Individual differences in second-language vowel learning

Angelos Lengeris

A thesis submitted in fulfilment of requirements for the degree of
Doctor of Philosophy

to
Department of Speech, Hearing and Phonetic Sciences
Division of Psychology and Language Sciences
University College London (UCL)

2009

Declaration

I confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Angelos Lengeris

Abstract

Adults often have difficulty in acquiring non-native vowels especially when the vowel inventories in first (L1) and second language (L2) are very different. However, even when testing L2 groups with similar profiles, there are great individual differences in the perception and production of non-native sounds. Similarly, computer-based training studies of L2 sounds report that improvement after training can range greatly across individuals. This thesis explores possible sources of individual differences in Greek native speakers' perception and production of Southern British English vowels.

Study 1 examined the perceived relationship between English vowels (in /bVb/ and /bVp/ contexts) and Greek vowels along with English vowel discrimination by the same participants. Greek speakers were found to perceive English vowels via both spectral and temporal assimilation to their L1 categories despite the fact that Greek does not use duration in L1 vowel distinctions. Study 2 defined the endpoints for the synthetic vowel continua to be used in Study 3 using a best exemplars experiment. In study 3, Greek speakers from a homogenous population (in terms of L1 background, age of L2 learning, amount and quality of L2 input) were tested on a large test battery before and after receiving 5 sessions of high-variability perceptual training. The test battery examined their perception of natural and synthetic vowels in L1 (Greek) and L2 (English) and their frequency discrimination ability (F2 only) as well as their production of L2 vowels. Group results showed significant improvement in the trainees' perception of natural L2 vowels and their L2 vowel production. However, large individual differences were evident both before and after training. Vowel processing in L2 was found to relate to individual variability in vowel processing in L1 and, importantly, to frequency discrimination acuity, a finding that favours an *auditory processing* hypothesis for L1 and L2 speech perception of vowels.

Acknowledgments

First of all I would like to sincerely thank my two supervisors, Valerie Hazan and Paul Iverson for their guidance, input and support throughout the course of my PhD. Valerie has always been there when needed and I cannot even count the times I dropped by her office asking for advice and help; no matter how busy she was, until the last minute before the submission of my thesis she managed somehow to find time to answer all my questions and to solve any problems. Paul kindly provided me not only with feedback at various stages of the thesis, but also with the speech materials and software he and Bronwen Evans have developed over the years to use in my experiments. Two out of the three studies in this thesis have greatly benefited from his and Bronwen's previous work on L2 vowel perception.

I should also like to thank a number of people in the Department of Speech, Hearing and Phonetic Sciences at the University College London. Stuart Rosen and Souhila Messaoud-Galusi for their help with the software used to run and analyze the synthetic speech perception tasks. Stephen Nevard and David Cushing for resolving many technical issues throughout the years, Michael Ashby for his valuable input in the selection of most of the English stimuli used in this thesis and Mark Huckvale for tutoring opportunities. Many thanks to Kayoko, Kota, Melanie, Sophie, Chierh, Jeong, Katharine and Sam for making the PhD room such a pleasant environment to work in.

This thesis would not have been completed without the financial support from the A.G. Leventis Foundation. I am also obliged to the UCL Graduate School, the UCL Old Students' Association Trust and the UCL Phonetics and Linguistics Department for funding my participation to various conferences and workshops.

Words cannot describe how thankful I am to my family whose love and support mean everything to me. We have been through a lot these last years and without their support on all possible levels I would not have been able to complete my thesis. This is a result of our combined efforts. Last but not least I would like to thank Evia. Not only she is responsible for my decision to study Phonetics but she made sure I stayed on the right track by being with me ever since!!

In memory of my dear father

Μπαμπά μου θα σε θυμάμαι για πάντα

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

This thesis is concerned with the acquisition of vowels in a second language (L2) by adult native speakers of another language. The perception and production of L2 sounds is not an easy task and several theories have been proposed to explain the difficulties a learner may be faced with when acquiring an L2. Well-documented factors in determining the success of L2 phoneme learning include the relationship between the segmental inventory of the first language (L1) and the L2 (e.g. Best, 1995; Flege, 1995a; Kuhl, 2000), the age of learning an L2 (e.g. Flege *et al.*, 1999a), the length of residence in an L2 setting (e.g. Flege *et al.*, 1997a) and the degree of ongoing L1 use (e.g. Flege & Mackay, 2004). Other factors such as motivation to learn and language learning aptitude have not received much attention in the L2 phoneme literature.

It is clear, however, that even after attempting to control for the factors mentioned above, large individual differences are often found in cross-sectional or longitudinal studies examining the perception and production of non-native sounds. Similar results are reported in studies that involve intensive computer-based training of L2 sounds in laboratory conditions where, additionally to large between-subject variability before training, improvement after training can range from no improvement to very significant gains across individuals (e.g. Bradlow *et al.*, 1997; Hazan *et al.*, 2005), which suggests that there are idiosyncratic differences in the effectiveness of training. Current cross-language/L2 models such as Perceptual

Assimilation Model (PAM: Best *et al.*, 1988; Best, 1995; Best & Tyler, 2007), the Speech Learning Model (SLM: Flege, 1995a, 2002), and the Native Language Magnet model (NLM: Kuhl *et al.*, 1992; Kuhl, 2000; Kuhl *et al.*, 2008) offer no explanation for individual variability found within L2 populations with similar profiles. There are three possible explanations for the existence of such variability: first, participants may not have been well matched on those factors; second, these factors may be causally related to each other, i.e. a factor may be confounded with other factors; and third, there are other factors that were not controlled in the experiment. This thesis examines two unexplored sources of individual differences in L2 vowel learning by testing the perception and production of Southern British English vowels by native speakers of Standard Modern Greek. The main question to be addressed is whether L2 vowel processing is related to individual variability in L1 vowel processing and/or frequency discrimination acuity.

1.1 L2 vowel perception and production

In the early months of life, infants appear to be able to discriminate all sounds that are used to signal contrasts in any language (Eimas *et al.*, 1971; Lasky *et al.*, 1975; Streeter, 1976; Trehub, 1976; Aslin *et al.*, 1981). However, by the end of their first year infants fail to discriminate non-native consonant contrasts (Werker *et al.*, 1981; Werker & Tees, 1983, 1984; Werker & Lalonde, 1988). Sensitivity to non-native vowel contrasts appears to decline somewhat earlier, at around six months of age (Kuhl *et al.*, 1992; Polka & Werker, 1994; Bosch & Sebastian-Galles, 2003), although one study by Polka & Bohn (1996) failed to show age or language influences on infant discrimination of non-native vowel contrasts suggesting that not all non-native sounds are affected similarly by language experience *(*see also *Best et al.*, 1988 for a similar lack of age effect found on English infants' perception of IsiZulu click consonants).

A great body of research using both natural and synthetic stimuli has examined adult learners' perception and production of non-native vowels. These studies have generally looked at difficulties in L2 vowel acquisition for speakers whose L1 vowel system differs considerably from the target L2 system or where there is a significant mismatch in the vowel categories themselves (for an excellent review of cross-language infant and adult studies, see Strange, 1995). In line with studies examining infant vowel perception mentioned above, adult L2 learners are generally found to face difficulties when perceiving and producing non-native vowels. For example, Gottfried (1984) found that native French speakers were more accurate in identifying French vowels than American English learners of French. In the same study it was also shown that native French speakers were better in discriminating French vowels than both American English learners of French and monolingual American English speakers with the American English learners of French outperforming the monolingual American English speakers when the vowels were embedded in a /tVt/ context (but not when the vowels were presented in isolation). Polka (1995) found that monolingual Canadian English speakers showed native-like performance for the German tense vowel contrast /u/-/y/ but not for the lax contrast /ʊ/-/ʏ/ which was attributed to differences in English speakers' assimilation of the German vowels to English vowel categories. Spanish learners of English have difficulty in English /iː/-/ɪ/ discrimination because they lack such a contrast in their L1, having a single vowel category in the F1/F2 vowel space occupied by the two English vowels (Flege *et al.*, 1997a; Escudero, 2005). Højen & Flege (2006) used a categorical AXB discrimination task to further explore the perception of English vowels by Spanish speakers. Their subjects obtained near chance scores (<60% correct) in three difficult English contrasts at an inter-stimulus interval (ISI) of 0 ms and 1000 ms.

Similar language effects have been reported even in the case of highly proficient bilinguals as shown in a series of studies testing Spanish-Catalan bilinguals' perception of Catalan vowels (Pallier *et al.*, 1997; Sebastián-Gallés & Soto-Faraco, 1999; Bosch *et al.*, 2000; Pallier *et al.*, 2001) even though, as will be discussed in following sections, other studies have demonstrated that native-like performance can be at times achieved by early L2 learners as in the case of Italian learners of English (Flege & Mackay, 2004). Still, Italian late learners of English often differ from native English speakers in perceiving English vowels (Flege *et al.*, 1999a). The perception of English vowels is also problematic for other language groups with fewer contrastive vowels in L1, such as Korean (Flege *et al.*, 1997a; Ingram & Park, 1997) and Catalan (Cebrian, 2006) learners of English.

Studies examining the production of vowels by L2 learners generally mirror the results concerning the perception of L2 vowels. Rochet (1995) used a synthetic /i/-/y/-/u/ vowel continuum to examine L1 language effects on the perception and production of non-native vowels. Native Portuguese speakers tended to label French /y/ tokens as /i/ whereas English native speakers tended to label the same tokens as /u/. According to Rochet (1995), this was due to differences in articulation between Portuguese /u/ and English /u/ with the latter being articulated in a more fronted position than the former. An imitation task showed that Portuguese speakers produced French /y/ as /i/ whereas English speakers produced French /y/ as /u/, suggesting, according to the author, that production errors may have a perceptual basis. Munro *et al.* (1996) compared English vowel productions of English native speakers and of 240 native Italian speakers who had arrived in Canada at ages ranging from 2 to 22 years. English vowel productions for native and non-native speakers were rated for degree of accent by native English speakers. Late learners obtained lower ratings than early learners across English vowels. Further, English vowel productions by Italian late learners of English were identified correctly by English native speakers less often than English vowel productions by monolingual English speakers. Similar results concerning the production of English vowels are reported in several other studies (Mcallister *et al.*, 2002; Piske *et al.*, 2002; Flege *et al.*, 2003) with some of those studies being discussed in more detail in following sections.

1.2 The use of duration in L2 vowel perception and production

L2 learners often have difficulties in weighting the acoustic cues that signal L2 contrasts. A common example demonstrating such a difficulty is the problematic for Japanese native speakers perception of the English /r/-/l/ contrast (e.g. Goto, 1971). Iverson *et al.* (2003) showed that Japanese speakers pay attention to the non-critical second formant (F2) frequency when trying to distinguish English /r/ from /l/ instead of focusing on the third formant (F3) onset frequency that is used by native speakers of English, the reason being that Japanese speakers mainly use the former when perceiving the single Japanese sound perceptually related to English /r/ and /l/, namely /ɾ/. In a vowel study, McAllister *et al.* (2002) tested the hypothesis that category formation is difficult when based on a phonetic feature not used contrastively in the L1 ('feature hypothesis'). The hypothesis, implied in the SLM (Flege, 1995a) predicts that success in using durational cues when acquiring L2 vowels will be related to previous experience with duration in L1 vowel distinctions. McAllister *et al.* (2002) compared the perception and production of the Swedish vowel length contrasts by native speakers of Estonian, American English and Latin American Spanish. The results in perception showed that the Estonian speakers who are extremely experienced with duration distinctions in their L1 outperformed the American English speakers who use duration as a secondary cue in L1 vowel distinctions. American English speakers in turn outperformed the Spanish speakers who do not use duration at all in perceiving L1 vowels. The cross-language differences in the production of Swedish vowels were fewer than the perception differences with the Spanish speakers being consistently less successful in producing the vowels than any other group. These results were seen as confirming the importance of L1 transfer when learning the vowels of an L2.

However, other studies have demonstrated that listeners remain sensitive to novel acoustic features when perceiving L2 vowels. Bohn (1995) examined the perception of American English vowels by native speakers of German, Spanish and Mandarin. The stimuli were synthetic vowel continua (/ɛ/-/æ/ and /i/-/ɪ/) that tested the learners' reliance on spectral and durational cues. Bohn (1995) found that duration was predominantly utilized in the perception of these vowels not only by native German speakers who make use of both spectral and durational cues in L1 to distinguish vowels but also by native Spanish and Mandarin speakers, neither of which group uses duration in contrasting L1 vowels. To explain this finding, Bohn (1995) proposed a 'desensitization hypothesis'. Bohn hypothesized that, when spectral information is not available (hence the term 'desensitization'), L2 learners will use durational information irrespective of whether duration is used in their L1, as duration is a cue that is acoustically salient and easy to access. The overreliance on durational cues to differentiate the English /iː/-/ɪ/ contrast shown by Spanish speakers has been reported in several subsequent studies (Flege *et al.*, 1997a; Morrison, 2002; Escudero & Boersma, 2004; Escudero, 2005; Cebrian, 2006). Similar results have been reported for native speakers of Korean who are also inexperienced with the duration feature in L1 vowel distinctions (Flege *et al.*, 1997a). Escudero & Boersma (2004) offered a different explanation for Spanish speakers' preference for durational cues when distinguishing English /iː/ from /ɪ/. They proposed that since Spanish does not employ duration contrastively to signal vowel contrasts, it is easier for Spanish speakers to create a new category (duration) than splitting their already-existing (spectral) Spanish /i/ category. Finally, Iverson & Evans (2007b) provided further evidence for the availability of durational cues in L2 vowel perception for speakers that do not use duration in L1 vowel distinctions; their results comparing Spanish, French, German and Norwegian speakers' perception of the Southern British English vowel system suggested that L2 vowel learning shows a high degree of uniformity in the use of secondary acoustic cues such as duration and intrinsic formant movement irrespective of L1 background.

1.3 Factors affecting L2 speech processing

Research on L2 acquisition has identified a variety of factors that are related to the subject's background and may affect success in acquiring a second language[1]. These factors can be assigned to three broad categories: (1) factors concerned with the learner's first or second language experience such as age of L2 learning, relationship between the L1 and L2 sound inventories, length of residence in an L2-speaking environment and amount of ongoing L1 use; (2) factors concerned with the learner's language aptitude such as phonological memory and working/short-term memory; and (3) factors concerned with the learner's attitudes towards language learning such as motivation. Cross-language speech perception and production studies have mainly focused on the first category and research in the cross-language/L2 phonetics literature on the other two categories is limited.

1.3.1 Experience-related factors

The age of first exposure to the L2, usually indexed by the age of immigrants' arrival (AOA) in an L2 setting (or equally by the age of L2 learning (AOL)), is by far the most frequently examined factor in the L2 literature. The idea that age may restrict language learning gained interest after Lenneberg (1967) published the *Biological Foundations of Language* introducing the concept of a Critical Period of language acquisition, according to which, the ability to acquire language successfully is biologically linked to age. In an attempt to explain the empirical observation that young learners are usually better in acquiring an L2 than older learners are, the Critical Period hypothesis was before long extended to the field of L2 acquisition. It has therefore been claimed that biologically determined maturational constraints exist when learning the L2 grammar (Johnson & Newport, 1989), syntax (Patkowski, 1980) and pronunciation (Patkowski, 1990). Regarding L2 phoneme learning,

[1] Given the scope of this thesis which particularly concerns individual differences under the same experimental conditions, other factors related to the task or the stimuli employed to assess cross-language speech perception (see for example the excellent review in Beddor & Gottfried, 1995) will not be discussed.

significant age effects are reported with respect to L2 vowel perception (Flege *et al.*, 1999a) and production (Flege *et al.*, 1999a; Piske *et al.*, 2001), and L2 consonant perception (Mackay *et al.*, 2001b) and production (Mackay *et al.*, 2001a).

However, to support the view that age effects are due to a maturational-based loss in neural plasticity, evidence is needed that a) there is a sharp drop-off in the ability to learn a second language, b) all early L2 learners can achieve native-like performance and c) all late L2 learners fail to achieve native-like performance. On the contrary, a great body of evidence suggests that the perceptual system remains plastic and that there is no discontinuity in L2 learning ability but rather a gradual decline with age (e.g. Flege *et al.*, 1999a); discrimination of non-native contrasts can be improved through natural or laboratory training (e.g. Logan *et al.*, 1991; Lively *et al.*, 1993; Lively *et al.*, 1994; Bradlow *et al.*, 1997; Lambacher *et al.*, 2005; Iverson & Evans, 2007a); not all early bilinguals perform equally well and not all late bilinguals perform equally poorly on perceptual tasks (e.g. Bongaerts *et al.*, 1995; Bongaerts *et al.*, 1997).

An alternative explanation for the advantage of early over late L2 learners that has more or less dominated the field the past years is that experience with the ambient language rather than maturational constraints impede L2 acquisition. Flege (1987) was the first to propose that adults might be more eager than children to accept an L2 sound as an instance of an already established L1 category and hence their difficulty in L2 learning compared to children. Recent work in L1 speech perception suggests that infants acquire their L1 categories through distribution-based learning (Maye *et al.*, 2002; Maye & Weiss, 2003). This process may sharpen L1 perception but unavoidably *interferes* with L2 learning (Iverson *et al.*, 2003; Kuhl *et al.*, 2006; Kuhl *et al.*, 2008). The effect of L1 tuning on L2 perception has been discussed by the three current cross-language models previously mentioned, the Perceptual Assimilation Model (PAM: Best *et al.*, 1988; Best, 1995; Best & Tyler, 2007), the Speech Learning Model (SLM: Flege, 1995a, 2002), and the Native Language Magnet model (NLM: Kuhl *et al.*, 1992; Kuhl, 2000; Kuhl *et al.*, 2008).

PAM was originally proposed to account for naïve listeners' perception of *non-native* sounds and has recently been applied to L2 learning (Guion *et al.*, 2000; Best & Tyler, 2007). The model is based on direct realism principles (Fowler, 1986; Browman & Goldstein, 1989) and posits that non-native contrasts are perceived in terms of their articulatory/gestural similarity/dissimilarity to the native categories. When perceiving a foreign sound a listener can (a) assimilate it to a native category, (b) perceive it as an uncategorized sound (i.e. a sound that falls between two native categories), or (c) perceive it as a non-speech sound. According to PAM, discrimination of a non-native contrast depends on how each member of the contrast is assimilated to the native categories. There are several possible assimilation types and for each assimilation type there is a specific discrimination prediction: Two-Category assimilation (TC) where each member of a contrast assimilates to a different native category and discrimination is predicted to be excellent; Single-Category assimilation (SC) where both members assimilate equally well or poorly to a single native category and discrimination is predicted to be very poor, perhaps at near-chance levels; Category-Goodness assimilation (CG) where both members assimilate to the same native category but with one member being a closer match to that category than the other and discrimination is predicted to be moderate to very good (with the magnitude of the difference in category goodness defining the degree of difficulty); Uncategorized-Categorized assimilation (UC) where one non-native phone is Categorized while the other is Uncategorized (as described above) and discrimination is predicted to be very good; Uncategorized-Uncategorized assimilation (UU) where both non-native phones are uncategorized and discrimination is predicted to vary from fair to good according to how similar these sounds are to each other and to native categories; and Non-Assimilable (NA) where both non-native phones are perceived as non-speech sounds and discrimination is predicted to be very good.

SLM is concerned with *L2 learning* and particularly with experienced L2 learners. It posits that speech-learning mechanisms remain intact across the life span. The advantage of early over late L2 learners (Flege *et al.*, 1999a; Flege & Mackay, 2004) is attributed to the fact that as the L1 categories develop with age (Lee *et al.*, 1999;

Hazan & Barrett, 2000), they become stronger attractors of L2 categories (Flege *et al.*, 2003). L2 categories are initially classified in terms of L1 categories (‘equivalence qualification’) based on the perceived phonetic similarity/dissimilarity between the L1 and L2 categories; the formation of a new category thus requires from the learner to detect phonetic differences between the L2 category and the closest L1 category. One important tenet of SLM is that most (although not all) production errors have a perceptual basis, in other words perception accuracy can limit production accuracy. Another important principle of SLM is that the L2 learner possesses a single phonological system; the L1 and L2 phonetic categories interact and thus interference is bidirectional (Flege *et al.*, 2003). The first important difference between SLM and PAM is the level of L2 language experience the two models are addressing; however, as previously mentioned, the latest version of PAM has been extended to L2 speech learning (Guion *et al.*, 2000; Best & Tyler, 2007). SLM also differs from PAM in that it does not specify the nature of the cross-language perceived similarity/dissimilarity. Finally, while PAM provides discrimination predictions for pairs of non-native sounds, SLM predicts the difficulty listeners will face when learning individual L2 sounds.

NLM aims at explaining the development of speech perception from infancy to adulthood, namely how infants start their life as language-general perceivers (Eimas *et al.*, 1971), to become by the end of first year, language-specific perceivers (Werker & Tees, 1984). According to the model, during that period, infants develop prototypes for native categories. The prototypes act as perceptual ‘magnets’ shrinking the perceptual space around the prototypes. That means that two tokens near the prototype are more difficult to discriminate than two tokens equally apart from each other but further apart from the prototype in the perceptual space. Kuhl *et al.* (1992) showed that Swedish infants demonstrated a magnet effect for Swedish /y/ but not for American English /i/ at around sixth month whereas English infants at the same age demonstrated the reverse pattern. Regarding non-native perception, this reduced ‘discriminability’ around the prototype is predicted to interfere with the ability of an L2 learner to discriminate two L2 sounds close to the L1 category/prototype. In a recent revised version of NLM, the NLM-e (expanded), Kuhl *et al.* (2008) describe five new principles incorporated in the model;

i. *Distributional patterns and infant-directed speech are agents of change*. Building on the infants' sensitivity to L1 distribution patterns (Maye *et al.*, 2002; Maye & Weiss, 2003) this principle points out how the exaggerated acoustic cues contained in infant-directed speech compared to adult-directed speech facilitates statistical learning (Liu *et al.*, 2003).
ii. *Language exposure produces neural commitment that affects future learning*. It is suggested that neural networks become committed to L1 distributional patterns which are difficult to overcome in adulthood. Japanese speakers' lack of sensitivity to the third formant (F3) onset frequency when perceiving the English /r/-/l/ contrast already mentioned is probably the most well-known example of this difficulty caused by L1 language commitment (Iverson *et al.*, 2003).
iii. *Social interaction influences early language learning at the phonetic level*. It is proposed that statistical input may not be enough if not provided during natural learning situations. For example, exposure to input from a television or an audiotape does not promote learning while exposure to the same input provided during social interaction does (Kuhl *et al.*, 2003).
iv. *The perception-production link is forged developmentally*. Perceptual learning precedes and therefore guides production as infants attempt to relate the sounds they produce with the sounds stored in their memory.
v. *Early speech perception predicts language growth*. Native and non-native performance of phonetic perception can predict future language abilities (Tsao *et al.*, 2004)

With respect to age effects on L2 learning, the position taken by NLM-e could be summarized as '...phonetic learning causes a decline in neural flexibility, suggesting that experience, not simply time, is a critical factor driving phonetic learning and perception of a second language' (Kuhl *et al.*, 2008, p. 993).

Another factor frequently examined in the L2 literature is length of residence (LOR) in an L2 setting. As Piske *et al.* (2001) note in their comprehensive review of factors affecting degree of foreign accent in an L2, LOR has been found to affect L2 pronunciation in some but not all studies. Flege *et al.* (1997a) examined the effect of English-language experience on L2 learners' perception and production of English vowels. Four groups of non-native speakers (German, Spanish, Mandarin and Korean) who were first exposed to English when they arrived to the US and differed in terms of the years they had spent in the US were tested. The results showed that the more experienced non-native speakers were better than the less experienced ones both in perceiving and producing English vowels. Similar results supporting the significance of experience with an L2 are reported in Asher & García (1969) and Flege & Fletcher (1992).

However, other studies report no effect of experience (Oyama, 1976; Tahta *et al.*, 1981; Piper & Cansin, 1988). Flege & Liu (2001) suggest that the lack of an effect of LOR in some studies may have been due to the quality of the L2 input the sampling population received; in their study, an effect of LOR on L2 learning (assessed by means of a consonant identification task, a grammaticality judgment task, and a listening comprehension task) was found for Chinese immigrants in the US who had been enrolled as students at an American University but not for Chinese immigrants who had worked full-time during their stay in the US. *Piske et al.* (2001) suggest that the effect of LOR is more likely to be found if the difference in years of residence between the groups to be compared is relatively large and that the effect of LOR may depend on the subject's stage of learning.

> '...for highly experienced subjects, additional years of experience in the L2 appear to be unlikely to lead to a significant decrease in degree of L2 foreign accent. In the early phases of L2 learning, on the other hand, additional experience in the L2 may well lead to less foreign-accented L2 speech' (Piske *et al.*, 2001, p. 199)

The degree of ongoing L1 use is a factor that has received attention relatively recently in the L2 with all evidence so far supporting its importance in acquiring an L2. Flege & MacKay (2004) examined the discrimination of English vowels by four groups of native speakers of Italian that differed in terms of their AOA in Canada (early vs. late) and their ongoing use of Italian while living in Canada (high L1 use vs. low L1 use). The results showed that early learners of English outperformed late learners and that low-L1-use learners outperformed high-L1-use learners. In addition, only the early low-L1-use learners achieved native-like performance. Similar results are reported in MacKay *et al.* (2001b) for English consonant perception in noise and in Piske *et al.* (2002) for English vowel production (both studies tested Italian immigrants in Canada) and in studies testing degree of foreign accent in sentence production (Flege *et al.*, 1997b; Piske *et al.*, 2001) and recognition of English words (Meador *et al.*, 2000).

1.3.2 Language aptitude-related factors

Language aptitude can be described as 'the learner's overall capacity to master a foreign language' (Dörnyei, 2005, p. 33-34). Traditionally, two test batteries have been used for assessing language aptitude, namely the *Modern Language Aptitude Test* (MLAT) (Carroll & Sapon, 1959) and the *Pimsleur Language Aptitude Battery* (PLAB) (Pimsleur, 1966) with the former test being more popular than the latter. Carroll (1981) identified the following four aspects of language aptitude and noted that MLAT measures the first three: (1) phonetic coding ability; (2) grammatical sensitivity; (3) rote learning ability; and (4) inductive learning ability. After reviewing previous research, Carroll (1981) concluded that language aptitude is a strong predictor of success in learning a second language. Skehan (1998) proposed a three component model where Carroll's grammatical sensitivity and inductive learning ability form part of a single language analytic ability component with the other two components being phonetic coding ability and memory ability. In general, research conducted both before and after Carroll's (1981) review article has confirmed the importance of language aptitude in learning a second language in classroom settings (Gardner, 1980; Horwitz, 1987; Ehrman & Oxford, 1995; Hummel, 2009).

With respect to L2 phoneme learning, one study attempted to relate one aspect of learning aptitude, namely phonological short-term memory (PSTM) and L2 perception accuracy. MacKay *et al.* (2001b) examined the identification of English consonants by native speakers of Italian as a function of chronological age, AOA in Canada, L1 (Italian) use and PSTM scores. PSTM was assessed by asking subjects to repeat Italian pseudo-words. There was a negative correlation between PSTM scores and percentage of errors in both word-initial and word-final English consonant identification; those Italian speakers who obtained higher pseudo-word repetition scores made fewer errors in English consonant identification. Further, PSTM scores independently accounted for 8% and 15% of the variance in word-initial and word-final consonant identification scores respectively which is noteworthy considering that AOA was found to independently account for about the same amount of variance in identification scores (although AOL accounted for more variance in word-initial than in word-final scores, i.e. 18% and 9% respectively). This is in line with work showing that PSTM influences success in children's (Service, 1992) and in adults' (O'brien *et al.*, 2007; Hummel, 2009) L2 learning in classroom settings and in immersion settings (Sleve & Miyake, 2006).

A few studies have attempted to relate L2 learning and musical ability in a controlled manner reporting very little if any evidence for a link between musical ability and L2 learning (Tahta *et al.*, 1981; Thompson, 1991; Flege *et al.*, 1999b). However, in a recent study examining the relation between musical ability and L2 proficiency (measured in four areas, namely receptive phonology, productive phonology, syntax and lexical knowledge) for Japanese late learners of English, Sleve & Miyake (2006) found that musical ability predicted perceptive and productive phonology but not syntax and lexical knowledge. As the authors note, one important difference from previous studies was that musical ability was assessed via objective psychometrically validated tests rather than relying on the participants' subjective self-ratings (e.g. Thompson, 1991; Flege *et al.*, 1995). Alexander *et al.* (2005) provided further evidence for a connection between musical ability and L2 speech perception. Their results showed that a group of American English musicians (musical experience was defined by eight or more years of continuous private piano or voice lessons) with no previous exposure to Mandarin Chinese were more successful in identifying and discriminating the four lexical tones of Mandarin than a group of American English non-musicians with no previous exposure to Mandarin. However, given that pitch is a shared acoustic feature of music and lexical tone perception, it is very difficult to say whether the link between the two is specific to the acquisition of tone languages or may relate to general L2 learning abilities.

1.3.3 Affective factors

With respect to affective factors, research has mainly focused on whether motivation affects degree of foreign accent. Bongaerts & Schils (1995) and Bongaerts *et al.* (1997) tested highly motivated and successful Dutch learners of British English. The participants in both studies were late learners that had been exposed to spoken English after entering the university (from the age of around 18). Bongaerts & Schils (1995) found that all 10 Dutch participants were indistinguishable from native English controls; Bongaerts *et al.* (1997) found that 5 out of 11 participants (9 of whom had also participated in the first study) met a criterion of 'nativelikeness', i.e. their English sentence productions received a mean rating that fell within 2 standard deviations of the mean rating given to the English controls. Moyer (1999) tested English learners of German who were employed in Germany to teach undergraduate students in a variety of tasks assessing their degree of their foreign accent. Moyer (1999) found that professional motivation was a significant factor in degree of predicting foreign accent. Finally, Elliott (1995) examined English students' pronunciation in Spanish. Production accuracy was measured on 4 tasks: (1) mimicking pronunciation at a word level, (2) mimicking pronunciation at a sentence level, (3) reading words, and (4) communicating spontaneously. The results showed that among several variables tested, including cognitive, affective and instructional ones, motivation/attitude was the most significant predictor of production accuracy in the three out of four tasks tested (no variable could predict production accuracy in task 1). It therefore seems that motivation is related to pronunciation accuracy in an L2, although it has to be noted that there are a few studies that report no such effect (Oyama, 1976; Thompson, 1991; Flege *et al.*, 1999b; Yeni-Komshian *et al.*, 2000). As noted in Piske *et al.* (2001), it is not always easy to quantify motivation in a precise manner which may explain the results of those studies that found no effect of motivation on degree of foreign accent.

1.3.4 'X factor'?

As discussed in previous sections, research on L2 phoneme learning has primarily focused on four factors related to the subject's language experience at the time of testing, namely the relation between the L1 and L2 phonetic systems, the age of L2 acquisition, the duration of L2 immersion and the degree of ongoing L1 use. Very little work has been done concerning factors related to the subject's speech or non-speech auditory abilities, such as phonological short-term memory and musical ability. At the same time, there is both anecdotal and experimental evidence for large individual differences in L2 performance even after the above-mentioned factors were controlled (e.g. Hazan *et al.*, 2006). The following two sections will discuss two unexplored sources of individual differences in L2 phoneme learning, namely individual differences in L1 vowel processing and non-speech (auditory) processing of sounds.

1.4 Factors affecting L1 speech processing

Individual variability in L1 speech perception research has been traditionally considered as a variable that should be removed from the data. For example, early studies of categorical perception reported only on mean identification and discrimination functions and sometimes the 'poor' performers were entirely eliminated from the dataset (e.g. Liberman *et al.*, 1961). Although individual differences in L1 speech perception are expected not to be easily noticed in everyday life given the redundancy of the speech signal (for example, we normally expect people to have no difficulty in identifying words or sentences presented at a positive signal-to-noise ratio), such differences become evident when examining people with hearing loss (Crandell, 1991) or, in the case of normal hearing listeners, when the system is stressed, for example when listening to speech under adverse conditions (Surprenant & Watson, 2001) or when analytical tasks are used, for example in acoustic cue weighting for consonants (Hazan & Rosen, 1991) or in discrimination accuracy for vowels, which is of most relevance for this work and will be discussed in the following section.

1.4.1 Perception of vowels

Detection thresholds, also referred as difference limens (DL) or just-noticeable differences (jnd) for spectral changes in vowels have been extensively tested in a series of studies by Kewley-Port and colleagues (Kewley-Port & Watson, 1994; Kewley-Port & Zheng, 1999; Kewley-Port, 2001; Liu & Kewley-Port, 2004b, a). Although these studies aim mainly at establishing thresholds for frequency discrimination under different experimental conditions such as changes in stimulus uncertainty, consonantal context and training of the subjects (Kewley-Port, 2001), background noise (Liu & Kewley-Port, 2004a) and quality of synthesized vowels (Liu & Kewley-Port, 2004b), they also demonstrate that individuals differ greatly in their ability to discern subtle changes in formant frequencies for vowels.

