

Appendix O

Stimuli used in the Perceptual Training Tasks of the Control Group

Table 1

Tokens Used in the First Block of the Perceptual Training

	Minimal Pairs		**Minimal Pairs**	
	[d]	[ð]	[s]	[θ]
	breed	breathe	bass	bath
	dare	there	force	forth
	den	then	mass	math
	dough	though	miss	myth
	fodder	father	mouse	mouth
	ladder	lather	pass	path
	letter	leather	sank	thank
	mutter	mother	seem	theme
	riding	writhing	sin	thin
	seed	seethe	sing	thing
	skating	scathing	sink	think
	sued	soothe	sinker	thinker
	udder	other	sought	thought
	wetter	weather	sum	thumb
	ride	writhe	truce	truth
	Other Words		**Other Words**	
	day	clothe	scene	health
	dime	bathe	seat	teeth
	do	these	seed	three
	added	they	sew	with
Total nº	**19**	**19**	**19**	**19**

Table 2

Tokens Used in the Second Block of the Perceptual Training

	Minimal Pairs		
	[m]	[n]	[ŋ]
		fan	fang
		lawn	long
		sin	sing
		stun	stung
		thin	thing
	dime	dine	
	hem	hen	
	lame	lane	
	mere	near	
	might	night	
	them	then	
	rum	run	rung
	sam	san	sang
	simmer	sinner	singer
	sum	sun	sung
	whim	win	wing
	rim		ring
	swim		swing
	swimmer		swinger
	Other Words		
	comb	know	strong
	climb	sane	tongue
	camp		going
	from		hanger
			singing
Total nº	**18**	**18**	**18**

Table 3

Organization of the Stimuli in the AX Discrimination Tasks of the First Block of Training

[d]	[ð]	[s]	[θ]
breed	**breathe**	bass	**bath**
breathe	breed	**bath**	bass
breed	breed	bass	bass
breathe	**breathe**	**bath**	**bath**
dare	**there**	mass	**math**
there	dare	**math**	mass
dare	dare	mass	mass
there	**there**	**math**	**math**
den	**then**	sank	**thank**
then	den	**thank**	sank
den	den	sank	sank
then	**then**	**thank**	**thank**
dough	**though**	sin	**thin**
though	dough	**thin**	sin
dough	dough	sin	sin
though	**though**	**thin**	**thin**
sued	**soothe**	sink	**think**
soothe	sued	**think**	sink
sued	sued	sink	sink
soothe	**soothe**	**think**	**think**
ride	**writhe**	sought	**thought**
writhe	ride	**thought**	sought
ride	ride	sought	sought
writhe	**writhe**	**thought**	**thought**
seed	**seethe**	sum	**thumb**
seethe	**seed**	**thumb**	sum
seed	seed	sum	sum
seethe	**seethe**	**thumb**	**thumb**
udder	**other**	pass	**path**
other	udder	**path**	pass
udder	udder	pass	pass
other	**other**	**path**	**path**
riding	**writhing**	miss	**myth**
writhing	riding	**myth**	miss
riding	riding	miss	miss
writhing	**writhing**	**myth**	**myth**
ladder	**lather**	seem	**theme**
lather	ladder	**theme**	seem
ladder	ladder	seem	seem
lather	**lather**	**theme**	**theme**
fodder	**father**	mouse	**mouth**
father	fodder	**mouth**	mouse
fodder	fodder	mouse	mouse
father	**father**	**mouth**	**mouth**

Table 4

Organization of the Stimuli in the AX Discrimination Tasks of the Second Block of Training

[n]	[ŋ]	[m]	[n]	[m]	[ŋ]
fan	**fang**	dime	**dine**	rim	**ring**
fang	fan	**dine**	dime	**ring**	rim
fang	**fang**	dime	dime	rim	rim
fan	fan	**dine**	**dine**	**ring**	**ring**
lawn	**long**	them	**then**	swimmer	**swinger**
long	lawn	**then**	them	**swinger**	swimmer
lawn	lawn	them	them	swimmer	swimmer
long	**long**	**then**	**then**	**swinger**	**swinger**
sin	**sing**	mere	**near**	swim	**swing**
sing	sin	**near**	mere	**swing**	swim
sin	sin	mere	mere	swim	swim
sing	**sing**	**near**	**near**	**swing**	**swing**
stun	**stung**	might	**night**	whim	**wing**
stung	stun	**night**	might	**wing**	whim
stun	stun	might	might	whim	whim
stung	**stung**	**night**	**night**	**wing**	**wing**
thin	**thing**	lame	**lane**	rum	**rung**
thing	thin	**lane**	lame	**rung**	rum
thin	thin	lame	lame	rum	rum
thing	**thing**	**lane**	**lane**	**rung**	**rung**
run	**rung**	run	**rum**	sam	**sang**
rung	run	**rum**	run	**sang**	sam
run	run	run	run	sam	sam
rung	**rung**	**rum**	**rum**	**sang**	**sang**
san	**sang**	san	**sam**	simmer	**singer**
sang	san	**sam**	san	**singer**	simmer
san	san	san	san	simmer	simmer
sang	**sang**	**sam**	**sam**	**singer**	**singer**
sinner	**singer**	simmer	**sinner**	sum	**sung**
singer	sinner	**sinner**	simmer	**sung**	sum
sinne**r**	sinner	simmer	simmer	sum	sum
singer	**singer**	**sinner**	**sinner**	**sung**	**sung**
sun	**sung**	sum	**sun**		
sung	sun	**sun**	sum		
sun	sun	sum	sum		
sung	**sung**	**sun**	**sun**		
win	**wing**	win	**whim**		
wing	win	**whim**	win		
win	win	win	win		
wing	**wing**	**whim**	**whim**		

Appendix P

Timeline of the Research Study

Second semester 2011/2012: 27th February – 22nd June 2012

Week	Date	Experiment Procedures	EG	CG
1st	02/28	**Background Questionnaire**	*n*=22	*n*=12
2nd-3rd	02/28 – 03/13	**Pretest: Production**	*n*=22	*n*=12
3rd	03/13	**Pretest: Perception**	n=22	n=12
		7 AFC Identification Test		
4th	03/20	**Training session 1**: Front vowels /i/-/ɪ/	n=22	-
		2 AFC ID task		
		AX discrimination task		
		Training session 1: Voiceless fricatives /θ/-/s/	-	n=12
		2 AFC ID task		
		AX DI task		
5th	03/27	**Training session 2:** Back vowels /u/-/ʊ/	n=22	-
		2 AFC ID task		
		AX DI task		
		Training session 2: Voiced fricative and stop /ð/-/d/	-	n=12
		2 AFC ID task		
		AX DI task		
		Dental fricatives /θ/-/ð/		
		2 AFC ID task		
6th		INTERRUPTION – Easter holidays		
7th	04/10	**Training session 3**: Mid and low front vowels /ɛ/-/æ/	n=22	-
		2 AFC ID task		
		AX DI task		
		Low vowels /æ/-/ʌ/		
		2 AFC ID task		
		AX DI task		

		Training session 3: Nasals /m/-/n/ 2 AFC ID task AX DI task	-	n=12
8th	04/17	**Training session 4:** /i/, /ɪ/, /ɛ/, /æ/, /u/, /ʊ/, /ʌ/ 7 AFC ID task Oddity DI task	n=22	-
		Training session 4: Nasals /n/-/ŋ/ 2 AFC ID task AX DI task	-	n=12
9th	04/24	**Training session 5:** /i/, /ɪ/, /ɛ/, /æ/, /u/, /ʊ/, /ʌ/ 7 AFC ID task Oddity DI task **Follow-up Questionnaire**	n=22	-
		Training session 5: Nasals /m/, /n/, /ŋ/ 3 AFC ID task	-	n=12
9th	04/24-27	**Posttest: Production**	n=22	n=12
9th	04/26	**Posttest: Perception** 7 AFC ID Test 2 Generalization ID Tests	n=22	n=12
17th	06/18-22	**Delayed Posttest: Production**	n=19	n=11
17th	06/22	**Delayed Posttest: Perception** 7 AFC ID Test	n=19	n=11

Note. AFC=alternative-forced-choice; ID=identification; DI=discrimination.

Appendix Q

Recording Booth and Equipment

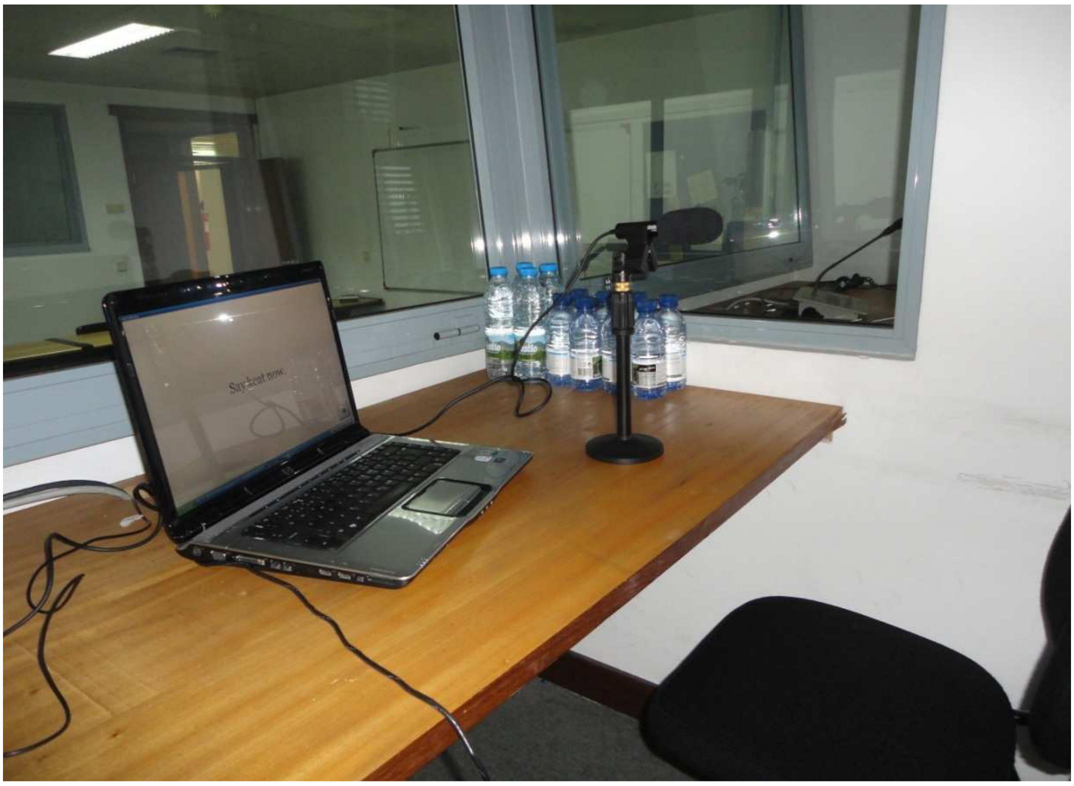


*Figure 1***.** Photographs of the conference room booth and the recording equipment.

Appendix R

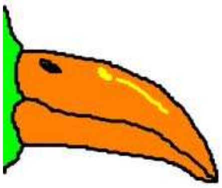
Intensity Normalization Script

```
# Written by Paul Boersma
# Modified by Andreia Rauber

Create Strings as file list... list *.wav
n = Get number of strings

for i to n
   select Strings list
   file$ = Get string... 'i'
   Read from file... 'file$'
   obj$ = selected$("Sound")
   Scale peak... 0.99
   Write to WAV file... 'obj$'.wav
endfor

#select all
#Remove
```

Appendix S

Pictures used in the EP Production Test

bico [i]

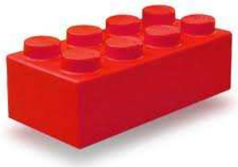


copo [ɔ]

dedo [e]

côco [o]

lego [ɛ]



cubo [u]

pato [a]

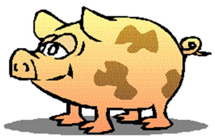
Appendix T

Pictures used in the AE Production Test

feet [i]

boot [u]

pig [ɪ]



book [ʊ]

bed [ɛ]

cup [ʌ]

cat [æ]

Appendix U

Instructions for the Identification Test

PERCEPTION TEST

Instructions for the Vowel Identification Experiment (Trial and Test)

In this experiment you will be hearing a series of monosyllabic consonant-vowel-consonant words with one of 7 vowels. The words will be similar to the ones listed below:

[i] heed (rhymes with "feet")
[ɪ] hid (rhymes with "pig")
[E] head (rhymes with "bed")
[æ] had (rhymes with "cat")
[U] hood (rhymes with "book")
[u] who'd (rhymes with "boot")
[ʌ] hud (rhymes with "cup")

Your task is to click a button on the screen to indicate what vowel you think was spoken.
The 7 buttons on the screen are labeled with the phonetic symbol for the vowel, and with one of the 7 words listed above.
After choosing the vowel, you have to click on one of the buttons of the scale from 1 (poor) to 3 (good) to rate how good your perception of the vowel was.

The experiment consists of 238 trials. When you complete the experiment, a window will appear on the screen and your results will be displayed.

When you've finished reading these instructions, please click on "*Aplicação*", write your first and last names, and then press "*Iniciar Teste de Identificação*" to begin the experiment.

Be careful not to exit the experiment before you finish it. If you do that, you have to restart it from the beginning.

Appendix V

Instructions for the Generalization Test

Instructions for the Vowel Identification Experiment – Generalization Test

In this experiment you will be hearing a series of words with one of 7 vowels. The words will be similar to the words listed below:

[i] heed (rhymes with "feet")
[ɪ] hid (rhymes with "pig")
[E] head (rhymes with "bed")
[æ] had (rhymes with "cat")
[U] hood (rhymes with "book")
[u] who'd (rhymes with "boot")
[ʌ] hud (rhymes with "cup")

Your task is to click a button on the screen to indicate what vowel you think was spoken.
The 7 buttons on the screen are labeled with the phonetic symbol for the vowel, and with one of the 7 words listed above.

After choosing the vowel, you have to click on one of the buttons of the scale from 1 (poor) to 3 (good) according to how good your perception of the vowel was.

The experiment consists of 112 trials. When you complete the experiment, a window will appear on the screen and your results will be displayed.

When you've finished reading these instructions, please click on "*Aplicação*", write your first and last names, and then press "*Iniciar Teste de Identificação*" to begin the experiment.

Appendix W

Articulatory Descriptions of the Trained Phonetic Segments

PERCEPTUAL TRAINING – Experimental Group
SESSION 1: 20/03/2012

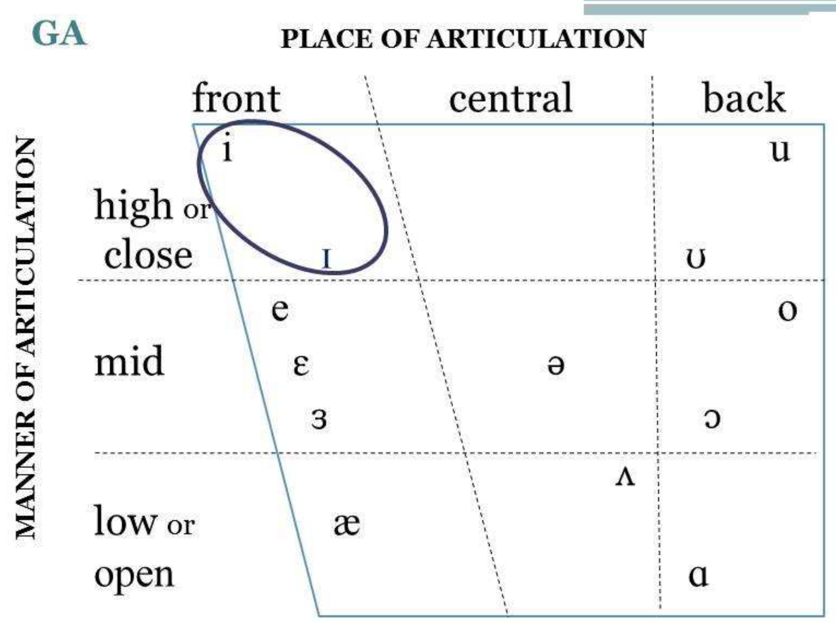
- Vowels are classified according to four questions:

1) How **high** is the tongue? (the vertical distance between the upper surface of the tongue and the palate)

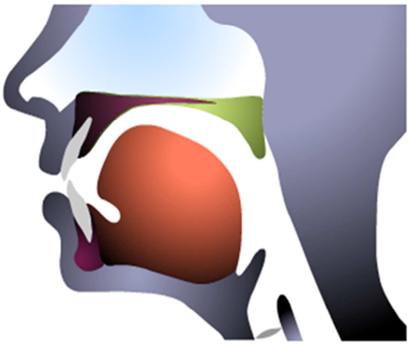
2) What **part of the tongue** (between front and back) is involved; that is, what part is raised? What part is lowered?