Kewley-Port & Watson (1994) examined discrimination thresholds for increments and decrements in formant frequency for the first (F1) and second (F2) formants for isolated synthetic steady-state English vowels (all ten English monophthongs). Thresholds were obtained using adaptive procedures (Levitt, 1971) in a two-alternative, forced-choice task with feedback where a standard stimulus was followed by two stimuli, one identical to the standard and the other from a set of synthesized stimuli for each vowel. To obtain thresholds, optimal conditions were employed, i.e., stimuli were presented under minimal-stimulus-uncertainty conditions and highly trained subjects were used. Kewley-Port & Watson (1994) report relatively similar thresholds for F1 across subjects but large individual differences in thresholds for F2. The authors note that although previous studies of formant-frequency discrimination do not explicitly examine individual differences, such differences are apparent in the data provided in two studies (Mermelstein, 1978; Gagne & Zurek, 1988). Kewley-Port (2001) examined the effect of stimulus uncertainty, consonantal context and training on discrimination thresholds for vowels. Of relevance here are two findings of the study. First, discrimination thresholds (Barks) for 37 participants before and after 1 hour of training varied greatly among individuals both in terms of initial vowel discrimination ability and degree of improvement after this short period of training. Second, although performance for these 37 listeners was initially about 230% worse when compared to listeners that had received over 47 hours of training (highly-trained listeners achieving asymptotic performance in the same tasks), about 20% of them showed, after just 1 hour of training, vowel discrimination thresholds which fell within the distribution of the highly-trained listeners.

Individual variability in the perception of L1 vowels is also reported in Gerrits & Schouten (2004). The study examined the extent to which vowels are *categorically* perceived, in other words the extent to which discrimination of synthetic vowel stimuli is predicted from classification of the same stimuli. Consonants (especially stops) are said to be more categorically perceived than vowels (Fry *et al.*, 1962; Repp, 1981, 1984) and it has been suggested that this is due to differences in the availability of auditory short-term memory traces in discrimination between consonants and vowels (Pisoni, 1973, 1975; Schouten & Hessen, 1992). Gerrits & Schouten (2004) tested 19 Utrecht University students' identification and discrimination of a synthetic vowel continuum spanning from Dutch /i/ to Dutch /u/ embedded in a /pVp/ context in two experiments. In both experiments, the /pVp/ stimuli were presented for identification and discrimination in two conditions, one in isolation and one where the stimuli were embedded in a passage. In experiment 1, discrimination was assessed via a four-interval, two-alternative forced choice task (4I2AFC) task (AABA/ABAA) with an interstimulus interval (ISI) of 200 ms while in the second experiment a two-interval, two-alternative forced choice task (2IAFC) was used (AB/BA) with an ISI of 500 ms. According to Gerrits & Schouten (2004), during the former task the participants were not referring to criteria external to the stimuli (established phoneme categories) and were thus functioning in an 'auditory' or 'psychoacoustic' mode while during the latter task the participants were functioning in a 'phonemic labelling' mode. Importantly, experiment 1 revealed large differences in discrimination acuity among individuals whereas no differences were observed among individuals in the classification tasks, which, according to the authors, coincides with studies showing that when subjects are operating in an auditory mode they may differ widely in discrimination performance (Repp, 1981; Rosen & Howell, 1987).

1.4.2 Perception in noise

A few studies have examined individual performance in speech-in-noise perceptual tasks in L1 (Rupp & Phillips, 1969; Middelweerd *et al.*, 1990; Surprenant & Watson, 2001; Kidd *et al.*, 2007) and all report large differences among normal-hearing listeners. Surprenant & Watson (2001) examined, among others, the recognition of CV syllables, words and sentences embedded in speech-shaped noise by 93 Indiana University students. While all participants were tested within normal-hearing limits (<20 dB HL at frequencies from 250 to 8000 Hz), percent correct scores for all tasks ranged about 30 percentage points. Surprenant & Watson (2001) also note that the best 10% of listeners had a signal-to-noise ratio (SNR) threshold (50% correct recognition) of -6.1 dB HL for recognition of words in sentences while the worst 10% of listeners had an SNR threshold (50% correct recognition) of 0.8 dB HL in the same task, a difference of about 7dB HL. Comparable, although slightly smaller between-subject differences in syllable, word and sentence recognition in noise are reported in Kidd *et al.* (2007) where the same tasks were given to a larger group of adults (340 students and nonstudents) in the area of Indiana University.

1.5 Factors affecting general auditory processing

In his comprehensive review of factor-analytic studies of human cognitive abilities, Carroll (1993) identified 8 linearly independent factors of individual differences in auditory receptive ability based on 38 studies published before 1993 that include hearing acuity, speech sound discrimination, and discrimination of tones with respect to pitch, intensity, duration and rhythm among others. However, as discussed in Johnson *et al*. (1987), since these studies have used different test batteries and types of subjects it is difficult to compare their findings. In an attempt to explore such differences, Watson and colleagues developed the Test of Basic Auditory Capabilities (TBAC: Watson *et al*., 1982a; Watson *et al.*, 1982b) which has been used in several studies since. The first version of TBAC included eight subtests, six using single tones and tone sequences and two with speech sounds. Surprenant & Watson (2001) added the three speech subtests mentioned in the previous section (identification of CVs, words and sentences in noise) to the original TBAC. Their results indicate large individual differences across all eleven subtests. Kidd *et al.* (2007) added another eight subtests consisting of more tests on spectral and temporal acuity and a task testing perception of familiar environmental sounds. The nineteen subtests included in this final version of TBAC were the following: (1) *pitch discrimination*; (2) *single-tone intensity discrimination*; (3) *single-tone duration discrimination*; (4) *pulse-train discrimination*; (5) *embedded test-tone loudness*; (6) *temporal order for tones*; (7) *temporal order for syllables*; (8) *syllable identification*; (9-12) *Sinusoidal amplitude modulation* at four rates, 8, 20, 60 and 200 Hz; (13) *ripple noise discrimination*; (14) *gap detection*; (15) *gap-duration discrimination*; (16) *nonsense syllable identification* (17) *word identification* (18) *sentence identification* (19) *environmental sound identification*. Factor analysis revealed four factors, one for loudness and duration, a second for amplitude modulation, a third for familiar sounds and a fourth for pitch and time. Performance across subtests was normally distributed with examples of exceptional performance at both ends of the distributions. Section 1.7 will discuss what the implications of these findings might be in terms of second language phoneme learning but first the literature on laboratory perceptual training for second language learning will be reviewed.

1.6 Training in the laboratory

One of the arguments against the notion of a critical period for language acquisition comes from a number of laboratory training studies conducted the past years. These studies have consistently shown that a short period of intensive perceptual training in laboratory conditions can significantly improve the perception and production of L2 sounds and that learning generalizes beyond stimuli and speakers heard in training, which supports the view that the perceptual system remains plastic over the life span (it is true though that, to my knowledge, we are still lacking studies showing that this learning has an impact on communicative ability in conversational speech). This section will discuss some methodological issues on laboratory training of L2 sounds and will review past training studies focusing mainly on vowel studies. Laboratory training studies reporting on individual differences found before and after training will also be discussed.

1.6.1 Discrimination vs. identification training

The main distinction in the training literature is that between discrimination and identification training. In discrimination training, subjects hear two stimuli in each trial and are asked to decide whether the two stimuli are the same or different. In identification training, subjects hear a single stimulus in each trial and are asked to label the sound using a number of given L2 categories. In both types of training, feedback is usually provided.

The first studies that attempted to modify perception of sounds in the laboratory adopted discrimination training (e.g. Carney *et al.*, 1977; Edman, 1980; Pisoni *et al.*, 1982). Carney *et al.* (1977) successfully trained American English speakers in discriminating small within-category differences along the English Voice Onset Time (VOT) /p/-/b/ continuum. Although Carney *et al.*'s (1977) study aimed at testing whether monolingual adults' perceptual sensitivities within L1 categories can improve under laboratory conditions, its success motivated Strange & Dittmann (1984) to adopt the same protocol to test whether it is possible to improve adults' perception of non-native sounds. Strange & Dittmann (1984) recruited native speakers of Japanese in their 30s (range = 25-33 years) who had resided in the US 5 to 30 months prior to testing. All participants reported difficulty in perceiving and producing English /r/ and /l/ and were eager to improve their English. The pre/post-tests included a minimal pair identification task with natural minimal pairs contrasting /r/ and /l/ in initial, medial and final position as well as identification and discrimination tasks on two synthetic /r/-/l/ series ('rock' vs. 'lock' and 'rake' vs. 'lake'). Training consisted of 14-18 sessions on the synthetic rock-lock series using an AX (same/different) discrimination task with immediate feedback. Strange & Dittmann (1984) found that training improved the trainees' perception of the synthetic rock-lock series and that learning transferred to the untrained rake-lake series (although performance was slightly worse than for the trained rock-lock series). However, the trainees did not improve in their ability to identify the naturally-produced minimal pairs from pre-test to post-test.

The lack of evidence that learning can transfer to natural tokens in Strange & Dittmann's (1984) study was attributed partly to the use of a discrimination training procedure and partly to the low variability of the training stimuli. Regarding the former issue, it has been claimed that discrimination training tends to tailor learners' attention to within-category differences rather than focusing on the crucial for identification between-category differences (Jamieson & Morosan, 1986). Regarding the latter issue, it is believed that the use of a single talker and a single context impedes transfer of learning to other talkers and contexts. An alternative approach to training is a high-variability identification training technique that has dominated the field the past 20 years (Logan *et al.*, 1991; Lively *et al.*, 1993; Bradlow *et al.*, 1997; Bradlow *et al.*, 1999; Hazan *et al.*, 2005; Lambacher *et al.*, 2005; Iverson & Evans, 2007a; Nishi & Kewley-Port, 2007b, 2008; Iverson & Evans, 2009). The particular technique emphasizes the importance of exposure to natural minimal pairs contrasting the target sounds in multiple environments spoken by multiple talkers in a situation that resembles more real-world communication with native speakers and promotes the learning of 'robust' phoneme categories.

Logan *et al.* (1991) trained 6 native speakers of Japanese in perceiving the English /r/-/l/ contrast. The subjects were Indiana University students who had lived in the US from 6 months to 3 years at the time of testing. The pre/post-tests were the same 16 minimal pairs used in Strange & Dittmann (1984). Training consisted of 15 training sessions using a two-alternative forced choice identification task instead of discrimination tasks used in early training studies. The training stimuli were 68 minimal pairs that contrasted /r/ and /l/ in multiple positions and differed from those used in the pre-test. The post-test included the same minimal pairs as in the pre-test and two tests of generalization. The first test of generalization consisted of 98 novel words from minimal pairs contrasting /r/ and /l/ produced by one of the speakers used in training. The second test of generalization consisted of 96 novel words from minimal pairs contrasting /r/ and /l/ produced by a speaker not used in either the pre-test or training. Results showed improvement in minimal pair identification from pre-test to post-test as well as transfer of learning to both tests of generalization (new words spoken by a talker used in training and new words spoken by a talker heard for the first time in the post-test).

In a follow-up study Lively *et al.* (1993) attempted to separate the contribution of two sources of variability in the training stimuli used in their 1991 study; variability introduced by the use of different talkers and variability due to the use of different phonetic environments. The need of examining the relative effects of talker and phonetic environment variability was pointed out by Pruitt (1993) in his critique of Logan *et al.*'s (1991) study. In experiment 1, Lively *et al.* (1993) trained Japanese native speakers with tokens that contrasted /r/ and /l/ in three different environments (initial singleton and intervocalic positions and initial consonant clusters) produced by 5 native English talkers. In experiment 2, subjects were trained with tokens from a wider variety of phonetic environments but this time the training stimuli were produced by a single English talker. In both experiments improvement from pre-test to post-test was found, however, learning generalized to new words spoken by a new talker only in experiment 1. Lively *et al.* (1993) concluded that talker variability is a critical factor in obtaining generalization in cross-language training studies. In another study of this series of studies, Lively *et al.* (1994) tested long-term retention of learning for Japanese native speakers who had never lived in an English-speaking country. Using the same training method, it was shown that learning was retained for at least 6 months after the completion of training.

The high variability training approach used in the Japanese /r/-/l/ studies has been adopted to improve the perception of other segmental and suprasegmental L2 contrasts. These include the perception of English word final /t/ and /d/ by Chinese native speakers (Flege, 1995b), the perception of Hindi dental and retroflex stops by native speakers of English and Japanese (Pruitt, 1995; Pruitt *et al.*, 2006), the perception of Japanese vowel length contrast by English native speakers (Hirata, 2004; Hirata *et al.*, 2007; Tajima *et al.*, 2008) and the perception of Mandarin lexical tones (Wang *et al.*, 1999; Wang *et al.*, 2003) by native English speakers.

1.6.2 Vowel training studies

Research on vowel training is relatively scarce compared to research on consonant training with most studies focusing on improving Japanese native speakers' perception of American English vowels (Lambacher *et al.*, 2005; Sperbeck *et al.*, 2005; Nishi & Kewley-Port, 2007b). One difference between consonant and vowel training studies is that the former usually train a binary L2 contrast (e.g. /r/-/l/, or /t/-/d/) whereas the latter usually attempt to train several L2 vowels at the same time.

Nishi & Kewley-Port (2007a) compared the effectiveness of two sets of training stimuli on Japanese native speakers' perception of American English vowels, one with 9 target vowels (full set) and one with just 3 target vowels (subset). Seventeen native speakers of Japanese in their 30s were recruited to participate in the study. None of the participants had lived outside Japan for more than one year. The participants were assigned to one of three experimental groups; six were assigned to the group that received the full set training (all 9 American English monophthongs), six were assigned to the group that received the subset training (only the three most difficult American English monophthongs), and five served as controls, i.e. received no training. The listeners in the two trained groups were students in the Intensive

English Program at Indiana University and family members of Japanese graduate students. The listeners in the control group were graduate students who had recently arrived in the United States.

The experimental protocol followed the commonly used pre-test, training, post-test procedure and a test of retention (3 months after the post-test). The pre/post-test stimuli contained nonsense CVCə words (embedded in six different consonantal contexts). Real CVC words (multiple consonantal contexts, thirty-six words in total) were used to test generalization of learning. The training stimuli contained the same nonsense words used in pre/post-tests. Training consisted of 9 sessions of identification with feedback; when an incorrect answer was given listeners could decide whether they wanted to hear the sound of the correct or the incorrect stimulus up to 10 times in any order, or to proceed to the next trial. The results showed that listeners in both training groups improved in their perception of English vowels and that improvement generalized to new talkers and words and was maintained after 3 months. However, the group that was trained on three vowels never improved on untrained vowels suggesting that full set protocol was more effective than the subset protocol. According to Nishi & Kewley-Port (2007a), this might be due to the fact that when trained on a large set of vowels covering the entire target vowel space learners experience a wider range of spectral and temporal combinations and allophonic variability than when trained on just a few vowels even if these are the most difficult ones.

In a follow-up study, Nishi & Kewley-Port (2008) trained Korean native speakers (mean age = 23 years, range = 19-30 years), all students in the Intensive English Program Music school or business school at Indiana University, again in perceiving American English vowels. None of the participants had lived outside Korea for more than one year. This time three training protocols were compared. One was the same full set training protocol as in Nishi & Kewley-Port (2007a) and the other two were ‘hybrid’ protocols; one where the first 6 days used the full set stimuli and the last 3 days the subset stimuli (9V-3V protocol) and one where the order of the sets was reversed (3V-9V protocol). The results showed that all three protocols were successful in improving Korean listeners' perception of American English vowels; however, no advantage was found for the hybrid protocols over the original full set protocol. In fact, results for the 3V-9V protocol showed that early focused training on a smaller and more difficult vowel set may have had detrimental effects on the subsequent learning of L2 vowels; participants trained on the 3V-9V protocol did not improve on one of the three difficult vowels (/ʊ/).

Iverson & Evans (2009) trained Spanish and German native speakers on an even larger set of 14 Southern British English vowels, including monophthongs and diphthongs. To increase the range of phonetic variability and the naturalness in the training stimuli, only natural minimal pairs were used. The procedure was partly adaptive in terms of the contrasts the trainees were most exposed to during training, i.e. half of the trials were chosen adaptively based on the subject's errors (for more details on the adaptive part of training, see Section 4.3.2). The results showed that after five sessions of high variability training both groups improved their perception of English vowels with German speakers improving to a larger degree than Spanish speakers (around 20 vs. around 10 percentage points of improvement respectively) and that improvement was retained 4-5 months after training for both language groups. After receiving ten additional sessions of training, the same Spanish speakers showed additional improvement reaching the level German speakers had achieved after five sessions of training. The more crowded vowel space of German which contains 15 monophthongal vowels thus facilitated learning compared to the less crowded vowel space of Spanish which contains just 5 monophthongal vowels. Further, despite vowel identification improvement, listeners did not improve their best exemplar locations (i.e. perceptual representations) for English vowels which, according to Iverson & Evans (2009), suggests that high variability training may aid listeners in applying more successfully their already existing knowledge about L2 vowel categories to L2 identification instead of changing the representation of these categories.

1.6.3 Perceptual training improves production

The relationship between perception and production of non-native phonetic contrasts is an important question for our understanding of the processes involved the acquisition of an L2. There are two ways to study this relationship. The first way is to perform a cross-sectional analysis, i.e. to examine the relationship between perception and production at a single point in time. Previous cross-sectional studies have revealed significant albeit modest correlations between perception and production of vowels (Flege *et al.*, 1997a; Flege, 1999; Flege *et al.*, 1999a; Mcallister *et al.*, 2002) and consonants (Flege & Schmidt, 1995; Schmidt & Flege, 1995). The second way is to examine how changes in one domain (perception) affect changes in the other domain (production).

Bradlow *et al.* (1997) were the first to investigate the effects of high variability identification training with feedback on Japanese native speakers' perception and production of the English /r/-/l/ distinction. The participants were 11 adult Japanese speakers ranging in age from 19 to 22 years who had never lived in an English-speaking environment and their knowledge of English was restricted to English instruction in Japan. A comparable group of Japanese speakers served as controls, i.e., received no training. Production improvement was assessed using two tasks with native English speakers. The first task tested whether native English speakers could reliably discriminate between the trainees' pre-test and post-test productions of English /r/ and /l/. In the second task, native English speakers identified the trainees' pre-test and post-test productions of /r/ and /l/. Perception results replicated the findings of previous studies concerning the effectiveness of auditory training. Importantly, improvement in English /r/-/l/ perception transferred to gains in production in both tasks employed. Bradlow *et al.* (1999) replicated these results and also demonstrated that improvement in both domains was retained three months after training confirming that high variability training results in long-term modifications in perception and production and that the two domains are closely linked. In a study examining the effect of audiovisual training on the perception and production of English consonants by native Japanese speakers, Hazan *et al.* (2005) also showed that perceptual training significantly improved Japanese native speakers' /r/-/l/ production (Experiment 3 of the study) although, as discussed in the following section, there were substantial between-subject differences in the effectiveness of training.

Similar results with respect to vowel training are found in Lambacher *et al.* (2005). The study trained Japanese speakers on American English vowels and examined gains in English vowel perception and production. The participants were all university students and none had lived in an English-speaking country. In the pre/post-tests, Japanese speakers performed a 5-alternative forced-choice identification task that included the five mid and low English vowels /æ, ɑ, ʌ, ɔ, ɝ/ embedded in a varied CVC context as well as a vowel production task whereby subjects were asked to produce the 5 target vowels. English vowel production was assessed by having native English speakers identify the English vowels produced by the Japanese speakers in a 5-alternative forced-choice identification task and an acoustic analysis of those vowels. Both English vowel perception and production improved after perceptual training replicating the results of consonant training studies.

1.6.4 Individual differences in pre- and post-training performance

Most of the above studies report on mean group improvement in perception and production of L2 sounds after perceptual training. However, when individual performance is looked at, there is variability in pre-training performance among subjects with similar profiles and in the effectiveness of training. For example, Hazan *et al.* (2005) report that the difference between pre-test and post-test /r/-/l/ identification scores ranged from -5% to +48% across individuals; similarly, the difference between pre-test and post-test scores in /r/-/l/ production (based on native English speakers' judgments) ranged between -11% to +20% across individuals. Further, despite the fact that the 11 Japanese speakers tested in Bradlow *et al.* (1997) constituted a homogenous L2 group (in terms of L1 background, age, experience with written and spoken English etc), pre-test accuracy in both perception and production varied considerably across individuals. Although subjects improved significantly in both domains after perceptual training, improvement in perception and production was not significantly correlated for individuals; some subjects improved slightly in perception but showed large gains in production, and others had the reverse. Similar results are reported in Bradlow *et al.* (1999). Pre-test identification and production performance for the 11 Japanese speakers that were trained ranged from 51.56% to 85.94% and from 55.95% to 98.50% respectively. Individual gains after perceptual training in perception and production ranged from +6.25% to +25% and from -0.57% to +17.05% respectively.

Wong and colleagues have recently addressed the issue of individual variability in computer-based training focusing on the learning of L2 suprasegmentals (Lee *et al*., 2007; Wong & Perrachione, 2007). Wong & Perrachione (2007) trained native speakers of American English in using Mandarin pitch patterns for lexical identification (English pseudowords superimposed with Mandarin tones) and examined the relationship between success in learning and two variables, namely a more general pitch auditory ability (perception of pitch patterns in a non-lexical task) and previous musical experience. Although only one speaker was used in training, the results showed that all 17 participants improved their lexical identification after training. Subjects' pre-training pitch pattern identification accuracy was a significant predictor of post-training lexical identification accuracy using Mandarin tones. Further, seven out of nine successful learners were amateur musicians (musical experience was defined by at least 6 years of formal private lessons in a musical instrument starting before the age of 10 years). As noted in Wong & Perrachione (2007) and was discussed briefly in Section 1.3.2, it remains to be shown whether pitch general auditory ability and/or musical ability facilitates learning segmental and other aspects of an L2.

Lee *et al.* (2007) examined the effectiveness of multi-talker training in American English speakers' use of Mandarin pitch contrasts in lexical identification and the interaction between training type (single vs. multiple talkers) and learners' pre-training non-lexical pitch identification ability. It was shown that pre-training pitch identification ability in non-lexical contexts predicted successful learning of lexical identification using pitch contrasts regardless of training type which is consistent with Wong & Perrachione's (2007) results. Interestingly, high-variability training was beneficial only for learners with high pre-training pitch ability whereas low-variability training was more beneficial for learners with low pre-training pitch ability, a finding that is not in line with the segmental training studies reviewed above.

1.7 This thesis: Research goals and hypotheses

This thesis examined the acquisition of Southern British English vowels by native speakers of Greek. Three separate studies were run. Since there are no previous data in the literature concerning this particular L1/L2 acquisition, Study 1 aimed, first, at providing a general sense of English vowel perception by Greek learners. To that end, two experiments were designed, namely a cross-language assimilation experiment and a discrimination experiment. Given that duration is not used in Greek vowel distinctions, these experiments also examined whether native speakers of Greek are sensitive to durational cues when perceiving the English vowel system thus testing the two competing hypotheses in the literature, the feature hypothesis (Mcallister *et al.*, 2002) and the desensitization hypothesis (Bohn, 1995). The final goal of Study 1 was to help in selecting specific English and Greek vowels that would test Greek speakers' perception of synthetic L2 and L1 vowels in Study 3. Study 2 defined the endpoints of the Greek continua using an experiment designed to find the best exemplar locations for the relevant vowels in the perceptual space; the endpoints for the English vowel continua were taken from another study that used the same software and testing procedures to find the best exemplar locations of Southern British English vowels (Iverson & Evans, 2007b).

Study 3 aimed mainly at exploring the sources of individual differences in the learning of English vowels by Greek native speakers. Learning was assessed by means of training (using the high-variability approach) a group of Greek speakers in perceiving the Southern British English vowels and examining gains in both perception and production. As reviewed in previous sections, past research and theoretical models in cross-language/L2 speech perception and production have identified several factors that may affect success in L2 phoneme learning. However, these accounts cannot explain differences among individuals with similar profiles found in cross-sectional and training studies. At the same time, there is strong evidence that individuals show a large degree of variability when tested on analytical tasks in their L1 and on non-speech auditory/psychoacoustic tasks. It was hypothesized that individual differences in L2 vowel perception and production may be attributed to individual differences in L1 vowel perception (*L1 phonetic* hypothesis) and/or individual differences in non-speech perception (*auditory processing* hypothesis*)*.

The *L1 phonetic* hypothesis was deduced from the assumption shared by current cross-language/L2 models that experience with the ambient language interferes with L2 learning. For example, as seen before, the SLM attributes age effects to age-changes in robustness of L1 phonetic categories. That is, as children grow up they become more 'committed' to their L1 categories which results in the difficulty adult learners are faced with when learning an L2 compared to children (e.g. Walley & Flege, 1999; Flege *et al.*, 2003). If this L1-L2 perception 'trade-off' is extended to adult L2 learners, individuals with relatively poorly defined L1 vowel categories (shallower identification slopes, better within-category discrimination) might prove to be better at retuning their L1 system and learning L2 vowels. Maye (2007) predicts, similarly, that if there are long-term differences between individuals in L1 attentional weights, that is if some people encode more veridical episodic representations for speech than others then these individuals may be able to tune their L1 system more easily than those with stronger L1 filters.

The *auditory processing* hypothesis predicts that success in retuning the L1 system and learning L2 vowel categories will depend on individuals' auditory abilities measured in non-speech psychoacoustic tasks. Such a prediction seems consistent with the findings in Wong & Perrachione (2007) and Lee *et al.* (2007) that an auditory pitch ability can predict success in the use of pitch patterns in lexical identification by L2 learners. Importantly, the *auditory processing* hypothesis would predict that this auditory ability underlies not only L2 but also L1 vowel perception.

To test these hypotheses a large pre/post battery was used that included several tasks with synthetic and natural vowels in quiet and noise testing Greek native speakers' perception of L1 (Greek) and L2 (English) vowels, their perception of a non-speech continuum (F2 formant frequency only) and their production of English vowels. The use of the L1 vowel tasks and the non-speech task aimed at testing whether individual differences in L2 vowel perception and production were due to individual differences in L1 vowel perception and/or individual differences in spectral auditory acuity. L1 and L2 perception of natural vowels was assessed not only in quiet but also in noise for two reasons. First, it served as another test of category robustness; better performance in speech-in-noise identification might mean more flexible vowel categories which in turn might mean successful L2 learning if the *L1 phonetic* hypothesis is correct. Second, for the first time it was tested whether high-variability auditory training would improve L2 vowel identification not only in quiet as done in previous studies but also in noise conditions. Given that multi-talker babble was used as noise, if training was indeed successful that would mean that learning transferred to a situation that is closer to real-world communication than in any other study. Examining L1, L2 and non-speech perception before and after training aimed at testing whether pre-training performance could predict post-training performance in any way, that is whether successful learners could be identified before training. The main questions addressed in Study 3 were:

- What is the effect of auditory phonetic training on Greek native speakers' *perception* of Southern British English vowels in quiet and noise?
- What is the effect of auditory phonetic training on Greek native speakers' *production* of Southern British English vowels?

- Are individual differences in pre-test L2 vowel *perception* related to individual differences in L1 vowel and/or non-speech perception?
- Are individual differences in pre-test L2 vowel *production* related to individual differences in L1 vowel and/or non-speech perception?
- Are individual gains in post-test L2 vowel *perception* related to individual differences in pre-training L1 vowel, L2 vowel and/or non-speech perception?
- Are individual gains in post-test L2 vowel *production* related to individual differences in pre-training L1 vowel, L2 vowel and/or non-speech perception?

1.8 Overview

The thesis is structured as follows: **Chapter 2** starts with a short description of the two vowel systems examined, the Standard Modern Greek vowel system and the Southern British English vowel system. The chapter then presents the results of two experiments investigating the perceptual assimilation and discrimination of Southern British English vowels by native speakers of Greek and discusses the use of durational cues in both experiments by Greek speakers. Finally, the specific vowel pairs chosen to test Greek speakers' perception of synthetic L1 and L2 vowels in Study 3 are presented.

Chapter 3 reports the results of the best exemplars experiment designed to find the location of the Greek vowels in a multi-dimensional space that included F1 and F2 formant movement and duration. These locations would serve as endpoints for the synthetic Greek vowel continua used in Study 3.

Chapter 4 presents all perception and production tasks employed and all procedures used in testing Greek native speakers' processing abilities in L1 (Greek), L2 (English) and non-speech before and after perceptual training as well as the training stimuli and procedures themselves.

Chapter 5 discusses the results obtained in all tasks employed in the pre-test as well as the relations between L1, L2 and non-speech performance for individuals.

Chapter 6 discusses the results obtained in all tasks employed in the post-test, the relations between L1, L2 and non-speech performance for individuals in the post-test as well as the relations between pre-test and post-training performance for individuals.

Finally, **Chapter 7** summarizes and discusses all results and presents some limitations of this work and directions for future research.

Chapter 2

Perceptual assimilation & discrimination of English vowels

This chapter reports on two experiments investigating the perception of English vowels by native speakers of Greek[2]. The goal of these experiments was threefold: first, to explore the cross-language relationships between Greek and English vowels in order to select specific vowel pairs in Greek and English that would test Greek speakers' perception of synthetic L1 and L2 vowels in Study 3; second, to examine whether Greek speakers have access to durational cues when perceiving English vowels thus evaluating the main competing hypotheses in the literature, namely the feature hypothesis (Mcallister *et al.*, 2002) and the desensitization hypothesis (Bohn, 1995); and third, to test whether perceptual assimilation patterns can predict discrimination performance as hypothesized by Best's PAM. Apart from some impressionistic observations and general predictions concerning the perception of English vowels by native speakers of Greek (based on a phonemic account of the Greek vowel system) there is no study in the literature to experimentally test how Greek speakers perceive the English vowel system. Rather than asking participants to identify English vowels from a synthetic continuum varying in durational and spectral cues, which is a common technique for assessing the use of duration as a perceptual cue in the L2 literature (Bohn, 1995; Flege *et al.*, 1997a; Cebrian, 2006),

[2] This chapter is a modified version of Lengeris & Hazan (2007) and Lengeris (2009).

two perceptual tasks using natural English vowels were employed: (1) a cross-language perceptual assimilation task, and (2) a categorical oddity discrimination task. In order to obtain representative data on how Greek speakers perceive the entire English vowel space, all eleven English monophthongs were used as perceptual stimuli. The vowels were placed in two contexts, namely /bVb/ and /bVp/. That way, it was possible to evaluate how the shortening of the vowels before a voiceless stop consonant would affect their perception. Before presenting the procedures used and the results obtained, a short description of the Greek and the English vowel system will be provided.

2.1 Greek vs. English vowels

The phonemic inventory of Standard Modern Greek consists of five vowels /i, e, a, o, u/ and employs no tense-lax or long-short distinctions (Arvaniti, 1999; Fourakis *et al.*, 1999; Sfakianaki, 2002; Nicolaidis, 2003; Baltazani, 2007). Figure 2.1 displays mean first (F1) and second (F2) formant frequencies of the five Greek vowels as reported in Fourakis *et al.* (1999) and Nicolaidis (2003). The values from Fourakis *et al.* (1999) are taken from 5 male speakers who read /ˈpVsV/ words in slow tempo and in focus position with the target vowels being in the first syllable. The values from Nicolaidis (2003) are taken from conversational speech produced by 2 male speakers and are pooled over stress and position in the word. Taking into consideration the methodological differences between the two studies, Figure 2.1 shows a fairly similar positioning of the vowels relative to each other in the vowel space. Nicolaidis' (2003) data show of course centralized F1 and F2 values compared to the values from Fourakis *et al.* (1999) which is expected for vowels in conversational speech compared to vowels in read speech (for a comprehensive review of studies on Greek vowels, see Arvaniti, 2007). According to Fourakis *et al.*'s (1999) data, Greek /i/ and /u/ are high front and back vowels respectively, Greek /e/ and /o/ are between high-mid and low-mid front and back vowels respectively, and Greek /a/ is a low (or low-mid) central vowel. Greek has a simple syllable structure that takes the form of C(0-3)VC(0-1). Open syllables are much

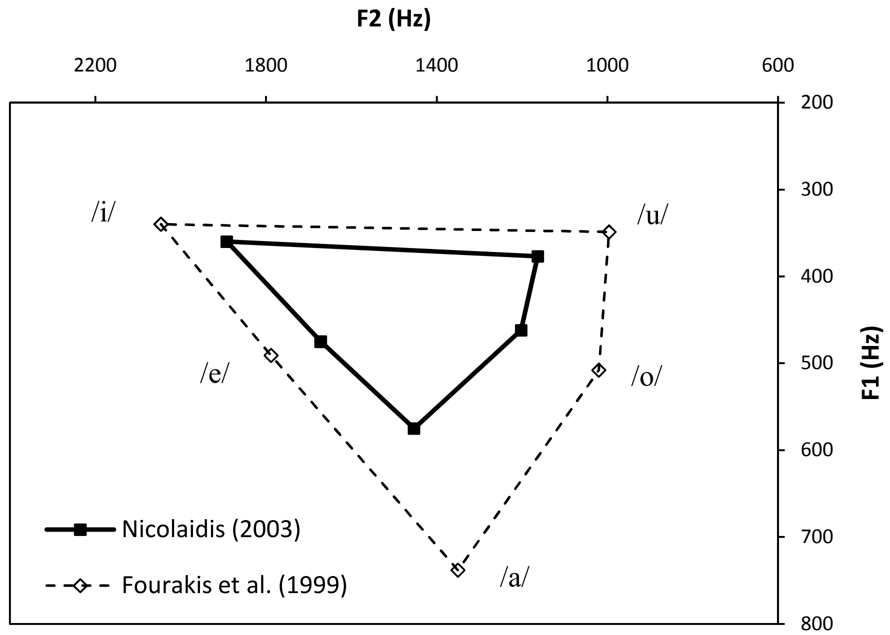


Figure 2.1: Mean F1 and F2 formant frequencies of the five Greek vowels from Fourakis *et al.* (1999) and Nicolaidis (2003), see text for details.

more common than closed ones and the consonants in word-final position are limited to /s/ and /n/ (except in loan words and words from Katharevousa, a 'purified' form of the Greek language mainly used in official and formal documents until 1976 when Dimotiki, the 'popular' Greek language became the official language). The Southern British English vowel system, the target system in this work, is more complex than the Greek one. It consists of eleven monophthongs that can take stress /iː, ɪ, e, æ, ʌ, ɑː, ɜː, ɒ, ɔː, ʊ, uː/ with some vowels being inherently longer than others (e.g. Giegerich, 1992) and eight diphthongs /eɪ, aɪ, ɔɪ, əʊ, aʊ, ɪə, eə, ʊə/. Vowels in all varieties of English are longer before voiced than before voiceless consonants with no change of quality (e.g. Peterson & Lehiste, 1960 for American English; House, 1961; Giegerich, 1992 for Southern British English).

2.2 Recordings of Greek and English vowels

Native speakers of English and Greek recorded productions of their L1 vowels. English vowels were used as perceptual stimuli. Greek vowels were used in order to explain the results concerning the cross-language assimilation patterns (Experiment 1) as well as the Greek native speakers' discrimination performance (Experiment 2).

2.2.1 English vowel stimuli

Three native speakers of Southern British English, all female, (*mean* = 26.4 years, *range* = 23-30 years) recorded the 11 English monophthongs (/iː, ɪ, e, æ, ʌ, ɑː, ɜː, ɒ, ɔː, ʊ, uː/) in the sentence *I read* ___ *on the screen* (speakers were instructed to use the present tense of *read* /riːd/). Vowels were produced in a /bVb/ and a /bVp/ context. The speakers read each vowel four times at a normal speaking rate, giving a total of 264 tokens (3 speakers × 11 vowels × 2 contexts × 4 repetitions). The recordings took place in an anechoic chamber at UCL with a sampling rate of 44.1 kHz, using a Sony 60ES DAT recorder with a B&K Sound Level Meter Type 2231 fitted with a 4165 microphone cartridge. The author and a very experienced phonetician, a native speaker of Southern British English, chose the best three tokens for each English vowel (in almost all cases these were the first three tokens). The final number of stimuli was 198 (3 speakers × 11 vowels × 2 contexts × 3 repetitions). Duration and F1 and F2 measurements were taken for each vowel. All measurements were made manually using the SFS speech analysis software (Huckvale, 2008). Duration was measured from spectrograms, from the onset to the offset of periodic energy in F2. F1 and F2 frequencies were measured by placing the cursors at the centre of the relatively steady-state region of each vowel. Spectral peaks were then estimated from an LPC analysis with 12 coefficients below 5 kHz, and the selection of peaks corresponding to F1 and F2 were verified by visual examination of the spectrogram and an average FFT spectrum of the interval. The process was checked by moving the cursors by small amounts to ensure that the peak frequencies were not strongly influenced by selecting a specific time interval. The

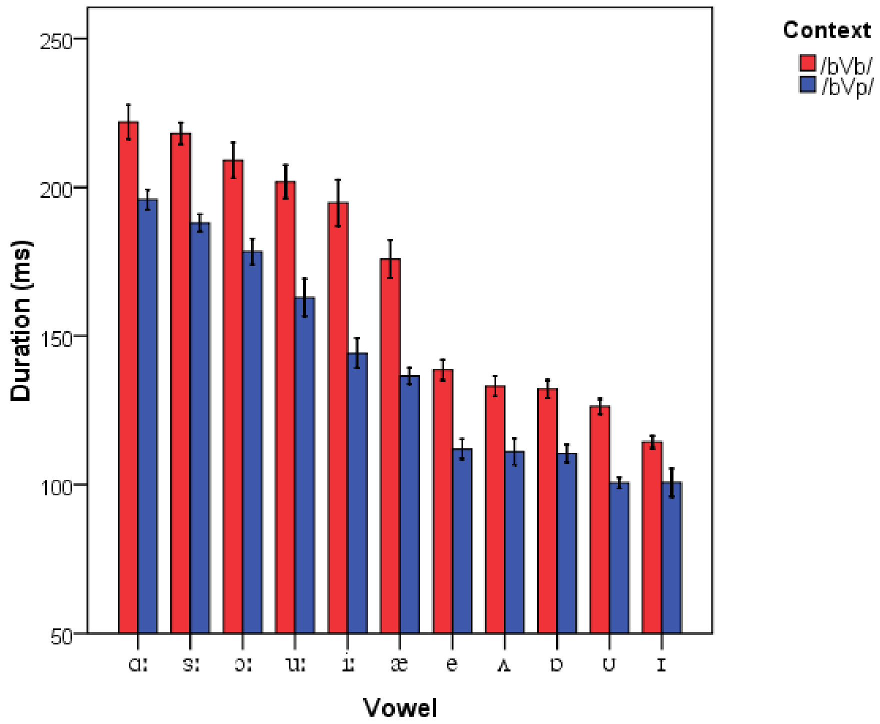


Figure 2.2: Mean durations (ms) of the English vowel stimuli used. Error bars represent standard errors of the mean.

decision to perform all acoustic analyses manually was motivated from the fact that Greek vowels were between nasal consonants thus making duration and formant estimation less reliable. In order to be consistent across languages it was therefore decided to measure English vowels manually too although that was not necessary as in the case of Greek vowels. Mean vowel durations for English vowels in two consonantal contexts, averaged across speakers and repetitions, are displayed in Figure 2.2. Mean F1 and F2 frequencies for English vowels in two consonantal contexts, again averaged across speakers and repetitions are plotted in Figure 2.3. A visual inspection of the two figures indicates that the duration of English vowels is clearly affected by consonantal context whereas their F1 and F2 frequencies are very similar across contexts. Vowel durations were submitted to a two-way ANOVA with Vowel (eleven levels) and Context (two levels) as factors. The ANOVA yielded a significant main effect of Vowel [$F(10,176) = 149.5$; $p < 0.001$], confirming that

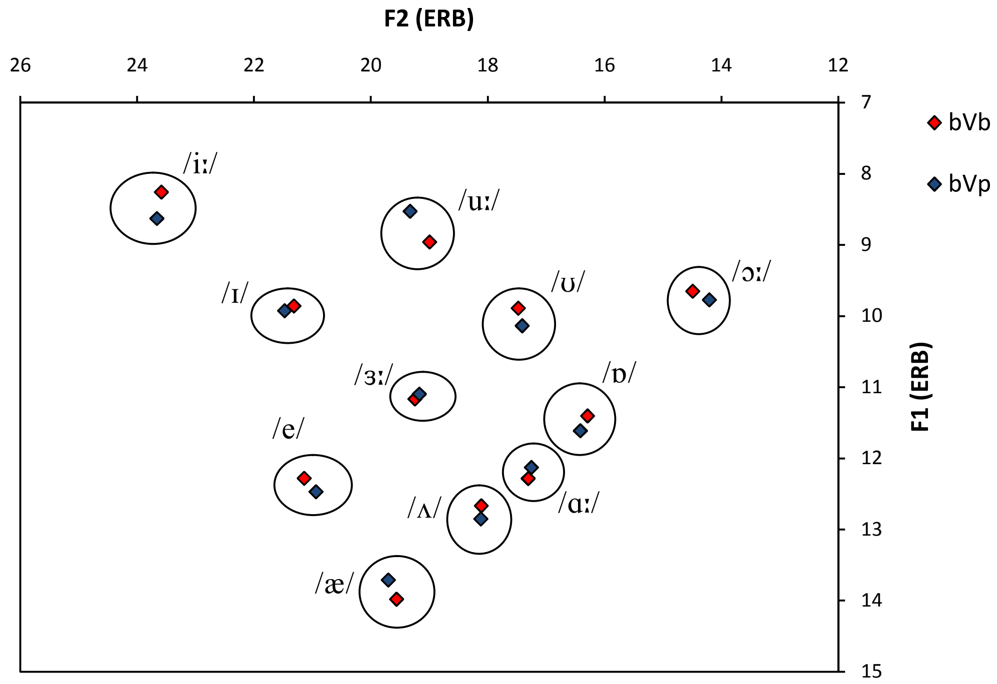


Figure 2.3: Mean F1 and F2 frequencies for 11 English vowels averaged across speakers and repetitions, in /bVb/ and /bVp/ contexts. The ellipses surrounding English vowels are for illustration purposes only and have no statistical status.

English vowels differ in intrinsic duration and a significant main effect of Context [$F(1,176) = 243.7$; $p < 0.001$], confirming that English vowels are shorter in /bVp/ than in /bVb/ context (*mean* = 142 ms vs. *mean* = 171 ms). The ANOVA also yielded a significant Vowel × Context interaction [$F(10,176) = 2.6$; $p < 0.05$], which indicated that the shortening of duration as a result of context was greater for some vowels than for others (Giegerich, 1992). Table 2.1 (Section 2.3.4) presents mean durations for English vowels in both consonantal contexts averaged across speakers and repetitions (standard deviations in parentheses).

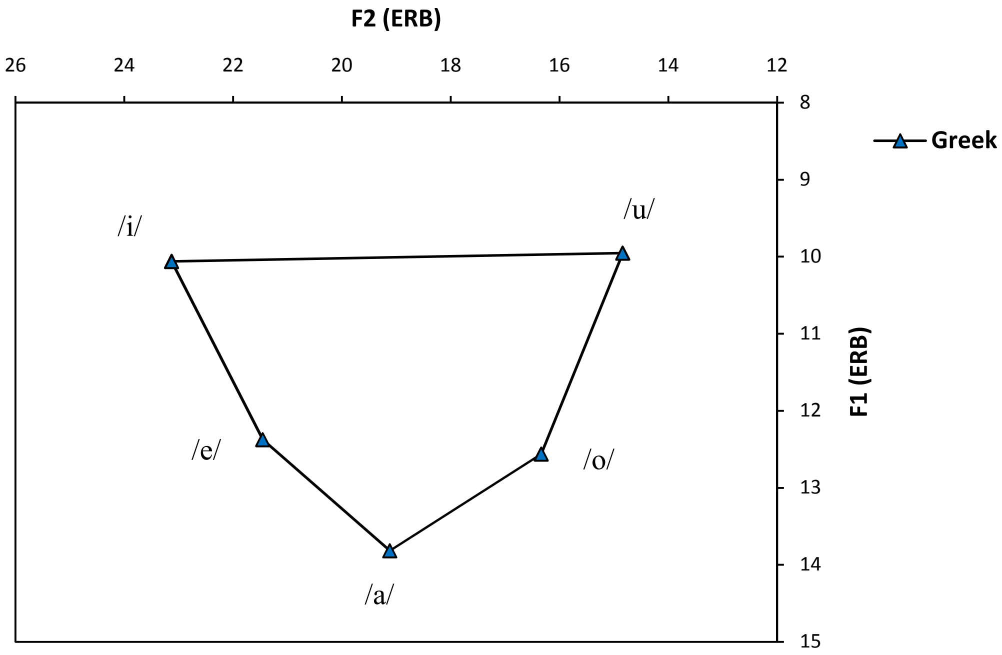


Figure 2.4: Mean F1 and F2 frequencies for 5 Greek vowels averaged across speakers and repetitions.

2.2.2 Greek vowels

Three native speakers of Greek (*mean* = 27 years, *range* = 26-28 years) recorded their L1 vowels /i, e, a, o, u/ in the sentence [ðʝaˈvazo ___ stin oˈθoni] ('I read ___ on the screen') in a quiet room in Athens. All speakers were female in order to be consistent across languages. Greek vowels were uttered in a /mVn/ context. That differed from the /bVb/ and /bVp/ contexts used to elicit the English perceptual stimuli. Ideally, consonantal context should be kept constant across languages; however, given that Greek vowels were recorded in order to collect data on vowel duration in Greek (i.e. these vowels would serve as reference points), the use of a phonologically permissible structure in Greek which would elicit more natural data and which was not expected to affect considerably the conclusions to be drawn was preferred over matching for context across L1 and L2[3]. The speakers read each vowel four times at a normal speaking rate giving 60 tokens for Greek (3 speakers × 5 vowels × 4 repetitions). Recordings were made using a digital recorder (MicroTrack 24/96) in a quiet room at a sampling rate of 44.1 kHz. The first three repetitions for each vowel were selected for acoustic analysis. The author, a native speaker of Greek, judged whether the Greek speakers had correctly produced the tokens. Duration and F1 and F2 measurements were taken for each vowel using the SFS speech analysis software. Duration was measured from spectrograms, taking as vowel onset and offset points the clearly visible changes in the amplitude of upper formants. F1 and F2 were measured in the same way it was done for English vowels. Mean F1 and F2 frequencies for all Greek vowels averaged across speakers and repetitions are plotted in Figure 2.4. Mean durations and standard deviations are displayed in Table 2.1 (Section 2.3.4), averaged across speakers and repetitions.

2.3 Experiment 1: Cross-language perceptual assimilation

The purpose of this experiment was to assess how the vowels of English and Greek are perceptually related. As discussed in Flege *et al.* (1997a), although a variety of techniques have been used in the past for that purpose (e.g. comparisons of the phonetic symbols representing the vowels in question, comparisons of the positioning of the vowels in the vowel space represented by F1 and F2 measurements), the most successful way to directly assess perceived phonetic similarity so far is through a cross-language mapping task (Best, 1995; Schmidt, 1996; Flege & Mackay, 2004; Cebrian, 2006). If L2 learners have access only to cues used contrastively in their L1, as proposed by the feature hypothesis, Greek listeners' perceptual assimilation of English vowels to their L1 vowel categories should not be affected by the context of the vowel stimuli (/bVb/ or /bVp/). If on the other hand, the use of duration is a language-independent perceptual strategy based

[3] In Lengeris (2009), Japanese learners of English were also tested on the same English stimuli and their performance was compared to that of Greek speakers. Data on vowel duration in Japanese were collected using a /mVn/ context which is the only CVC context that is phonologically permissible in both Greek and Japanese (although /n/ would be a separate mora in Japanese).

on the salience of duration, as proposed by the desensitization hypothesis, Greek listeners' perceptual assimilation of English vowels to their L1 vowel categories should be affected by the context of the vowels.

2.3.1 Participants

Eighteen adult learners of English, all university students (*mean* = 23.3 years, *range* = 18-25 years), were tested. All speakers were from Athens, spoke Standard Modern Greek and were tested in Greece. Subjects had received formal English instruction in Greece for 10-15 years by L1-accented language instructors. Their class level was rather high and relatively uniform across individuals (e.g. Cambridge First Certificate in English, Cambridge Certificate in Advanced English) but they had very little, if any, interaction with native speakers of English and none had spent a period of more than one month in an English-speaking environment as shown in a language questionnaire completed by the participants before testing. All of the listeners tested reported normal hearing and no language impairments.

2.3.2 Stimuli

The eleven English vowels (in /bVb/ and /bVp/ contexts) described in Section 2.2.1 were used as perceptual stimuli.

2.3.3 Procedure

Participants were tested individually in quiet rooms using a laptop computer. They were presented the 198 English /bVb/ and /bVp/ tokens at a comfortable intensity level over high quality headphones and completed two tasks: a forced-choice cross-language identification task, and a goodness-rating task. They first heard an English token and identified which of their L1 vowel categories sounded closest to that token by clicking on a label on a screen. The labels were given in Greek orthography "I" /i/, "E" /e/, "A" /a/, "O" /o/, "OY" /u/. Then, they heard the same token again and rated its goodness-of-fit to the chosen L1 vowel category using a scale from 1 (totally different) to 7 (identical). The 198 stimuli were blocked by context, with order of context counterbalanced across listeners. Before the test began, a 33-trial practice session (3 speakers × 11 vowels) was presented to familiarize listeners with the procedure. Consonantal context in the practice session was different from the context to be tested first. Written instructions were given in Greek before testing.

2.3.4 Results

The frequency with which an L1 (Greek) category was selected by the listeners to classify each English vowel was converted to a percentage of total presentations and the mean goodness rating that vowel received as an example of Greek category was estimated. Mean percentage classification and goodness rating were combined into a single metric unit (i.e. the two numbers were multiplied) expressing a 'fit index' of each English vowel to an L1 vowel category (Halle *et al.*, 1999; Guion *et al.*, 2000; Iverson & Evans, 2007b). Table 2.1 presents the L1 vowel that was judged to be perceptually most like each English vowel (as indicated by a higher fit index, see also Table I in Appendix A for the most frequent and the second most frequent L1 classification with the relevant goodness ratings). As can be seen, fit indexes (for the modal response) in both /bVb/ and /bVp/ contexts varied from 2.2 to 5.4. Some English vowels were consistently assimilated to a single Greek category, i.e. were judged to be 'good' examples of that category while others were judged to be 'poor' examples of a Greek category or even heard as falling between two Greek categories (Uncategorized sounds in PAM terminology). In a number of cases, Greek native speakers assimilated more than one English vowel to the same Greek category although with varying degrees of fit: both English /iː/ and /ɪ/ were assimilated to Greek /i/; both English /e/ and /ɜː/ were assimilated to Greek /e/; both English /æ/ and /ʌ/ were assimilated to Greek /a/; English /ɑː/, /ɒ/ and /ɔː/ were assimilated to Greek /o/; and, finally, both English /ʊ/ and /uː/ were assimilated to Greek /u/.

Table 2.1: Percent classification, goodness rating and overall assimilation fit of English vowels to L1 vowel categories for Greek listeners, and *t* test results indicating whether English vowels fitted better to L1 categories in the context where the mean duration in L2 was closer to the mean duration in L1. Mean vowel durations (ms) in L2 (in both /bVb/ and /bVp/ contexts) and L1 are given. Standard deviations are also given in parentheses.

English vowel	Mean duration	L1 closest vowel	Mean duration in L1	Identification and goodness rating	Fit index
1. biːb	195 (22)	i	107 (8)	100 (5.0)	5.0
biːp	145 (9)			100 (5.4)	5.4 [a]
2. bɪb	114 (6)	i	107 (8)	100 (5.4)	5.4 [a]
bɪp	98 (12)			100 (5.0)	5.0
3. beb	138 (10)	e	116 (09)	97 (5.0)	4.8
bep	112 (9)			92 (5.2)	4.9
4. bɜːb	213 (10)	e	116 (09)	87 (3.0)	2.6
bɜːp	188 (8)			77 (3.2)	2.5
5. bæb	179 (18)	a	122 (10)	95 (4.7)	4.5
bæp	140 (7)			95 (5.2)	5.0 [a]
6. bʌb	133 (10)	a	122 (10)	62 (4.2)	2.6
bʌp	111 (13)			66 (4.1)	2.7
7. bɑːb	219 (16)	o	119 (9)	57 (4.0)	2.2
bɑːp	191 (10)			54 (4.1)	2.2
8. bɒb	132 (7)	o	119 (9)	97 (5.0)	4.8
bɒp	110 (6)			97 (5.0)	4.8
9. bɔːb	211 (17)	o	119 (9)	55 (4.1)	2.3
bɔːp	178 (13)			52 (4.1)	2.3
10. bʊb	132 (9)	u	112 (7)	92 (4.0)	3.7
bʊp	104 (5)			84 (4.2)	3.5
11. buːb	202 (16)	u	112 (7)	82 (3.8)	3.1
buːp	166 (21)			92 (3.8)	3.5 [a]

[a] $p < 0.05$ higher fit index in the context where vowel duration in L2 was closer to vowel duration in L1.

To examine whether context affected the assimilation of English vowels to Greek vowel categories, the fit indexes derived for the 11 English vowels were submitted to a two-way repeated-measures analysis of variance (ANOVA) with Vowel (11 levels) and Context (2 levels) as factors. The ANOVA yielded a significant main effect of Vowel [$F(10,170) = 25.3$; $p < 0.001$] and a significant Vowel × Context [$F(10,170) = 10.1$; $p < 0.001$] interaction. The significant effect of Vowel confirmed that English vowels varied in their overall fit to L1 categories (range = 2.2 to 5.4). The significant Vowel × Context interaction indicated that context affected how well English vowels fitted to L1 categories but this effect was not uniform across contexts. This initial analysis suggests that Greek listeners attend to both spectral and durational cues when perceiving English vowels.

To further analyze the effect of context on assimilation patterns, paired samples *t* tests (each with df = 17) compared the fit indexes derived for each English vowel in two consonantal contexts (significance level set to $p < 0.005$ to correct for multiple comparisons). The *t* tests showed that four English vowels differed in their fit to L1 categories as a function of context; English /iː, æ, uː/ fitted better in /bVp/ context while English /ɪ/ fitted better in /bVb/ context. Greek listeners preferred (as indicated by a higher fit index) these four English vowels in the context where the mean vowel duration in L2 was closer to the mean vowel duration in L1; English /iː, uː, æ/ in their 'short' version and English /ɪ/ in its 'long' version. Although context did not affect the fit indexes for all English vowels it is important to note that most of the vowels that fitted equally well to Greek vowel categories across consonantal contexts had either a mean duration that was equally close in either context to the mean duration in L1 (e.g. English /ʌ/) or were generally judged as being 'poor' examples of an L1 category (e.g. English /ɑː/ and /ɔː/). Regarding the latter case, it seems that if an L2 vowel did not spectrally match an L1 category well, a better fit in duration did not significantly change the listener's identification and/or goodness rating judgement.

2.4 Experiment 2: L2 discrimination

The purpose of this experiment was to examine Greek listeners' discrimination of English vowels in /bVb/ and /bVp/ contexts. Discrimination was assessed by means of a categorical discrimination test often used in L2 perception studies (e.g. Guion *et al.*, 2000; Aoyama *et al.*, 2004; Flege & Mackay, 2004). According to the feature hypothesis, Greek listeners should not have access to durational cues in L2 vowel perception and hence their discrimination of English vowels should not be affected by the context in which vowels are presented to the listeners. According to the desensitization hypothesis on the other hand, Greek listeners should be able to use duration in L2 vowel perception and hence their discrimination should be affected by consonantal context. Given the cross-language perceptual data obtained in Experiment 1, an additional question addressed in this experiment was whether cross-language perceptual assimilation patterns predicted L2 discrimination as proposed by Best's PAM.

2.4.1 Participants

Participants were the same as in Experiment 1. Ten English university students (mean = 25.3 years, range = 18-28 years) all born in London were also tested as controls.

2.4.2 Stimuli

The eleven English vowels described in Section 2.2.1 were combined to create nine contrastive vowel pairs /iː/-/ɪ/, /ɪ/-/e/, /æ/-/ʌ/, /æ/-/ɑː/, /æ/-/ɜː/, /ʌ/-/ɑː/, /ɒ/-/ɔː/, /ʊ/-/uː/ and /ɔː/-/uː/. Contrast selection was based on previous findings for Spanish learners of English whose system is very similar to the Greek one (Flege *et al.*, 1995; Flege *et al.*, 1997a; Iverson & Evans, 2007b). An effort was made to use contrasts that would vary in degree of discrimination difficulty.

2.4.3 Procedure

Greek listeners participated in Experiment 2 after completing Experiment 1 using the same laptop and headphones. In each trial of the categorical discrimination test, listeners were presented with three items, each spoken by a different native English speaker. Each contrast was tested by eight 'different' trials that contained an odd vowel category and eight 'catch' trials that contained three tokens of the same vowel category. The participants were instructed to identify the odd item out by clicking '1', '2' or '3' (in the 'different' trials) or 'same' when all the vowel instances were judged to belong to the same category. They were also asked to ignore differences in speakers' voices and to focus on vowel identity. The inter-stimulus interval (ISI) was 1.2 sec and the inter-trial interval (ITI) was 3 sec. To minimize response bias, A' scores (Snodgrass *et al.*, 1985) were computed for each contrast based on *hits*, when the odd item was correctly selected in 'different' trials and *false alarms*, when an item was incorrectly selected in 'catch' trials. If H (hit) = FA (false alarm) then A' = 0.5. If $H > FA$ then A' = $0.5 + [(H\text{-}FA)*(1+H\text{-}FA)]/[(4*H)*(1\text{-}FA)]$ and if $H < FA$ then A' = $0.5 - [(FA\text{-}H)*(1+FA\text{-}H)]/[(4*FA)*(1\text{-}H)]$. A' score of 1.0 indicates perfect discrimination of a contrast, whereas A' score of 0.5 indicates discrimination at chance level. Before the experiment began, a 20-item practice session (20 trials randomly selected) was presented to familiarize listeners with the procedure in a context that was different from the context to be tested first. As in experiment 1, written instructions were given in Greek before testing.

2.4.4 Results

Table 2.2 shows the accuracy with which Greek speakers discriminated the nine English vowel contrasts in two contexts. A' scores were firstly submitted to a two-way repeated-measures ANOVA with Contrast (9 levels) and Context (2 levels) as factors. The ANOVA yielded a significant main effect of Contrast [$F(8,136) = 13.5$; $p < 0.001$], demonstrating that discrimination scores varied considerably across English contrasts and a significant effect of Context [$F(1,17) = 25.2$; $p < 0.001$], demonstrating Greek listeners' sensitivity to vowel duration changes. A significant

Table 2.2: Mean duration of vowels (ms) in each English contrast, vowel duration ratio (longer-to-shorter), duration difference between vowels (ms) and mean discrimination scores obtained by Greek listeners.

English contrast	Mean duration (ms)	Vowel duration ratio (longer-to-shorter)	Duration difference (ms)	A' score
1. biːb vs. bɪb	195 vs. 114	1.71	81	0.79
biːp vs. bɪp	145 vs. 98	1.48	47	0.80
2. bɪb vs. beb	114 vs. 138	1.21	26	0.99
bɪp vs. bep	98 vs. 112	1.14	14	0.97
3. bæb vs. bʌb	179 vs. 133	1.35	46	0.74*
bæp vs. bʌp	140 vs. 111	1.26	29	0.51
4. bæb vs. bɑːb	179 vs. 219	1.22	40	0.83
bæp vs. bɑːp	140 vs. 191	1.36	51	0.80
5. bæb vs. bɜːb	179 vs. 213	1.19	34	0.92
bæp vs. bɜːp	140 vs. 188	1.34	48	0.96
6. bʌb vs. bɑːb	133 vs. 219	1.65	86	0.71*
bʌp vs. bɑːp	111 vs. 191	1.72	80	0.56
7. bɒb vs. bɔːb	132 vs. 211	1.60	79	0.83
bɒp vs. bɔːp	110 vs. 178	1.62	68	0.84
8. bʊb vs. buːb	132 vs. 202	1.53	70	0.70*
bʊp vs. buːp	104 vs. 166	1.60	62	0.61
9. bɔːb vs. buːb	211 vs. 202	1.04	9	0.74
bɔːp vs. buːp	178 vs. 166	1.07	12	0.75

* $p < 0.05$ higher discrimination than that obtained in the other context.

Contrast × Context [$F(8,136) = 6.6$; $p < 0.001$] interaction showed that the effect of context on discrimination was not uniform across contrasts. These results are consistent with the results of Experiment 1 and indicate that Greek listeners make use of both spectral and durational cues when discriminating L2 vowels.

To further explore the effect of context on English vowel discrimination, paired samples *t* tests were used comparing the A' scores obtained for each English contrast in /bVb/ and /bVp/ context (significance level set to $p < 0.005$ to correct for multiple comparisons). These comparisons showed that context significantly affected the discrimination of three out of nine English contrasts for Greek listeners (Table 2.2); Greek listeners showed a better discrimination for English /æ/-/ʌ/, /ʌ/-/ɑː/, and /ʊ/-/uː/ in /bVb/ than in /bVp/ context.

The next step was to see whether context-induced changes in the duration difference between the vowels of those pairs could explain the observed differences in discrimination, in other words whether the context that showed higher discrimination also provided the listeners with more temporal information than the other context did. An inspection of Table 2.2 reveals that in two of those contrasts (English /ʌ/-/ɑː/ and /ʊ/-/uː/) that was not the case since the difference in duration between the vowels in those pairs was relatively constant across contexts, i.e. less than 10 ms which is unlikely to be noticed by listeners in syllables whose vowels have the durations of those in this work (in fact the longer-to-shorter vowel ratio in these contrasts is smaller in the context that showed better discrimination). Additionally, there was at least one English contrast (/iː/-/ɪ/) that did provide the listeners with considerably more temporal information in /bVb/ than in /bVp/ context, i.e. 81 ms vs. 47 ms respectively (longer-to-shorter duration ratio 1.71 vs. 1.48) but showed similar discrimination accuracy across contexts. Greek listeners' discrimination was about as accurate in this case as for the /ɒ/-/ɔː/ contrast, which only differed by 11 ms (79 ms vs. 68 ms) and had similar duration ratios (1.60 vs. 1.62). These examples demonstrate that listeners do not simply compare the duration of the two vowels in a pair when trying to distinguish between the two vowels. If that was the case, more temporal information should always facilitate discrimination performance.

Finally, it was investigated whether perceptual assimilation patterns (Experiment 1) predicted discrimination accuracy (Experiment 2). The nine English contrasts tested in Experiment 2 were assigned to PAM categories based on the perceptual assimilation data obtained in Experiment 1. First, the cross-language identification percentages defined whether an English vowel was consistently identified with a

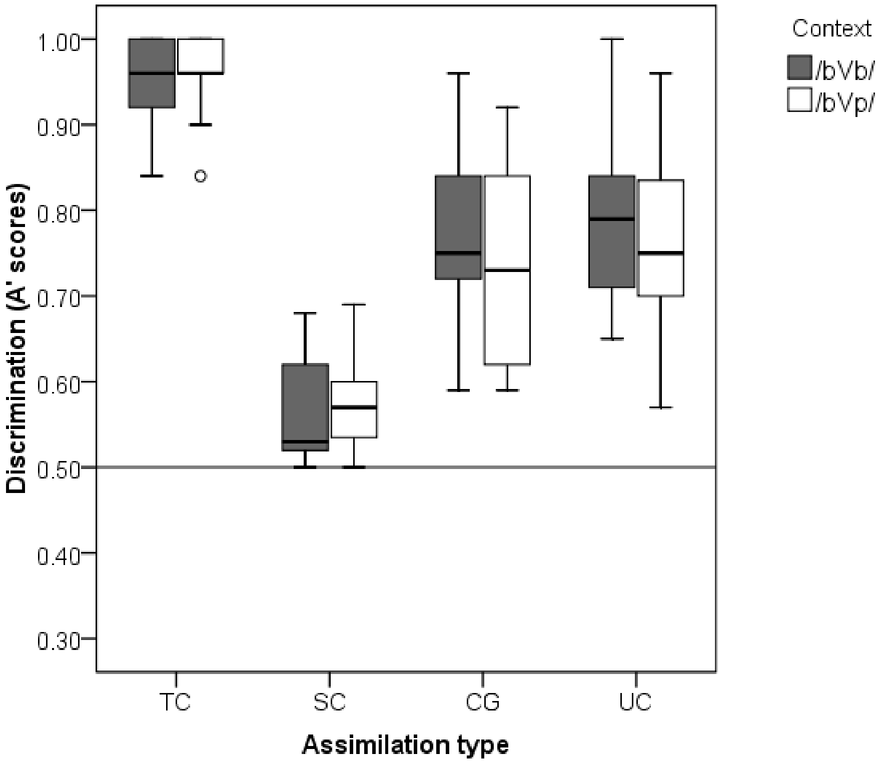


Figure 2.5: Boxplots of English vowel discrimination accuracy for Single-Category (SC), Category Goodness (CG), Uncategorized-Categorized (UC), and Two-Category (TC) assimilation types averaged over nine English contrasts in two consonantal contexts by Greek listeners. Whiskers extend to at most 1.5 times the interquartile range of the box. A' score of 0.5 indicates discrimination at chance level.

single Greek category or was heard as falling between two Greek categories (an Uncategorized sound according to PAM). A 60% identification criterion was adopted (Harnsberger, 2001 discusses the much higher identification criterion of 90%, but this results in most non-native sounds being classed as uncategorized). When both English vowels in a contrast were identified with the same Greek vowel, paired sample *t* tests defined whether that contrast would qualify as a Category Goodness contrast (i.e. the fit indexes of the two vowels differed significantly) or a Single-Category contrast (i.e. the fit indexes of the two vowels did not differ significantly, with significance level set to $p < 0.005$).