3) What is the **position of the lips**? (rounded or unrounded, neutral)

4) How **long** does the articulation of the vowel take?



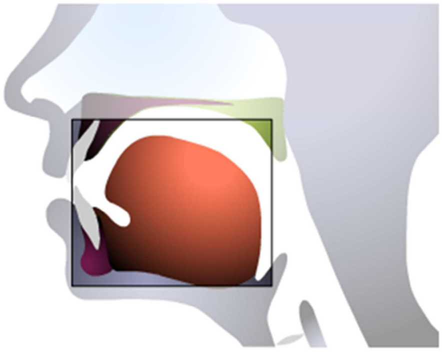
Front Vowels [i] [ɪ]



/i/ The tongue is positioned forward and high in the oral cavity with the sides in contact with the teeth laterally and the tip positioned behind the lower teeth.

[i]

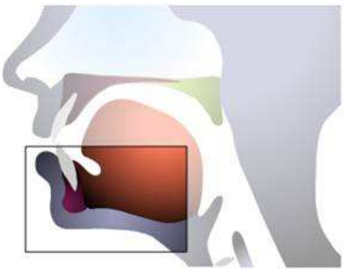
ea**t, **wee**k, tr**ee



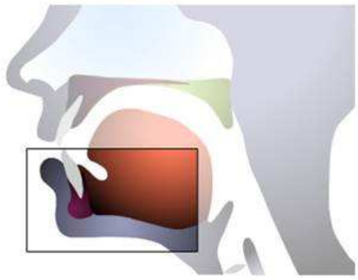
/ɪ/ The tongue is positioned forward and slightly lower in the oral cavity than for /i/, with the sides in contact with the teeth laterally and the tip positioned behind the lower teeth.

[ɪ]

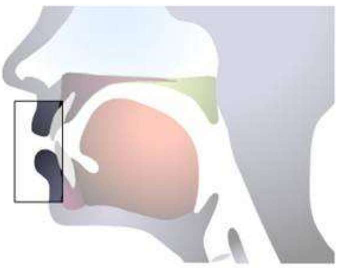
***i**nvite, p**i**n*



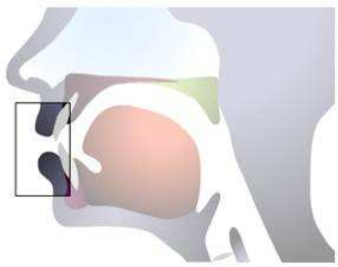
/i/ The mandible is elevated.



/ɪ/ The mandible is slightly lower than for /i/.



/i/ The lips are unrounded, and may be retracted.



/ɪ/ The lips are unrounded.

SESSION 2: 27/03/2012

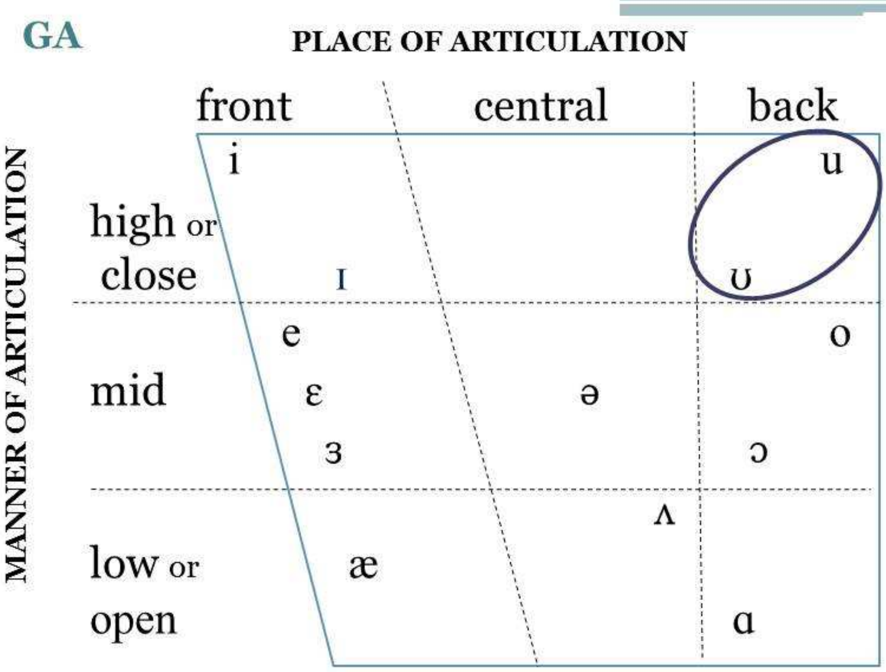
- Vowels are classified according to four questions:

1) How **high** is the tongue? (the vertical distance between the upper surface of the tongue and the palate)

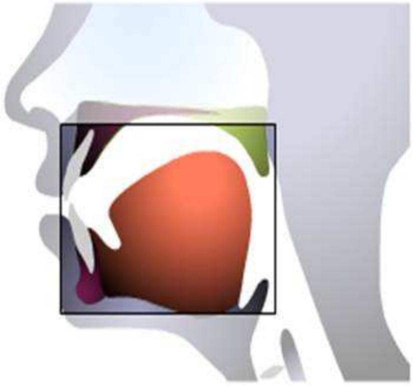
2) What **part of the tongue** (between front and back) is involved; that is, what part is raised? What part is lowered?

3) What is the **position of the lips**? (rounded or unrounded, neutral)

4) How **long** does the articulation of the vowel take?



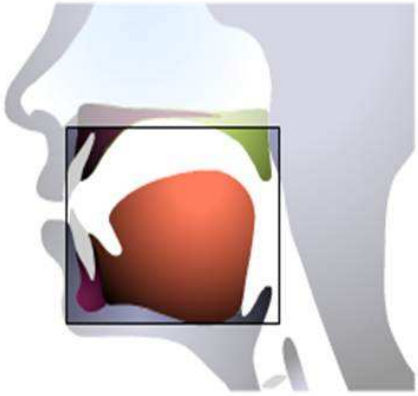
Back Vowels [u] [ʊ]



[u]

*b**oo**t, f**oo**d, dr**ew***

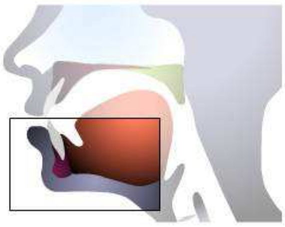
The tongue body is elevated into a **high** and **back** position with contact against the upper molars.



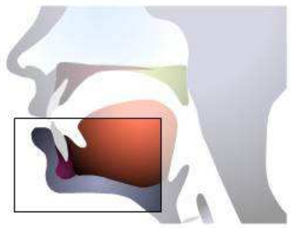
[ʊ]

*f**oo**t, b**oo**k, p**u**t*

The tongue body is **back** and elevated into a **mid-high position** with contact against the upper molars.

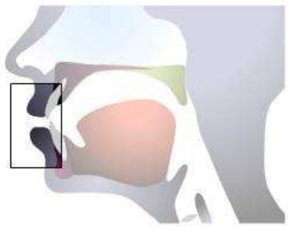


[u] The mandible is elevated.



[ʊ] The mandible is elevated but may **lower slightly**.

[u] The lips are rounded and protruded.



[ʊ] The lips are rounded and protruded.

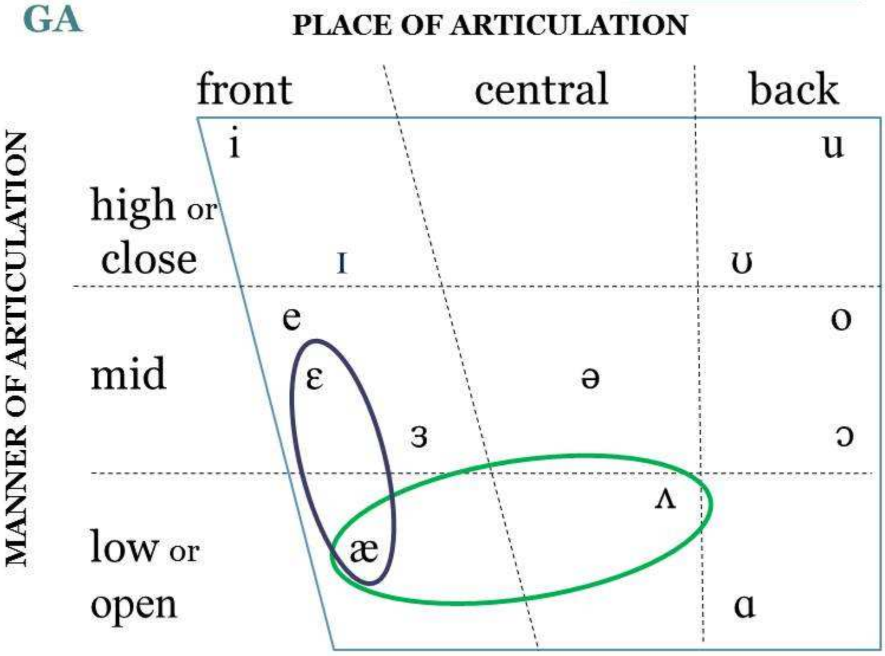
- Vowels are classified according to four questions:

1) How **high** is the tongue? (the vertical distance between the upper surface of the tongue and the palate)

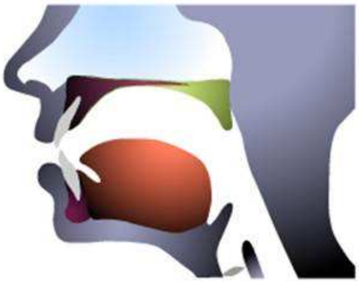
2) What **part of the tongue** (between front and back) is involved; that is, what part is raised? What part is lowered?

3) What is the **position of the lips**? (rounded or unrounded, neutral)

4) How **long** does the articulation of the vowel take?



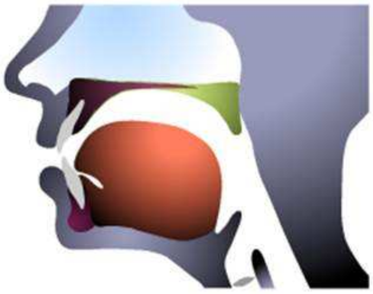
Front Vowels [æ] [ɛ]



[æ]

***a**t, gl**a**ss*

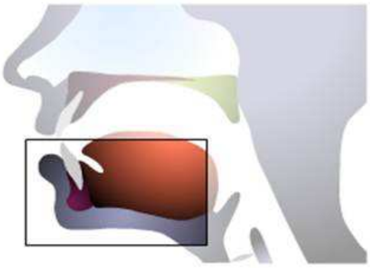
The tongue is positioned **slightly forward** and **low** in the oral cavity, with the tip positioned behind the lower teeth.



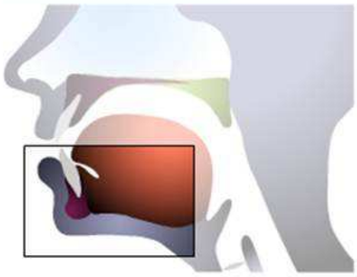
[ɛ]

***e**xtra, b**e**t*

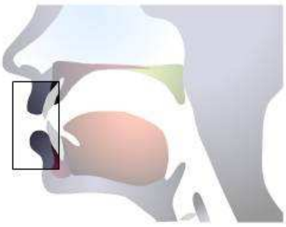
The tongue is positioned **forward** and **high** in the oral cavity with the sides in contact with the teeth laterally and the tip positioned behind the lower teeth,



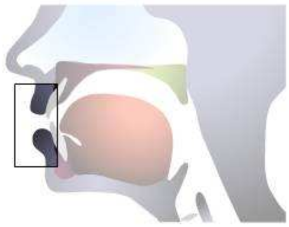
[æ] The mandible is **lowered** more than for any other front vowel.



[ɛ] The mandible is **elevated.**

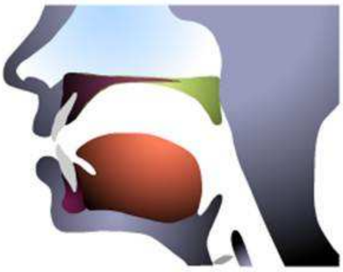


[æ] The lips are **unrounded** but may be retracted.



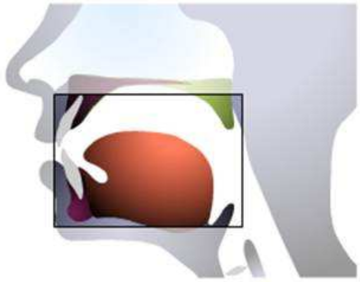
[ɛ] The lips are **unrounded** and may be retracted.

Front and Central Vowels [æ] [ʌ]



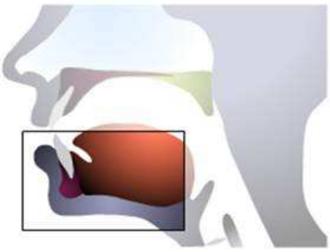
[æ]
***a**t, gl**a**ss*

The tongue is positioned **slightly forward** and **low** in the oral cavity, with the tip positioned behind the lower teeth.

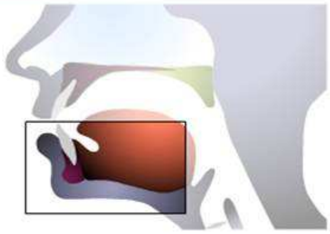


[ʌ]
***u**nder, s**u**n*

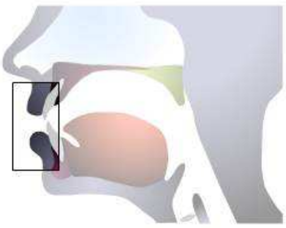
Although variable, the tongue is in the **center** of the oral cavity.



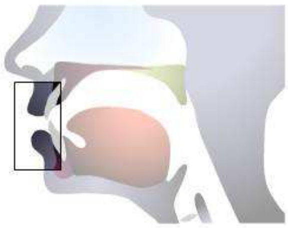
[æ] The mandible is **lowered** more than for any other front vowel.



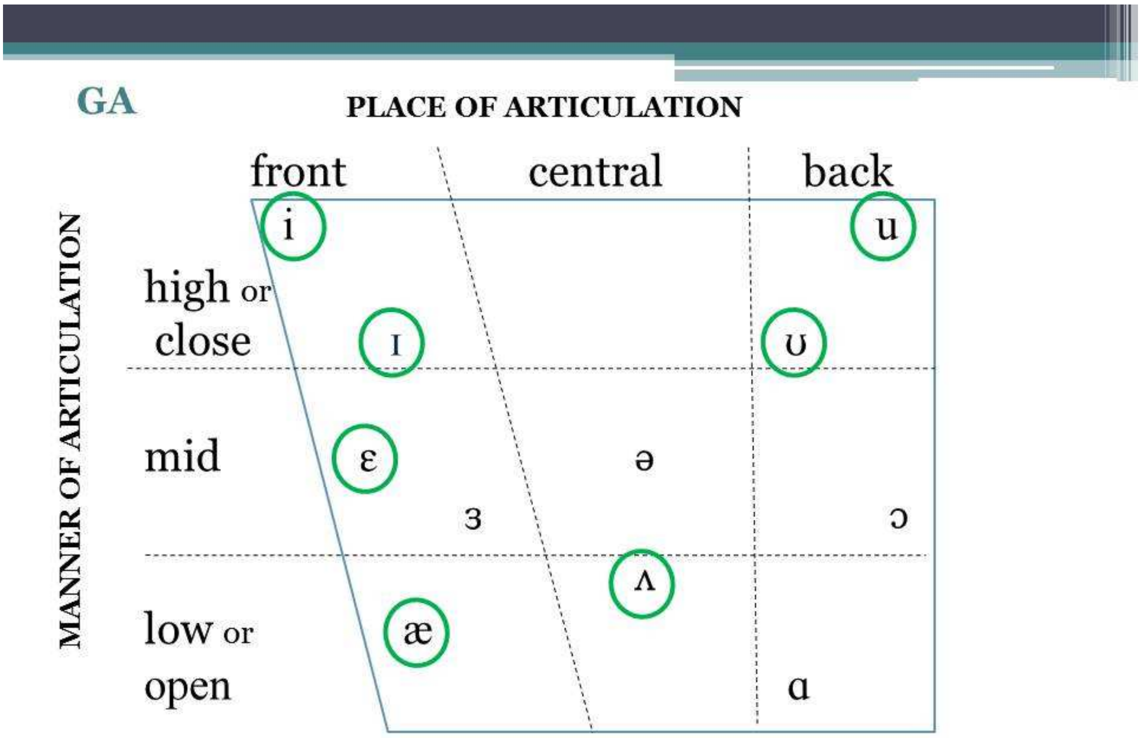
[ʌ] The jaw is **slightly lowered,** although its position varies depending on phonetic context.



[æ] The lips are **unrounded** but may be retracted.



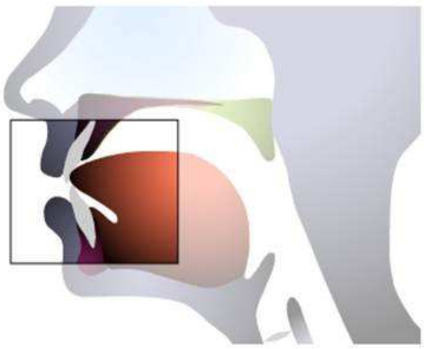
[ʌ] The lips are **unrounded.**


GA
PLACE OF ARTICULATION
front
central
back
MANNER OF ARTICULATION
high or
close
mid
low or
open
i
u
ɪ
ʊ
ɛ
ə
ɜ
ɔ
ʌ
æ
ɑ

PERCEPTUAL TRAINING – Control Group

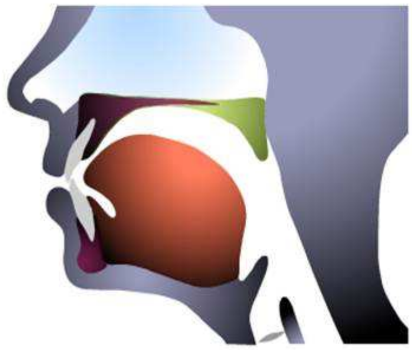
SESSION 1: 20/03/2012

Session 1: voiceless dental fricative [θ] and alveolar fricative [s]



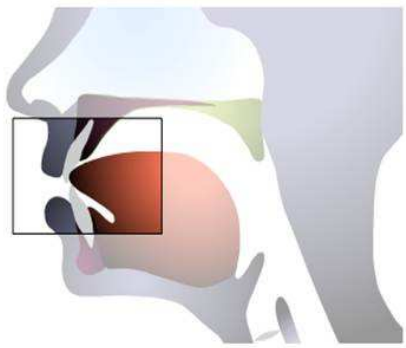
/θ/ Tongue tip is brought forward just below the upper teeth (interdental) or into slight contact with the back of the upper teeth (dental) to create a construction between tongue tip and upper teeth.

e.g. ***thing***, ***tooth****brush*, *wi****th***

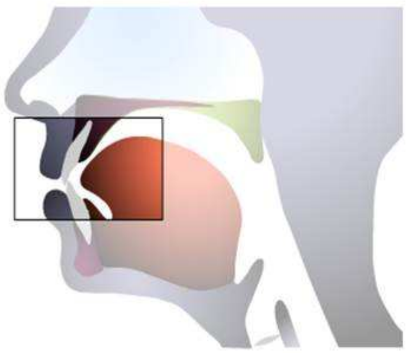


/s/ The apex and blade of the tongue are elevated into contact with the hard palate, leaving a narrow midline groove open.

e.g. ***s****oap*, *a****ss****ume*, *mi****ss***



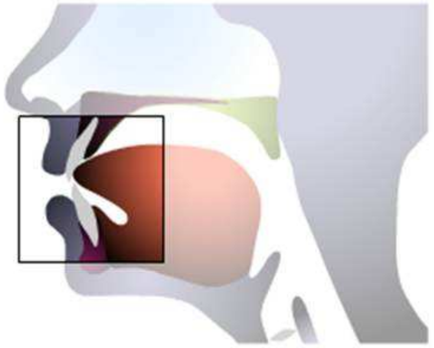
/θ/ Air forced between tongue surface and cutting edge of the upper teeth (interdental) or inside surface of the teeth (dental) creates audible frictional turbulence.



/s/ Air forced under pressure along midline groove creates audible frictional turbulence.

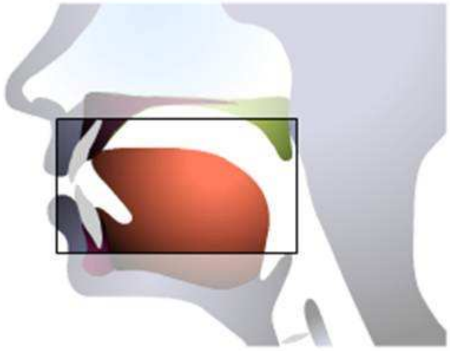
SESSION 2: 27/03/2012

Session 2: Voiced dental fricative [ð] and alveolar stop [d]



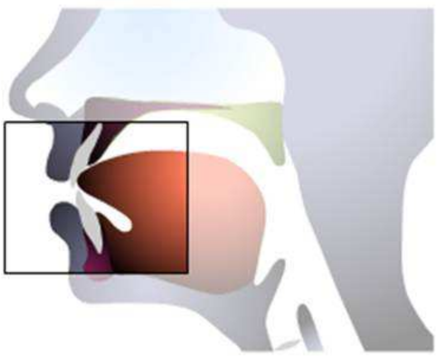
/ð/ Tongue tip is brought forward just below the upper teeth (interdental) or into slight contact with back of the upper teeth (dental) to create a constriction between the tongue tip and upper teeth.

e.g. ***th**at*, *wi**th**er*, *smoo**th***

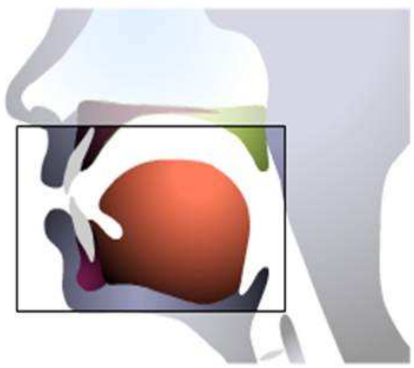


/d/ The front and sides of the tongue contact the alveolar ridge anteriorly and laterally.

e.g. ***d**eer*, *ra**d**ar*, *blee**d***

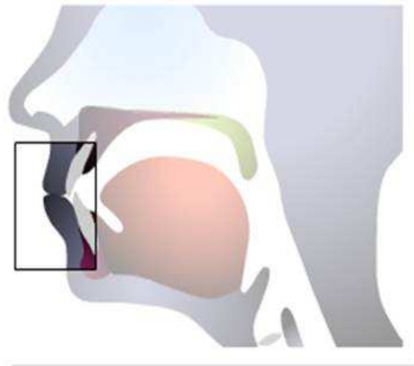


/ð/ Air forced between tongue surface and the cutting edge of the upper teeth (interdental) or inside surface of the teeth (dental) creates audible frictional turbulence.



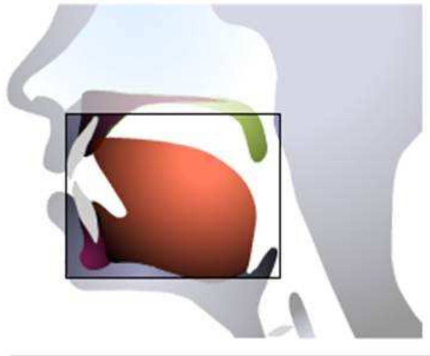
/d/ Air pressure built up behind obstruction is released by lowering the tongue, producing noise burst.

Session 3: bilabial nasal stop [m] and alveolar nasal stop [n]



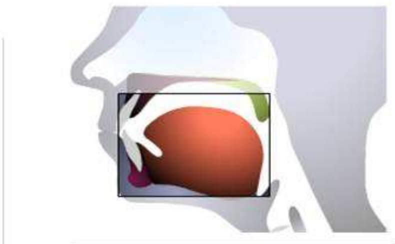
/m/ The lips are brought together to obstruct the oral cavity.

e.g. ***m**ask*, *a**m**ount*, *cal**m***

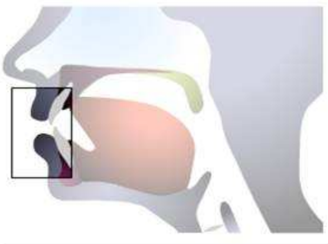


/n/ The front and sides of the tongue contact the alveolar ridge anteriorly and laterally to obstruct the oral cavity.

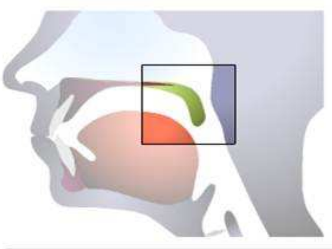
e.g. ***kn**ob*, *pi**n***



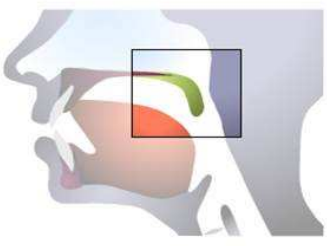
/m/ Tongue position may vary depending on phonetic context.



/n/ Lip configuration may vary depending on phonetic context.

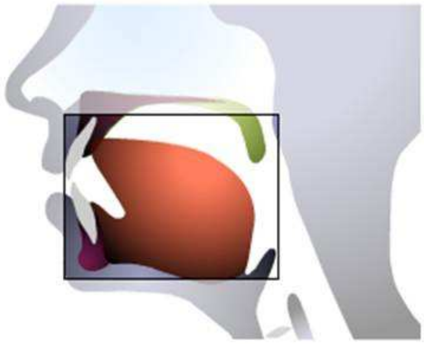


/m/ The velopharyngeal port is open, allowing acoustic energy and airflow to pass through the nose.



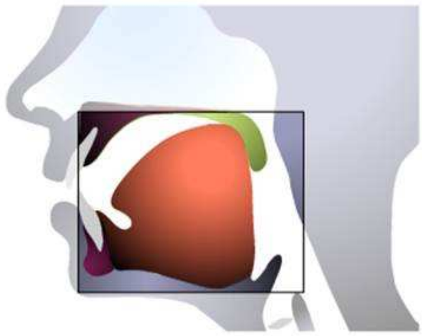
/n/ The velopharyngeal port is open, allowing acoustic energy and airflow to pass through the nose.

Session 4: alveolar nasal stop [n] and velar nasal stop [N]



/n/ The front and sides of the tongue contact the alveolar ridge anteriorly and laterally to obstruct the oral cavity.

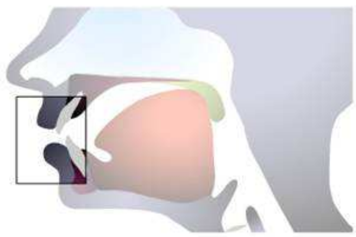
e.g. ***knob***, *pin*



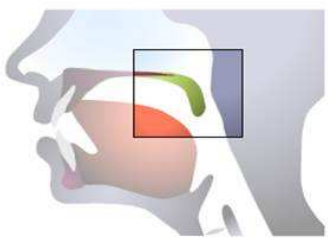
/ŋ/ The tongue dorsum is elevated and retracted to contact the soft palate, obstructing the oral cavity.

e.g. *singer, ring*

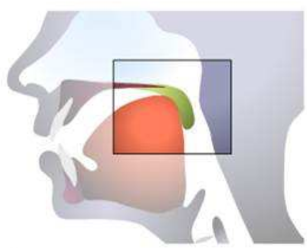
/n/ Lip configuration may vary depending on phonetic context.



/ŋ/ Lip configuration may vary depending on phonetic context.

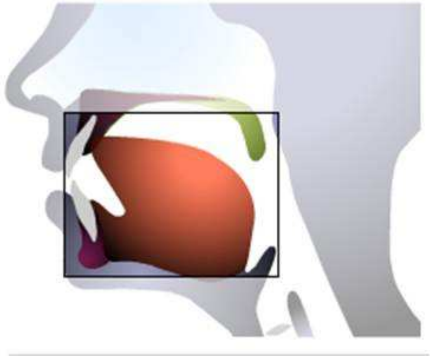


/n/ The velopharyngeal port is open, allowing acoustic energy and airflow to pass through the nose.



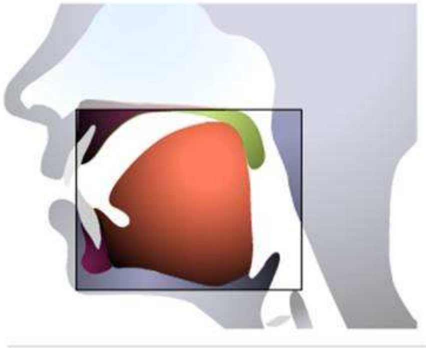
/ŋ/ The velopharyngeal port is open, allowing acoustic energy and airflow to pass through the nose.

Session 5: bilabial, alveolar and velar nasal stops: [m] [n] [N]



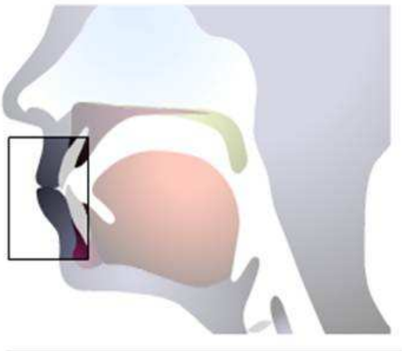
/n/ The front and sides of the tongue contact the alveolar ridge anteriorly and laterally to obstruct the oral cavity.

e.g. ***knob**, pin*



/ŋ/ The tongue dorsum is elevated and retracted to contact the soft palate, obstructing the oral cavity.

e.g. *singer, ring*



/m/ The lips are brought together to obstruct the oral cavity.

e.g. ***mask**, amount, calm*

Source: http://www.uiowa.edu/~acadtech/phonetics/

Appendix X

Instructions for the Perceptual Training Sessions

A) Experimental Group

PERCEPTUAL TRAINING - Session 1: Front Vowels

1) Instructions for the AX Discrimination Experiments

In this experiment you will be hearing sequences of two monosyllabic consonant-vowel-consonant words. Your task is to **decide whether the vowels** of the word pair **match or not**.

On the screen you will see two buttons labeled "**DIFFERENT**" and "**SAME**". Press the "DIFFERENT" button if the vowels of the word pair don't match or press the "SAME" button if they match. After you've made your decision, the correct answer will be shown on the screen. If your choice was correct, a "Next" button will be displayed so that you can move forward. If your choice was incorrect, you will have to click on the Replay button to listen to the sequence again and choose the correct answer.

The experiment consists of **72** trials. When you complete the experiment, your results will be displayed on the screen.

When you've finished reading these instructions, please click on "*Aplicação*", write your first and last names, and then press "*Iniciar Treino de Discriminação*" to begin the experiment.

2) Instructions for the Vowel Identification Experiments

In this experiment you will be hearing a series of monosyllabic consonant-vowel-consonant words with one of 2 vowels.