Figure 2.5 shows the mean discrimination scores obtained by Greek listeners for each assimilation type in two consonantal contexts, averaged over all English vowel contrasts. Native English control listeners obtained excellent discrimination scores (mean = 0.96 to 0.98) across vowel contrasts and hence their A' scores will not be discussed further. A one-way ANOVA examined the effect of assimilation type on A' scores obtained by Greek listeners. The effect of assimilation type was significant [$F(3,320) = 267.5$; $p < 0.001$]. Tukey post-hoc comparisons of the four assimilation types showed the following results: Two-Category contrasts were easier than Uncategorized-Categorized, Category Goodness and Single-Category contrasts, as PAM would predict with listeners obtaining generally very high scores in Two-Category contrasts. Uncategorized-Categorized and Category Goodness contrasts were easier than Single-Category contrasts again as expected with the latter being the most difficult contrasts to discriminate. Although there was a trend of discrimination scores for Uncategorized-Categorized contrasts being higher than Category Goodness contrasts, this difference was not significant (note the large variability in scores after averaging over contrasts).

2.5 Selection of vowel pairs in Greek and English

One of the goals of study 1 was to provide data on the perceived relationship between English and Greek vowels that would be used to select two English and two Greek vowel pairs. These vowel pairs would test Greek speakers' perception of synthetic L2 and L1 vowels in Study 3 thus evaluating the predictions of the *L1 phonetic* hypothesis. It was therefore important to select vowel pairs that would cover similar areas in the acoustic/perceptual space in English and Greek. Based on the perceptual assimilation and discrimination results from Experiments 1 and 2 respectively the following comparable contrasts were selected for English and Greek:

- English /iː/-/ɪ/ vs. Greek /i/-/e/
- English /æ/-/ʌ/ vs. Greek /a/-/o/

The English front tense-lax contrast /iː/-/ɪ/ is probably the most difficult contrast for learners with small vowel systems and no tense-lax or long-short vowel distinctions (e.g. Spanish, Italian) and has been widely tested in the L2 literature. According to the results of Experiment 1, both /iː/ and /ɪ/ were assimilated, across consonantal contexts, to Greek /i/ 100% of the time. Additionally, both /iː/ and /ɪ/ received high goodness ratings; /iː/ received a mean rating of 5 in /bVb/ context (hence an overall fit index of 5) and a mean rating of 5.4 in /bVp/ context (hence an overall fit index of 5.4) while for /ɪ/ the picture was reversed. Discrimination results for this contrast showed that Greek listeners were able to distinguish /iː/ from /ɪ/ in both consonantal contexts on the basis of duration. Since both English /iː/ and /ɪ/ were assimilated to Greek /i/ and the closest vowel to Greek /i/ is /e/, the Greek /i/-/e/ contrast was selected as the most comparable to English /iː/-/ɪ/.

The English low /æ/-/ʌ/ contrast is also considered a challenging one for learners with small vowel systems that contain a single /a/ category in their inventory. English /æ/ was assimilated to Greek /a/ 95% of the time across consonantal contexts and received fairly high goodness ratings across contexts (although Greek listeners preferred it when placed in a /bVp/ context, that is when its duration was closer to that of Greek /a/). English /ʌ/ was assimilated to Greek /a/, however, not as consistently as English /æ/ both in terms of percent identification (62% of the time in /bVb/ and 66% of the time in /bVp/ context) and goodness ratings (4.2 in /bVb/ and 4.1 in /bVp/ context). Since both English /æ/ and /ʌ/ were assimilated to Greek /a/ (with English /ʌ/ being assimilated about 30% of the time to Greek /o/, see Table I in Appendix A), the Greek /a/-/o/ contrast was selected as the most comparable to English /æ/-/ʌ/. Given the assimilation pattern for English /ʌ/, it was predicted that discrimination in synthetic /æ/-/ʌ/ will be easier than in /iː/-/ɪ/ for Greek listeners at least when discrimination is performed on the basis of spectral information.

2.6 Summary of results

Study 1 examined the perceptual assimilation and discrimination of English vowels by native speakers of Greek (Experiments 1 and 2 respectively). The same participants performed both experiments. The goal of the two experiments was threefold. The first goal was to explore, for the first time, the perceived relationship between Greek and English vowels. A practical aspect was to select specific vowel pairs in the two languages to be further examined in following experiments with synthetic vowels. The second goal was to examine whether Greek listeners have access to durational cues when perceiving the English vowels given that duration is not used in Greek vowel distinctions. The third goal was to test discrimination predictions for English vowel contrasts (Experiment 2) based on cross-language perceptual assimilation data (Experiment 1).

As discussed in the Introduction, the availability of durational cues to listeners with no such L1 experience has been a matter of debate over the past years with studies arriving at different conclusions with respect to this issue. The two main proposals in the literature are represented by the feature hypothesis, proposed by McAllister *et al.* (2002) and the desensitization hypothesis, proposed by Bohn (1995). The former hypothesis is based on the notion of L1 transfer when learning an L2 and posits that L2 learners do not have access to cues that are not used in L1 to signal contrasts. The latter hypothesis posits that L2 learners are sensitive to durational cues when perceiving L2 vowels irrespective of the status of duration in their L1 and that, in fact, learners tend to rely more on durational than spectral cues when faced with difficult L2 contrasts (Escudero, 2005; Cebrian, 2006 among others). To assess the predictions of the two hypotheses, the vowel stimuli in both experiments were embedded in /bVb/ and /bVp/ contexts. That way, the effect of vowel duration differentiations introduced by the voicing vs. voicelessness of the stop consonant following the vowel on Greek speakers' performance was tested.

The results of Experiment 1 showed that Greek speakers' perceptual assimilation of English vowels to L1 categories was affected by consonantal context. This finding seems to be in disagreement with the feature hypothesis and, at first glance, in support of the desensitization hypothesis. However, when looking more closely at between-context comparisons conducted for each English vowel separately it was found that English vowels generally fitted better to L1 categories in the context where they resembled more the duration of the spectrally closest L1 vowel. This suggests that L2 learners assimilate both temporally and spectrally L2 vowels to L1 categories and hence duration does not seem to have a special status as compared to spectral properties in L2 vowel perception (Bohn, 1995). The observed patterns of assimilation reflect a spectral and temporal 'matching' to the L1 categories irrespective of whether the L1 has a phonemic vowel length contrast or not.

Discrimination performance in Experiment 2 was generally consistent with the predictions made by Best's PAM (Best, 1995). Greek listeners had no difficulty with Two-Category contrasts, had some difficulty with Uncategorized-Categorized and Category Goodness contrasts and found Single-Category contrasts the most difficult to discriminate. The discrimination scores for Uncategorized-Categorized contrasts were somewhat lower than predicted and did not differ significantly from those obtained for Category Goodness contrasts. Guion *et al.* (2000) report on a similar finding in their data and propose a possible revision of PAM regarding the discriminability of Uncategorized-Categorized contrasts where the uncategorized sound is close in the perceptual space to the categorized one.

The results regarding the effect of context on the discrimination of English vowels showed that Greek listeners were sensitive to durational cues. Again, this seems to run contra the feature hypothesis and in favour of the desensitization hypothesis. However, paired comparisons conducted for each English contrast separately indicated that L2 learners were not simply comparing the durations of the two members in a pair when trying to distinguish one from another. There were contrasts which proved to be easier in one context than the other despite the fact that the duration difference between the two vowels was similar across contexts. There were also contrasts where context-induced changes in the duration difference between the two vowels did not result in changes in discrimination performance. Lengeris (2009) compared the perception of Southern British English vowels by Greek and Japanese native speakers. Both Greek and Japanese employ five relatively similar quality distinctions in their vowel system but differ greatly with regard to the use of duration in contrasting vowels; the Japanese vowel system contains five short (one-mora) and five long (two-morae) vowels /i, e, a, o, u/ and /iː, eː, aː, oː, uː/ respectively (in Standard Japanese the high back vowel is unrounded /ɯ/, see for example Shibatani, 1990). The short and the long vowels are almost identical in terms of spectral characteristics with the former being approximately 50% shorter (Shibatani, 1990; Hirata, 2004). Using the same perceptual stimuli and testing procedures as in here, it was found that Japanese speakers too perceived English vowels via spectral and temporal assimilation to their L1 categories. This confirms the uniformity in the use of durational cues by two language groups that differ fundamentally in the use of duration in L1 (Iverson & Evans, 2007b). The major difference between the two groups in terms of temporal assimilation patterns lay in the fact that Japanese speakers assimilated each English vowel to a short or long vowel category whereas Greek speakers assimilated it to their 'single' duration category.

Finally, based on the results of experiment 1 and 2 the following comparable contrasts were selected for English and Greek: (1) English /iː/-/ɪ/ vs. Greek /i/-/e/ and (2) English /æ/-/ʌ/ vs. Greek /a/-/o/. Since English /ʌ/ was occasionally assimilated (about 30% of the time) to Greek /o/ it is predicted that English /æ/-/ʌ/ will suffer less from L1 spectral interference than English /iː/-/ɪ/, an L2 contrast where both vowels are assimilated 100% of the time to a single L1 vowel category by Greek speakers (Greek /i/).

Chapter 3

Greek best exemplar locations

The following experiment was designed to find best exemplar locations of the five Greek vowels /i, e, a, o, u/ in a 5-dimensional space that included F1 and F2 formant movement (i.e. onset and offset of the F1 and F2 frequencies) and duration. The locations of all vowels except /u/ would serve as endpoints for the Greek vowel continua /i/-/e/ and /a/-/o/ that would test Greek listeners' identification and discrimination of synthetic L1 (Greek) vowels in Study 3 (Chapters 4-6). The selection of the endpoints for the English vowel continua /iː/-/ɪ/ and /æ/-/ʌ/ that would test Greek listeners' identification and discrimination of synthetic L2 (English) vowels was based on another study using the same software and method (Iverson & Evans, 2007b). For consistency and clarity reasons, the procedures followed for finding the best exemplar locations of both Greek and English vowels will be presented in this chapter.

3.1 Method

3.1.1 Participants

Twelve native speakers of Standard Modern Greek (5 male and 7 female) with a mean age of 26 years (range = 23-30 years) were tested in Greece. They were all from Athens and had no regional accent. Subjects were tested in a quiet room using a laptop computer and high quality headphones. All of the participants reported no hearing or language impairment. In Iverson & Evans (2007b), seventeen native speakers of English (median age = 28 years, range = 18-49 years) were tested in the United Kingdom. They performed the best exemplars task in quiet rooms using PCs (desktops, laptops, and pocket PCs) and high quality headphones.

3.1.2 Model speakers

A male native speaker of Standard Modern Greek[4] was recorded in an anechoic chamber at UCL with a sampling rate of 44.1 kHz. A Sony 60ES DAT recorder with a B&K Sound Level Meter Type 2231 fitted with a 4165 microphone cartridge was used. Iverson & Evans (2007b) recorded a male native speaker of Southern British English in his late 20s in the same anechoic chamber using the same recording procedures.

3.1.3 Stimuli

The Greek speaker uttered the five Greek vowels /i, e, a, o, u/ in the context /pVta/ (stressed on the first syllable) embedded in the carrier sentence *Πες* ___ *ξανά* /pes ___ ksana/ 'say ___ again'. The particular context created a minimally contrastive set of 3 words, i.e. *πίτα* /ˈpita/ 'pie', *πέτα* /ˈpeta/ 'fly', 'throw', *πάτα* /ˈpata/ 'press', 'step on' and two non-words, i.e. **πότα* /ˈpota/ and **πούτα* /ˈputa/. The two non-words /ˈpota/ and /ˈputa/ are phonotactically acceptable in Greek and the use of these contexts was not expected to affect vowel productions in any way. The Southern British English speaker uttered 13 English vowels in the context /bVt/ embedded in the carrier sentence *Say* ___ *again* /seɪ ___ əgen/. That context created a minimally contrastive set of 13 words, i.e., *beat* /biːt/, *bit* /bɪt/, *bet* /bet/, *Burt* /bɜːt/, *bat* /bæt/, *Bart* /bɑːt/, *but* /bʌt/, *bot* /bɒt/, *bought* /bɔːt/, *boot* /buːt/, *bait* /beɪt/, *bite* /baɪt/, *bout*

[4] The speaker was the author who was 28 years old at the time of the recordings.

/baʊt/, *boat* /bəʊt/. Both speakers read each word four times at a normal speaking rate. They also recorded *the North Wind and the Sun* translated in Greek and English respectively.

The /pVta/ and /bVt/ contexts were selected so that subjects performing the best exemplar locations experiments as well subjects performing the tasks with synthetic vowels presented in following chapters would be faced with very similar (and at the same time phonotactically acceptable) structures in the two languages. Greek voiceless stops /p, t, k/ are unaspirated in all positions (Botinis *et al.*, 2000; Arvaniti, 2001, 2007); English voiced stops /b, d, g/ are phonetically realized as voiceless in initial position, that is although they are phonologically described as voiced [+voice], vocal fold vibration starts after the release burst (Docherty, 1992). Thus, Greek /p/ and English /b/ (in initial position) are phonetically realized quite similarly.

Before synthesizing the vowels to be used in the Greek best exemplars experiment, the sentences and the passage produced by the Greek speaker were normalized to a 'model' speaker in terms of their formant frequencies and median pitch. This was done using signal processing in Praat (Boersma & Weenink, 2005) following the procedures used in Iverson & Evans (2007b) to normalize the sentences and the passage produced by their English speaker. The materials were normalized to reduce any effect the vocal tract differences between the Greek and the English speaker might have on locating the best exemplars in the two languages. The formant frequencies were scaled using the F2 of /i/ of each speaker (averaged across the speaker's four repetitions of this vowel in the carrier sentences) because F2 is consistently produced across speakers and can be measured reliably. The sampling rate was changed to match the speakers' F2 of /i/ to 2290 Hz, which is an average value for male speakers (Peterson & Barney, 1952). The Greek speaker had an average F2 for /i/ of 2138 Hz so the sampling rate was changed from 44100 to 47235 Hz (the English speaker in Iverson & Evans, 2007b had an average F2 for /i/ of 2473 Hz so the authors changed the sampling rate from 44100 to 40843 Hz). After manually correcting for any errors introduced by changing sampling rate and rescaling duration, the pitch was scaled to 112 Hz, which is an average value for male speakers (Hazan & Markham, 2004). Finally, the stimuli were re-synthesized using pitch synchronous overlap and add (PSOLA) in Praat and down-sampled to 11025 Hz.

The 'hybrid' stimuli were made up of a synthetic vowel embedded within a (signal processed) natural sentence of the native speaker for each language including the initial release /p/ burst and the final /ta/ for Greek and the initial release /b/ burst and the final /t/ for English. The synthetic vowels were created using a Klatt synthesizer (Klatt & Klatt, 1990) in cascade/parallel configuration with a sampling rate of 11025 Hz and matched the natural ones in terms of F0 and amplitude. The rest of the synthesis parameters were kept the same across vowels in each language. These were the F4 and F5 frequencies (3500 and 4500 Hz respectively), the formant bandwidths (B1=100, B2=180, B3=250, B4=300, B5=550), the tilt (TL=0 dB slope) and the open quotient (OQ=60%). The F1 and F2 frequencies changed in a linear way from the beginning to the end of the vowel. F1 formant frequency ranged between 5 and 15 Equal Rectangular Bandwidth (ERB) (Glasberg & Moore, 1990). F2 formant frequency started from 10 ERB, was at least 1 ERB higher than F1 and reached a limit that was defined by the equation F2 = 25 - (F1-5) /2. The synthetic vowels were 1 ERB apart from each other and their durations spanned logarithmically in 7 steps (54, 75, 104, 144, 200, 277 and 383 ms). Overall, 109,375 vowels were synthesized in each language.

3.1.4 Procedure

The Greek native speakers found best exemplar locations of all five Greek vowels /i, e, a, o, u/. Although Greek /u/ would not used as endpoint in any of the Greek synthetic vowel continua in Study 3, it was decided to map its location in the perceptual space for two reasons: first, since the participants would perform the task for four Greek vowels the whole process would be more balanced if Greek /u/ was also included; second, apart from providing values for the synthetic vowel continua endpoints, this experiment provided an excellent opportunity to collect perceptual data on the entire Greek vowel system.

Testing began with the participants listening to the Greek version of *The North Wind and the Sun* to familiarize themselves with the characteristics of the speaker's voice. During the experiment, the participants saw on the screen a /pVta/ structure written both in Greek orthography and Roman alphabet (e.g. ***πίτα*** and **pita** respectively) and heard a synthetic vowel stimulus embedded in the carrier sentence (*Πες* ___ *ξανά* /pes ___ ksana/ 'say ___ again'). They had to rate how close the vowel stimulus was to a good exemplar of the vowel displayed on the screen by clicking on a continuous bar (see Figure 3.1 for the experiment interface).

A goodness optimization method (Iverson & Evans, 2003; Evans & Iverson, 2004, 2007; Iverson & Evans, 2007b, 2009) was adopted to find best exemplar locations of the Greek vowels. During testing, an algorithm would search along 7 vectors (straight-line paths cutting through the five-dimensional space) so that the best exemplar on each vector would be found after 5 trials per vector. The whole process thus required just 35 trials for each vowel despite the large number of synthesized vowels available to listeners and was completed in about half an hour. Vector 1 was designed to locate an approximation of listeners' best exemplar by passing through the location of the natural production of the target vowel and the middle of the vowel space (F1 = 500 Hz, F2 = 1500 Hz) without varying duration; Vector 2 only varied duration while all other parameters were fixed; Vector 3 only varied the F1 and F2 onset frequencies keeping all other parameters fixed while Vector 4 was orthogonal to Vector 3 in the F1/F2 onset space; Vector 5 only varied the F1 and F2 offset frequencies while Vector 6 was orthogonal to Vector 5 in the F1/F2 offset space; finally, Vector 7 simultaneously varied all five dimensions and listeners were able to fine-tune their best exemplar location for each vowel. On the first two trials for each vector, listeners would hear the most extreme stimuli synthesized along the vector. On the remaining three trials, stimuli were selected based on the listeners' previous judgments, i.e. by weighting listeners' best exemplar locations and goodness ratings thus far (for a full description of the procedure, see Iverson & Evans, 2007b).

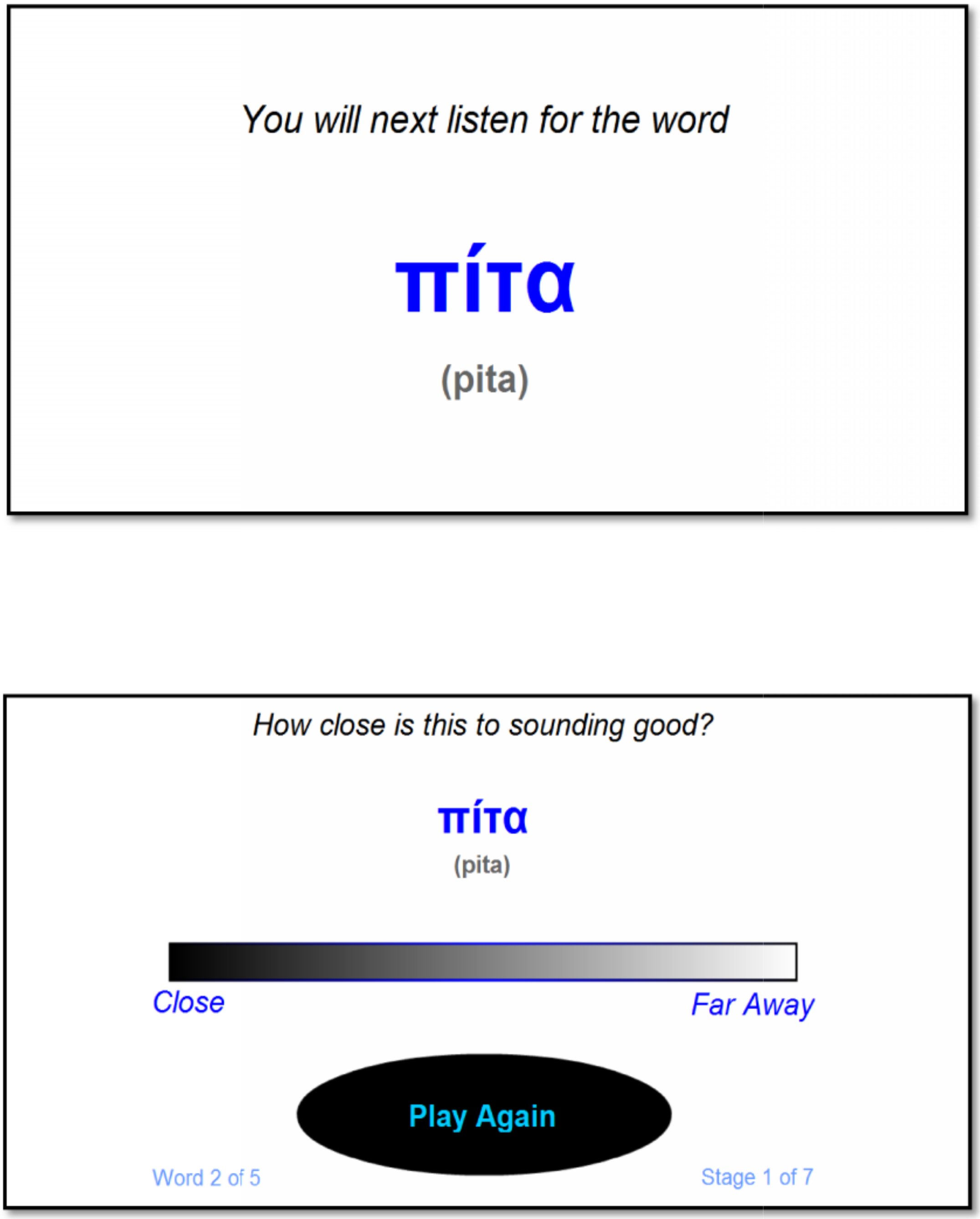


Figure 3.1: Screen shots of the experiment interface in the Greek best exemplar locations experiment.

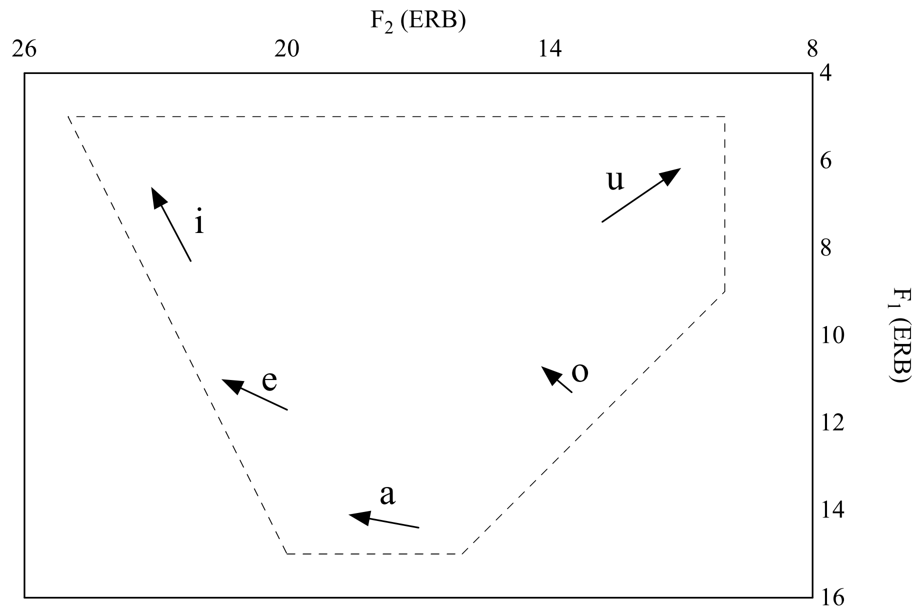


Figure 3.2: Mean best exemplar locations of the five Greek vowels. The arrows show the direction of the F1 and F2 formant movement for each vowel, i.e. the onset and offset of the F1 and F2 formant frequencies.

3.2 Results

Figure 3.2 shows the mean best exemplar locations of the five Greek vowels. As can be seen, Greek vowels are well separated in the perceptual space with no overlap between vowels (Haws and Fourakis, 1995; Botinis *et al.*, 1997). Interestingly, there is some evidence of formant movement although that is not particularly large; all vowels except /u/ moved towards the /i/ corner of the vowel space. This formant movement was unexpected given that Greek vowels are traditionally described as monophthongs. Table 3.1 presents the best exemplar locations (ERB and Hz) of Greek vowels. Table 3.2 presents mean F1 and F2 values (Hz) for Greek vowels taken from Fourakis *et al.* (1999) and Nicolaidis (2003) for comparison. Greek best exemplar locations are clearly hyper-articulated compared to production data which is typical for exemplar/prototypical values (e.g. Johnson *et al.*, 1993).

Table 3.1: Mean best exemplar locations (ERB and Hz) of the five Greek vowels.

	F1 onset		F1 ending		F2 onset		F2 ending	
Vowel	ERB	Hz	ERB	Hz	ERB	Hz	ERB	Hz
/i/	8.3	330	6.6	236	22.2	2265	23.1	2518
/e/	11.7	577	11.0	518	20.0	1739	21.5	2084
/a/	14.4	848	14.1	814	17.0	1196	18.6	1464
/o/	11.3	543	10.7	494	13.5	749	14.2	825
/u/	7.4	278	6.2	215	12.8	678	11.0	518

Table 3.2: Mean F1 and F2 values (Hz) from Fourakis *et al.* (1999) and Nicolaidis (2003).

	Fourakis *et al.* (1999)		Nicolaidis (2003)	
	Read speech		Spontaneous speech	
	F1 (Hz)	F2 (Hz)	F1 (Hz)	F2 (Hz)
/i/	340	2046	360	1892
/e/	491	1788	475	1672
/a/	738	1350	575	1453
/o/	508	1020	462	1202
/u/	349	996	377	1163

3.3 Summary of results

Study 2 aimed at defining the endpoints for the two Greek vowel continua /i/-/e/ and /a/-/o/ that would test Greek listeners' perception of synthetic L1 (Greek) vowels in the following chapters of this thesis. This was done using a best exemplars experiment (Iverson & Evans, 2003; Evans & Iverson, 2004, 2007; Iverson & Evans, 2007b, 2009). The endpoints for the English vowel continua that would test Greek listeners' perception of synthetic L2 (English) vowels were based on Iverson & Evans (2007b). Iverson & Evans (2007b) used the exact same procedures to find English native speakers' best exemplar locations of English vowels. For consistency reasons and, most importantly, in order to obtain perceptual data of the entire Greek vowel system (for a discussion on the need to study the perception of Greek vowels using a variety of experimental protocols, see Arvaniti, 2007) it was decided to include Greek /u/ in the experiment although it would not be used in any of the synthetic continua employed in following experiments.

Participants mapped their best exemplars in a 5-dimensional space that included F1 and F2 formant movement and duration using a goodness optimization method (Iverson & Evans, 2003; Evans & Iverson, 2004, 2007; Iverson & Evans, 2007b, 2009). The method located best exemplars of the Greek vowels after 35 trials despite the large set of synthesized vowels (109,375 vowels in total). It was found that Greek vowels are well separated in the perceptual space confirming previous research (Haws & Fourakis, 1995; Botinis *et al*., 1997). There was some evidence of formant movement although that was not particularly large. All vowels were, as expected, hyper-articulated compared to typical productions of Greek vowels (Iverson & Evans, 2003 also found that best exemplar locations for Southern British English vowels were hyper-articulated compared to the natural productions of their 'model' speaker).

Chapter 4

Training experiment

This chapter presents the test battery completed by Greek native speakers participating in Study 3. The tasks employed evaluated the predictions of the *L1 phonetic* hypothesis and the *auditory processing* hypothesis by testing the participants' perception of natural and synthetic vowels in L1 (Greek) and L2 (English) and of non-speech as well as their production of L2 vowels. One group (trained group) performed the same tasks before and after receiving high-variability auditory training, also described in this chapter. Another group (control group) performed the same tasks before and after a period of time similar to that required for the trained group to complete the training (about two weeks). The use of the control group aimed at assessing not only the effectiveness of high-variability training but also the effect of mere learning in pre/post-test that would come from test repetition. The test battery included the following 9 tasks (all tasks except for tasks 3 and 7 were run in quiet conditions):

1) Identification of synthetic Greek vowels
2) Discrimination of synthetic Greek vowels
3) Identification of natural Greek vowels in noise
4) Identification of synthetic English vowels
5) Discrimination of synthetic English vowels
6) Identification of natural English vowels
7) Identification of natural English vowels in noise

8) English vowel production
9) Discrimination of a non-speech continuum

Tasks 1-2 tested Greek speakers' processing of Greek vowel categories (identification boundaries and slopes and discrimination accuracy) using analytical test procedures in an attempt to reveal individual differences in L1 vowel perception. Performance in these tasks would help in evaluating the *L1 phonetic* hypothesis. Task 3 was used as another test of vowel category robustness that might reveal individual differences in the processing of natural L1 vowels. The use of noise was necessary to avoid ceiling effects in performance. Tasks 4-5 tested Greek speakers' processing of English vowel categories using the exact same analytical procedures as with Greek vowels (identification boundaries and slopes and discrimination accuracy) in an attempt to reveal individual differences in L2 vowel perception. Task 6 served as a baseline measure of L2 vowel identification accuracy. Task 7 served as a baseline measure of L2 vowel identification accuracy under adverse listening conditions, resembling more L2 perception in naturalistic settings. Task 8 served a baseline measure of L2 vowel production accuracy. Finally, task 9 tested Greek speakers' frequency discrimination accuracy using a signal that was a non-speech analog of second formant frequency, thus evaluating the *auditory processing* hypothesis. Since all 9 tasks were employed before and after auditory vowel training, it was possible to evaluate the effects of training on all 9 tasks, taking into consideration pre/post performance in the control group, who received no intermediate tuition.

4.1 Participants

Twenty eight adult native speakers of Greek participated in total. Eighteen were trained and ten served as controls, i.e. performed the pre/post tests but received no training. The trainees (8 male and 10 female) had a mean age of 23 years (range = 18-35 years) and the controls (4 male and 6 female) had a mean age of 26 years (range = 18-42 years). All participants spoke Standard Modern Greek with no regional accent and were tested in Athens. The majority (24/28) were recruited from two language schools; their English proficiency level was rather high and relatively uniform across individuals (Cambridge FCE, Cambridge CAE). Participants had 10-12 years of formal English instruction in Greece by L1-accented instructors and had very little, if any, interaction with native speakers of English. None of the participants had spent a period of more than one month in an English-speaking environment as shown in a language questionnaire completed by all participants before testing. The participants passed a pure-tone hearing screening at frequencies from 250 to 4000 Hz at 20 dB SPL. They were all paid for their participation with the fee being proportionate to the time spent on the study; the trainees received £20 while the controls received £7.

4.2 Perceptual stimuli

4.2.1 Pre/post materials

All perception tasks and English vowel production elicitation procedures were the same in pre/post tests with one exception: post-training identification of natural English vowels in quiet and in noise included a new English speaker. The inclusion of an English speaker who had not been used in either the pre-test or the training materials tested generalization of learning.

4.2.1.1 Natural vowels

Two native speakers of Standard Modern Greek (1 male, 1 female) in their 20s produced the five Greek vowels in a /pVs/ context. The particular context created a minimally contrastive set, i.e. *πεις* /pis/ 'to say', *πες* /pes/ 'say' (imperative), *πας* /pas/ 'to go', πως /pos/ 'how', πους /pus/ 'foot'. The duration of each word was around 400 ms (the exact duration varied according to the intrinsic duration of the vowel). The speakers produced three repetitions of the five Greek words, for a total of 30 stimuli. The best two repetitions for each vowel produced by each speaker were chosen by the author to be included in the experiment.

Two native speakers of Southern British English (1 male, 1 female) in their 20s produced ten English monophthongal vowels (all monophthongs except /ʊ/) in a /bVt/ context. The particular context created a minimally contrastive set, i.e. *beat* /biːt/, *bit* /bɪt/, *bet* /bet/, *bat* /bæt/, *but* /bʌt/, *Bart* /bɑːt/, *Burt* /bɜːt/, *bot* /bɒt/, *bought* /bɔːt/, *boot* /buːt/. The duration of each word was around 550 ms (the exact duration varied according to the intrinsic duration of the vowel). The speakers produced three repetitions of the ten English words, for a total of 60 stimuli. Another female speaker of Southern British English in her 20s produced the same materials to be used only in the generalization test. The best two repetitions for each vowel produced by each speaker were chosen by the author and a native speaker of Southern British English to be included in the experiment. Ideally, consonantal context should be kept constant across languages; however, the /pVs/ context is one of the very few contexts that creates a minimally contrastive set of real words containing all 5 vowels in Greek.