Your task is to click a button on the screen to **indicate what vowel you think was spoken**.
The 2 buttons on the screen are labeled with the phonetic symbol for the vowel, and with one of the 2 words listed below:

[i] **heed** (rhymes with "feet")
[ɪ] **hid** (rhymes with "pig")

If you wish to hear the word again before making up your mind, click the Replay button.
After you've made your decision, the correct answer will be shown on the screen. If your choice was correct, a "Next" button will be displayed so that you can move forward. If your choice was incorrect, you will have to click on the “Replay” button to listen to the word again and choose the correct vowel.

The experiment consists of **96** trials. When you complete the experiment, your results will be displayed on the screen.

When you've finished reading these instructions, please click on "*Aplicação*", write your first and last names, and then press "*Iniciar Treino de Identificação*" to begin the experiment.

PERCEPTUAL TRAINING - Session 2: Back Vowels

1) Instructions for the AX Discrimination Experiments

In this experiment you will be hearing sequences of two monosyllabic consonant-vowel-consonant words. Your task is to **decide whether the vowels** of the word pair **match or not**.

On the screen you will see two buttons labeled "**DIFFERENT**" and "**SAME**". Press the "DIFFERENT" button if the vowels of the word pair don't match or press the "SAME" button if they match. After you've made your decision, the correct answer will be shown on the screen. If your choice was correct, a "Next" button will be displayed so that you can move forward. If your choice was incorrect, you will have to click on the Replay button to listen to the sequence again and choose the correct answer.

The experiment consists of **72** trials. When you complete the experiment, your results will be displayed on the screen.

When you've finished reading these instructions, please click on "*Aplicação*", write your first and last names, and then press "*Iniciar Treino de Discriminação*" to begin the experiment.

2) Instructions for the Vowel Identification Experiments

In this experiment you will be hearing a series of monosyllabic consonant-vowel-consonant words with one of 2 vowels.

Your task is to click a button on the screen to **indicate what vowel you think was spoken**.
The 2 buttons on the screen are labeled with the phonetic symbol for the vowel, and with one of the 2 words listed below:

[U] hood (rhymes with "book")
[u] who'd (rhymes with "boot")

If you wish to hear the word again before making up your mind, click the Replay button.
After you've made your decision, the correct answer will be shown on the screen. If your choice was correct, a "Next" button will be displayed so that you can move forward. If your choice was incorrect, you will have to click on the “Replay” button to listen to the word again and choose the correct vowel.

The experiment consists of **96** trials. When you complete the experiment, your results will be displayed on the screen.

When you've finished reading these instructions, please click on "*Aplicação*", write your first and last names, and then press "*Iniciar Treino de Identificação*" to begin the experiment.

PERCEPTUAL TRAINING - Session 3

<u>Part 1: Front Vowels</u>

1) Instructions for the AX Discrimination Experiment

In this experiment you will be hearing sequences of two monosyllabic consonant-vowel-consonant words. Your task is to **decide whether the vowels** of the word pair **match or not**.

On the screen you will see two buttons labeled "**DIFFERENT**" and "**SAME"**. Press the "DIFFERENT" button if the vowels of the word pair don't match or press the "SAME" button if they match. After you've made your decision, the correct answer will be shown on the screen. If your choice was correct, a "Next" button will be displayed so that you can move forward. If your choice was incorrect, you will have to click on the Replay button to listen to the sequence again and choose the correct answer.

The experiment consists of **72 trials**. When you complete the experiment, your results will be displayed on the screen.

When you've finished reading these instructions, please click on "*Aplicação*", write your first and last names, and then press "*Iniciar Treino de Discriminação*" to begin the experiment.

2) Instructions for the Vowel Identification Experiment

In this experiment you will be hearing a series of monosyllabic consonant-vowel-consonant words with one of 2 vowels.
Your task is to click a button on the screen to **indicate what vowel you think was spoken**.

The 2 buttons on the screen are labeled with the phonetic symbol for the vowel, and with one of the 2 words listed below:

[æ] had (rhymes with "cat")
[E] head (rhymes with "bed")

If you wish to hear the word again before making up your mind, click the Replay button. After you've made your decision, the correct answer will be shown on the screen. If your choice was correct, a "Next" button will be displayed so that you can move forward. If your choice was incorrect, you will have to click on the "Replay" button to listen to the word again and choose the correct vowel.

The experiment consists of **96 trials.** When you complete the experiment, your results will be displayed on the screen.

When you've finished reading these instructions, please click on "*Aplicação*", write your first and last names, and then press "*Iniciar Treino de Identificação*" to begin the experiment.

Part 2: Front and Central Vowels

1) Instructions for the AX Discrimination Experiment

In this experiment you will be hearing sequences of two monosyllabic consonant-vowel-consonant words. Your task is to **decide whether the vowels** of the word pair **match or not**.

On the screen you will see two buttons labeled "**DIFFERENT**" and "**SAME"**. Press the "DIFFERENT" button if the vowels of the word pair don't match or press the "SAME" button if they match. After you've made your decision, the correct answer will be shown on the screen. If your choice was correct, a "Next" button will be displayed so that you can move forward. If your choice was incorrect, you will have to click on the Replay button to listen to the sequence again and choose the correct answer.

The experiment consists of **48 trials**. When you complete the experiment, your results will be displayed on the screen.

When you've finished reading these instructions, please click on "*Aplicação*", write your first and last names, and then press "*Iniciar Treino de Discriminação*" to begin the experiment.

2) Instructions for the Vowel Identification Experiment

In this experiment you will be hearing a series of monosyllabic consonant-vowel-consonant words with one of 2 vowels.

Your task is to click a button on the screen **to indicate what vowel you think was spoken**.

The 2 buttons on the screen are labeled with the phonetic symbol for the vowel, and with one of the 2 words listed below:

[æ] had (rhymes with "cat")
[ʌ] hud (rhymes with "cup")

If you wish to hear the word again before making up your mind, click the Replay button. After you've made your decision, the correct answer will be shown on the screen. If your choice was correct, a "Next" button will be displayed so that you can move forward. If your choice was incorrect, you will have to click on the “Replay” button to listen to the word again and choose the correct vowel.

The experiment consists of **64 trials.** When you complete the experiment, your results will be displayed on the screen.

When you've finished reading these instructions, please click on "*Aplicação*", write your first and last names, and then press "*Iniciar Treino de Identificação*" to begin the experiment.

TRAINING SESSIONS 4 & 5: Vowels

1) Instructions for the ABX Discrimination Experiment

In this experiment you will be hearing sequences of three monosyllabic consonant-vowel consonant words with one of 7 vowels. The words will be similar to the ones listed below:

[i] heed (rhymes with "feet")
[ɪ] hid (rhymes with "pig")
[E] head (rhymes with "bed")
[æ] had (rhymes with "cat")
[U] hood (rhymes with "book")
[u] who'd (rhymes with "boot")
[ʌ] hud (rhymes with "cup")

On each sequence there will be either a word with a different vowel, or the three words will have the same vowel. Your task is to **decide whether the word with the odd vowel is in position 1, 2 or 3, or whether the three words have the same vowel.**

In the center of the screen you will see buttons labeled "**1", "2", "3"** and "**SAME**". Press button "1", "2", "3" to identify the position of the odd vowel or "SAME" if the three vowels match.

After you've made your decision, the correct answer will be shown on the screen. If your choice was correct, a "Next" button will be displayed so that you can move forward. If your choice was incorrect, you will have to click on the "Replay" button to listen to the sequence again and choose the correct answer.

The experiment consists of **64** trials. When you complete the experiment, your results will be displayed on the screen.

When you've finished reading these instructions, please click on "*Aplicação*", write your first and last names, and then press "*Iniciar Treino de Discriminação*" to begin the experiment.

2) Instructions for the Vowel Identification Experiment

In this experiment you will be hearing a series of monosyllabic consonant-vowel-consonant words with one of 7 vowels. The words will be similar to the ones listed below:

[i] heed (rhymes with "feet")
[ɪ] hid (rhymes with "pig")
[E] head (rhymes with "bed")
[æ] had (rhymes with "cat")
[U] hood (rhymes with "book")
[u] who'd (rhymes with "boot")
[ʌ] hud (rhymes with "cup")

Your task is to click a button on the screen **to indicate what vowel you think was spoken**.
The 7 buttons on the screen are labeled with the phonetic symbol for the vowel, and with one of the 7 words listed above.
After you've made your decision, the correct answer will be shown on the screen. If your choice was correct, a "Next" button will be displayed so that you can move forward. If your choice was incorrect, you will have to click on the "Replay" button to listen to the sequence again and choose the correct answer.

The experiment consists of **168** trials. When you complete the experiment, a window will appear on the screen and your results will be displayed.

When you've finished reading these instructions, please click on "*Aplicação*", write your first and last names, and then press "*Iniciar Treino de Identificação*" to begin the experiment.

B) Control Group

PERCEPTUAL TRAINING - Session 1: Voiceless dental and alveolar fricatives
1) Instructions for the AX Discrimination Experiments

In this experiment you will be hearing sequences of two words. Your task is to **decide whether the voiceless fricatives of the word pair match or not.**

On the screen you will see two buttons labeled "**DIFFERENT**" and "**SAME**". Press the "DIFFERENT" button if the voiceless fricatives of the word pair don't match or press the "SAME" button if they match. After you've made your decision, the correct answer will be shown on the screen. If your choice was correct, a "Next" button will be displayed so that you can move forward. If your choice was incorrect, you will have to click on the Replay button to listen to the sequence again and choose the correct answer.

The experiment consists of **96** trials. When you complete the experiment, your results will be displayed on the screen.

When you've finished reading these instructions, please click on "*Aplicação*", write your first and last names, and then press "*Iniciar Treino de Discriminação*" to begin the experiment.

2) Instructions for the Consonant Identification Experiments

In this experiment you will be hearing a series of words with one of two voiceless fricative consonants.

Your task is to click a button on the screen to **indicate what consonant you think was spoken**.
The 2 buttons on the screen are labeled with the phonetic symbol for the consonant, and with one of the 2 words listed below:

[s] sink
[th] think

If you wish to hear the word again before making up your mind, click the Replay button.
After you've made your decision, the correct answer will be shown on the screen. If your choice was correct, a "Next" button will be displayed so that you can move forward. If your choice was incorrect, you will have to click on the “Replay” button to listen to the word again and choose the correct vowel.

The experiment consists of **72** trials. When you complete the experiment, your results will be displayed on the screen.

When you've finished reading these instructions, please click on "*Aplicação*", write your first and last names, and then press "*Iniciar Treino de Identificação*" to begin the experiment.

PERCEPTUAL TRAINING - Session 2: Voiced dental fricative and alveolar stop
1) Instructions for the AX Discrimination Experiments

In this experiment you will be hearing sequences of two words. Your task is to **decide whether the voiced consonants of the word pair match or not.**

On the screen you will see two buttons labeled "**DIFFERENT**" and "**SAME**". Press the "DIFFERENT" button if the voiceless fricatives of the word pair don't match or press the "SAME" button if they match. After you've made your decision, the correct answer will be shown on the screen. If your choice was correct, a "Next" button will be displayed so that you can move forward. If your choice was incorrect, you will have to click on the Replay button to listen to the sequence again and choose the correct answer.

The experiment consists of **96** trials. When you complete the experiment, your results will be displayed on the screen.

When you've finished reading these instructions, please click on "*Aplicação*", write your first and last names, and then press "*Iniciar Treino de Discriminação*" to begin the experiment.

2) Instructions for the Consonant Identification Experiments

In this experiment you will be hearing a series of words with one of two voiced consonants.

Your task is to click a button on the screen to **indicate what consonant you think was spoken.**
The 2 buttons on the screen are labeled with the phonetic symbol for the consonant, and with one of the 2 words listed below:

[d] day
[dh] they

If you wish to hear the word again before making up your mind, click the Replay button.
After you've made your decision, the correct answer will be shown on the screen. If your choice was correct, a "Next" button will be displayed so that you can move forward. If your choice was incorrect, you will have to click on the “Replay” button to listen to the word again and choose the correct vowel.

The experiment consists of **72** trials. When you complete the experiment, your results will be displayed on the screen.

When you've finished reading these instructions, please click on "*Aplicação*", write your first and last names, and then press "*Iniciar Treino de Identificação*" to begin the experiment.

3) Instructions for the Consonant Identification Experiments

In this experiment you will be hearing a series of words with one of two dental fricative consonants.

Your task is to click a button on the screen to **indicate what consonant you think was spoken**.
The 2 buttons on the screen are labeled with the phonetic symbol for the consonant, and with one of the 2 words listed below:

[th] thigh
[dh] thy

If you wish to hear the word again before making up your mind, click the Replay button.
After you've made your decision, the correct answer will be shown on the screen. If your choice was correct, a "Next" button will be displayed so that you can move forward. If your choice was incorrect, you will have to click on the “Replay” button to listen to the word again and choose the correct vowel.

The experiment consists of **72 trials**. When you complete the experiment, your results will be displayed on the screen.

When you've finished reading these instructions, please click on "*Aplicação*", write your first and last names, and then press "*Iniciar Treino de Identificação*" to begin the experiment.

TRAINING SESSION 3: bilabial nasal stop [m] and alveolar nasal stop [n]

1) Instructions for the AX Discrimination Experiments

In this experiment you will be hearing sequences of two words. Your task is to **decide whether the nasal stops of the word pair match or not.**

On the screen you will see two buttons labeled **"DIFFERENT**" and **"SAME".** Press the "DIFFERENT" button if the nasal stops of the word pair don't match or press the "SAME" button if they match. After you've made your decision, the correct answer will be shown on the screen. If your choice was correct, a "Next" button will be displayed so that you can move forward. If your choice was incorrect, you will have to click on the Replay button to listen to the sequence again and choose the correct answer.

The experiment consists of **80 trials**. When you complete the experiment, your results will be displayed on the screen.

When you've finished reading these instructions, please click on "*Aplicação*", write your first and last names, and then press "*Iniciar Treino de Discriminação*" to begin the experiment.

2) Instructions for the Consonant Identification Experiments

In this experiment you will be hearing a series of words with one of two nasal stops. Your task is to click a button on the screen to **indicate what consonant you think was spoken.**
The 2 buttons on the screen are labeled with the phonetic symbol for the consonant, and with one of the 2 words listed below:

[m] sum
[n] sun

If you wish to hear the word again before making up your mind, click the Replay button.
After you've made your decision, the correct answer will be shown on the screen. If your choice was correct, a "Next" button will be displayed so that you can move forward. If your choice was incorrect, you will have to click on the “Replay” button to listen to the word again and choose the correct vowel.

The experiment consists of **72 trials**. When you complete the experiment, your results will be displayed on the screen.

When you've finished reading these instructions, please click on "*Aplicação*", write your first and last names, and then press "*Iniciar Treino de Identificação*" to begin the experiment.

TRAINING SESSION 4: alveolar nasal stop [n] and velar nasal stop [N]

1) Instructions for the AX Discrimination Experiments

In this experiment you will be hearing sequences of two words. Your task is to **decide whether the nasal stops of the word pair match or not.**

On the screen you will see two buttons labeled **"DIFFERENT"** and **"SAME".** Press the "DIFFERENT" button if the nasal stops of the word pair don't match or press the "SAME" button if they match. After you've made your decision, the correct answer will be shown on the screen. If your choice was correct, a "Next" button will be displayed so that you can move forward. If your choice was incorrect, you will have to click on the Replay button to listen to the sequence again and choose the correct answer.
The experiment consists of **80 trials**. When you complete the experiment, your results will be displayed on the screen.

When you've finished reading these instructions, please click on "*Aplicação*", write your first and last names, and then press "*Iniciar Treino de Discriminação*" to begin the experiment.

2) Instructions for the Consonant Identification Experiments

In this experiment you will be hearing a series of words with one of two nasal stops. Your task is to click a button on the screen to **indicate what consonant you think was spoken.**
The 2 buttons on the screen are labeled with the phonetic symbol for the consonant, and with one of the 2 words listed below:

[n] sun
[N] sung

If you wish to hear the word again before making up your mind, click the Replay button.
After you've made your decision, the correct answer will be shown on the screen. If your choice was correct, a "Next" button will be displayed so that you can move forward. If your choice was incorrect, you will have to click on the "Replay" button to listen to the word again and choose the correct vowel.

The experiment consists of **72 trials**. When you complete the experiment, your results will be displayed on the screen.

When you've finished reading these instructions, please click on "*Aplicação*", write your first and last names, and then press "*Iniciar Treino de Identificação*" to begin the experiment.

TRAINING SESSION 5: bilabial, alveolar and velar nasal stops [m] [n] [N]

1) Instructions for the Consonant Identification Experiment

In this experiment you will be hearing a series of words with one of three nasal stops. Your task is to click a button on the screen to **indicate what consonant you think was spoken**.
The 3 buttons on the screen are labeled with the phonetic symbol for the consonant, and with one of the 3 words listed below:

[m] sum
[n] sun
[N] sung

If you wish to hear the word again before making up your mind, click the Replay button.
After you've made your decision, the correct answer will be shown on the screen. If your choice was correct, a "Next" button will be displayed so that you can move forward. If your choice was incorrect, you will have to click on the "Replay" button to listen to the word again and choose the correct vowel.

The experiment consists of **162 trials**. When you complete the experiment, your results will be displayed on the screen.

When you've finished reading these instructions, please click on "*Aplicação*", write your first and last names, and then press "*Iniciar Treino de Identificação*" to begin the experiment.

Appendix Y

Computer Screens of the Training Tasks

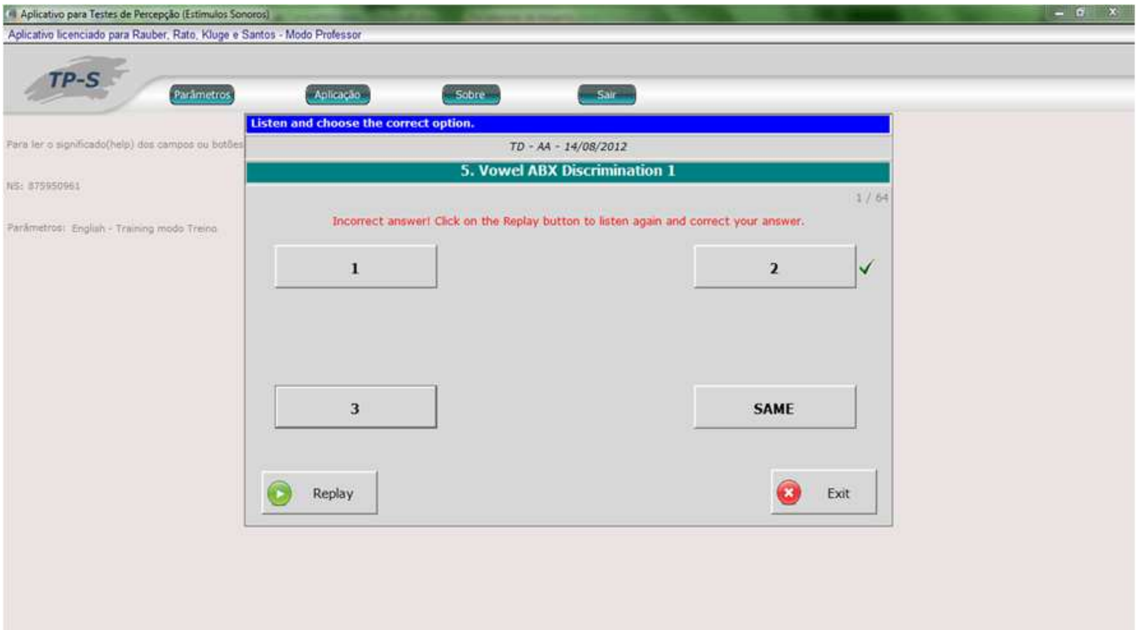


Figure 1. Example of an oddity discrimination training task with immediate feedback.

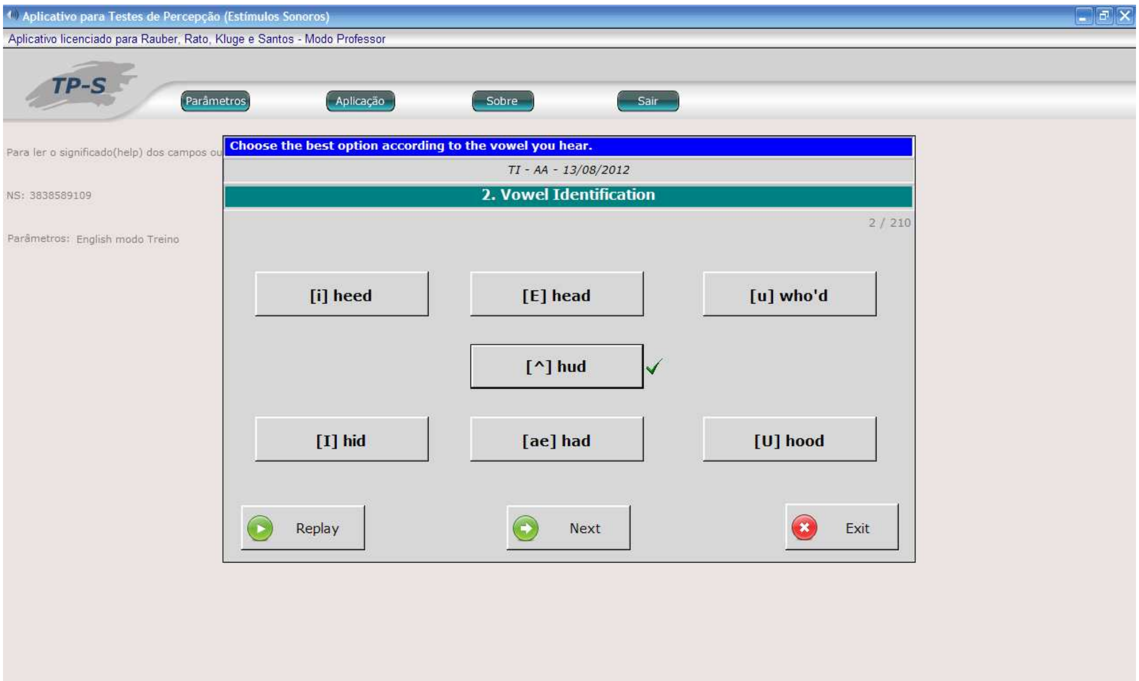


Figure 2. Example of an identification training task with immediate feedback.

Appendix Z

Training Session in the Computer Lab

*Figure 1***.** Photographs of a training session in the computer lab.

Appendix AA

Vowel Normalization Script

```
#By Ricardo Bion
form normalize
comment normalize each participant to a new max and min value
integer nmaxF1: 675
integer nminF1: 266
integer nmaxF2: 2157
integer nminF2: 1109
endform

select all

Collapse rows... speaker "" F1 "" ""
participants = Get number of rows
select all
tablex = selected("Table")
  for i from 1 to participants
  select tablex
  Extract rows where column (number)... speaker "equal to" i
  t'i' = selected("Table")
  endfor
count = 0
for y from 1 to participants
  select t'y'
  call other
    for i from 1 to 63
    count = count + 1
    f1v'count' = Get value... 'i' F1
    f2v'count' = Get value... 'i' F2
    vo'count'$ = Get value... 'i' vowel
    endfor
endfor
select tablex
count = 0
for y from 1 to participants
  for i from 1 to 63
  count = count + 1
  vo$ = vo'count'$
  Set string value... 'count' vowel 'vo$'
  f1v = f1v'count'
  Set numeric value... 'count' F1 'f1v'
  f2v = f2v'count'
  Set numeric value... 'count' F2 'f2v'
  endfor
endfor
```

```
select all
minus tablex
Remove
procedure other
nvalues = Get number of rows
for formant to 2
   for i from 1 to nvalues
   v'i' = Get value... i F'formant'
   endfor

# define max e min based on the max and min vowel_mean-+vowel_SD
table1 = selected("Table")
Collapse rows... vowel "" "F1 F2" "" ""
nrows = Get number of rows
table2 = selected("Table")
   for i from 1 to nrows
   select table2
   label$ = Get value... i vowel
   printline 'label$'
   value'i' = Get value... i F'formant'
   t = value'i'
   printline 't'
   select table1
   Extract rows where column (text)... vowel "is equal to" 'label$'
   sd'i' = Get standard deviation... F'formant'
   t = sd'i'
   printline 't'
   endfor
# define max e min
min = 999999
max = 0
   for i from 1 to nrows
   # max
   temp = value'i' + sd'i'
   if temp > max
   max = temp
   endif
   # min
   temp = value'i' - sd'i'
   if temp < min
   min = temp
   endif
   endfor
################### defined max and min
# convert these values to a scale from 0 to 1
   for i from 1 to nvalues
   normalized'i' = (v'i' - min)/(max-min)
   endfor

printline 'newline$'
```

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

Appendix AB

Script to calculate the Euclidean distance

```
#Written by Paul Boersma
#Modified by Andreia Rauber

clearinfo
select all
table1 = selected("Table")
Collapse rows... speaker "" F1 "" ""
participants = Get number of rows
Remove

#printline participant'tab$'pair'tab$'ED

for participant to participants
   select table1

  # extrair os dados de um participante
   Extract rows where column (number)... speaker "equal to" participant

   # calcular todos os means de F1 e F2 de todas as vogais
   Collapse rows... vowel "" "F1 F2" "" ""

   #organizar alfabeticamente por vogal
   Sort rows... vowel

   # mean de F1 e F2 de cada vogal
   for formant to 2
     for vowel to 7
     f'formant''vowel'= Get value... 'vowel' F'formant'
     endfor
   endfor

# calcular as EDs e imprimir na tela
iI= sqrt(((f14 - f16)^2)+((f26 - f24)^2))
printline 'participant''tab$'i-I'tab$''iI'

eae= sqrt(((f11 - f12)^2)+((f22 - f21)^2))
printline 'participant''tab$'E-ae'tab$''eae'

vU= sqrt(((f15 - f17)^2)+((f27 - f25)^2))
printline 'participant''tab$'v-U'tab$''vU'

uU= sqrt(((f13 - f17)^2)+((f27 - f23)^2))
printline 'participant''tab$'u-U'tab$''uU'

vae= sqrt(((f11 - f15)^2)+((f25 - f21)^2))
printline 'participant''tab$'v-ae'tab$''vae'
```

```
pause

# limpar para comecar novamente
select all
minus table1
Remove

endfor
```

Appendix AC

Formant and Duration Measurements Script

```
# Written by Paul Boersma, April 25, 2006
# Modified by Andreia Rauber

Read Table from table file... tableVogais2.txt
numberOfRows = Get number of rows
assert numberOfRows = 21
previousSpeaker$ = ""
for row to numberOfRows
	speaker$ = Get value... row speaker
	gender$ = Get value... row gender
	start = Get value... row start
	end = Get value... row end
	#
	if speaker$ <> previousSpeaker$
		if previousSpeaker$ <> ""
			select Sound 'previousSpeaker$'
			Remove
		endif
		Read from file... 'speaker$'.wav
		previousSpeaker$ = speaker$
	else
		select Sound 'speaker$'
	endif
	#
	# Formant analysis.
	#
	formantCeiling = if gender$ = "M" then 5000 else 5500 fi
	duration = end - start
	mid = start + duration / 2
	startpart = mid - duration / 5
	endpart = mid + duration / 5
	Extract part... startpart endpart Rectangular 1.0 no
	Rename... segment
	windowLength = Get total duration
	To Formant (burg)... 0 5 formantCeiling windowLength 50
	for iformant to 3
		f'iformant' = Get value at time... iformant windowLength/2 Hertz Linear
		b'iformant' = Get bandwidth at time... iformant windowLength/2 Hertz
Linear
	endfor
	plus Sound segment
	Remove
	#
	# Save results in tableVogais2.
```

```
	#
	select Table tableVogais2
	for iformant to 3
		formant = if f'iformant' = undefined then 0 else f'iformant' fi
		bandwidth = if b'iformant' = undefined then 0 else b'iformant' fi
		Set string value... row F'iformant' 'formant:3'
		Set string value... row B'iformant' 'bandwidth:3'
	endfor
endfor
Write to table file... tableVogais2.txt
select Sound 'previousSpeaker$'
Remove
```

Appendix AD

Script to Create Table

```
# Paul Boersma, April 25, 2006
# Modified by Andreia Rauber

Create Table with column names... tableVogais2 63
... speaker dialect gender vowel start end dur
... F1 B1 F2 B2 F3 B3

row = 0
call measureSpeakers PE F 1
#call measureSpeakers PE M 1
assert row = 63 ; 'row'
select Table tableVogais2
Write to table file... tableVogais2.tx
procedure measureSpeakers dialect$ gender$ numberOfSpeakers
	for speaker to numberOfSpeakers
		speaker$ = "'dialect$'_'gender$'_'speaker'"
		Read from file... 'speaker$'.TextGrid
		numberOfIntervals = Get number of intervals... 2
		assert numberOfIntervals = 127 ; 'speaker$'
		for iinterval to numberOfIntervals
			label$ = Get label of interval... 2 iinterval
			if label$ <> ""
				start = Get starting point... 2 iinterval
				end = Get end point... 2 iinterval
				duration = end - start
				assert duration > 0.010 ; 'speaker$' 'start'
				#
				# Get the vowel: one of i, e, E, a, O, o, u.
				#
				vowel$ = mid$ (label$, 1, 1)

				if vowel$ = "\"
					vowel$ = mid$ (label$, 1, 3)
				endif
				# Store results in tableVogais2.
				#
				select Table tableVogais2
				row += 1
				Set string value... row speaker 'speaker$'
				Set string value... row dialect 'dialect$'
				Set string value... row gender 'gender$'
				Set string value... row vowel 'vowel$'
				Set string value... row start 'start:3'
				Set string value... row end 'end:3'
				Set string value... row dur 'duration:3'
```

```
                    #
                    select TextGrid 'speaker$'
                endif
            endfor
            Remove
        endfor
endproc
```

Appendix AE

Script to Plot Vowels Produced by a Group of Speakers

```
# Script written by Ricardo Bion, November 2006
# Modified September, 2010

clearinfo

form PARTICIPANT
comment put 0 for all participants
integer plot_participant: 0
integer max_F2: 3200
integer min_F2: 700
integer max_F1: 1000
integer min_F1: 250
endform

##############################################################
#Erase all
Select outer viewport... 0 8 0 6
Black
Line width... 1
Plain line
Font size... 12
Axes... log10(max_F2) log10(min_F2) log10(max_F1) log10(min_F1)

#One logarithmic mark bottom... 600 yes yes no
One logarithmic mark bottom... 800 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... 3200 yes yes no
One logarithmic mark left... 300 yes yes no
One logarithmic mark left... 400 yes yes no
One logarithmic mark left... 500 yes yes no
One logarithmic mark left... 600 yes yes no
One logarithmic mark left... 800 yes yes no
One logarithmic mark left... 1000 yes yes no

Draw inner box

Text left... yes %F_%1 %(%H%e%r%t%z%)
Text bottom... yes %F_%2 %(%H%e%r%t%z%)
########################################################
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$ = "Blue"
line_of_the_sd$ = "Dashed line"


table1 = selected("Table")
Collapse rows... 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 AF

Script to Plot Vowels Produced by Groups of Speakers

```
# Paul Boersma, 28 December 2008
# Modified by Andreia Rauber

select all
table280 = selected("Table")
#table280 = Read from file... table280_Medianas_log.Table

dialect1$ = "CG"
dialect2$ = "EG"
gender1$ = "N"
gender2$ = "M"
vowel1$ = "i"
vowel2$ = "\ic"
vowel3$ = "\ef"
vowel4$ = "ae"
vowel5$ = "\vt"
vowel6$ = "\hs"
vowel7$ = "u"
f1min = 250
f1max = 850
f2min = 1100
f2max = 2400
lineType1$ = "Solid line"
lineType2$ = "Dashed line"
fontSize1 = 15
fontSize2 = 12
fontColor1$ = "Black"
fontColor2$ = "Blue"
letterType1$ = "##"
letterType2$ = ""