4.2.1.2 Synthetic vowel continua

Two Greek and three English synthetic vowel continua were created using the exact same synthesis method and parameters described in previous chapter for creating the set of synthetic vowels used in the best exemplars experiments; the Greek continua spanned from Greek /i/ to Greek /e/ and from Greek /a/ to Greek /o/ and the English continua spanned from English /iː/ to English /ɪ/ and from English /æ/ to English /ʌ/. As previously mentioned, the endpoints of the synthetic Greek vowel continua were based on the Greek best exemplars experiment in Study 2 and the endpoints of the synthetic English vowel continua were based on the English best exemplars experiment in Iverson & Evans (2007b). The synthetic vowels were embedded within natural consonantal contexts; the Greek /i/-/e/ continuum was embedded in a natural /ˈpVta/ context (the initial release /p/ burst and the final /ta/ were taken from the natural sentence recorded for the best exemplars experiment); the Greek /a/-/o/

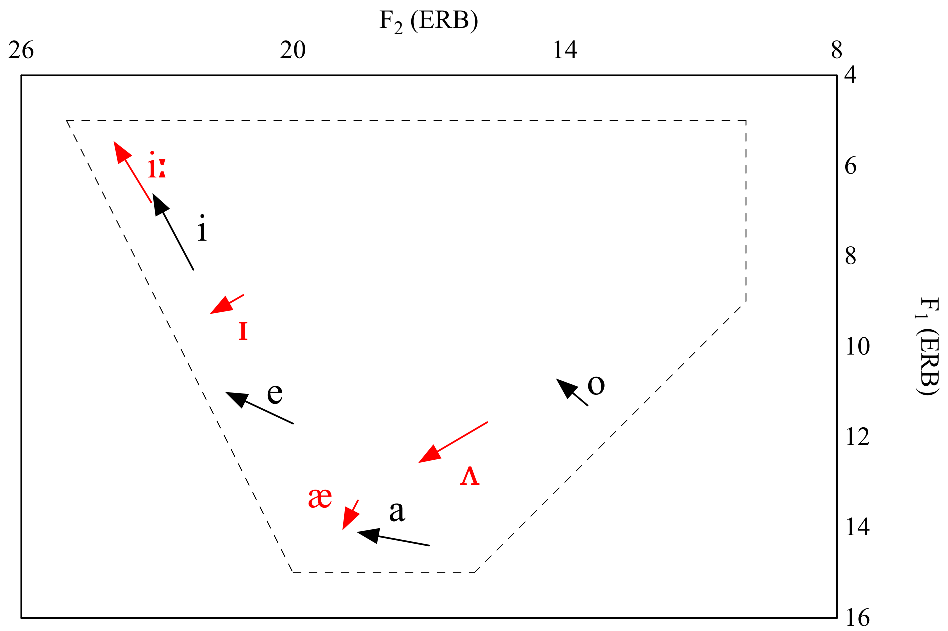


Figure 4.1: Location of Greek synthetic vowels (in black) and English synthetic vowels (in red) in the vowel space.

continuum was embedded in a natural /ˈpVte/ context (the initial release /p/ burst and the final /te/ were taken from a natural sentence recorded for this purpose and signal-processed following the procedure described in the previous chapter). The use of two contexts was necessary since there is no minimal pair in Greek contrasting all 4 vowels in either a /ˈpVta/ or a /ˈpVte/ context. The two contexts yielded two minimal pairs, i.e. /pita/ 'pie' - /peta/ 'throw' (/i/-/e/ continuum), and /pate/ 'to go' - /pote/ 'when' (/a/-/o/ continuum). The English /iː/-/ɪ/ and /æ/-/ʌ/ continua were embedded in a natural /bVt/ context (the initial release /b/ burst and the final /t/ were taken from the natural sentence recorded for the best exemplars experiment in Iverson & Evans, 2007b).

Based on the duration values obtained in the Greek best exemplars experiment, vowel duration in the Greek /i/-/e/ continuum was set to 55 ms and that in the Greek

Table 4.1: F1 and F2 beginning and end values and duration for the endpoints of the 5 vowel continua used (two Greek and three English continua).

Continuum	Endpoint	F1 beg (Hz)	F1 end (Hz)	F2 beg (Hz)	F2 end (Hz)	Duration (ms)
1. /i/-/e/	/i/	330	236	2265	2518	55
	/e/	577	518	1739	2084	
2. /a/-/o/	/a/	848	814	1196	1464	65
	/o/	543	494	749	825	
3. /iː/-/ɪ/	/iː/	247	182	2527	2785	110
natural	/ɪ/	364	392	1986	2170	70
4. /iː/-/ɪ/	/iː/	247	182	2527	2785	90
neutralized	/ɪ/	364	392	1986	2170	
5. /æ/-/ʌ/	/æ/	701	809	1458	1521	85
	/ʌ/	574	651	1011	1233	

Natural = natural duration
Neutralized = neutralized duration

/a/-/o/ continuum was set to 65 ms. There were two versions of the English /iː/-/ɪ/ continuum: in the 'natural duration' condition, /iː/ had a duration of 110 ms and /ɪ/ had a duration of 70 ms while in the 'neutralized duration' condition the duration was set to 90 ms, a duration intermediate to that of the 'natural duration' condition. A comparison of performance in the two duration conditions would show the weight given by each listener to the vowel duration cue. Vowel duration in the English /æ/-/ʌ/ continuum was set to 85 ms. Duration values in English continua were somewhat different from those in Iverson & Evans (2007b), however, they were closer to the English vowel production data obtained in Study 1. Further, a slight modification was made to the location of the /ʌ/ endpoint compared to its location in Iverson & Evans (2007b). Iverson & Evans (2007b) included some Northern British English speakers in their study which resulted in /ʌ/ being somewhat higher in the vowel space than we would expect for Southern British English speakers. The final endpoint stimuli were judged to be excellent examples of Greek and English vowels by the author and a native Southern British English phonetician respectively. Vowel endpoint stimuli are plotted in the vowel space in Figure 4.1. There were 51 stimuli in each vowel continuum varying in 50 equal steps in Hertz in terms of differences in F1 and F2 formant frequencies (and duration for the /iː/-/ɪ/ *natural duration* continuum). Table 4.1 presents F1 and F2 and duration values for the endpoints in all vowel continua used. Figure 4.2 and Figure 4.3 show the endpoint stimulus 0 /pita/ and the endpoint stimulus 50 /peta/ in the Greek /i/-/e/ continuum. Figure 4.4 and Figure 4.5 show the endpoint stimulus 0 /biːt/ and the endpoint stimulus 50 /bɪt/ in the English /iː/-/ɪ/ (*neutralized duration*) continuum.

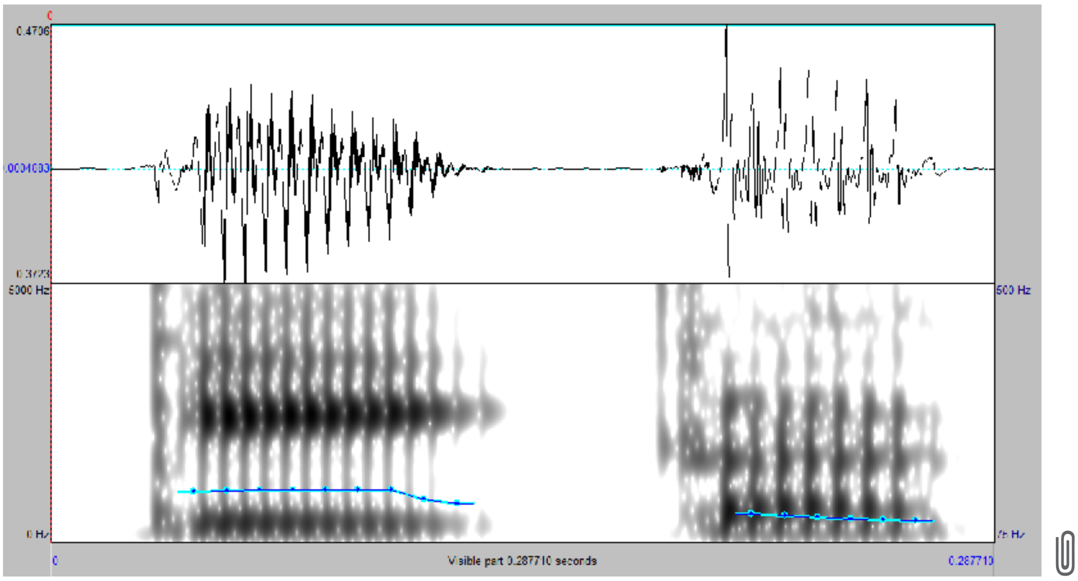


Figure 4.2: Waveform and spectrogram of the 'hybrid' stimulus 0 /pita/ in the Greek /i/-/e/ continuum. The first vowel was synthetic while the rest of the phones were natural.

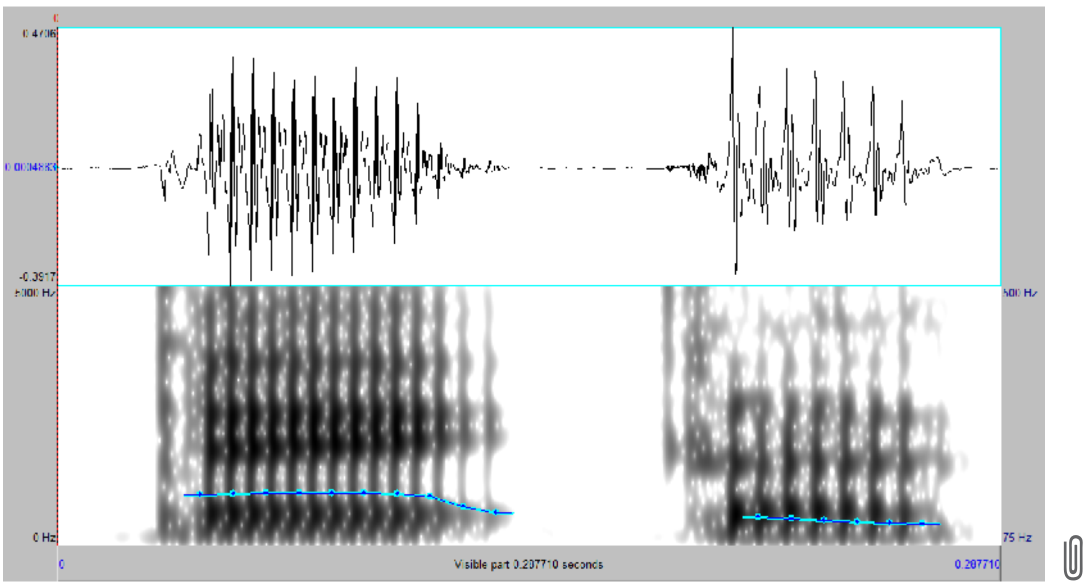


Figure 4.3: Waveform and spectrogram of the 'hybrid' stimulus 50 /peta/ in the Greek /i/-/e/ continuum. The first vowel was synthetic while the rest of the phones were natural.

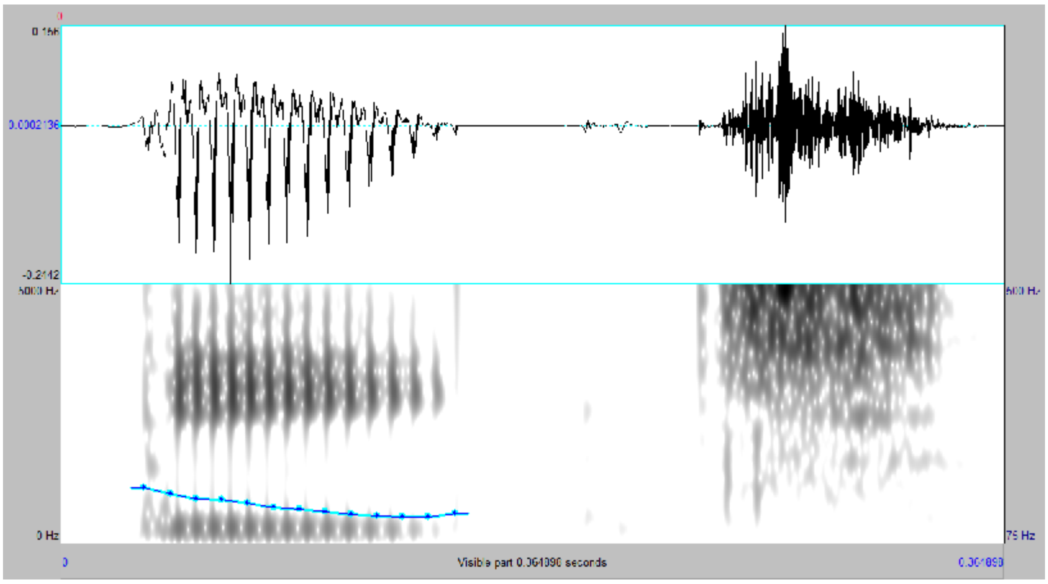


Figure 4.4: Waveform and spectrogram of the ‘hybrid’ stimulus 0 /biːt/ in the English /iː/-/ɪ/ (*neutralized duration*) continuum. The first vowel was synthetic while the rest of the phones were natural.

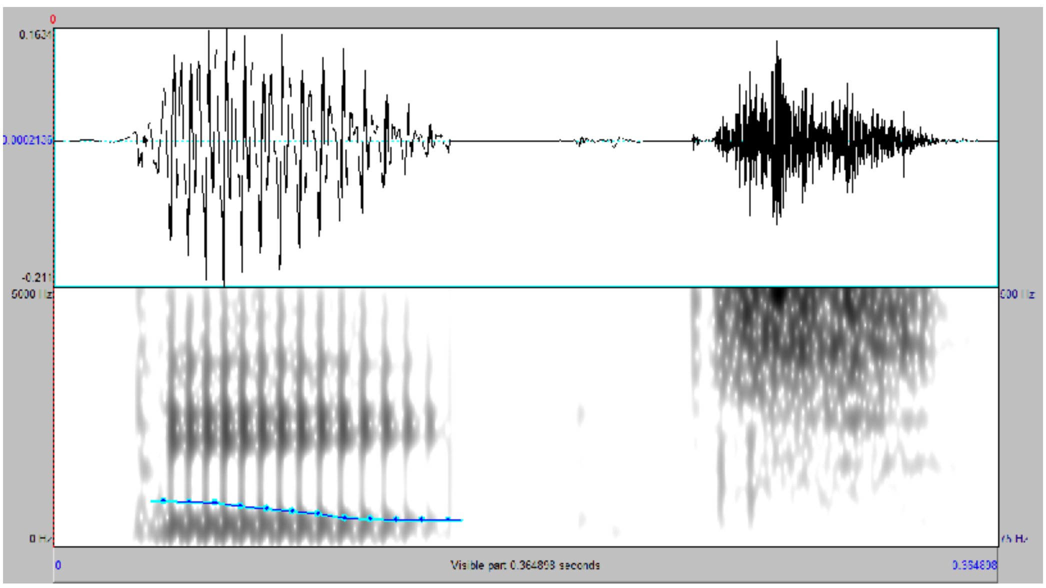


Figure 4.5: Waveform and spectrogram of the ‘hybrid’ stimulus 50 /bɪt/ in the English /iː/-/ɪ/ (*neutralized duration*) continuum. The first vowel was synthetic while the rest of the phones were natural.

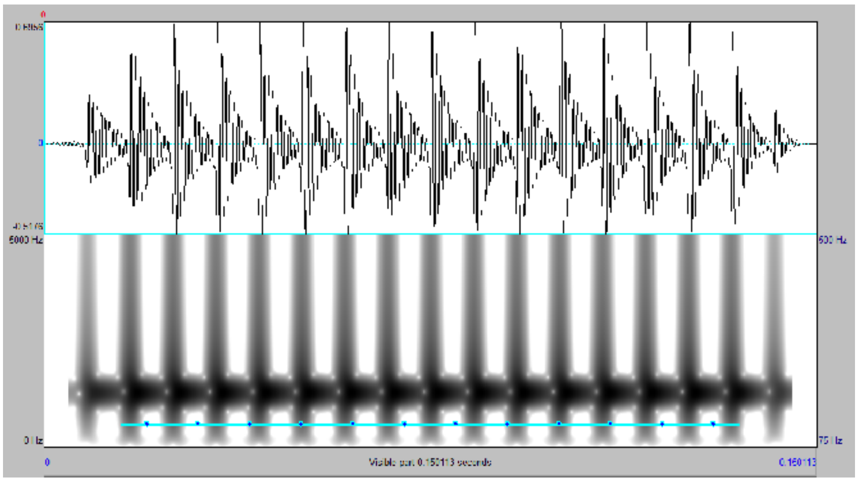


Figure 4.6: Waveform and spectrogram of non-speech endpoint stimulus 0 (F2 = 1250 Hz).

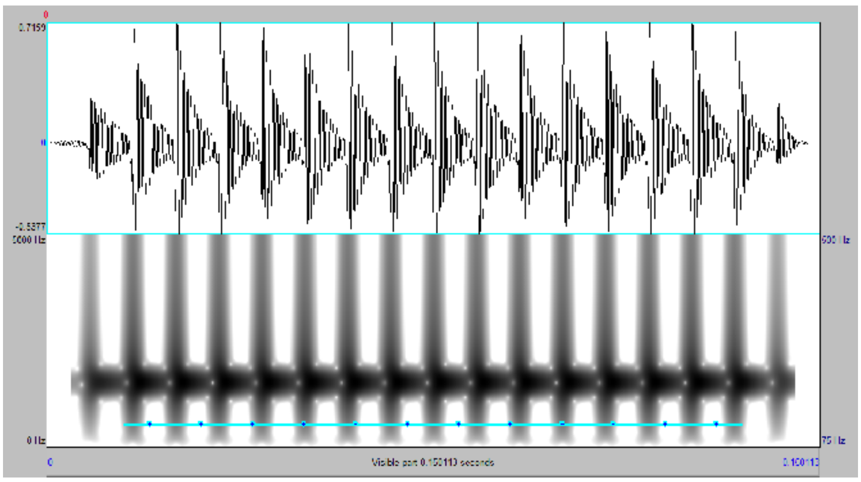


Figure 4.7: Waveform and spectrogram of non-speech endpoint stimulus 50 (F2 = 1500 Hz).

4.2.1.3 Non-speech continuum

The non-speech continuum had a single formant frequency that spanned from 1250 to 1500 Hz (and thus was a non-speech analogue to a vowel second formant). Its duration was 150 ms and its pitch was set constant at 120 Hz (thus resembling the pitch of a male speaker). It was therefore decided to use a non-speech continuum that would share similar acoustic properties with the speech continua (i.e. harmonic structure, similar duration and pitch) without being speech. To reduce the possibility of the particular type of non-speech being treated as speech, the non-speech task was the first task completed by Greek participants (see also Section 4.3). Waveforms and spectrograms of non-speech endpoint stimuli 0 and 50 are shown in Figure 4.6 and Figure 4.7 respectively.

4.2.2 Training

The training software, stimuli and procedures were exactly the same as in Iverson & Evans (2007a) and Iverson & Evans (2009). These were real English words containing 14 English vowels spoken by five native speakers of Southern British English (2 male, 3 female). The training stimuli included all 10 vowels used in the pre/post tests and four diphthongs that were not used in the pre/post tests. Given that very few minimal pair sets in English contrast all 14 vowels, words were arranged in four minimal-pair groups /iː ɪ aɪ eɪ/ (e.g. *heat*, *hit*, *height*, *hate*), /uː aʊ ɜː/ (e.g. *blues*, *blouse*, *blurs*), /ɒ əʊ ɔː/ (e.g. *stock*, *stoke*, *stork*), and /e æ ɑː ʌ/ (e.g. *mesh*, *mash*, *marsh*, *mush*). Iverson & Evans (2007a) and Iverson & Evans (2009) arranged the 14 vowels in the above groups after conducting a hierarchical cluster analysis on identification data by native Spanish and German speakers obtained in Iverson & Evans (2007a); the first three groups contained vowels which were problematic for both Spanish and German speakers and the last group contained the remaining vowels. Given the similarity of the Greek and the Spanish vowel systems and the discrimination results for Greek speakers obtained in Study 1, Greek speakers were expected to face similar difficulties with these four groups of vowels. For each of the four groups, Iverson & Evans (2007a) and Iverson & Evans (2009) selected ten sets of minimal pair words, giving a total number of 140 different words which ensured the large variability of the training stimuli. These vowels were recorded twice by each English speaker in an anechoic chamber at UCL with a sampling rate of 44.1 kHz and then down sampled to 11.025 kHz.

4.3 Procedure (pre/post tests)

Participants were tested in Greece in quiet rooms using a laptop and high-quality headphones. Each participant carried out all tasks in a single session lasting about 1½ hours. As previously mentioned, the non-speech discrimination task was employed first. Further, all tasks with English vowels preceded those with Greek vowels. Identification tasks with natural vowels were run using Praat. Identification and discrimination tasks with synthetic vowels were run using Glimpse and Sparedux respectively (both programs were developed at UCL Department of Human Communication Science and Department of Phonetics and Linguistics). The test battery was presented with the following order:

1. Discrimination of a non-speech continuum
2. Identification of natural English vowels in quiet
3. Identification of natural English vowels in noise
4. English vowel production
5. Identification and discrimination of synthetic English vowels; each vowel pair was first identified and then discriminated, with the following order used: (1) English /iː/-/ɪ/ *natural duration*, (2) English /æ/-/ʌ/, and (3) /iː/-/ɪ/ *neutralized duration*
6. Identification of natural Greek vowels in noise
7. Identification and discrimination of synthetic Greek vowels; each vowel pair was first identified and then discriminated, first Greek /i/-/e/ and then Greek /a/-/o/

4.3.1 Pre/post materials

4.3.1.1 English vowel perception in quiet and in noise

The natural English /bVt/ words described in section 4.2.1.1 were presented within a forced-choice identification task. Participants heard an English word through headphones at a comfortable listening level and chose one of the ten options as displayed on a computer screen (using English orthographic labels, e.g. *beat*, *bit*, *bet*). Each English vowel option included a common English word containing the same vowel. In the noise condition, multi-talker babble (mixed recordings from 20 different speakers at approximately equal levels) was played simultaneously with the natural English /bVt/ words at an SNR of -4 dB. The level of noise was decided after running a short pilot test with 4 native speakers of Greek who had just moved to London to study and whose level of English experience was comparable to the speakers that would be tested in Greece. After trying different SNRs (from -2 to -6 dB), an SNR of -4 dB was selected yielding percent correct accuracy of about 40%. The noise started about 200 ms before the beginning of the word and ended about 100 ms after the end of the word. For each condition, the total number of presentations was 40 in the pre-test (2 speakers × 10 vowels × 2 repetitions) and 60 in the post-test (3 speakers × 10 vowels × 2 repetitions). Vowel presentations were blocked by speaker and in each block vowels were fully randomized. Before performing the identification task in quiet, listeners heard all of the words spoken by one speaker once together with their orthographic labels; the same was done before identification in noise.

4.3.1.2 Greek vowel perception in noise

The natural Greek /pVs/ stimuli described in section 4.2.1.1 were presented within a forced-choice identification task. Participants heard a Greek word through headphones at a comfortable listening level and chose one of the five options as displayed on a computer screen (using Greek orthographic labels, e.g. *πεις* /pis/, *πες* /pes/, *πας* /pas/) for a total of 20 presentations (2 speakers × 5 vowels × 2 repetitions). Vowel presentations were blocked by speaker and in each block vowels were fully randomized. The multi-talker babble was played simultaneously with the natural Greek /pVs/ words at an SNR of -10 dB. The level of noise was decided after running a short pilot test with 5 native speakers of Greek where different SNRs were tried (from -4 to -12 dB). Although individual differences were observed even within this small sample and were expected to occur in the actual test, at an SNR of -10 dB mean percent correct accuracy for these 5 Greek speakers was about 75%. That was

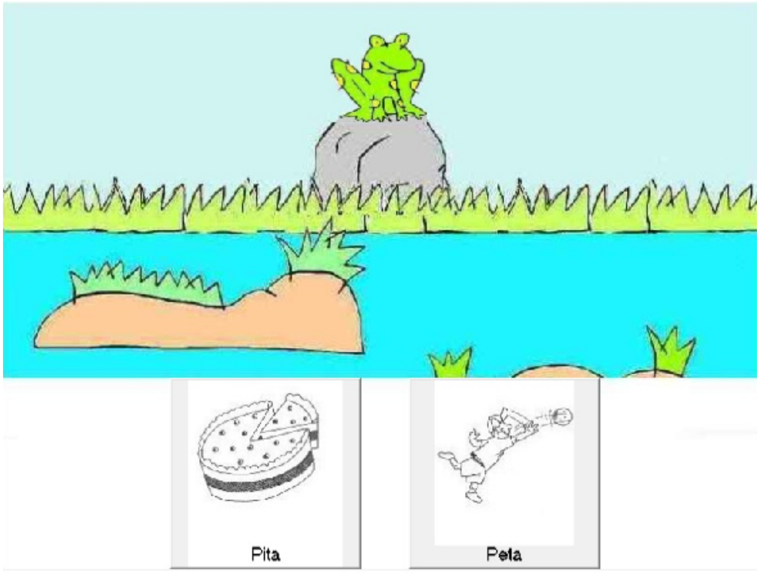


Figure 4.8: Experimental interface for the 2AFC identification task showing the two alternatives, /pita/ 'pie' (left picture) and /peta/ 'throw' (right picture) in the Greek /i/-/e/ continuum.

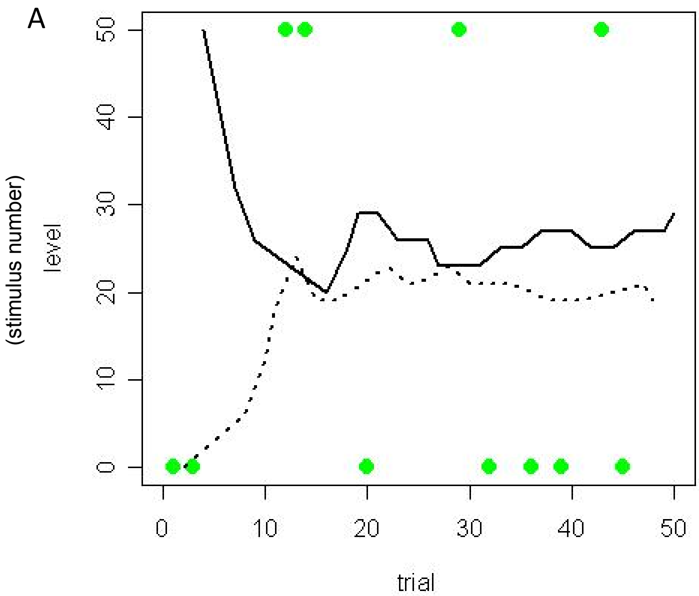
a different level of accuracy to that selected for English. However, the level of noise required to obtain an intelligibility level of 40% in Greek would be so high that it would increase the possibility of the two tasks tapping into different processing abilities. The noise started about 200 ms before the beginning of the word and ended about 100 ms after the end of the word. Before testing, listeners heard all of the words spoken by one speaker once together with their orthographic labels.

4.3.1.3 Identification tasks

The participants' identification boundaries and slopes were tested on the five synthetic vowel continua (2 Greek and 3 English) described in section 4.2.1.2, /pita/-/peta/, /pate/-/pote/, /biːt/- /bɪt/ *natural duration*, /biːt/-/bɪt/ *neutralized duration*, and /bæt/-/bʌt/. Identification boundary defines the point in the continuum where the two vowel responses are equally probable, i.e. the phoneme boundary; identification slope measures the consistency with which a listener is categorizing the continuum.

Identification was assessed by means of a two-alternative forced-choice (2AFC) task. One frog appeared on the screen 'saying' one stimulus from the continuum. The participants were asked to identify the vowel by clicking on a button displaying the stimulus word as well as a picture of that word (see Figure 4.8 for the experimental interface used in the identification of the /pita/-/peta/ continuum and Appendix B for all pictures used). Pictures were used together with printed words to reduce any effects from orthography on vowel perception. The stimuli were presented using an interleaved adaptive procedure in order to focus presentations in the region of most interest, i.e. near the phoneme boundary. Two independent adaptive tracks started at opposite ends of the continuum and estimated the point on the continuum where the stimuli were labeled as a given word (either /pita/ or /peta/ to use the same example as before) 71% of the time using a 2-down/1-up rule (Levitt, 1971). To prevent listeners from continuously hearing ambiguous stimuli when performing the task, 20% of the trials were stimuli taken from the endpoints of the continuum. The test ended after 7 reversals or 50 trials. Figure 4.9 shows a complete run of an individual on the /pita/-/peta/ identification task. For each listener and vowel pair, logistic regression was used to obtain a best-fit sigmoid function from all test trials and estimates of the identification boundary and slope were calculated from the fitted coefficients.

A



B

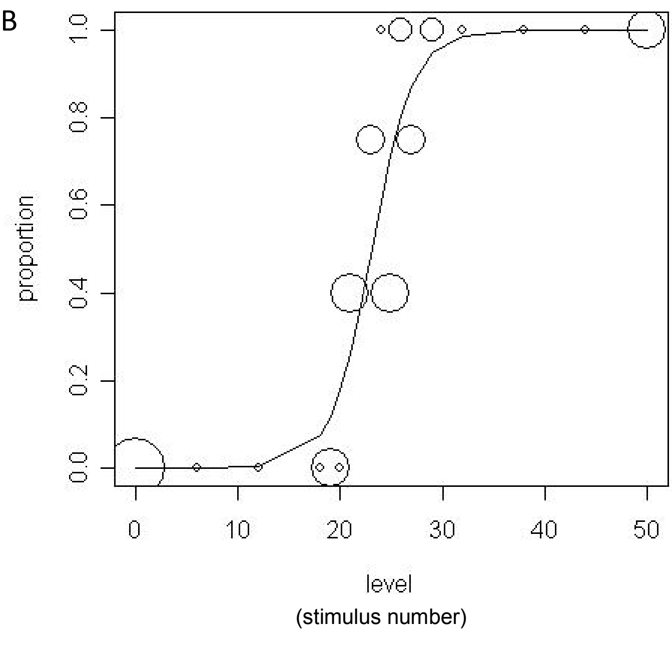


Figure 4.9: Complete run by an individual on the /pita/-/peta/ identification task showing the two independent adaptive tracks (A) and the identification function which is derived from the tracks (B). In A, the green dots indicate successful labelling of endpoint stimuli. In B, the size of the circle at a particular step shows the total number of stimuli presentations at that step.

4.3.1.4 Discrimination tasks

The participants' discrimination was tested on the five synthetic vowel continua (2 Greek and 3 English) described in section 4.2.1.2, /pita/-/peta/, /pate/-/pote/, /biːt/-/bɪt/ *natural duration*, /biːt/-/bɪt/ *neutralized duration*, and /bæt/-/bʌt/ and on the non-speech continuum described in section 4.2.1.3. Discrimination was assessed by means of a three-alternative forced-choice (3AFC) task. The experimental interface was similar to that presented in the previous section for identification, however, this time three frogs appeared on the screen with each frog 'saying' one stimulus from the continuum. The participants were told that two of the words were the same while one was different from the two and that they should indicate the different one by clicking the appropriate frog which could be in any of the three positions. Feedback was provided in the form of a tick or an x mark above the selected frog, indicating a correct or a wrong answer respectively.

A method of 'standard' was used against which the other stimuli were compared. The standard was one endpoint of the continuum (the first vowel in each of the five continua and the 1250 Hz endpoint in the non-speech continuum). A 3-down/1-up rule was used (Levitt, 1971) which found the just noticeable difference (jnd), i.e. the

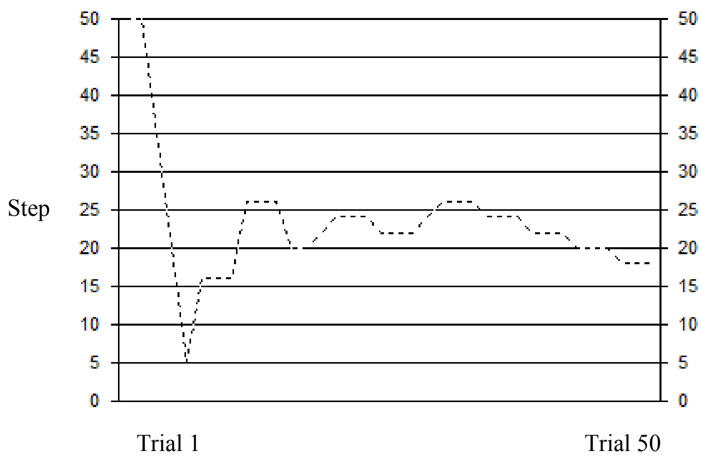


Figure 4.10: Complete run by an individual on the /pita/-/peta/ discrimination task. Step 0 is the /pita/ endpoint and step 50 is the /peta/ endpoint. The jnd is relative to step 0.

stimulus that could be discriminated from the standard 79% of the time. The test ended after 7 reversals or 50 trials. The mean of the last four reversals defined the jnd. The inter-stimulus interval (ISI) was 250 ms. Figure 4.10 shows a complete run of an individual on /pita/-/peta/ discrimination.