Erase all
Viewport... 0 6 0 4
Axes... log10(f2max) log10(f2min) log10(f1max) log10(f1min)
Line width... 1
One logarithmic mark bottom... 1100 yes yes yes
One logarithmic mark bottom... 1200 yes yes yes
One logarithmic mark bottom... 1350 yes yes yes
One logarithmic mark bottom... 1500 yes yes yes
One logarithmic mark bottom... 1700 yes yes yes
One logarithmic mark bottom... 1900 yes yes yes
One logarithmic mark bottom... 2100 yes yes yes
One logarithmic mark bottom... 2300 yes yes yes
One logarithmic mark left... 300 yes yes yes
One logarithmic mark left... 450 yes yes yes
```

```
One logarithmic mark left... 600 yes yes yes
One logarithmic mark left... 750 yes yes yes
#One logarithmic mark left... 900 yes yes yes
#One logarithmic mark left... 1000 yes yes yes
Text bottom... yes F2 (Hz)
Text left... yes F1 (Hz)
Draw inner box
for dialect to 2
	dialect$ = dialect'dialect'$
	select table280
	table140 = Extract rows where column (text)... dialect "is equal to" 'dialect$'
	for gender to 1
		gender$ = gender'gender'$
		select table140
		table70 = Extract rows where column (text)... gender "is equal to"
'gender$'
		numberOfRows = Get number of rows
		#assert numberOfRows = 70
		for vowel to 7
			vowel$ = vowel'vowel'$
			select table70
			Extract rows where column (text)... vowel "is equal to" 'vowel$'
			numberOfRows = Get number of rows
			#assert numberOfRows = 10
			for iformant to 2
				f'iformant'_'vowel'_'dialect'_'gender' = Get mean...
F'iformant'
			endfor
			Remove
		endfor
		select table70
		Remove
	endfor
	select table140
	Remove
endfor
Line width... 2
for dialect to 2
	lineType$ = lineType'dialect'$
	'lineType$'

	fontColor$ = fontColor'dialect'$
	'fontColor$'


	for gender to 1
		for vowel to 6
			nextVowel = vowel + 1
			Draw line... log10(f2_'vowel'_'dialect'_'gender')
log10(f1_'vowel'_'dialect'_'gender')
```

```
			... log10(f2_'nextVowel'_'dialect'_'gender')
log10(f1_'nextVowel'_'dialect'_'gender')
		endfor
	endfor
endfor
Line width... 1
Solid line
for dialect to 2
	for gender to 1
		for vowel to 7
			Paint circle (mm)... white log10(f2_'vowel'_'dialect'_'gender')
log10(f1_'vowel'_'dialect'_'gender') 3
		endfor
	endfor
endfor
for dialect to 2
	letterType$ = letterType'dialect'$
	for gender to 1
		for vowel to 7
			vowel$ = vowel'vowel'$
			Text special... log10(f2_'vowel'_'dialect'_'gender') Centre
			... log10(f1_'vowel'_'dialect'_'gender')-0.003 Half Times
fontSize'gender' 0 'letterType$''vowel$'
		endfor
	endfor
endfor


for dialect to 2
	fontColor$ = fontColor'dialect'$
	'fontColor$'

	for gender to 1
		for vowel to 7
			vowel$ = vowel'vowel'$
			Text special... log10(f2_'vowel'_'dialect'_'gender') Centre
			... log10(f1_'vowel'_'dialect'_'gender')-0.003 Half Times
fontSize'gender' 0 'letterType$''vowel$'
		endfor
	endfor
endfor
```

Appendix AG

Identification Tests' Scores per Participant

Participant	Pretest			Posttest			Delayed Posttest			Generalization Test		
	/ æ- ɛ /	**/ i - ɪ /**	**/u - ʊ /**	**/ æ- ɛ /**	**/ i - ɪ /**	**/u - ʊ /**	**/ æ- ɛ /**	**/ i - ɪ /**	**/u - ʊ /**	**/ æ- ɛ /**	**/ i - ɪ /**	**/u - ʊ /**
E1	66.67	63.34	76.67	83.34	98.34	98.34	91.67	100	98.33	83.34	97.22	86.11
E2	71.67	50	36.67	93.34	95	51.67	91.67	95	67.5	100	91.67	50
E3	61.67	80	73.34	70	90	88.34	75	98.34	95.84	83.34	91.67	91.67
E4	75	41.67	76.67	78.34	83.34	80	70	90	87.5	91.67	88.89	66.67
E5	71.67	45	61.67	81.67	88.34	81.67	71.67	83.34	85.84	91.67	91.67	80.56
E6	71.67	60	70	83.34	78.33	78.34	no data	no data	no data	94.45	80.56	38.89
E7	83.34	65	55	86.67	78.34	68.34	90	93.34	83.34	97.22	88.89	63.89
E8	56.67	43.34	50	68.34	85	63.34	70	83.34	69.17	80.56	94.45	58.34
E9	56.67	45	55	88.33	85	91.67	95	91.67	94.17	100	97.22	72.22
E10	66.67	45	68.34	81.67	95	91.67	80	95	88.33	88.89	100	75
E11	56.67	23.34	65	66.67	100	78.33	91.67	98.34	94.17	77.78	97.22	94.45
E12	66.67	38.33	60	88.34	56.67	78.33	85	63.33	73.33	86.11	80.56	72.23
E13	63.33	88.34	68.34	80	98.34	85	90	100	93.34	94.45	91.67	83.34
E14	60	48.34	45	91.67	88.33	91.67	86.67	96.67	93.34	94.45	100	86.11
E15	51.67	56.67	43.34	71.67	51.67	51.67	no data	no data	no data	86.11	75	38.89
E16	80	70	60	91.67	96.67	80	no data	no data	no data	83.34	91.67	86.11
E17	61.67	46.67	71.67	70	85	81.67	78.33	85	84.17	91.67	91.67	66.67
E18	68.33	66.67	51.67	68.34	80	61.67	70	73.33	70.83	80.56	83.33	44.45
E19	73.34	61.67	70	88.34	68.34	71.67	88.34	78.33	77.5	91.67	80.56	50
E20	73.34	41.67	80	83.34	63.34	81.67	95	70	75.83	97.22	69.45	58.33
E21	65	71.67	75	91.67	78.34	68.34	93.34	71.67	74.17	91.67	83.33	55.56
E22	63.33	95	90	85	98.34	80	78.33	98.34	91.67	100	97.22	63.89
C23	85	91.67	88.34	85	95	93.33	83.34	100	100	94.44	97.22	91.67
C24	60	50	65	63.34	53.34	73.34	no data	no data	no data	47.23	50	41.67
C25	56.67	70	40	70	73.33	16.67	73.33	61.67	40	72.23	61.11	25
C26	68.34	53.34	66.67	71.67	55	71.67	70	56.67	63.34	75	52.78	66.67
C27	81.67	68.34	50	73.34	66.67	48.34	70	43.33	47.5	86.11	50	38.89
C28	75	73.34	63.34	78.33	70	50	81.67	68.34	48.33	69.44	72.22	47.22
C29	56.67	98.34	78.33	63.34	96.67	81.67	58.34	96.67	89.17	75	83.34	88.89
C30	78.33	65	70	91.67	80	26.67	85	78.34	59.17	80.56	94.44	55.56
C31	78.34	61.67	76.67	55	76.67	75	66.67	66.67	73.34	55.56	58.34	75
C32	73.34	68.34	28.33	75	56.67	45	73.34	63.34	60	72.22	66.67	55.56
C33	70	48.34	45	75	45	51.67	83.34	38.34	43.33	75	58.34	41.67
C34	65	50	81.67	60	61.67	81.67	61.67	61.67	71.67	66.67	72.23	58.34

Appendix AH

Identification Tests' Scores and Percentage Differences per Participant

Table 1

Identification Pretest and Posttest Scores and % Differences per Participant

Participant	Pretest / æ- ɛ /	Posttest / æ- ɛ /	% Dif	Pretest / i - ɪ /	Posttest / i - ɪ /	% Dif	Pretest /u - ʊ /	Posttest /u - ʊ /	% Dif
E1	66.67	83.34	16.67	63.34	98.34	35	76.67	98.34	21.67
E2	71.67	93.34	21.67	50	95	45	36.67	51.67	15
E3	61.67	70	8.33	80	90	10	73.34	88.34	15
E4	75	78.34	3.34	41.67	83.34	41.67	76.67	80	3.33
E5	71.67	81.67	10	45	88.34	43.34	61.67	81.67	20
E6	71.67	83.34	11.67	60	78.33	18.33	70	78.34	8.34
E7	83.34	86.67	3.33	65	78.34	13.34	55	68.34	13.34
E8	56.67	68.34	11.67	43.34	85	41.66	50	63.34	13.34
E9	56.67	88.33	31.66	45	85	40	55	91.67	36.67
E10	66.67	81.67	15	45	95	50	68.34	91.67	23.33
E11	56.67	66.67	10	23.34	100	76.66	65	78.33	13.33
E12	66.67	88.34	21.67	38.33	56.67	18.34	60	78.33	18.33
E13	63.33	80	16.67	88.34	98.34	10	68.34	85	16.66
E14	60	91.67	31.67	48.34	88.33	39.99	45	91.67	46.67
E15	51.67	71.67	20	56.67	51.67	-5	43.34	51.67	8.33
E16	80	91.67	11.67	70	96.67	26.67	60	80	20
E17	61.67	70	8.33	46.67	85	38.33	71.67	81.67	10
E18	68.33	68.34	0.01	66.67	80	13.33	51.67	61.67	10
E19	73.34	88.34	15	61.67	68.34	6.67	70	71.67	1.67
E20	73.34	83.34	10	41.67	63.34	21.67	80	81.67	1.67
E21	65	91.67	26.67	71.67	78.34	6.67	75	68.34	-6.66
E22	63.33	85	21.67	95	98.34	3.34	90	80	-10
C23	85	85	0	91.67	95	3.33	88.34	93.33	4.99
C24	60	63.34	3.34	50	53.34	3.34	65	73.34	8.34
C25	56.67	70	13.33	70	73.33	3.33	40	16.67	-23.33
C26	68.34	71.67	3.33	53.34	55	1.66	66.67	71.67	5
C27	81.67	73.34	-8.33	68.34	66.67	-1.67	50	48.34	-1.66
C28	75	78.33	3.33	73.34	70	-3.34	63.34	50	-13.34
C29	56.67	63.34	6.67	98.34	96.67	-1.67	78.33	81.67	3.34
C30	78.33	91.67	13.34	65	80	15	70	26.67	-43.33
C31	78.34	55	-23.34	61.67	76.67	15	76.67	75	-1.67
C32	73.34	75	1.66	68.34	56.67	-11.67	28.33	45	16.67
C33	70	75	5	48.34	45	-3.34	45	51.67	6.67
C34	65	60	-5	50	61.67	11.67	81.67	81.67	0

Table 2
Identification Posttest and Delayed Posttest Scores and % Differences per Participant

Participant	Posttest / æ- ɛ /	Delayed Post / æ- ɛ /	% Dif	Posttest / i - ɪ /	Delayed Post / i - ɪ /	% Dif	Posttest /u - ʊ /	Delayed Post /u - ʊ /	% Dif
E1	83.34	91.67	8.33	98.34	100	1.66	98.34	98.33	-0.01
E2	93.34	91.67	-1.67	95	95	0	51.67	67.5	15.83
E3	70	75	5	90	98.34	8.34	88.34	95.84	7.5
E4	78.34	70	-8.34	83.34	90	6.66	80	87.5	7.5
E5	81.67	71.67	-10	88.34	83.34	-5	81.67	85.84	4.17
E6	83.34	no data	no data	78.33	no data	no data	78.34	no data	no data
E7	86.67	90	3.33	78.34	93.34	15	68.34	83.34	15
E8	68.34	70	1.66	85	83.34	-1.66	63.34	69.17	5.83
E9	88.33	95	6.67	85	91.67	6.67	91.67	94.17	2.5
E10	81.67	80	-1.67	95	95	0	91.67	88.33	-3.34
E11	66.67	91.67	25	100	98.34	-1.66	78.33	94.17	15.84
E12	88.34	85	-3.34	56.67	63.33	6.66	78.33	73.33	-5
E13	80	90	10	98.34	100	1.66	85	93.34	8.34
E14	91.67	86.67	-5	88.33	96.67	8.34	91.67	93.34	1.67
E15	71.67	no data	no data	51.67	no data	no data	51.67	no data	no data
E16	91.67	no data	no data	96.67	no data	no data	80	no data	no data
E17	70	78.33	8.33	85	85	0	81.67	84.17	2.5
E18	68.34	70	1.66	80	73.33	-6.67	61.67	70.83	9.16
E19	88.34	88.34	0	68.34	78.33	9.99	71.67	77.5	5.83
E20	83.34	95	11.66	63.34	70	6.66	81.67	75.83	-5.84
E21	91.67	93.34	1.67	78.34	71.67	-6.67	68.34	74.17	5.83
E22	85	78.33	-6.67	98.34	98.34	0	80	91.67	11.67
C23	85	83.34	-1.66	95	100	5	93.33	100	6.67
C24	63.34	no data	no data	53.34	no data	no data	73.34	no data	no data
C25	70	73.33	3.33	73.33	61.67	-11.66	16.67	40	23.33
C26	71.67	70	-1.67	55	56.67	1.67	71.67	63.34	-8.33
C27	73.34	70	-3.34	66.67	43.33	-23.34	48.34	47.5	-0.84
C28	78.33	81.67	3.34	70	68.34	-1.66	50	48.33	-1.67
C29	63.34	58.34	-5	96.67	96.67	0	81.67	89.17	7.5
C30	91.67	85	-6.67	80	78.34	-1.66	26.67	59.17	32.5
C31	55	66.67	11.67	76.67	66.67	-10	75	73.34	-1.66
C32	75	73.34	-1.66	56.67	63.34	6.67	45	60	15
C33	75	83.34	8.34	45	38.34	-6.66	51.67	43.33	-8.34
C34	60	61.67	1.67	61.67	61.67	0	81.67	71.67	-10

Table 3

Identification Posttest and Generalization Test Scores and % Differences per Participant

Participant	Posttest **/ æ- ɛ /**	GT **/ æ- ɛ /**	**% Dif**	Posttest **/ i - ɪ /**	GT **/ i - ɪ /**	D **% Dif**	Posttest **/u - ʊ /**	GT **/u - ʊ /**	**% Dif**
E1	83.34	83.34	0	98.34	97.22	-1.12	98.34	86.11	-12.23
E2	93.34	100	6.66	95	91.67	-3.33	51.67	50	-1.67
E3	70	83.34	13.34	90	91.67	1.67	88.34	91.67	3.33
E4	78.34	91.67	13.33	83.34	88.89	5.55	80	66.67	-13.33
E5	81.67	91.67	10	88.34	91.67	3.33	81.67	80.56	-1.11
E6	83.34	94.45	11.11	78.33	80.56	2.23	78.34	38.89	-39.45
E7	86.67	97.22	10.55	78.34	88.89	10.55	68.34	63.89	-4.45
E8	68.34	80.56	12.22	85	94.45	9.45	63.34	58.34	-5
E9	88.33	100	11.67	85	97.22	12.22	91.67	72.22	-19.45
E10	81.67	88.89	7.22	95	100	5	91.67	75	-16.67
E11	66.67	77.78	11.11	100	97.22	-2.78	78.33	94.45	16.12
E12	88.34	86.11	-2.23	56.67	80.56	23.89	78.33	72.23	-6.1
E13	80	94.45	14.45	98.34	91.67	-6.67	85	83.34	-1.66
E14	91.67	94.45	2.78	88.33	100	11.67	91.67	86.11	-5.56
E15	71.67	86.11	14.44	51.67	75	23.33	51.67	38.89	-12.78
E16	91.67	83.34	-8.33	96.67	91.67	-5	80	86.11	6.11
E17	70	91.67	21.67	85	91.67	6.67	81.67	66.67	-15
E18	68.34	80.56	12.22	80	83.33	3.33	61.67	44.45	-17.22
E19	88.34	91.67	3.33	68.34	80.56	12.22	71.67	50	-21.67
E20	83.34	97.22	13.88	63.34	69.45	6.11	81.67	58.33	-23.34
E21	91.67	91.67	0	78.34	83.33	4.99	68.34	55.56	-12.78
E22	85	100	15	98.34	97.22	-1.12	80	63.89	-16.11
C23	85	94.44	9.44	95	97.22	2.22	93.33	91.67	-1.66
C24	63.34	47.23	-16.11	53.34	50	-3.34	73.34	41.67	-31.67
C25	70	72.23	2.23	73.33	61.11	-12.22	16.67	25	8.33
C26	71.67	75	3.33	55	52.78	-2.22	71.67	66.67	-5
C27	73.34	86.11	12.77	66.67	50	-16.67	48.34	38.89	-9.45
C28	78.33	69.44	-8.89	70	72.22	2.22	50	47.22	-2.78
C29	63.34	75	11.66	96.67	83.34	-13.33	81.67	88.89	7.22
C30	91.67	80.56	-11.11	80	94.44	14.44	26.67	55.56	28.89
C31	55	55.56	0.56	76.67	58.34	-18.33	75	75	0
C32	75	72.22	-2.78	56.67	66.67	10	45	55.56	10.56
C33	75	75	0	45	58.34	13.34	51.67	41.67	-10
C34	60	66.67	6.67	61.67	72.23	10.56	81.67	58.34	-23.33

Appendix AI

Normalized F1 and F2 values of AE Vowels

Table 1

Normalized Mean and Median F1 and F2 Values of AE Vowels Produced at Pretest

	Vowel	**/i/**		**/ɪ/**		**/ɛ/**		**/æ/**		**/u/**		**/ʊ/**		**/ʌ/**	
		Mean (SD)	**Mdn**	**Mean (SD)**	**Mdn**	**Mean (SD)**	**Mdn**	**Mean (SD)**	**Mdn**	**Mean (SD)**	**Mdn**	**Mean (SD)**	**Mdn**	**Mean (SD)**	**Mdn**
EG	**F1 (Hz)**	295	291	319	305	594	595	629	634	321	314	345	332	559	559
	F2 (Hz)	2094	2108	2013	2050	1770	1765	1739	1748	1286	1241	1232	1217	1508	1523
		/i/		**/ɪ/**		**/ɛ/**		**/æ/**		**/u/**		**/ʊ/**		**/ʌ/**	
		Mean (SD)	**Mdn**	**Mean (SD)**	**Mdn**	**Mean (SD)**	**Mdn**	**Mean (SD)**	**Mdn**	**Mean (SD)**	**Mdn**	**Mean (SD)**	**Mdn**	**Mean (SD)**	**Mdn**
CG	**F1 (Hz)**	297	295	334	307	599	609	629	639	340	320	351	335	590	590
	F2 (Hz)	2096	2107	2006	2042	1776	1783	1745	1751	1254	1234	1242	1220	1547	1545

Table 2

Normalized Mean and Median F1 and F2 Values of AE Vowels Produced at Posttest

	Vowel	**/i/**		**/ɪ/**		**/ɛ/**		**/æ/**		**/u/**		**/ʊ/**		**/ʌ/**	
		Mean (SD)	**Mdn**	**Mean (SD)**	**Mdn**	**Mean (SD)**	**Mdn**	**Mean (SD)**	**Mdn**	**Mean (SD)**	**Mdn**	**Mean (SD)**	**Mdn**	**Mean (SD)**	**Mdn**
EG	**F1 (Hz)**	292	283	383	379	573	570	630	630	320	314	399	404	543	544
	F2 (Hz)	2097	2108	1878	1874	1749	1752	1702	1708	1273	1255	1277	1256	1486	1498
		/i/		**/ɪ/**		**/ɛ/**		**/æ/**		**/u/**		**/ʊ/**		**/ʌ/**	
		Mean (SD)	**Mdn**	**Mean (SD)**	**Mdn**	**Mean (SD)**	**Mdn**	**Mean (SD)**	**Mdn**	**Mean (SD)**	**Mdn**	**Mean (SD)**	**Mdn**	**Mean (SD)**	**Mdn**
CG	**F1 (Hz)**	301	294	333	323	595	602	628	631	320	310	345	348	551	560
	F2 (Hz)	2097	2106	2015	2042	1785	1813	1731	1758	1287	1272	1251	1247	1546	1552

Table 3
Normalized Mean and Median F1 and F2 Values of AE Vowels Produced at Delayed Posttest

Delayed Posttest	**Vowel**	**/i/**		**/ɪ/**		**/ɛ/**		**/æ/**		**/u/**		**/ʊ/**		**/ʌ/**	