4.3.1.5 English vowel production task

The participants read from a screen one at a time the 10 English words they had previously attempted to identify. The words were produced in isolation. Thus, although the participants had heard the English vowels to be produced before, the task was not a direct-imitation one. Recordings were made using a MicroTrack 24/96 digital recorder in a quiet room at a sampling rate of 44.1 kHz.

In order to get a quality rating for vowel production by L2 learners, two native speakers of Southern British English in their 20s first identified each vowel from a forced-choice set of 10 English categories (all English monophthongs except /ʊ/) and then, after hearing the same vowel once more, rated its goodness in a scale from 1 (very bad example) to 7 (excellent example). Each English speaker performed 560 judgments (28 Greek speakers × 10 vowels × 2 repetitions) with vowels fully randomized.

4.3.2 Training

The training programme consisted of a pre-test phase, a training phase and a post-test phase (e.g. Logan *et al.*, 1991). The training procedure was the same as in Iverson & Evans (2007a) and Iverson & Evans (2009). The trainees completed five sessions of high-variability auditory training (vowel identification with feedback) each consisting of 225 trials with a different talker each session. A short session consisting of 14 trials was given before training to familiarize trainees with the procedure. Training was partly adaptive; the first 70 trials were 5 random repetitions of the 14 English vowels, the next 85 were based on the participant's errors and the last 70 trials were again 5 random repetitions of the 14 English vowels. As discussed in Iverson & Evans (2009), in the adaptive trials the selection probability of a vowel was defined by combining the proportion of misses (i.e. the listener failed to select that vowel and instead chose another vowel) and false alarms (i.e. the listener chose incorrectly that vowel as a response). That way, a vowel that would prove difficult for the listener to perceive would be tested more times than a vowel that would prove easy. The participants could have a short break in the middle of each session (after trial 112). Training was conducted at the participants' homes; the training software was installed on their laptops and they were asked to do the training in a quiet room using headphones and to complete the 5 sessions in 2 weeks. Training information (subject details, training session, and date) was monitored without the participants having access to that information to ensure that participants completed all sessions in two weeks as asked.

During training the participants heard an English word and chose one of 3 or 4 candidates as displayed on a computer screen (see Figure 4.11 for the training interface). For each candidate a more common word was given in case the participants did not know the particular word. If the target word was correctly identified "Yes!" was displayed on the screen, a cash register sound was heard and the target word was repeated once. If the target word was not identified correctly "Wrong" was displayed on the screen, two beeps were heard and both the target and the (incorrectly) chosen word were repeated twice. That way the participants received immediate feedback which helped in learning the target vowel. Percent correct identification was displayed at the end of each session.

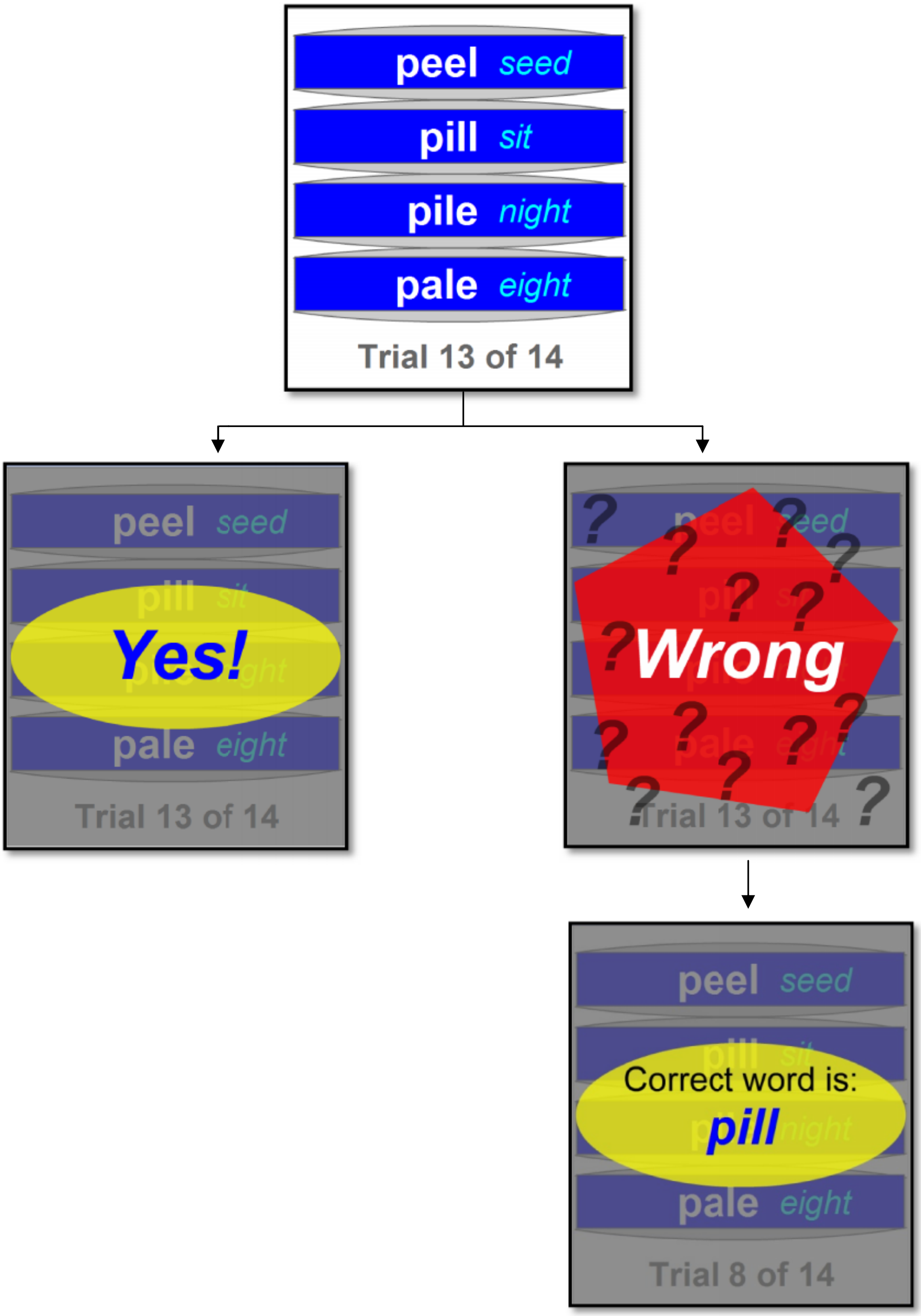


Figure 4.11: Screenshots of training interface.

Chapter 5

Pre-training results

This chapter presents the results for all tasks completed by Greek speakers in the pre-test. To ensure that the trained and the control group were well-matched prior to training, the two groups of Greek speakers are compared in terms of performance on those tasks. In order to evaluate the predictions of two hypotheses advanced in the Introduction, the *L1 phonetic* hypothesis and the *auditory processing* hypothesis, after reporting on group data, the chapter examines the relationships between and within tasks for all 28 participants tested together with individual data for those participants. The *L1 phonetic* hypothesis predicts that L1 and L2 vowel processing will be related for individuals and that frequency discrimination acuity will not relate to either L1 or L2 vowel processing. The *auditory processing* hypothesis predicts that frequency discrimination acuity, assessed by means of a non-speech discrimination task, will underlie L2 and, most likely, L1 vowel processing.

5.1 Natural vowels

5.1.1 Identification of English vowels in quiet and in noise

Figure 5.1 displays the interquartile range of percent identification scores obtained by the trained and the control group of Greek speakers in quiet and in noise (SNR =

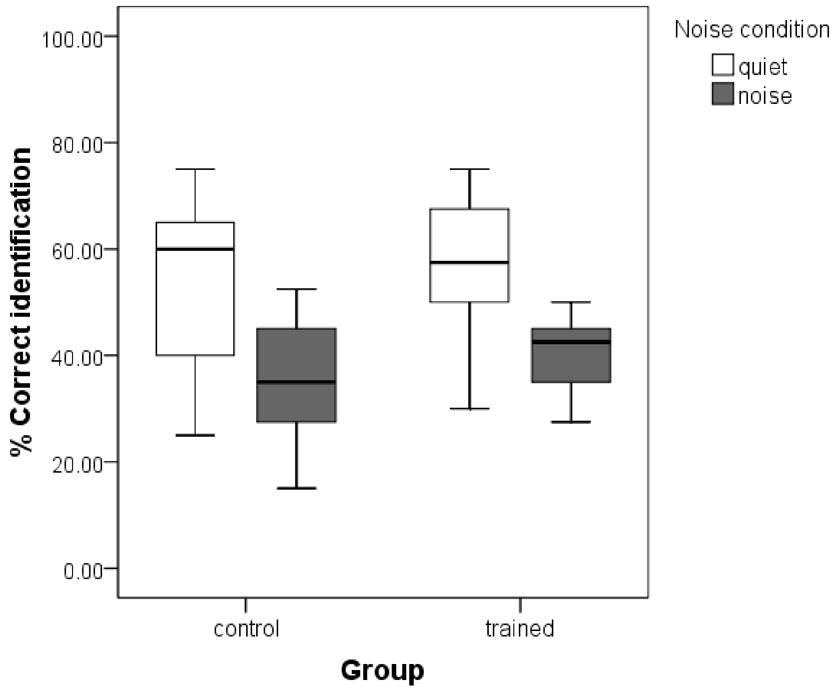


Figure 5.1: Boxplots showing the interquartile range of percent correct identification scores for natural English vowels by the control and the trained group of native speakers of Greek in quiet and in noise (SNR = -4 dB).

-4 dB) averaged across 10 English vowels. The two groups had very similar scores in quiet (trained: M = 56.9%; control: M = 55%) and in noise (trained: M = 40.3%; control: M = 35%) and for both groups identification accuracy was higher in quiet than in noise. Identification scores were submitted to a three-way repeated-measures ANOVA with Group (trained, control) as a between-subject factor and Noise condition (quiet, noise) and Vowel (10 vowels) as within-subject factors. The ANOVA revealed no main effect of Group [$F(1,26) = 0.64$, $p > 0.05$] which, combined with the fact that Group did not interact with any other factor, suggested that the two groups were well-matched (i.e. the control group was a good match for the group that received training in terms of pre-test identification accuracy both in quiet and in noise). The ANOVA also revealed significant main effects of Vowel [$F(9,234) = 4.74$, $p < 0.001$] and Noise condition [$F(1,26) = 84.36$, $p < 0.001$] and a significant Vowel × Noise condition interaction [$F(9,234) = 2.46$, $p < 0.05$], suggesting that the effect of Noise was not the same across English vowels; all English vowels were identified correctly at lower rates in noise than in quiet, however, a series of post hoc *t* tests revealed that this effect was significant ($p < 0.01$) for all vowels except English /ɪ/, /æ/, and /ɑː/.

Table 5.1 shows percentage identification responses for each English vowel in quiet averaged across the trained and control group. Correct identification ranged from 86% for /uː/ to 33% for /ɑː/. English /uː/ was in fact the only vowel that was identified correctly quite successfully by Greek speakers but this was probably due to the lack of a strong competitor /ʊ/ in the perceptual stimuli. With respect to vowel confusions, /iː/ was mostly confused with /ɪ/ and vice versa (/ɪ/ was also confused with /e/); /æ/ was mostly confused with /ʌ/ and vice versa; and /ɒ/ was mostly confused with /ɔː/ and vice versa. Table 5.2 shows percentage identification responses for each English vowel in noise averaged across the trained and control group. Correct identification ranged from 45% for /ɔː/ and /æ/ to 23% for /ɑː/. With respect to vowel confusions, /iː/ was mostly confused with /ɪ/ and vice versa; /æ/ was mostly confused with /ʌ/ and vice versa (/ʌ/ was also confused with /ɑː/); and /ɒ/ was mostly confused with /ɔː/ and vice versa. English vowel confusions were therefore fairly similar in quiet and noise with one exception: in noise condition, /uː/ was no longer the easiest vowel for Greek speakers to identify; its identification accuracy was severely affected by noise (from 86% correct in quiet to 39% correct in noise). Probably due to the absence of a strong competitor, /uː/ was confused with four English vowels /iː/, /ɪ/, /ɒ/, and /ɔː/ at rates between 10% and 14% (the confusion with /iː/ and /ɪ/ may be explained by the fact that English /uː/ is quite fronted compared to Greek /u/).

Table 5.1: Confusion matrix for English vowels in quiet identified by native speakers of Greek. Percentages of responses have been pooled over the trained and the control group. Identification responses <3% are not shown.

	Response									
Stimulus	iː	ɪ	e	ɜː	æ	ʌ	ɑː	ɒ	ɔː	uː
iː	**60**	38								
ɪ	26	**47**	23							
e	8		**56**	7	13	10	3			
ɜː	5		10	**57**		10	6			
æ				8	**54**	23	13			
ʌ				4	44	**42**	9			
ɑː				17	12	9	**33**	7	19	
ɒ							3	**54**	38	3
ɔː								22	**66**	7
uː		3							5	**86**

Table 5.2: Confusion matrix for English vowels in noise (SNR = -4 dB) identified by native speakers of Greek. Percentages of responses have been pooled over the trained and the control group. Identification responses <3% are not shown.

	Response									
Stimulus	iː	ɪ	e	ɜː	æ	ʌ	ɑː	ɒ	ɔː	uː
iː	**26**	33	3	6	4	9	6	3	4	6
ɪ	33	**43**	17				4			3
e	7		**40**	8	15	18	4		4	
ɜː	6		13	**40**	4	7	8	10	10	
æ			5	13	**45**	23	14			
ʌ				5	37	**27**	21	4	5	
ɑː			3	14	13	24	**23**	6	14	3
ɒ			5	6		3		**36**	40	6
ɔː				7		3	3	23	**45**	18
uː	12	13		4			3	10	14	**39**

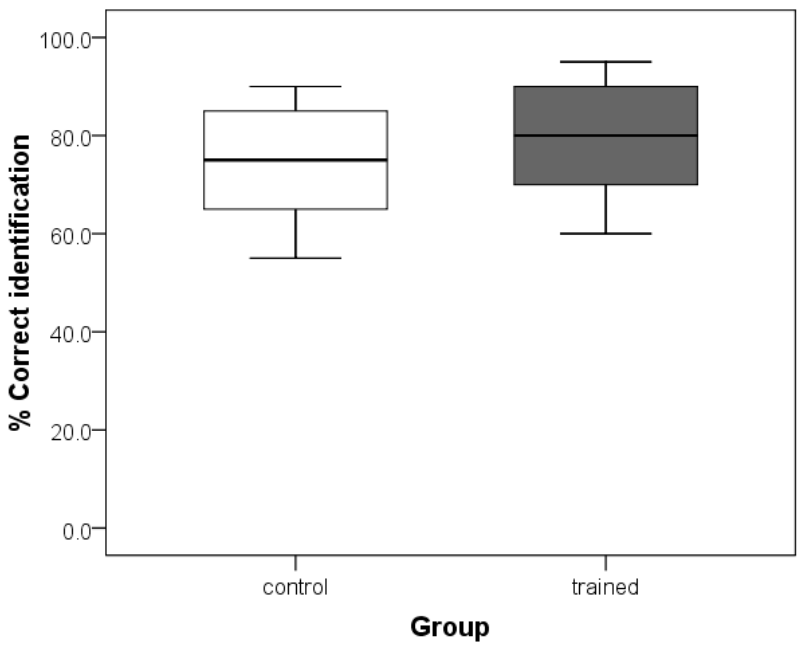


Figure 5.2: Boxplots showing the interquartile range of percent correct identification scores for Greek vowels by the control and the trained group of native speakers of Greek in noise (SNR = -10dB).

5.1.2 Identification of Greek vowels in noise

Figure 5.2 displays the interquartile range of percent identification scores obtained by the trained and the control group of native Greek speakers in noise (SNR = -10 dB) averaged across 5 Greek vowels. As can be seen, the two groups had very similar identification scores (trained: $M = 76.4\%$; control: $M = 72.5\%$). Identification scores were submitted to a two-way repeated-measures ANOVA with Group (trained, control) as a between-subject factor and Vowel (5 vowels) as a within-subject factor. The ANOVA revealed no main effect of Group ($F(1,26) = 0.89$ $p > 0.05$] which, combined with the fact that Group did not interact with Vowel [$F(4,104) = 0.66$, $p > 0.05$], suggested that the two groups were well matched prior to training. The ANOVA yielded a significant effect of Vowel [$F(4,104) = 9.46$, $p < 0.001$]. Pairwise comparisons (Bonferroni adjusted) showed that /i/ showed the best

Table 5.3: Confusion matrix for Greek vowels identified by native speakers of Greek in noise (SNR of -10 dB). Percentages of responses have been pooled over the trained and the control group. Identification responses <3% are not shown.

	Response				
Stimulus	i	e	a	o	u
i	**97**	3			
e	10	**73**	9	7	
a	4	13	**80**		
o	6	5	19	**63**	7
u	3	4	7	24	**62**

identification, followed by all other 4 vowels (other differences between vowels were not significant due to a large variability in the data). Table 5.3 shows percentage identification responses for each Greek vowel at an SNR of -10 dB averaged across the trained and the control group. Correct identification ranged from 62% for Greek /u/ to 97% for Greek /i/. In general, vowels were mostly confused with those closer in the perceptual/acoustical space; /e/ was confused with /i/, /a/ with /e/, /o/ with /a/ and /u/ with /o/.

5.2 Synthetic vowels and non-speech

5.2.1 Identification boundaries and slopes

Figure 5.3 displays the location of identification boundaries for five vowel continua averaged across groups (see following statistical analysis justifying why pooled data can be presented). As can be seen, Greek native speakers placed the identification boundary around the centre of each vowel continuum (i.e. between stimulus 20 and 30). The length of the whiskers indicates that there were large individual differences in phoneme boundary locations across continua with differences being somewhat smaller for the Greek vowel continua than the English ones. Identification boundary

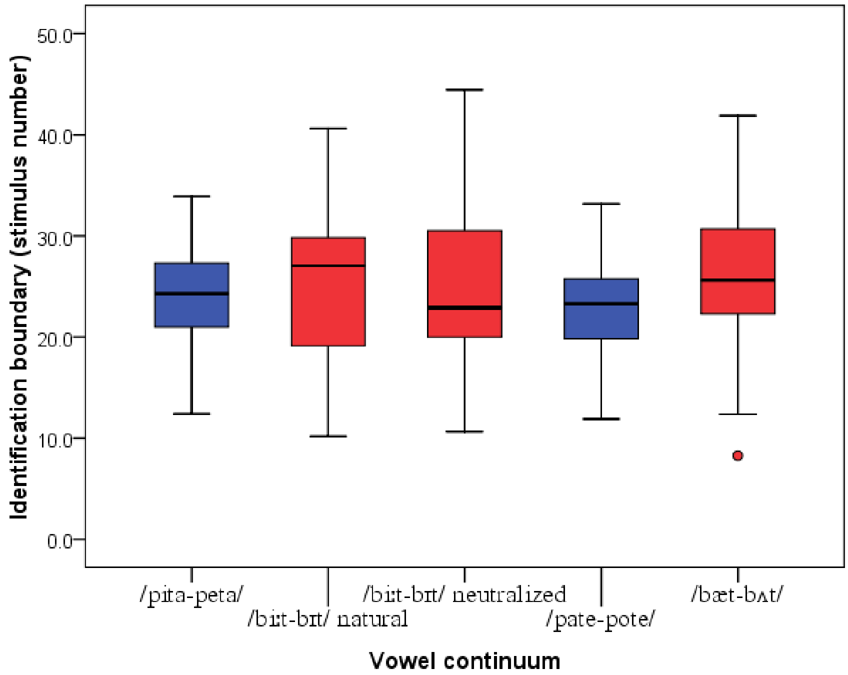


Figure 5.3: Boxplots showing identification boundaries for five vowel continua. The two Greek vowel continua are shown in blue and the three English vowel continua are shown in red. Stimulus 0 is always the first vowel in each continuum and stimulus 50 the second vowel in each continuum.

locations for five vowel continua were submitted to a repeated-measures ANOVA with Group (trained, control) as a between-subject factor and Vowel continuum (5 vowels) as a within-subject factor. There were no main effects of Group [$F(1,26) = 1.7$, $p > 0.05$] or Vowel continuum [$F(4,104) = 1.81$, $p > 0.05$] and no Group × Vowel continuum interaction [$F(4,104) = 1.91$, $p > 0.05$], confirming that the control and the trained group placed the identification boundary at around the same position across vowel continua.

Figure 5.4 displays the consistency with which the five vowel continua were identified averaged across groups (again see following statistical analysis justifying why pooled data can be presented). In this figure, the outcome number indicates the steepness of the identification slope; the bigger the number, the more consistently a

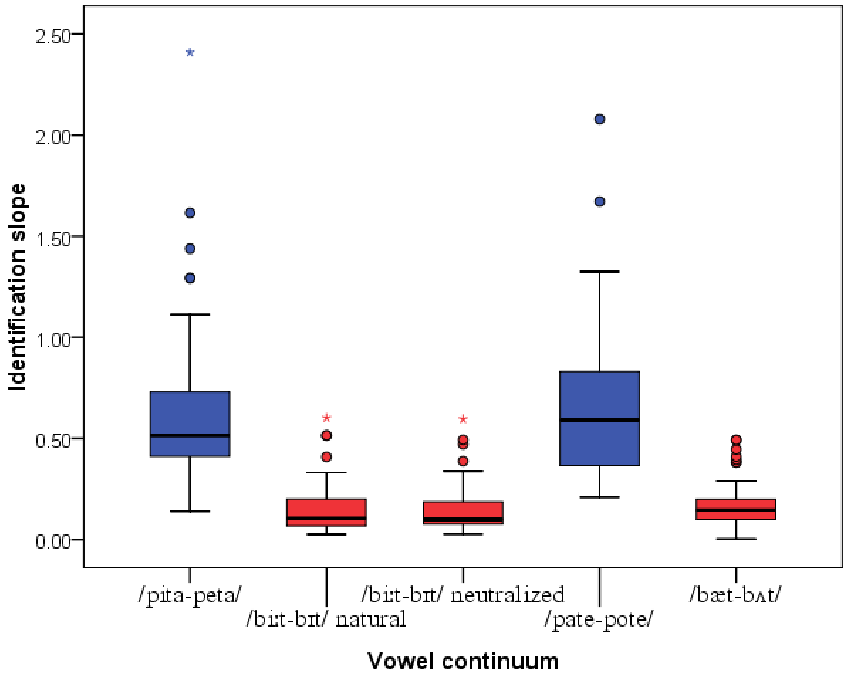


Figure 5.4: Boxplots showing identification consistency for five vowel continua (see text for details). The two Greek vowel continua are shown in blue and the three English vowel continua are shown in red.

vowel is categorized by the listener. A clear effect of language experience can be seen when comparing the consistency with which Greek speakers identified their L1 vowels and the English vowels. Identification slopes for five vowel continua were submitted to a repeated-measures ANOVA with Group (trained, control) as a between-subject factor and Vowel continuum (5 vowels) as a within-subject factor. The ANOVA showed a significant effect of Vowel continuum [$F(4,100) = 22.74$, $p < 0.001$] and no effect of Group [$F(1,25) = 1.4$, $p > 0.05$] or a Group × Vowel interaction [$F(4,100) = 0.6$, $p > 0.05$], suggesting that the two groups performed similarly across vowels. Greek speakers' identification slopes were steeper ($p < 0.05$) for Greek vowels (/i/-/e/: $M = 0.61$; /a/-/o/ $M = 0.67$) than for English vowels (/iː/-/ɪ/ *natural duration*: $M = 0.116$; /iː/-/ɪ/ *neutralized duration*: $M = 0.121$; /æ/-/ʌ/: $M = 0.17$). Paired samples t tests conducted for English vowels showed that Greek speakers' identification slopes did not differ for /iː/-/ɪ/ *natural* and /iː/-/ɪ/ *neutralized*

$t(26) = -0.36$, $p > 0.05$ suggesting that duration did not affect their performance. Greek speakers' identification slopes were significantly steeper in /æ/-/ʌ/ than they were in /iː/-/ɪ/ *natural* $t(26) = -2.21$, $p = 0.37$ and (marginally) significantly steeper than they were in /iː/-/ɪ/ *neutralized* $t(26) = -1.97$, $p = 0.51$. It is worth noting the large degree of individual differences in Greek vowel identification slopes; for English, the range was narrower probably due to limitations posed by L1 experience; Greek speakers were unable to show steep identification slopes for English vowels. To give a better sense of the participants' labelling ability and to demonstrate the effect of L1 experience on labelling performance, Figure 5.5 and Figure 5.6 show the identification functions, i.e. the proportion correct of the two alternatives in each continuum for the same individual in Greek /pita/-/peta/ and in English /biːt/-/bɪt/ *natural duration* continuum respectively. These identification functions translate to identification consistency of 0.50 and 0.10 respectively (very close to the average

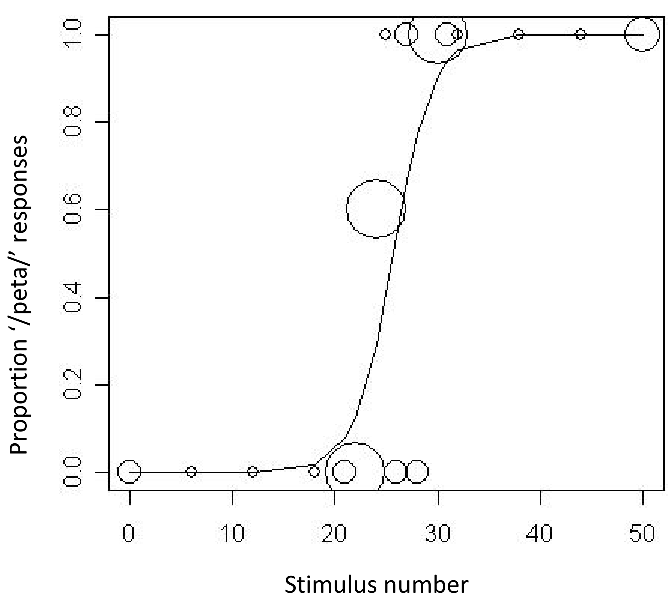


Figure 5.5: Identification function for one individual in the Greek /pita/-/peta/ continuum showing the proportion of /pita/ and /peta/ identification for endpoint stimuli. The size of the circle at a particular step shows the total number of stimuli presentations at that step. Logistic regression is used to obtain a best-fit sigmoid function from the data. The identification function shown translates to identification consistency of .50.

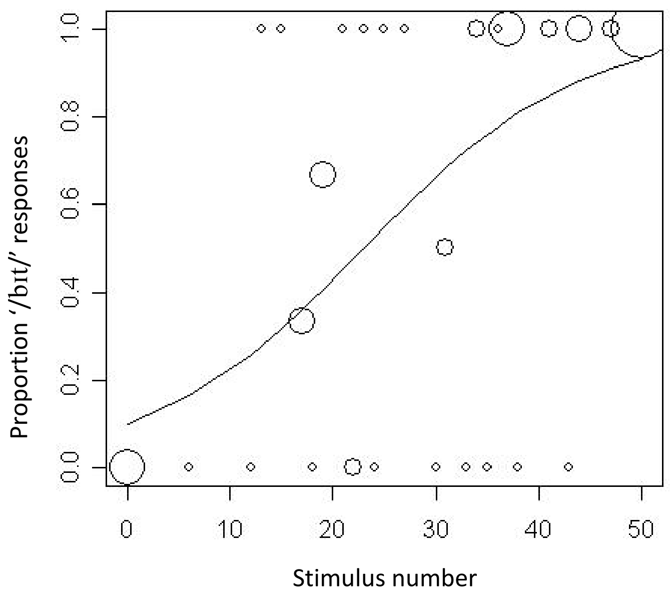


Figure 5.6: Identification function for the same individual in the English /biːt/-/bɪt/ *natural duration* continuum showing the proportion of /biːt/ and /bɪt/ identification for endpoint stimuli. The size of the circle at a particular step shows the total number of stimuli presentations at that step. Logistic regression is used to obtain a best-fit sigmoid function from the data. The identification function shown translates to identification consistency of .10.

values for Greek and English vowel continua) and show that the particular individual was quite consistent in labelling the Greek /pita/-/peta/ continuum but much less consistent in labelling the English /biːt/-/bɪt/ *natural duration* continuum.

5.2.2 Discrimination

The outcome measure of the adaptive discrimination task was the stimulus in the continuum that was just 'discriminable' from the endpoint of the continuum which was the fixed reference. Given that four different continua were employed (hence the acoustical/perceptual difference between the endpoints of those continua was not the same), before making any comparisons between continua it was necessary to define the acoustic/perceptual distance between endpoints for each continuum. To that end, the Euclidean distance (Hz) between the two endpoints (x_1, y_1) and (x_2, y_2) in each vowel continuum was calculated using the following equation:

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

Since each vowel endpoint was actually represented by two points in the vowel space (i.e. endpoints were not static but entailed formant movement), before applying the above equation, the centre (mean) of F1 and F2 movement was taken for each endpoint. To illustrate the whole procedure, Figure 5.7 displays the endpoints for each vowel continuum (black and red arrows for Greek and English endpoints respectively) as well as the Euclidean distance between each Greek and English endpoint (black and red dashed lines respectively).

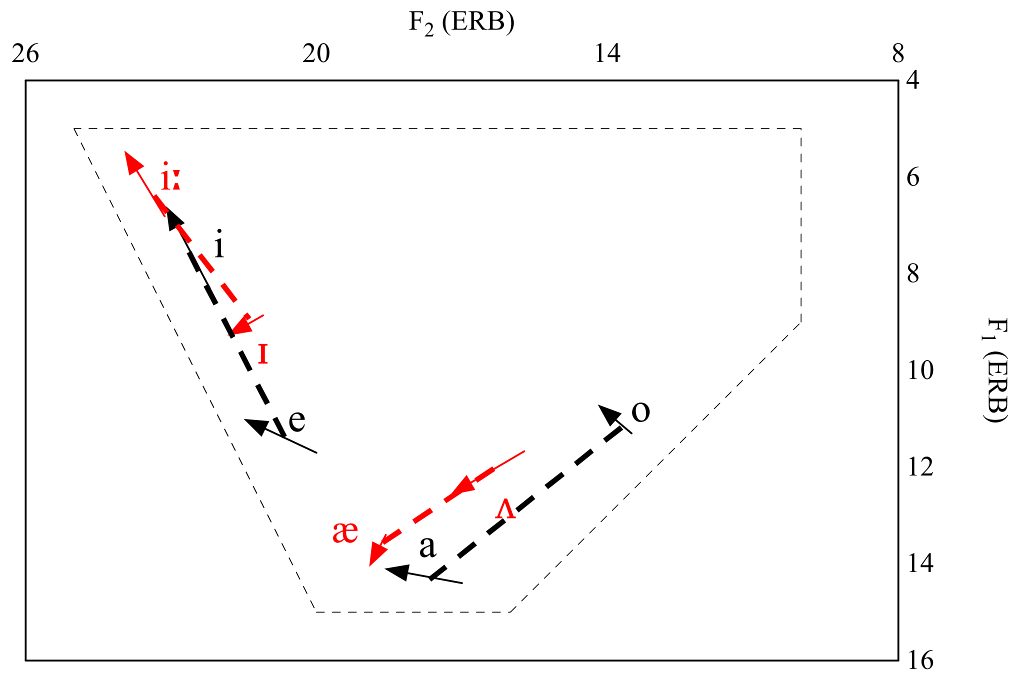


Figure 5.7: Endpoints of the synthetic Greek and English vowel continua (black and red arrows respectively) and the Euclidean distance between each Greek and English endpoint (black and red dashed lines respectively).