		Mean (SD)	**Mdn**	**Mean (SD)**	**Mdn**	**Mean (SD)**	**Mdn**	**Mean (SD)**	**Mdn**	**Mean (SD)**	**Mdn**	**Mean (SD)**	**Mdn**	**Mean (SD)**	**Mdn**
EG	**F1 (Hz)**	289	283	396	409	571	571	632	639	322	315	408	421	543	536
	F2 (Hz)	2106	2113	1874	1856	1761	1765	1709	1721	1271	1241	1277	1265	1497	1524
		/i/		**/ɪ/**		**/ɛ/**		**/æ/**		**/u/**		**/ʊ/**		**/ʌ/**	
		Mean (SD)	**Mdn**	**Mean (SD)**	**Mdn**	**Mean (SD)**	**Mdn**	**Mean (SD)**	**Mdn**	**Mean (SD)**	**Mdn**	**Mean (SD)**	**Mdn**	**Mean (SD)**	**Mdn**
CG	**F1 (Hz)**	301	293	355	344	590	600	637	638	345	327	377	380	558	565
	F2 (Hz)	2101	2110	2003	2019	1764	1772	1722	1713	1267	1232	1250	1235	1543	1544

Appendix AJ

Plots of Normalized Mean F1 and F2 Values of AE Vowels

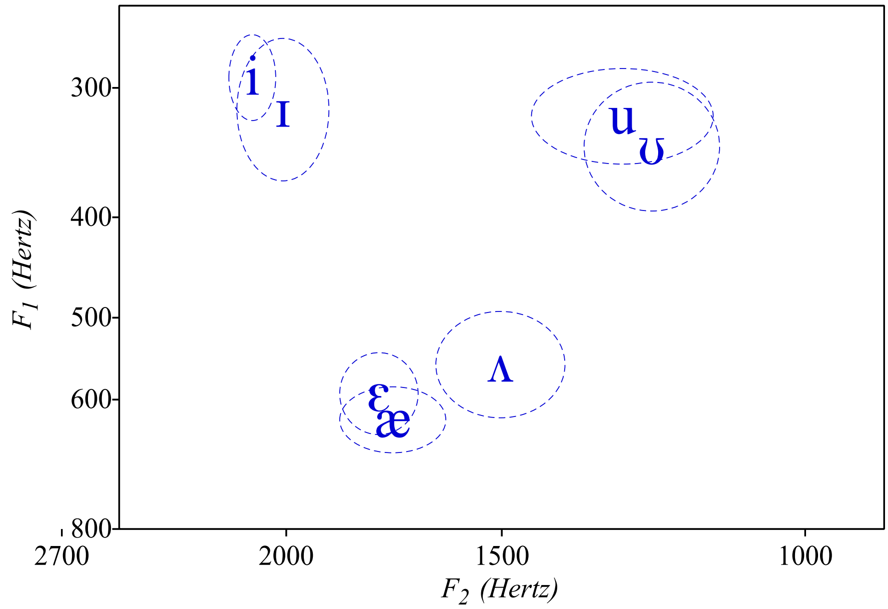


Figure 1. Vowel space of the experimental group at pretest.

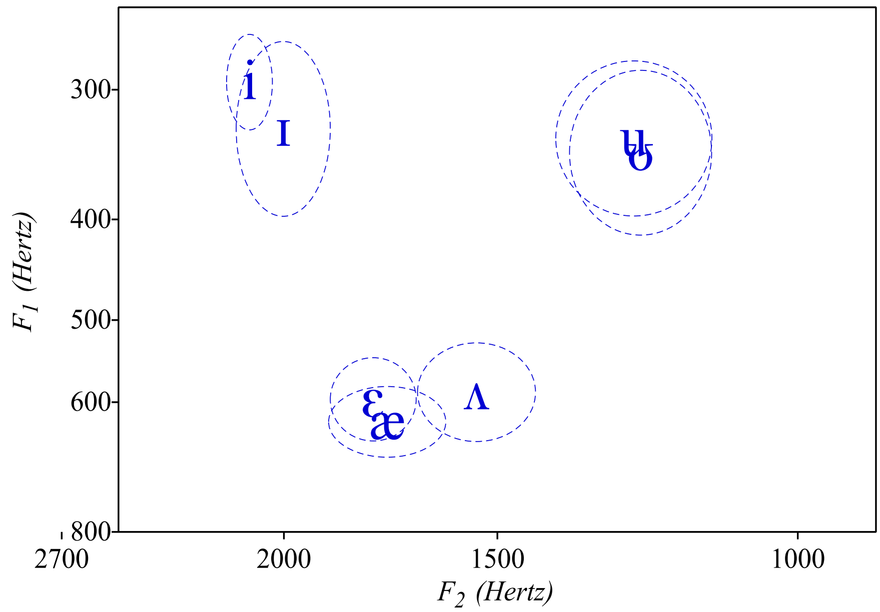


Figure 2. Vowel space of the control group at pretest.

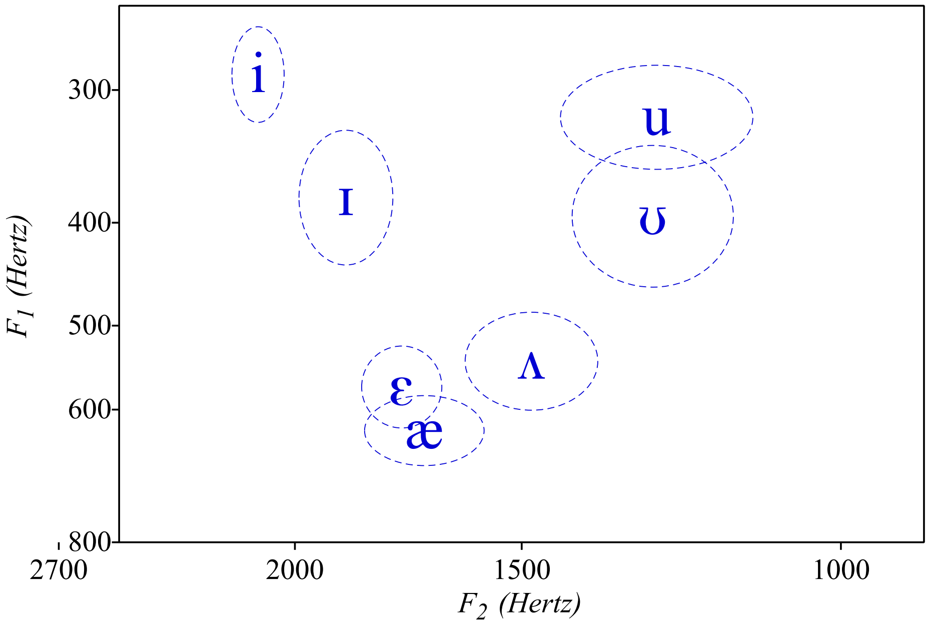


Figure 3. Vowel space of the experimental group at posttest.

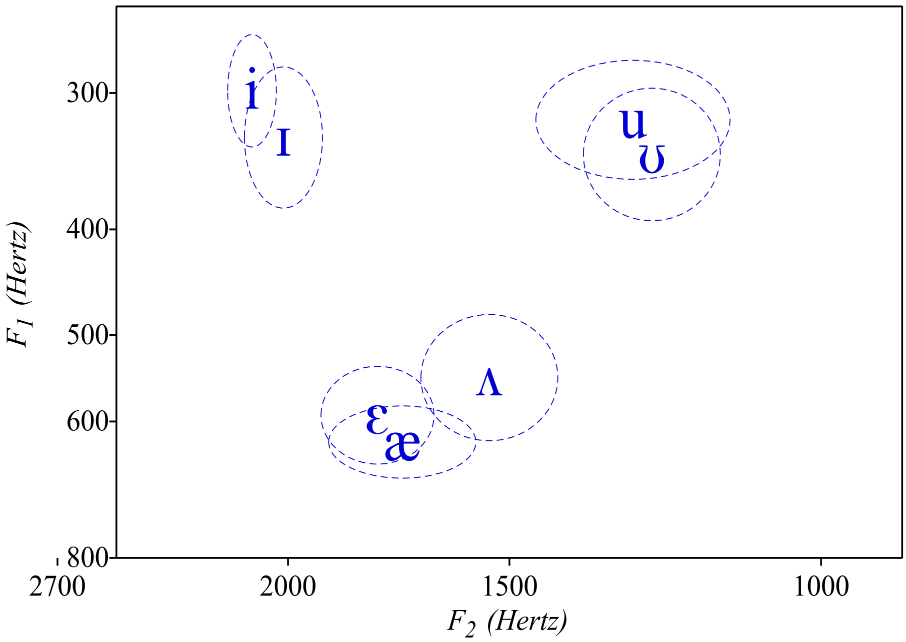


Figure 4. Vowel space of the control group at posttest.

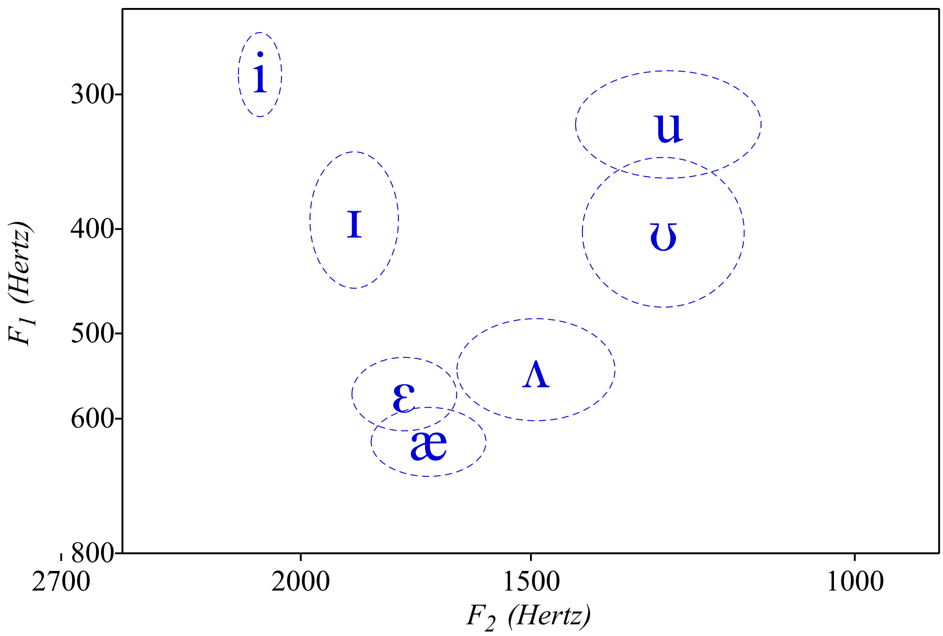


Figure 5. Vowel space of the experimental group at delayed posttest.

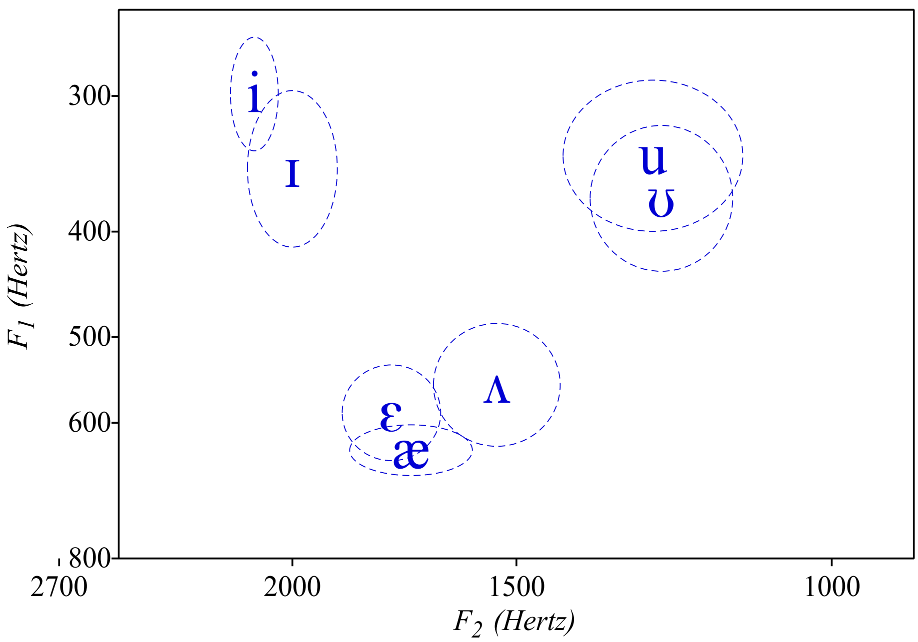


Figure 6. Vowel space of the control group at delayed posttest.

Appendix AK

Euclidean Distances and Duration Ratios per L2 Participant

Table 1

Individual Euclidean Distances

	Pretest			Posttest			Delayed Posttest		
Participant	**/ æ- ɛ /**	**/ i - ɪ /**	**/u - ʊ /**	**/ æ- ɛ /**	**/ i - ɪ /**	**/u - ʊ /**	**/ æ- ɛ /**	**/ i - ɪ /**	**/u - ʊ /**
E1	22.48	40.31	108.02	47.67	344.35	135.36	46.34	308.64	146.12
E2	133.60	297.04	131.33	101.04	314.66	96.14	138.51	164.12	17.13
E3	19.85	25.03	28.73	44.36	57.66	72.86	265.74	245.63	245.61
E4	11.77	62.61	22.80	22.98	141.37	58.18	94.52	260.31	237.66
E5	15.19	68.08	3.99	22.97	244.16	35.20	56.38	313.80	159.66
E6	52.63	20.69	7.34	219.44	111.52	25.93	no data	no data	no data
E7	27.73	8.37	90.98	63.98	337.53	199.05	56.90	222.65	115.45
E8	338.90	316.34	184.40	332.72	222.71	229.24	81.04	324.24	220.97
E9	38.30	7.57	157.87	123.22	280.06	75.40	98.34	314.31	40.87
E10	45.48	95.57	163.03	56.51	292.31	153.81	44.84	317.53	179.87
E11	29.75	19.22	9.44	18.22	177.68	114.32	221.50	177.27	281.91
E12	45.05	284.53	76.52	71.37	302.41	136.79	115.59	309.33	104.62
E13	25.70	20.65	85.33	90.95	316.36	237.57	79.52	421.43	169.63
E14	89.72	184.32	76.42	114.48	271.05	68.56	147.06	327.67	293.88
E15	43.09	44.22	99.00	117.56	191.71	28.11	no data	no data	no data
E16	18.67	45.26	86.41	87.23	311.56	60.37	no data	no data	no data
E17	8.84	36.07	64.33	20.56	283.75	243.03	51.18	198.74	96.08
E18	137.37	9.01	16.00	115.44	124.65	383.16	21.20	208.53	110.82
E19	30.33	288.81	67.00	38.95	404.30	138.53	34.77	195.94	85.66
E20	29.40	20.53	57.78	35.52	150.83	75.22	73.54	350.94	190.50
E21	15.80	56.98	42.89	20.88	75.55	71.10	66.04	266.66	94.01
E22	23.92	217.46	74.31	130.39	323.00	176.65	32.29	18.35	67.33
C23	20.84	28.86	32.78	39.69	66.64	74.18	131.07	235.87	76.59
C24	39.25	289.48	23.71	166.43	206.54	102.36	no data	no data	no data
C25	89.90	380.55	103.97	67.47	309.90	137.20	82.68	358.50	128.53
C26	236.43	206.17	58.49	298.56	221.96	100.52	265.04	223.48	100.69
C27	11.46	103.73	12.68	14.50	32.09	69.69	39.14	132.76	89.38
C28	49.28	30.32	79.98	14.13	18.18	61.79	35.37	29.49	52.56
C29	13.72	11.4	36.31	26.20	33.33	39.16	31.75	30.60	22.56
C30	6.76	40.53	37.17	33.30	33.91	30.46	22.37	53.33	49.05
C31	40.94	72.21	12.51	63.59	39.92	44.50	39.14	74.66	37.56
C32	38.04	16.14	16.79	19.78	12.93	15.72	140.92	40.16	78.49
C33	132.31	27.49	49.17	155.39	68.89	22.93	25.14	43.45	18.98
C34	33.59	3.24	34.64	21.41	41.53	56.29	12.58	36.32	33.64

Table 2
Individual Duration Ratios

Participant	Pretest			Posttest			Delayed Posttest		
	/ æ- ɛ /	**/ i - ɪ /**	**/u - ʊ /**	**/ æ- ɛ /**	**/ i - ɪ /**	**/u - ʊ /**	**/ æ- ɛ /**	**/ i - ɪ /**	**/u - ʊ /**
E1	1.03	1.05	1.11	1.10	0.93	1.03	1.07	0.95	0.94
E2	1.01	0.98	0.98	1.12	0.97	1.01	1.09	0.95	0.97
E3	1.14	1.10	0.98	1.35	1.02	1.02	1.44	1.02	1.05
E4	1.06	0.99	1.03	1.07	0.90	1.05	1.06	0.92	1.16
E5	1.05	0.95	0.98	0.97	0.90	0.96	1.02	0.92	0.89
E6	1.05	1.08	1.12	1.11	1.06	1.03	no data	no data	no data
E7	1.06	1.02	1.02	0.96	1.00	1.07	1.09	0.99	0.99
E8	0.97	1.19	1.05	1.10	0.98	0.94	1.02	0.95	0.90
E9	1.05	1.14	1.08	1.09	1.16	1.07	1.14	1.14	0.85
E10	1.04	1.05	1.01	1.05	1.23	1.14	1.20	0.96	1.18
E11	0.96	1.07	1.09	1.06	0.92	1.04	1.04	0.94	1.00
E12	1.04	1.03	1.02	1.10	0.88	1.03	1.06	0.87	1.00
E13	1.02	0.94	1.08	1.02	0.92	1.12	1.03	0.99	1.19
E14	1.07	1.07	1.15	1.01	1.02	1.08	0.99	1.20	1.06
E15	1.07	1.08	1.25	0.95	1.03	1.13	no data	no data	no data
E16	1.10	0.86	1.00	0.99	0.84	0.97	no data	no data	no data
E17	1.03	1.00	1.04	1.03	0.78	0.87	0.96	0.84	0.85
E18	1.03	1.09	1.10	0.98	1.00	0.93	1.02	1.12	1.01
E19	0.96	1.02	1.08	0.99	1.03	0.98	1.05	0.98	0.96
E20	1.06	1.06	1.03	1.04	0.90	0.92	1.04	0.95	0.88
E21	1.09	1.03	1.07	1.08	0.85	1.11	1.15	0.90	1.02
E22	1.11	1.06	1.14	1.10	1.17	1.28	1.12	1.07	0.96
C23	0.99	0.97	1.03	1.03	0.86	0.99	1.18	0.86	0.95
C24	1.00	1.05	1.15	1.01	1.19	1.16	no data	no data	no data
C25	1.03	1.11	0.97	0.96	1.02	1.05	0.96	1.05	1.09
C26	1.02	1.01	0.96	0.97	1.07	1.06	0.99	0.99	1.10
C27	1.03	1.09	1.07	0.99	1.06	1.02	0.90	1.00	1.09
C28	1.03	1.04	1.13	1.05	1.06	1.04	1.00	0.96	1.14
C29	1.03	1.05	1.22	1.26	0.87	0.98	1.11	0.93	1.02
C30	1.13	1.01	0.97	1.07	1.11	1.08	1.04	1.01	1.09
C31	1.00	1.07	1.07	1.00	1.07	0.93	0.95	1.10	1.04
C32	0.98	1.00	1.16	1.04	1.03	1.13	0.98	1.03	1.03
C33	1.06	1.05	1.02	1.02	1.07	1.00	1.01	1.03	1.00
C34	1.02	1.14	1.02	1.07	1.06	1.08	1.00	1.15	1.14

Appendix AL

Summary of the Statistical Descriptive Data of the Follow-up Questionnaire

	Mo	**Percent (%)**	**Label**
1. Easiest vowel contrast	1	54.5	/i/-/ɪ/
2. Most difficult vowel contrast	3	54.5	/u/-/ʊ/
3. Easiest vowel	7	45.5	/ʌ/
4. Most difficult vowel	6	54.5	/ʊ/
5. Easiest perception task	1	63.6	discrimination
6. Duration of training sessions	2	81.8	adequate
7. Duration of training program	2	81.8	sufficient
8. Tiredness	2	50	no
9. Motivation	3	68.2	motivated
10. Concentration	3	72.7	concentrated

Note. *Mo*=Mode.

Appendix AM

Individual Answers of the Follow-up Questionnaire

Part.	1	2	3	4	5	6	7	8	8.1	9	10
E1	[i]-[ɪ]	[æ]-[ɛ]	[i]	[æ]	DISCR	adequate	sufficient	yes	not much	motivated	concentrated
E2	[i]-[ɪ]	[u]-[ʊ]	[i]	[ʊ]	DISCR	adequate	sufficient	no	n/a	motivated	concentrated
E3	[i]-[ɪ]	[æ]-[ɛ]	[ʌ]	[ɛ]	ID	adequate	sufficient	no	n/a	highly motivated	concentrated
E4	[i]-[ɪ]	[u]-[ʊ]	[i]	[ʊ]	DISCR	adequate	sufficient	no	n/a	motivated	concentrated
E5	[u]-[ʊ]	[i]-[ɪ]	[ʌ]	[æ]	ID	adequate	sufficient	no	n/a	motivated	concentrated
E6	[æ]-[ɛ]	[u]-[ʊ]	[æ]	[ʊ]	DISCR	adequate	sufficient	no	n/a	motivated	concentrated
E7	[æ]-[ɛ]	[u]-[ʊ]	[ʌ]	[ʊ]	DISCR	adequate	sufficient	no	n/a	motivated	highly concentrated
E8	[æ]-[ɛ]	[u]-[ʊ]	[æ]	[ʊ]	ID	short	insufficient	yes	not much	highly motivated	concentrated
E9	[u]-[ʊ]	[æ]-[ɛ]	[ʌ]	[æ]	ID	adequate	sufficient	yes	not much	highly motivated	concentrated
E10	[i]-[ɪ]	[u]-[ʊ]	[i]	[ʊ]	DISCR	adequate	insufficient	yes	very	motivated	concentrated
E11	[i]-[ɪ]	[æ]-[ɛ]	[i]	[æ]	DISCR	adequate	sufficient	yes	not much	motivated	concentrated
E12	[i]-[ɪ]	[u]-[ʊ]	[ʌ]	[ʊ]	DISCR	adequate	sufficient	no	n/a	motivated	highly concentrated
E13	[i]-[ɪ]	[u]-[ʊ]	[ʌ]	[ʊ]	ID	adequate	sufficient	no	n/a	motivated	concentrated
E14	[i]-[ɪ]	[æ]-[ɛ]	[ʌ]	[æ]	DISCR	long	sufficient	yes	some	motivated	not much concentrated
E15	n/d	n/d	n/d	n/d	n/d	n/d	n/d	n/d	n/d	n/d	n/d
E16	n/d	n/d	n/d	n/d	n/d	n/d	n/d	n/d	n/d	n/d	n/d
E17	[i]-[ɪ]	[æ]-[ɛ]	[ʌ]	[ɛ]	DISCR	adequate	sufficient	yes	not much	highly motivated	concentrated
E18	[æ]-[ɛ]	[u]-[ʊ]	[ʌ]	[ʊ]	DISCR	adequate	sufficient	yes	some	highly motivated	highly concentrated
E19	[æ]-[ɛ]	[u]-[ʊ]	[ʌ]	[ʊ]	ID	adequate	sufficient	no	n/a	motivated	concentrated
E20	[æ]-[ɛ]	[u]-[ʊ]	[æ]	[ʊ]	DISCR	adequate	sufficient	yes	not much	motivated	concentrated
E21	[i]-[ɪ]	[u]-[ʊ]	[i]	[ʊ]	DISCR	adequate	sufficient	no	n/a	motivated	concentrated
E22	[i]-[ɪ]	[æ]-[ɛ]	[i]	[ʌ]	DISCR	adequate	sufficient	no	n/a	motivated	concentrated

Notes. Part = Participant. n/d=no data

1. Which one was the easiest vowel contrast to distinguish in the training? a) [i]-[ɪ] *heed-hid* b) [æ]-[ɛ] *had-head* c) [u]-[ʊ] *who'd-hood*

2. Which one was the most difficult vowel contrast to distinguish in the training? a) [i]-[ɪ] *heed-hid* b) [æ]-[ɛ] *had-head* c) [u]-[ʊ] *who'd-hood*

3. Which was the easiest vowel to identify in the training? a) [i] *heed* b) [ɪ] *hid* c) [æ] *had* d) [ɛ] *head* e) [u] *who'd* f) [ʊ] *hood* g) [ʌ] *hud*

4. Which was the most difficult vowel to identify in the training? a) [i] *heed* b) [ɪ] *hid* c) [æ] *had* d) [ɛ] *head* e) [u] *who'd* f) [ʊ] *hood* g) [ʌ] *hud*

5. Which one was the easiest training task? (a) discrimination tasks b) identification tasks)

6. The duration of the training sessions was: a) long b) adequate c) short

7. To learn the seven English vowels, the five training sessions were: a) insufficient b) sufficient c) more than sufficient

8. At the end of the training sessions, did you feel tired? a) yes b) no

8.1. If you answered affirmatively, indicate the degree of tiredness: a) very tired b) tired c) not very tired

9. Indicate your degree of motivation during the training program: 1) unmotivated; 2) not very motivated; 3) motivated 4) highly motivated

10. Indicate your degree of concentration during the training program: 1) not concentrated 2) not much concentrated 3) concentrated; 4) highly concentrated.

Appendix AN

Median F1 and F2 Values of EP and AE Vowels

		L1 EP		L2 AE		L1 AE	
		/i/	/e/	/i/	/ɪ/	/i/	/ɪ/
Vowel		**Mdn**	**Mdn**	**Mdn**	**Mdn**	**Mdn**	**Mdn**
Duration (ms)	**F**	82	100	95	92	130	103
	M	73	92	99	96	140	118
F1 (Hz)	**F**	398	487	426	438	308	501
	M	327	427	332	367	276	423
F2 (Hz)	**F**	2526	2106	2560	2445	2766	2121
	M	2134	1746	2187	2050	2331	1884

		L1 EP		L2 AE		L1 AE	
		/ɛ/	/a/	/ɛ/	/æ/	/ɛ/	/æ/
Vowel		**Mdn**	**Mdn**	**Mdn**	**Mdn**	**Mdn**	**Mdn**
Duration (ms)	**F**	114	119	116	124	116	167
	M	101	98	114	119	134	179
F1 (Hz)	**F**	652	859	714	734	704	820
	M	554	694	575	581	582	671
F2 (Hz)	**F**	2101	1525	2061	2002	1910	1808
	M	1696	1324	1723	1731	1729	1669

		L1 EP		L2 AE		L1 AE	
		/o/	/u/	/u/	/ʊ/	/u/	/ʊ/
Vowel		**Mdn**	**Mdn**	**Mdn**	**Mdn**	**Mdn**	**Mdn**
Duration (ms)	**F**	107	87	99	98	127	114
	M	98	78	108	97	135	128
F1 (Hz)	**F**	512	443	422	461	335	540
	M	449	359	350	388	306	454
F2 (Hz)	**F**	1076	916	1165	1151	1782	1554
	M	1012	906	1140	1049	1556	1371

Note. L1 EP= Portuguese vowels produced by native Portuguese speakers; L2 AE= English vowels produced by Portuguese speakers (namely, by the participants in the experimental group, at pretest); L1 AE= English vowels spoken by native American English speakers.