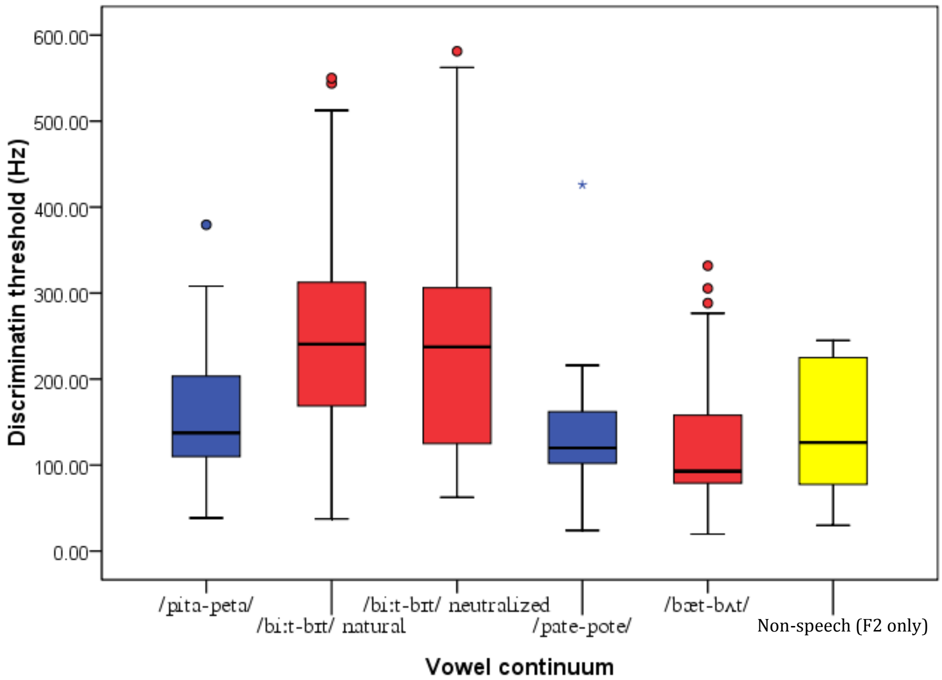


Figure 5.8: Discrimination thresholds (jnd from the fixed reference in Hz) for the five vowel continua and the non-speech continuum (F2 only). Greek vowel continua are shown in blue, English vowel continua are shown in red and the non-speech continuum is shown in yellow.

Having established the Euclidean distance between endpoints for each vowel continuum, it was then possible to calculate the jnd (Hz) each participant was able to detect from the endpoint of the continuum which was the fixed reference. For the non-speech continuum, only this last part of the procedure was followed, i.e. the equation for finding the Euclidean distance was applied and the jnd (Hz) from the fixed reference was calculated. The obtained jnds have been used in all statistical analyses for the rest of the thesis. Figure 5.8 presents discrimination thresholds for the five vowel continua and the non-speech continuum averaged across the trained and the control group. A repeated-measures ANOVA was carried out to evaluate the effect of Group (trained, control) and Continuum (5 vowel and 1 non-speech continuum) on discrimination thresholds (Hz). Although the non-speech continuum differed from the vowel continua in that only F2 frequency changed, it was decided to include it in the analysis. The ANOVA yielded a significant effect of Continuum [$F(5,130) = 24.58$, $p < 0.001$] and no effect of Group [$F(1,26) = 0.62$, $p > 0.05$] or Group × Continuum interaction [$F(5,130) = 0.35$, $p > 0.05$], suggesting that the control and the trained group performed similarly across continua. Pairwise comparisons showed that Greek speakers showed better discrimination for the two Greek vowel continua and the one English vowel continuum (Greek /i/-/e/: jnd = 166 Hz; Greek /a/-/o/: jnd = 144 Hz, and English /æ/-/ʌ/ jnd = 125 Hz) than for the two English vowel continua (English /iː/-/ɪ/ *natural*: jnd = 285 Hz; /iː/-/ɪ/ *neutralized*: jnd = 261 Hz). Greek speakers showed therefore an L1 advantage over the two duration versions of English /iː/-/ɪ/, an L2 contrast whose members assimilate to the same L1 vowel category /i/ and receive very high goodness ratings as shown in Study 1; however, no L1 advantage was found over English /æ/-/ʌ/, an L2 contrast whose members assimilate to the same L1 vowel category /a/ but with /æ/ being a much better spectral match to Greek /a/ than /ʌ/ (remember that the latter was heard as an instance of Greek /o/ around 30% of the time across consonantal contexts). Finally, discrimination threshold for the non-speech continuum (jnd = 154 Hz) was significantly lower than that for the English /iː/-/ɪ/ *natural duration* and /iː/-/ɪ/ *neutralized duration* continua but did not differ from Greek /i/-/e/ and /a/-/o/ and English /æ/-/ʌ/.

Table 5.4 shows mean identification boundaries (stimulus number) and discrimination thresholds (stimulus number and jnd in Hz) for five vowel continua. A jnd which is smaller than the identification boundary reflects within-category discrimination for the first vowel in each continuum. As can be seen, Greek speakers achieved on average clear within-category discrimination for the two Greek continua and the English /æ/-/ʌ/ continuum but not for the English /iː/-/ɪ/ *natural duration* and /iː/-/ɪ/ *neutralized duration* continua, which confirms that the English /æ/-/ʌ/ was an easier contrast than both duration versions of the English /iː/-/ɪ/ continuum.

Table 5.4: Identification boundaries (stimulus number) and discrimination thresholds (stimulus number and jnd in Hz) for five vowel continua. In bold, the vowel continua that showed clear within-category discrimination by Greek speakers. Standard deviations are given in parentheses.

Vowel continuum	Identification boundary (stimulus number)	Discrimination threshold (stimulus number)	Discrimination threshold (jnd in Hz)
Greek			
/i/-/e/	**25.24 (4.55)**	**15.11 (6.21)**	**166.19 (71.37)**
/a/-/o/	**23.93 (4.36)**	**11.99 (4.23)**	**144.06 (71.61)**
English			
/iː/-/ɪ/ *natural*	23.65 (8.24)	22.84 (10.03)	285.01 (112.17)
/iː/-/ɪ/ *neutralized*	25.07 (8.39)	21.84 (10.11)	260.59 (120.12)
/æ/-/ʌ/	**27.15 (5.55)**	**15.83 (7.78)**	**124.97 (64.18)**

5.3 English vowel production by Greek native speakers

5.3.1 Perceptual judgments

As mentioned in Chapter 4, English vowel production was assessed by 2 native listeners of Southern British English. The listeners performed a 10 AFC identification task and gave goodness judgments on the vowels produced by Greek speakers (each Greek speaker produced each English vowel twice). Figure 5.9 shows percent correct identification of English vowels produced by the control and the trained group of Greek speakers. Identification scores were submitted to a two-way repeated-measures ANOVA with Group (trained, control) as a between-subject factor and Vowel (10 vowels) as a within-subject factor. The ANOVA yielded a significant effect of Vowel [$F(9,234) = 3.84$, $p < 0.001$] and no effect of Group [$F(1,26) = 0.73$, $p > 0.05$] or Vowel × Group interaction [$F(9,234) = 1.14$, $p > 0.05$], which suggested that identification scores obtained for the vowels produced by the

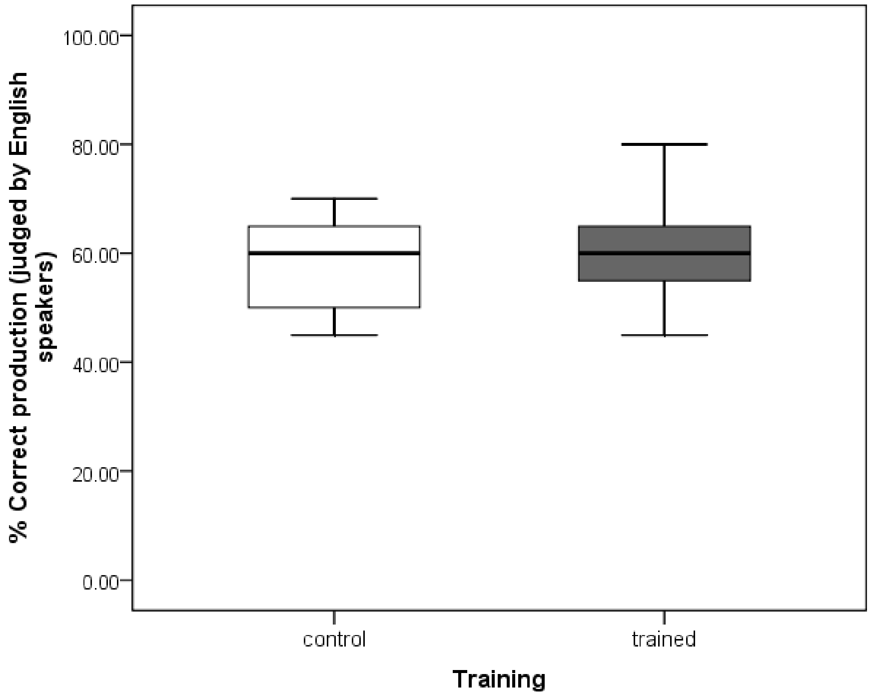


Figure 5.9: Boxplots showing the interquartile range of percent correct identification scores for English vowels produced by the control and the trained group of native speakers of Greek.

trained group (*M* = 61.9%) did not differ from those obtained for the vowels produced by the control group (*M* = 60.7%).

Table 5.5 shows percentage identification responses for each English vowel produced by Greek speakers pooled over the trained and the control group. Correct identification ranged from 87% for English /e/ to 37% for /ʌ/. English /iː/ was mostly confused with /ɪ/ and vice versa; /æ/ was mostly confused with /ʌ/ and vice versa; /ɒ/ was mostly confused with /ɔː/ and vice versa; and /uː/ was mostly confused with /ɔː/. An independent samples *t* test comparing goodness ratings given to English vowel productions by the control (*M* = 3.6) and the trained group (*M* = 3.4) revealed that the two groups did not differ prior to training $t(26) = 1.1$, $p > 0.05$.

Table 5.5: Percentage identification of English vowels produced by Greek speakers. Correct responses have been pooled over the control and the trained group. Identification responses <3% are not shown.

	Response									
Stimulus	iː	ɪ	e	ɜː	æ	ʌ	ɑː	ɒ	ɔː	uː
iː	**67**	32								
ɪ	48	**52**								
e			**87**	10						
ɜː			17	**72**		5	3			
æ					**58**	38	4			
ʌ					46	**37**	12			
ɑː				12	10	10	**66**			
ɒ								**63**	34	
ɔː								46	**50**	
uː									37	**60**

5.3.2 Acoustic analyses

The productions of English vowels by Greek speakers were analyzed acoustically in terms of duration and F1 and F2 frequencies using the SFS speech analysis software (Huckvale, 2008). Duration was measured from spectrograms, from the onset to the offset of periodic energy in F2. F1 and F2 frequencies were estimated automatically from an LPC analysis with 12 coefficients below 5 kHz and cross-checked from an average FFT spectrum when the LPC analysis failed to produce reasonable values. Figure 5.10 plots English vowels produced by Greek speakers pooled over the trained and the control group in the vowel space. English vowels seem to constitute the following five clusters: /iː/ and /ɪ/; /e/ and /ɜː/; /æ/, /ʌ/ and /ɑː/; /ɒ/ and /ɔː/; and /uː/ alone. This suggests that Greek native speakers were using their 5 Greek categories /i/, /e/, /a/, /o/, and /u/ respectively when asked to produce the vowels of English. These observations were examined statistically. Separate two-way repeated-

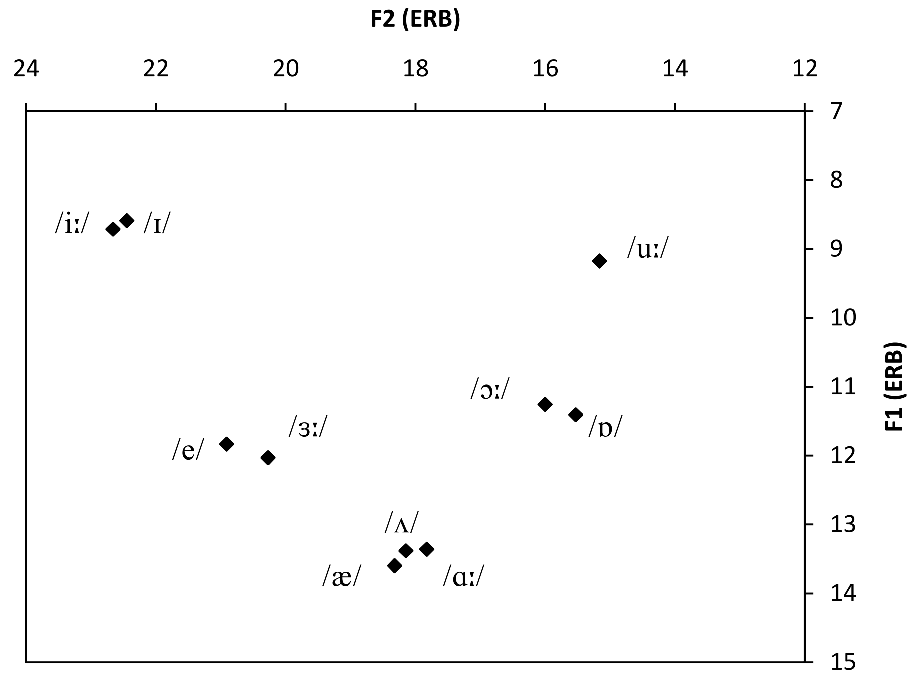


Figure 5.10: English vowel productions (ERB) by Greek speakers pooled over the trained and the control group.

measures ANOVAs with Group (trained, control) as a between-subject factor and Vowel (10 vowels) as a within-subject factor were carried out on F1 and F2 values. For F1, the ANOVA yielded a significant effect of Vowel [$F(9,234) = 221.15$, $p < 0.001$] and no effect of Group [$F(1,26) = 0.17$, $p > 0.05$] or Vowel × Group interaction [$F(9,234) = 0.41$, $p > 0.05$], suggesting that the two groups were well-matched prior to training. Pairwise comparisons (Bonferroni adjusted) separated the vowels as follows: /iː/, /ɪ/ and /uː/; /e/, /ɜː/, /ɒ/ and /ɔː/; and finally, /æ/, /ʌ/ and /ɑː/. For F2, the ANOVA yielded a significant effect of Vowel [$F(9,234) = 217.1$, $p < 0.001$] and no effect of Group [$F(1,26) = 0.69$, $p > 0.05$] or Vowel × Group interaction [$F(9,234) = 0.35$, $p > 0.05$] again suggesting that the two groups were well-matched prior to training. Pairwise comparisons (Bonferroni adjusted) separated the vowels as follows: /iː/ and /ɪ/; /e/ and /ɜː/; /æ/, /ʌ/ and /ɑː/; /ɒ/, /ɔː/ and /uː/. Acoustic analysis therefore confirmed that Greek listeners imposed, at least with

Table 5.6: Duration and F1 and F2 frequencies (ERB) for 10 English vowels produced by Greek speakers pooled over the trained and the control group.

Vowel	English speakers	Greek speakers		
	Duration (ms)	Duration (ms)	F1 (ERB)	F2 (ERB)
iː	145 (9)	137 (24)	8.59 (1.1)	22.54 (1.2)
ɪ	98 (12)	109 (21)	8.41 (1.1)	22.28 (1.4)
e	112 (9)	123 (26)	11.77 (0.8)	20.73 (1.2)
ɜː	188 (8)	161 (31)	11.97 (0.7)	20.33 (0.9)
æ	140 (7)	142 (29)	13.47 (0.6)	18.19 (1.0)
ʌ	111 (13)	143 (25)	13.32 (0.8)	18.12 (1.0)
ɑː	191 (10)	158 (34)	13.22 (0.9)	17.83 (1.1)
ɒ	110 (6)	140 (28)	11.29 (0.7)	15.39 (0.9)
ɔː	178 (13)	160 (36)	11.19 (0.9)	15.72 (1.2)
uː	166 (21)	143 (30)	9.11 (.07)	15.08 (0.9)

respect to spectral distinctions, their 5-vowel system on English vowel production. Table 5.6 shows mean F1 and F2 frequencies (ERB) for 10 English vowels pooled over trained and control group (standard deviations in parentheses).

Finally, a two-way repeated-measures ANOVA examined the effect of Group (trained, control) and Vowel (10 vowels) on the durations of English vowels produced by Greek speakers. The ANOVA yielded a significant effect of vowel [$F(9,234) = 14.68$, $p < 0.001$] and no effect of Group [$F(1,26) = 0.59$, $p > 0.05$] or Vowel × Group interaction [$F(9,234) = 0.34$, $p > 0.05$]. Pairwise comparisons (Bonferroni adjusted) distinguished the following groups: /ɪ/ and /e/ were the shortest vowels followed by /iː/, /æ/, /ʌ/, /ɒ/ and /uː/, with the longest vowels being /ɜː/, /ɑː/ and /ɔː/. This showed that Greek speakers attempted to differentiate English vowels (either because they knew that some vowels are longer than others or because of the effect of orthography when reading the /bVt/ words containing the vowels) using duration. Table 5.6 shows mean durations (ms) for English vowels pooled over the trained and the control group (standard deviations in parentheses). Mean vowel durations for 10 English vowels in /bVp/ context spoken by native English speakers from Study 1 are also given for comparison (standard deviations in parentheses).

5.4 Correlations across experimental measurements

Since all previous analyses showed that the trained group did not differ from the control group in any of the tasks employed, before examining correlations between experimental measurements it was decided, in order to gain more statistical power, to pool data over all 28 participants tested. A first set of analyses examined correlations across tasks that were expected to tap into similar processing abilities. As can be seen in Table 5.7, identification of natural English vowels in quiet was strongly correlated with identification of natural English vowels in noise (r = .728, p <.01). That is, individuals who were better at identifying English vowels in quiet were also better at identifying English vowels in noise. As expected, identification of natural Greek vowels in noise did not correlate with identification of natural English vowels in quiet (r = -.074, p >.05). Further, identification of natural Greek vowels in noise did not correlate either with identification of natural English vowel in noise (r = -.042, p >.05); given that different SNRs were used across languages it is possible that the two tasks were indeed tapping into different processing abilities (identification of L1 vowels at an SNR of -10dB vs. identification of L2 vowels that subjects assimilated to their L1 vowel categories at an SNR of -4dB).

Table 5.7: Correlations (*r*) among tasks using natural vowels.

	L1 identification in noise	L2 identification in quiet	L2 identification in noise
L1 identification in noise	1		
L2 identification in quiet	-.074	1	
L2 identification in noise	-.042	.728**	1

**p<0.01

Table 5.8: Correlations (*r*) among discrimination tasks for 6 synthetic continua (5 vowel continua and a non-speech continuum).

	Greek /i/-/e/	English /iː/-/ɪ/ *natural*	English /iː/-/ɪ/ *neutralized*	Greek /a/-/o/	English /æ/-/ʌ/	F2 only
Greek /i/-/e/	1					
English /iː/-/ɪ/ *natural*	.397*	1				
English /iː/-/ɪ/ *neutralized*	.325	.646**	1			
Greek /a/-/o/	.394*	.226	.425*	1		
English /æ/-/ʌ/	.478*	.580**	.728**	.462*	1	
F2 only	.553**	.673**	.590**	.507**	.752**	1

*p<0.05
**p<0.01

With respect to identification boundaries and slopes for synthetic Greek and English vowels, none of the correlations run either within or between L1 and L2 reached significance or was close to reaching significance. However, when looking at the correlations between L1, L2 and non-speech discrimination, the picture was different. As shown in Table 5.8, almost all pairs correlated with each other; even in a few cases where correlations failed to reach significance they were in the 'correct' direction. Importantly, non-speech discrimination accuracy correlated strongly with all vowel pairs in Greek and English. These results clearly demonstrate that, despite large individual differences in discrimination accuracy found in Section 5.2.2, individuals were consistently 'strong' or 'poor' discriminators across L1, L2 and non-speech.

Table 5.9: Correlations (*r*) among tasks tapping into different processing abilities.

L2 tasks	L1 natural id noise	L1 ID BOUNDARY	L1 ID SLOPE	L1 DISCRIMINATION	Non-speech discrimination
L2 Natural id quiet	-.074	-.128	-.163	-.550**	-.489**
L2 Natural id noise	-.017	.110	-.058	-.482**	-.401**
L2 ID BOUNDARY	-.256	.158	.088	.165	.222
L2 ID SLOPE	-.270	-.301	.078	-.130	-.293
L2 DISCRIMINATION	-.237	-.047	.237	.600**	.792**
L2 production	-.253	-.111	.029	-.297	-.123

**p*<0.05
***p*<0.01

The next set of analyses examined whether performance in the L2 (perception of natural vowels in quiet and in noise, perception of synthetic vowels and vowel production) correlated with L1 (perception of natural vowels in noise and perception of synthetic vowels) and non-speech performance. Before doing so, composite scores for perception tasks with synthetic vowels were calculated. For each participant, an L1 ID BOUNDARY, an L1 ID SLOPE and an L1 DISCRIMINATION score was calculated by averaging performance in two Greek vowel continua (/i/-/e/ and /a/-/o/). Similarly, an L2 ID BOUNDARY, an L2 ID SLOPE and an L2 DISCRIMINATION score was calculated by averaging performance in three English vowel continua (/iː/-/ɪ/ *natural duration,* /iː/-/ɪ/ *neutralized duration,* and /æ/-/ʌ/). As shown in Table 5.9, identification of natural English vowels in quiet correlated with L1 DISCRIMINATION (r = .550, p <.01) and non-speech discrimination (r = .489, p<.01); similarly, identification of natural English vowels in noise correlated with L1 DISCRIMINATION (r = .482, p <.01) and non-speech discrimination (r = .401, p <.01). That is, the successful discriminators were also successful at natural English vowel identification in quiet and in noise. Further, L2 DISCRIMINATION correlated with L1 DISCRIMINATION (r = .600, p <.01) and non-speech discrimination (r = .792, p <.01), which is consistent with previous results where vowel pairs were analyzed separately (see Table 5.8). It is interesting to note that L2 production was not correlated with either L1 DISCRIMINATION or non-speech discrimination while, as mentioned before, identification of natural English vowels in quiet and in noise correlated with these two measures. To confirm the lack of a link between L2 perception and production in the pre-test, no significant correlations were found between identification of English vowels in quiet and English vowel production (r = .242, p >.05) or between identification of English vowels in noise and English vowel production (r = .255, p >.05).

To give an idea of individual performance across tasks, Table 5.10 presents *z*-scores for all 28 participants (18 trainees and 10 controls) on eight tasks: identification of natural Greek vowels in noise, identification of natural English vowels averaged across quiet and noise, L1 SLOPE, L2 SLOPE, L1 DISCRIMINATION, L2 DISCRIMINATION, non-speech discrimination and L2 vowel production. Individuals who performed above 0.5 standard deviation of the mean performance in each task (i.e. approximately top 20th percentile), were considered as 'good' performers and are shown in bold; individuals who performed below 0.5 standard deviation of the mean performance in each task (approximately bottom 20th percentile), were considered as 'poor' performers and are shown in italics. Participants are ranked according to their accuracy in natural English vowel identification. Individual data confirm that participants were generally consistent across L1, L2 and non-speech discrimination; in particular, 7 participants can be described as 'good' performers across all 3 discrimination tasks (shown in shadowed cells) and 5 participants can be described as 'poor' performers across all 3 discrimination tasks (again shown in shadowed cells). Another 2 participants can be can be described as 'good' performers in 2 out of 3 discrimination tasks and 4 participants can be described as 'good' performers in 2 out of 3 discrimination tasks. The majority of 'good discriminators' were highly ranked on natural English vowel identification whereas the majority of 'poor discriminators' were low-ranked on natural English vowel identification. Participants were generally inconsistent in terms of natural Greek vowel identification in noise, L1 SLOPE, L2 SLOPE and L2 vowel production.

Table 5.10: Individual *z*-scores for trained (T) and control (C) participants on eight tasks prior to training. Individuals are ranked based on their natural English vowel identification scores. 'Good' performers are shown in bold and 'poor' performers are shown in italics (see text for details). Shadowed cells indicate individuals who performed consistently well or poorly across L1, L2 and non-speech discrimination.

	Perception of natural vowels		Perception of synthetic vowels				F2 discr.	L2 production
Case	L1	L2	L1 SLOPE	L2 SLOPE	L1 DISCR	L2 DISCR		
T5	0.01	**1**	-0.13	0.2	**0.83**	**0.78**	**0.56**	**0.65**
T8	-0.31	**0.74**	-0.16	-0.11	*-0.92*	*-0.53*	*-1.3*	*-0.65*
C25	**0.97**	**0.74**	0.06	**1.72**	**1.14**	**0.74**	**0.79**	**1.29**
T4	*-1.26*	**0.65**	-0.34	-0.16	**0.72**	0.31	0.33	*-0.86*
T10	*-1.9*	**0.65**	**2.18**	*-1.1*	**0.62**	**0.84**	**1.1**	0.22
T17	*-1.9*	**0.65**	-0.32	**1.61**	**0.62**	**0.84**	**1.1**	**1.94**
C20	**1.29**	**0.65**	*-0.93*	*-1.02*	**1.08**	**1.25**	**0.72**	-0.22
C27	**0.97**	**0.65**	*-0.78*	*-0.76*	0.21	0.44	-0.05	-0.43
T1	**1.29**	0.38	*-1.01*	0.27	**1.5**	**1.18**	**1.79**	0
T9	**0.65**	0.38	-0.12	-0.05	0	*-1.85*	*-1.5*	**0.65**
T2	**0.65**	0.3	*-1.24*	-0.46	0.36	**1.25**	**1.02**	*-1.08*
T3	0.01	0.21	*-0.74*	**1.05**	**0.93**	**1.18**	**1.6**	*-0.65*
C24	*-1.58*	0.21	0.06	0	*-1.13*	*-1.41*	*-1.32*	0.43
T12	**0.65**	0.12	**0.69**	*-0.82*	0.36	-0.16	0.41	**1.72**
T18	**0.65**	0.12	**0.69**	*-0.82*	0.36	-0.16	0.41	0.22
C22	*-0.94*	0.12	-0.32	**0.55**	0.16	*-1.51*	-0.36	*-1.08*
T7	0.33	0.03	0.45	0.39	0.06	-0.06	*-0.58*	0
C21	*-1.26*	0.03	-0.43	**2.71**	*-0.82*	0.21	*-0.61*	**0.86**
T14	**0.97**	-0.14	**2.03**	0.2	-0.1	**0.98**	0.33	*-0.86*
C19	0.01	-0.25	0.49	0.03	-0.1	0.44	**1.4**	*-0.86*
T6	**0.65**	*-0.5*	-0.33	0.28	**0.78**	**0.74**	-0.2	*-1.29*
T11	**0.97**	*-0.5*	*-0.76*	*-1.21*	0.47	0.1	**0.56**	**1.94**
C26	*-0.62*	*-0.76*	0.22	*-0.79*	*-1.69*	0.04	*-1.35*	**1.08**
C23	*-0.62*	*-1*	*-0.53*	*-1.28*	*-2.41*	*-1.88*	*-1.33*	*-1.08*
C28	0.33	*-1*	*-0.67*	**1.61**	*-0.82*	-0.06	*-1.04*	*-1.51*
T13	**0.97**	*-1.02*	-0.32	*-0.95*	*-1.9*	*-1.34*	*-0.62*	-0.22
T16	0.33	*-1.2*	**3.06**	*-0.58*	**0.93**	-0.74	*-0.66*	**0.65**
T15	*-0.94*	*-1.29*	*-0.81*	*-0.53*	*-1.23*	*-1.61*	*-1.2*	*-0.86*

5.5 Summary of results

This chapter examined Greek speakers' pre-training performance on a battery of perceptual tasks with natural and synthetic Greek and English vowels, a non-speech (F2 only) discrimination task and an English vowel production task. The relationships between tasks were also examined in an attempt to shed some light on the sources of individual variability in English vowel perception and production performance. Since the trained and the control group of Greek speakers were well-matched prior to training, i.e. they were not found to differ in any of the tasks employed, the following apply to all 28 Greek speakers tested in the pre-test.

First, Greek speakers' identification of natural English vowels in quiet and in noise was examined. Multi-talker babble (SNR = -4 dB) lowered English vowel identification accuracy by about 20 percentage points, from 56% correct to 38% correct. In both quiet and noise, /iː/ was mostly confused with /ɪ/, /æ/ was mostly confused with /ʌ/ (and to a lesser degree with /ɑː/), and /ɒ/ was mostly confused with /ɔː/. These results are similar to those reported in Study 1 and confirm the well-attested finding that L2 learners with 5 vowels in their system tend to struggle with the English high front /iː/-/ɪ/ pair, and the vowels /æ/-/ʌ/-/ɑː/. Greek native speakers showed the highest identification accuracy in quiet for English /uː/ (86% correct) but this was probably due to the lack of its strongest competitor /ʊ/ in the perceptual stimuli; in noise, English /uː/ suffered a severe drop in percent correct identification, from 86% to 39% correct. Mean correct identification of natural Greek vowels at an SNR of -10 dB was 74.5% with /e/ being confused with /i/, /a/ being confused with /e/, /o/ being confused with /a/ and /u/ being confused with /o/.

Regarding Greek speakers' identification of synthetic vowels in Greek and English, it was found that subjects placed the identification boundary around the centre of each continuum across vowels and languages. A clear effect of L1 experience was found on Greek speakers' identification consistency, expressed by the steepness of identification slopes, for Greek vowel continua when compared to English vowel continua. Importantly, Greek speakers' identification consistency did not differ for /iː/-/ɪ/ *natural duration* and /iː/-/ɪ/ *neutralized duration*, suggesting that subjects performed the same with and without the duration cue. Greek speakers showed steeper identification slopes in /æ/-/ʌ/ than in both /iː/-/ɪ/ *natural duration* and /iː/-/ɪ/ *neutralized duration*, which is consistent with the assimilation results of Study 1 and the prediction that /æ/-/ʌ/ would suffer less from L1 spectral interference than /iː/-/ɪ/. With respect to discrimination, Greek speakers' performance was better for Greek /i/-/e/ and /a/-/o/ than for English /iː/-/ɪ/ *natural* and /iː/-/ɪ/ *neutralized* but their discrimination for English /æ/-/ʌ/ did not differ from L1 discrimination, confirming that English /æ/-/ʌ/ was an easier L2 contrast than /iː/-/ɪ/ was. The lack of differences in discrimination accuracy between English /iː/-/ɪ/ *natural duration* and /iː/-/ɪ/ *neutralized duration* shows that Greek speakers did not benefit from the duration difference between English /iː/ and /ɪ/ in the /iː/-/ɪ/ *natural duration* continuum. It has to be noted though that by the time subjects reached their discrimination threshold for the /iː/-/ɪ/ *natural duration* continuum (mean jnd from the endpoint /iː/ was stimulus number 22.84, i.e. around the centre of the continuum), the duration difference between the two stimuli had been reduced by half, i.e. from 40 ms at the beginning of testing to around 20 ms.

English vowels produced by Greek speakers were correctly identified by English speakers at a rate of 61% and received a goodness rating of 3.5 in a 7-level Likert scale. The most frequent misidentifications included /iː/ vs. /ɪ/, /æ/ vs. /ʌ/ and vice versa; /ɒ/ vs. /ɔː/, and /uː/ vs. /ɔː/. Acoustic analyses performed on English vowels produced by Greek speakers showed that participants were using, at least in terms of spectral characteristics, their 5 Greek categories /i/, /e/, /a/, /o/, and /u/ to produce /iː/ and /ɪ/, /e/ and /ɜː/, /æ/, /ʌ/ and /ɑː/, /ɒ/ and /ɔː/ and /uː/ respectively. Greek speakers were found to produce some duration distinctions between English vowels but it is not clear whether this was due to their knowledge concerning English vowel durations or due to the effect of orthography when reading the /bVt/ words containing the vowels.

One common finding across all tasks employed was the existence of large individual differences in performance. Such differences are more striking in L1 vowel perception tasks if one considers that all participants were tested within normal-hearing thresholds yet in line with previous research on vowel thresholds (Kewley-Port & Watson, 1994; Kewley-Port, 2001; Gerrits & Schouten, 2004) and speech in noise (Surprenant & Watson, 2001; Kidd *et al.*, 2007). The existence of individual differences in non-speech discrimination is also consistent with previous research in psychoacoustics (Johnson *et al.*, 1987; Kidd *et al.*, 2007). To explore the relationships between individual differences in L1, L2 and non-speech processing, correlation analyses were run. It was found that individuals were consistent in performance across L1, L2 and non-speech discrimination, suggesting that a spectral acuity component underlies discrimination performance. Auditory spectral ability was also found to relate to natural English vowel identification in quiet and in noise but not to English vowel production. These results seem to support the *auditory processing* hypothesis over the *L1 phonetic* hypothesis; there was no evidence that individuals with less robust L1 categories were more flexible in terms of learning new categories, in other words there was no evidence of an inverse correlation between L1 ID SLOPE and natural English vowel perception or L2 ID SLOPE as the *L1 phonetic* hypothesis would predict.

Chapter 6

Post-training results

This chapter examines the effects of auditory phonetic training on the test battery completed by the trained group of Greek speakers and compares these results with the post-test results of the group of Greek speakers that received no training. As done in the previous chapter, after reporting on group data, individual data and correlations in performance within and between tasks in the post-test will be examined and the predictions of the *L1 phonetic* hypothesis and the *auditory processing* hypothesis will be evaluated. Correlations between pre-test and post-test performance will also be examined to see whether pre-test performance can predict post-test performance for individuals.

6.1 Natural vowels

6.1.1 Identification of English vowels in quiet and in noise

Figure 6.1 displays the interquartile range of percent identification scores obtained by the trained and the control group in quiet and in noise averaged across 10 English vowels in the pre-test, the post-test and the generalization test. Identification scores were first submitted to a three-way repeated-measures ANOVA with Group (trained, control) as a between-subject factor and Noise condition (quiet, noise) and Test

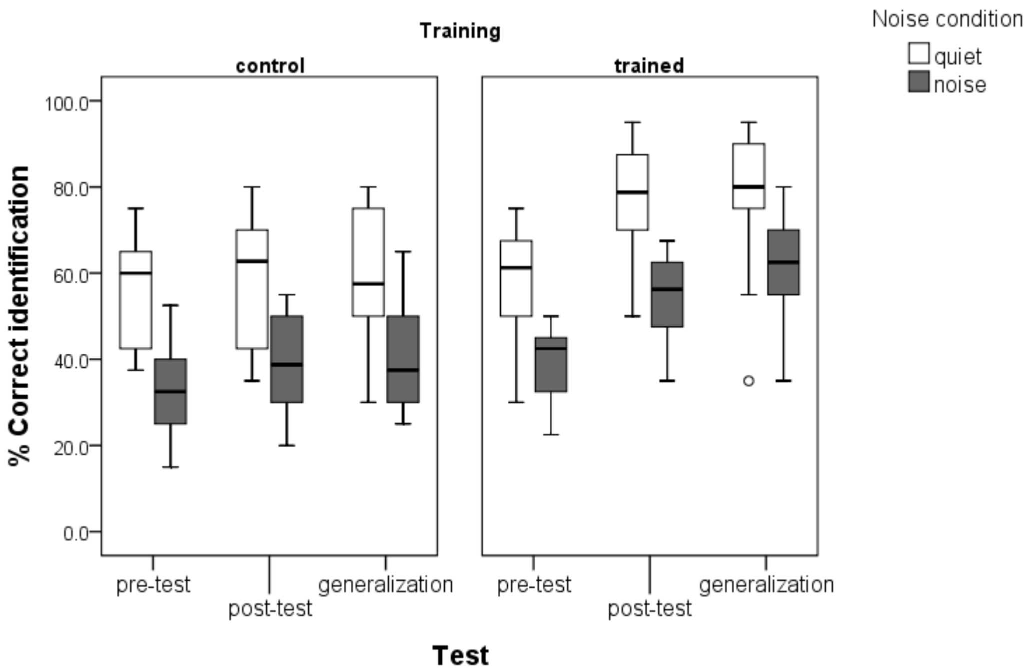


Figure 6.1: Boxplots showing the interquartile range of percent identification scores for English vowels by Greek speakers in quiet and in noise (SNR = -4dB) before and after auditory training. Whiskers extend to at most 1.5 times the interquartile range of the box.

(pre-test, post-test, generalization) as within-subject factors. There were significant main effects of Group [$F(1,26) = 10.885$, $p < 0.01$], Noise Condition [$F(1,26) = 229.38$, $p < 0.001$] and Test [$F(2,52) = 31.62$, $p < 0.001$]. There was also a significant Test × Group interaction [$F(2,52) = 13.94$, $p < 0.001$] which was explored through simple effect tests. The simple effect of Test was significant for the trained group [$F(2,34) = 48.642$, $p < 0.001$] but not for the control group [$F(2,16) = 2.22$, $p > 0.05$]. Pairwise comparisons (Bonferroni adjusted) showed that, across Noise conditions, the trained group improved from pre-test ($M = 48.6\%$) to post-test ($M = 65.9\%$) and generalization test ($M = 70.2\%$) whereas the control group did not improve from pre-test ($M = 47.7\%$) to either post-test ($M = 51.3\%$) or generalization test ($M = 51.9\%$). Figure 6.2 and Figure 6.3 show pre-test vs. post-test percent correct identification of English vowels by the trained group in *quiet* and in *noise* respectively. Auditory

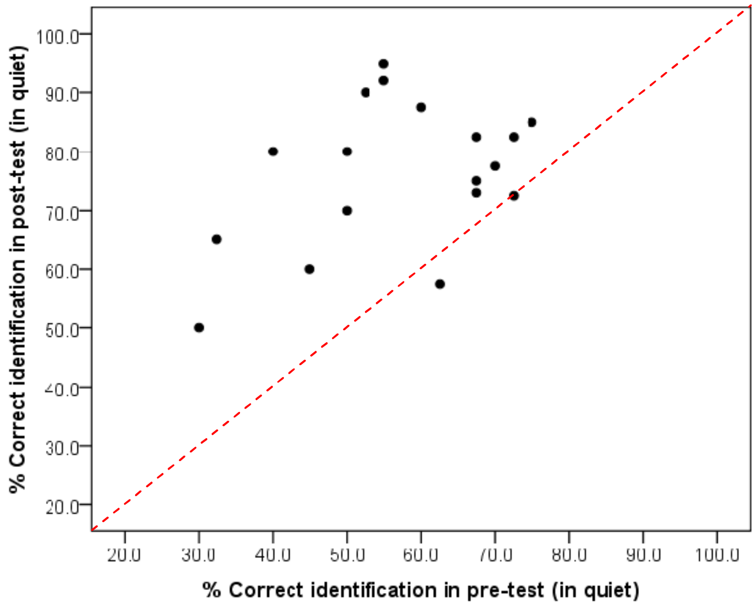


Figure 6.2: Scatterplot showing pre-test vs. post-test percent correct identification of English vowels in quiet by the trained group of Greek speakers. Performance along the red dotted line would represent no improvement between pre-test and post-test.

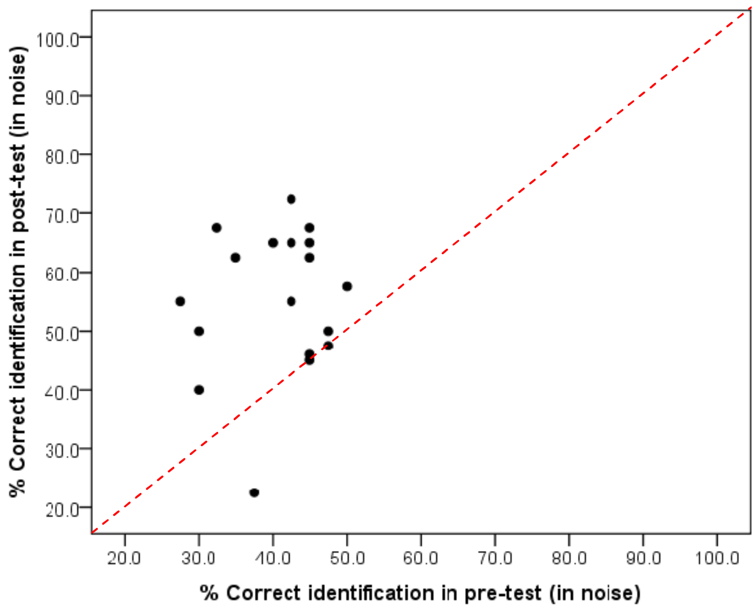


Figure 6.3: Scatterplot showing pre-test vs. post-test percent correct identification of English vowels in noise by the trained group of Greek speakers. Performance along the red dotted line would represent no improvement between pre-test and post-test.

Table 6.1: Confusion matrix for English vowels in quiet identified by the trained group of native speakers of Greek after auditory training (results before training are given in parentheses). Percentages of correct responses have been pooled over participants. Identification responses <3% are not shown.

	Response									
Stimulus	iː	ɪ	e	ɜː	æ	ʌ	ɑː	ɒ	ɔː	uː
iː	**84** **(58)**	16 (39)								
ɪ	23 (23)	**73** **(49)**	3 (22)				(3)			
e	(5)		**86** **(56)**	(9)	5 (14)	8 (8)	(5)			
ɜː	(5)		8 (8)	**83** **(66)**	(3)	(12)	9 (5)			
æ				3 (12)	**73** **(52)**	19 (22)	3 (12)			
ʌ				(4)	42 (47)	**44** **(36)**	9 (12)			
ɑː				12 (18)	6 (17)	3 (11)	**67** **(36)**	(3)	9 (14)	
ɒ							(3)	**64** **(48)**	34 (47)	
ɔː								16 (23)	**84** **(69)**	(6)
uː									3	**97** **(95)**

training clearly improved identification performance in both noise conditions for the vast majority of participants. At the same time, some participants improved to a much larger degree than others did. Table 6.1 and Table 6.2 show percent identification responses for each English vowel given by the trained group of Greek speakers in *quiet* and in *noise* respectively after perceptual training (in both tables results before training are given in parentheses).

Table 6.2: Confusion matrix for English vowels in noise (SNR = -4dB) identified by the trained group of native speakers of Greek after auditory training (results before training are given in parentheses). Percentages of correct responses have been pooled over participants. Identification responses <3% are not shown.

	Response									
Stimulus	iː	ɪ	e	ɜː	æ	ʌ	ɑː	ɒ	ɔː	uː
iː	**41**	23	3	5	8	9			6	
	(23)	(32)		(6)	(6)	(6)	(6)	(5)	(6)	(8)
ɪ	23	**61**	5	3						6
	(39)	**(47)**	(9)							
e			**59**	5	12	16	6			
	(5)		**(37)**	(9)	(17)	(19)	(3)		(5)	
ɜː			5	**67**		8	8	3	5	3
	(3)		(8)	**(45)**	(3)	(9)	(9)	(8)	(12)	
æ			3		**59**	23	9		3	
			(6)	(11)	**(52)**	(14)	(17)			
ʌ					28	**52**	16	3		
				(3)	(47)	**(22)**	(19)	(5)		
ɑː	3			3	27	19	**34**	8	6	
			(5)	(11)	(12)	(28)	**(27)**	(5)	(9)	(3)
ɒ								**59**	33	
			(6)	(3)		(5)		**(39)**	(41)	(5)
ɔː				3		3		30	**51**	14
				(6)		(5)	(3)	(20)	**(50)**	(14)
uː	3	5						13	9	**67**
	(12)	(9)		(5)				(12)	(19)	**(37)**

Identification scores for the trained group of Greek speakers were submitted to a three-way repeated-measures ANOVA with Test (pre-test, post-test), Noise condition (quiet, noise) and Vowel (10 vowels) as factors. This ANOVA yielded main effects of Test [$F(1,17) = 55.62$, $p < 0.001$], Noise condition [$F(1,17) = 130.64$, $p < 0.001$] and Vowel [$F(9,153) = 6.82$, $p < 0.001$]. There was also a Vowel × Noise condition interaction [$F(9,153) = 6.522$, $p < 0.001$] and a Test × Vowel × Noise condition interaction [$F(9,153) = 2.87$, $p < 0.01$], suggesting that, although noise lowered overall identification scores, for some vowels this was to a larger extent than for others; further, these were not the same vowels across pre-test and post-test. In the previous chapter (pre-test results), it was found that noise lowered identification performance for all vowels except /ɪ/, /æ/, and /ɑː/. In the post-test, post hoc *t* tests revealed that noise lowered identification performance for all vowels except /ʌ/.

6.1.2 Identification of Greek vowels in noise

Figure 6.4 displays the interquartile range of percent identification scores obtained by the trained and the control group of Greek native speakers in noise (SNR = -10dB) averaged across 5 Greek vowels in the pre-test and the post-test. Identification scores were submitted to a three-way repeated-measures ANOVA with Group (trained, control) as a between-subject factor and Test (pre-test, post-test) and

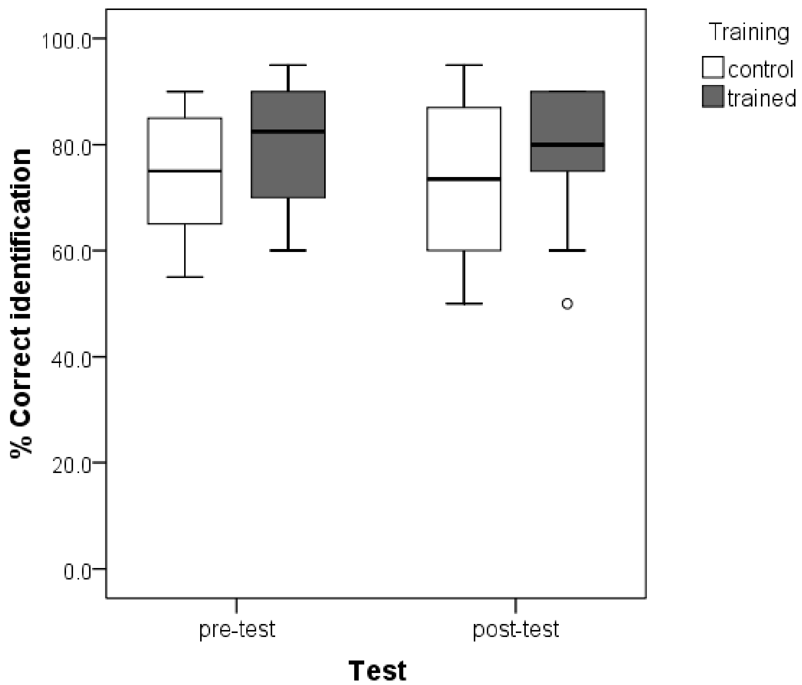


Figure 6.4: Boxplots showing the range of percent identification of Greek vowels in noise for the trained and the control group in pre-test and post-test.

Table 6.3: Confusion matrix for Greek vowels (SNR = -10 dB) identified by native speakers of Greek after auditory training (results before training are given in parentheses). Percentages of correct responses have been pooled over participants. Identification responses <3% are not shown.

	Response				
Stimulus	i	e	a	o	u
i	**94**	5			
	(96)	(3)			
e	16	**65**	11	5	3
	(12)	**(72)**	(8)	(6)	
a	5	5	**86**	4	
	(3)	(11)	**(83)**		
o			5	**76**	15
	(7)		(8)	**(71)**	(9)
u			6	19	**74**
	(3)		(6)	(23)	**(65)**

Vowel (5 vowels) as within-subject factors. The ANOVA showed a significant main effects of Vowel [$F(4,104) = 10.070$, $p < 0.001$] and no other significant main effects or interactions, suggesting that both groups performed similarly across tests. As expected, English vowel training did not change Greek speakers' identification of Greek vowels in noise. The results concerning the control group demonstrated no learning of the task itself. Pairwise comparisons (Bonferroni adjusted) revealed that, across groups and tests, Greek /i/ showed the best identification, followed by the other 4 vowels. Table 6.3 shows percent identification responses for each Greek vowel in noise given by the trained group after English vowel training (results before training are given in parentheses).

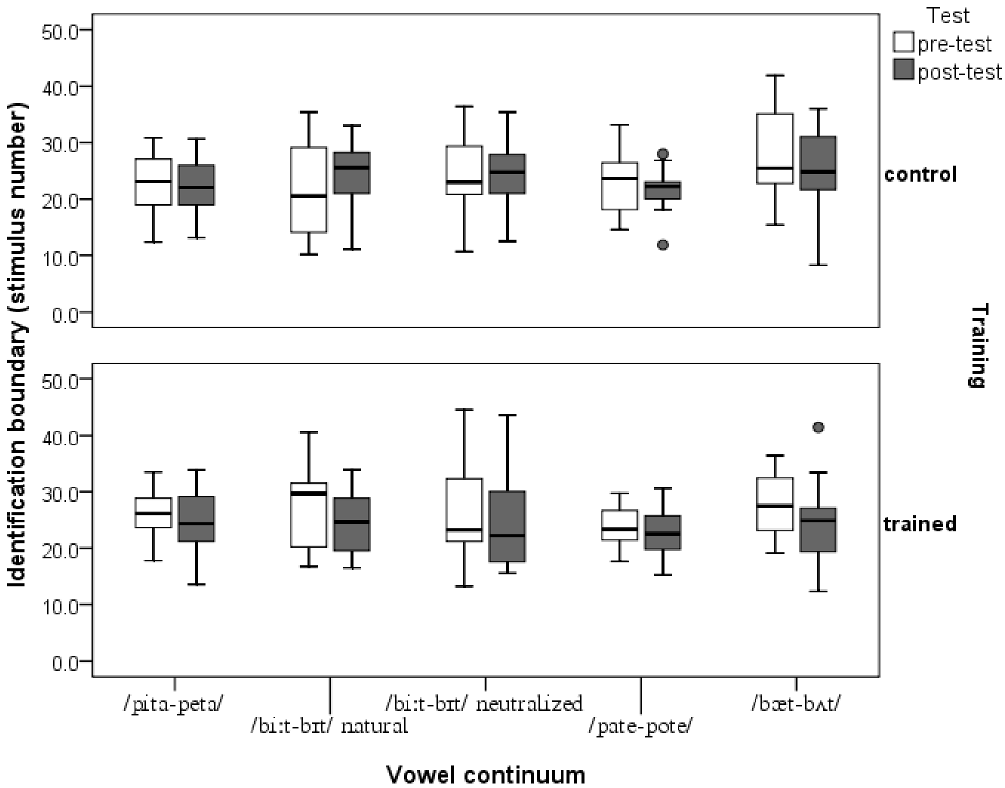


Figure 6.5: Boxplots showing identification boundaries for five vowel continua (two Greek and three English continua) for the trained and the control group in pre-test and post-test.

6.2 Synthetic vowels and non-speech

6.2.1 Identification boundaries and slopes

Figure 6.5 displays identification boundary locations for five vowel continua (two Greek and three English continua) for the trained and the control group in pre-test and post-test. Identification boundaries in the post-test seem very similar to those in the pre-test across vowel continua and groups. A repeated-measures ANOVA with Group (trained, control) as a between-subject factor and Test (pre-test, post-test) and Vowel continuum (5 levels) as within-subject factors confirmed this observation by showing no significant main effects or interactions.

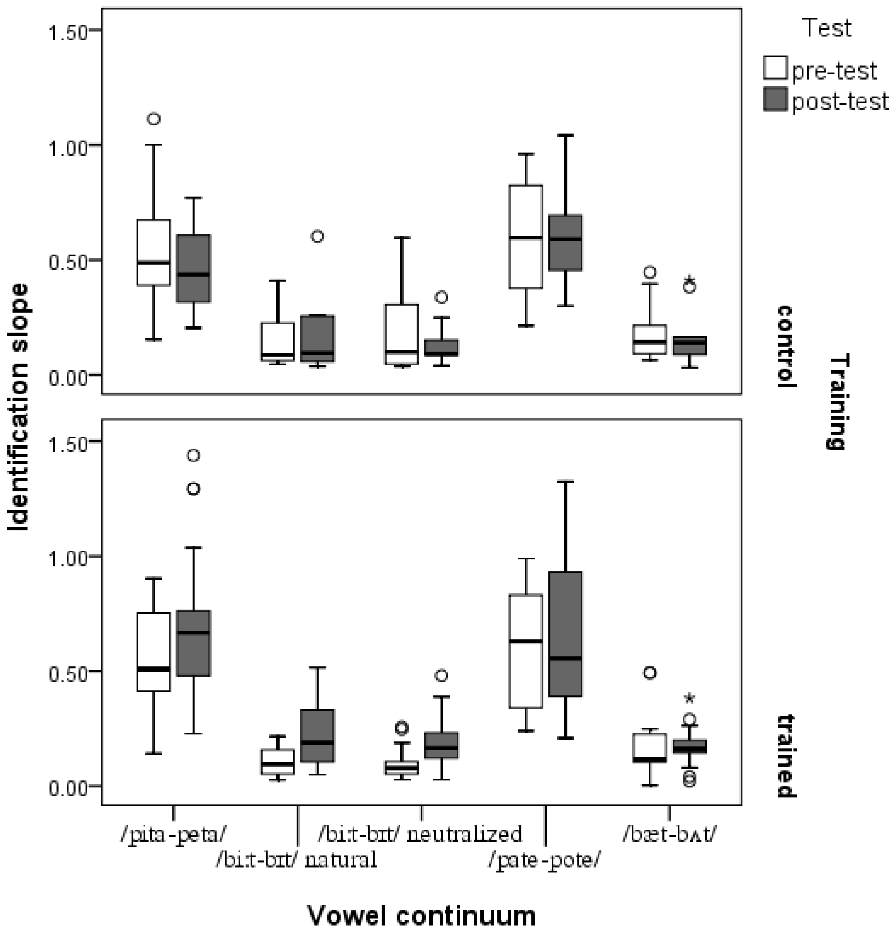


Figure 6.6: Boxplots showing identification slopes for the trained and the control group of Greek speakers on five vowel continua (two Greek and three English continua) in pre-test and post-test.

Figure 6.6 displays identification slopes for five vowel continua (two Greek and three English continua) in pre-test and post-test for the trained and the control group of native Greek speakers. Given the large differences across Greek and English vowel continua both in terms of identification consistency and of range of scores, it was decided to perform separate repeated measures ANOVA on identification slopes for each language. In each ANOVA, Group (trained, control) served as a between-subject factor and Test (pre-test, post-test) and Vowel continuum (2 levels for Greek and 3 levels for English) served as within-subject factors. For Greek, the ANOVA showed no significant main effects or interactions suggesting that, as expected, English vowel training did not change Greek speakers' L1 vowel identification consistency. For English, the ANOVA showed a significant main effect of Test [$F(1,26) = 7.77$, $p < 0.01$] and a significant Test × Group interaction [$F(1,26) = 6.39$, $p < 0.01$]. Post-hoc *t* tests showed that the trained group had significantly steeper identification slopes in the post-test than in the pre-test for /iː/-/ɪ/ *natural duration* and /iː/-/ɪ/ *neutralized duration* but not for /æ/-/ʌ/. Identification slopes for the control group did not change from pre-test to post-test for any of the three English vowel continua.

6.2.2 Discrimination

Figure 6.7 displays discrimination thresholds (jnd in Hz) for six continua (two Greek and three English vowel continua and a non-speech continuum) in pre-test and post-test for the trained and the control group. Discrimination thresholds were submitted to a repeated-measures ANOVA with Group (trained, control) as a between-subject factor and Test (pre-test, post-test) and Continuum (6 levels) as within-subject factors. There was a significant effect of Continuum [$F(5,130) = 26.32$, $p < 0.001$] and no effect of Test [$F(1,26) = 2.26$, $p > 0.05$] or Group [$F(1,26) = 0.74$, $p > 0.05$] or any interaction, suggesting that both groups performed similarly in the pre/post tests. Given that subjects were trained on English vowels, this is not surprising for the Greek vowel continua and the non-speech continuum but it is interesting to note the lack of any improvement in English vowel discrimination especially for /iː/-/ɪ/ *natural duration* and /iː/-/ɪ/ *neutralized duration* given the improvement in identification consistency for these pairs reported in the previous section. Pairwise comparisons showed that, across groups and tests, Greek native speakers showed better discrimination for the two Greek continua, one English vowel continuum and the non-speech pair (Greek /i/-/e/: jnd = 156 Hz; Greek /a/-/o/: jnd = 144 Hz, English /æ/-/ʌ/ jnd = 130 Hz) than for two English continua (English/iː/-/ɪ/ *natural duration*: jnd = 255 Hz; English /iː/-/ɪ/ *neutralized duration*: jnd = 262 Hz). As in the pre-test,

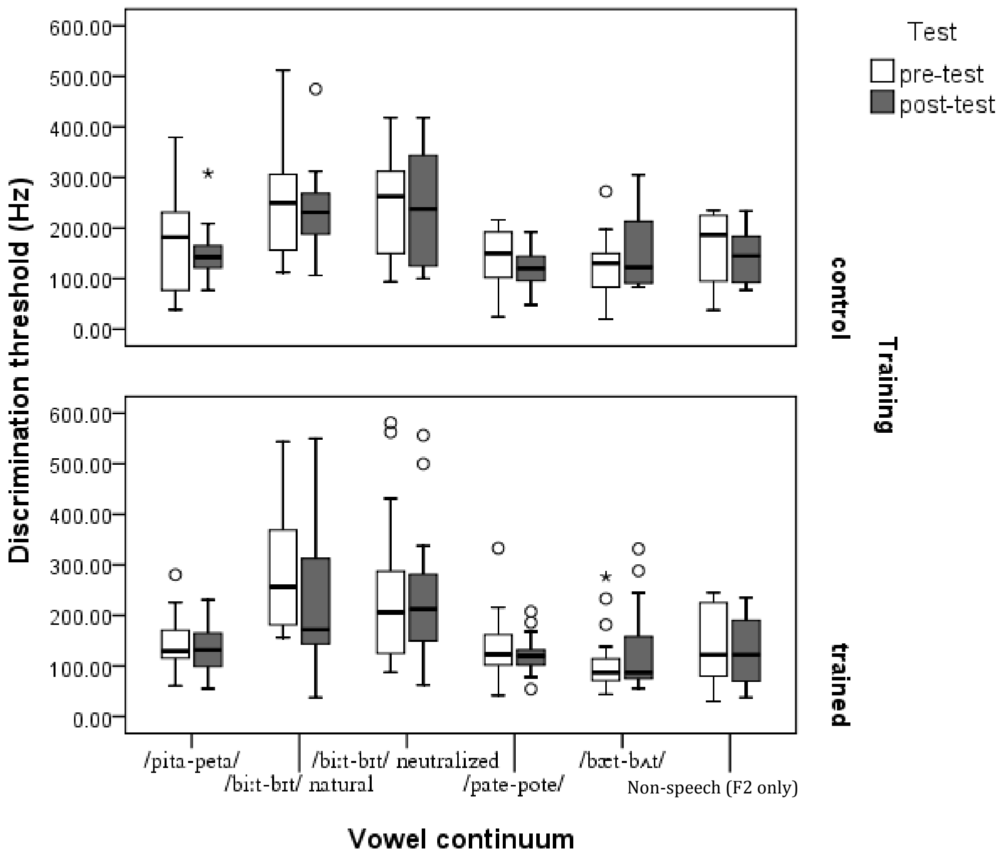


Figure 6.7: Boxplots showing discrimination thresholds (Hz) for the trained and the control group of Greek speakers on five vowel continua (two Greek and three English continua) and the non-speech continuum (F2 only) in pre-test and post-test.

Greek native speakers showed an L1 advantage over the two duration versions of the English /iː/-/ɪ/ continuum but no advantage over the English /æ/-/ʌ/ continuum and their discrimination threshold for the non-speech continuum was significantly lower than that for the English /iː/-/ɪ/ *natural duration* and /iː/-/ɪ/ *neutralized duration* continua but did not differ from discrimination accuracy for Greek /i/-/e/ and /a/-/o/ and English /æ/-/ʌ/.

6.3 English vowel production

6.3.1 Perceptual judgments

Figure 6.8 shows percent correct identification scores of English vowels produced by the trained and the control group in pre-test and post-test as judged by the English listeners (identification scores were pooled over 2 productions of each English vowel by each Greek speaker). A clear improvement after training can be seen for English vowels produced by the trained group whereas identification scores for English vowels produced by the control group do not seem to have changed from pre-test to post-test. Identification scores were submitted to a three-way repeated-measures ANOVA with Group (trained, control) as a between-subject factor and Test (pre-test,

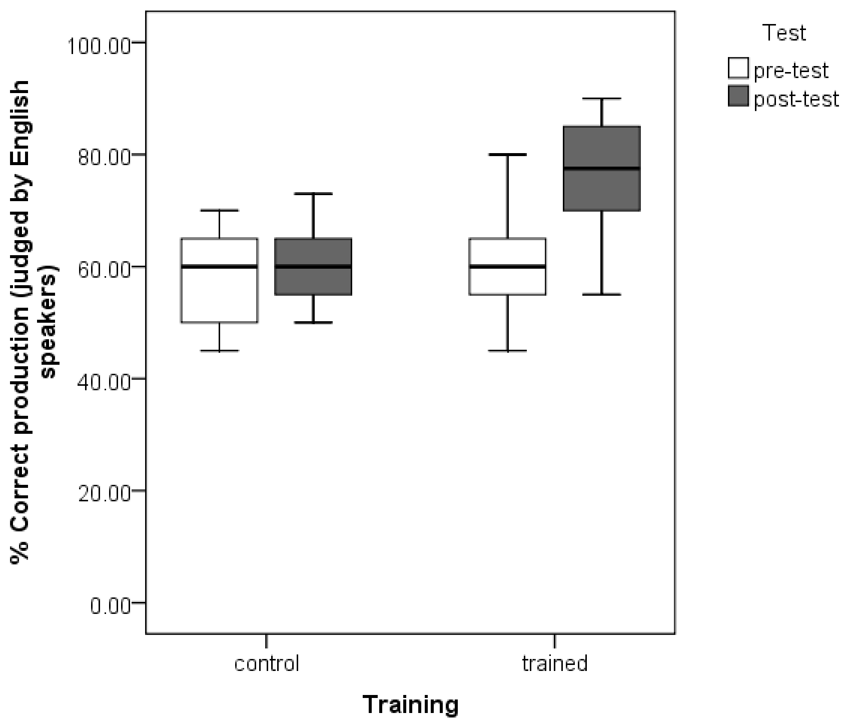


Figure 6.8: Boxplots showing the interquartile range of percent correct identification scores for English vowels produced by the control and the trained group of native speakers of Greek in pre-test and post-test.

Table 6.4: Percentage identification of English vowels produced by the trained group of Greek speakers after auditory training (results before training are given in parentheses) as judged by the native English listeners. Identification responses <3% are not shown.

	Response									
Stimulus	iː	ɪ	e	ɜː	æ	ʌ	ɑː	ɒ	ɔː	uː
iː	**90** **(70)**	10 (30)								
ɪ	32 (53)	**54** **(47)**	14							
e			**91** **(93)**	6 (7)						3
ɜː			7 (17)	**90** **(80)**	3	(3)				
æ			(3)	3	**80** **(50)**	13 (40)	3 (7)			
ʌ			(3)	3	30 (37)	**57** **(47)**	(13)	7	3	
ɑː				3 (17)	7 (13)	7 (7)	**77** **(60)**	3 (3)		
ɒ				3		(4)		**68** **(57)**	29	
ɔː				3			3	17 (40)	**77** **(57)**	
uː								(3)	37 (43)	**63** **(53)**

post-test) and Vowel (10 vowels) as within-subject factors. The ANOVA yielded significant main effects of Group [$F(1,26) = 5.5$, $p < 0.05$], Test [$F(1,26) = 6.07$, $p < 0.05$] and Vowel [$F(9,234) = 6.08$, $p < 0.001$] and a significant Test × Group interaction [$F(1,26) = 5.26$, $p < 0.05$]. The simple effect of Test was significant for the trained group [$F(1,17) = 10.49$, $p < 0.001$] but not for the control group [$F(1,9) = 0.31$, $p > 0.05$]; identification scores of English vowels produced by the trained group improved from 61.9% correct to 75.8% correct whereas identification scores of English vowels produced by the control group did not change from pre-test (M = 60.75 correct) to post-test (M = 61.2% correct). Table 6.4 shows percent identification responses for each English vowel produced by the trained group of Greek speakers after perceptual training (results before training are given in parentheses). All vowels except /e/ were identified correctly at higher rates in the post-test than in the pre-test although due to large variability in scores this effect was significant only for English /iː/, /æ/, /ʌ/, /ɔː/ and /uː/.

6.3.2 Acoustic analyses

In Figure 6.9 English vowels produced by the trained group of Greek speakers before and after auditory training are plotted in the vowel space. As already discussed, English vowels produced by Greek speakers were arranged into 5 clusters in the pre-test, suggesting that subjects were using their 5 spectral qualities /i/, /e/, /a/, /o/, and /u/ when asked to produce English vowels. After auditory training, there was much less overlap of English vowels than it was before training, especially in the high front area of /iː/ and /ɪ/, the mid front/central area of /e/ and /ɜː/ and the low area of /æ/, /ɑː/, and /ʌ/. Table 6.5 presents pre-test and post-test mean F1 and F2 frequencies (ERB) for English vowels produced by the trained group (standard deviations in parentheses).

A two-way repeated-measures ANOVA examined the effect of Test (pre-test, post-test) and Vowel (10 vowels) on the duration of English vowels produced by the trained group. The ANOVA yielded main effects of Test [$F(1,17) = 8.20$, $p < 0.01$] and Vowel [$F(9,153) = 23.51$, $p < 0.001$] and a significant Test × Vowel interaction [$F(9,153) = 8.23$, $p < 0.001$], suggesting that overall English vowels were longer in the post-test (M = 165 ms) than in the pre-test (M = 145 ms) and that this effect was not uniform across all vowels. Paired samples *t* tests showed that /iː/, /ɜː/, /æ/, /ɑː/, /ɔː/ and /uː/ were longer in the post-test than in the pre-test whereas /ʌ/ showed the opposite pattern, i.e. was shorter in the post-test than in the pre-test. Table 6.5 shows

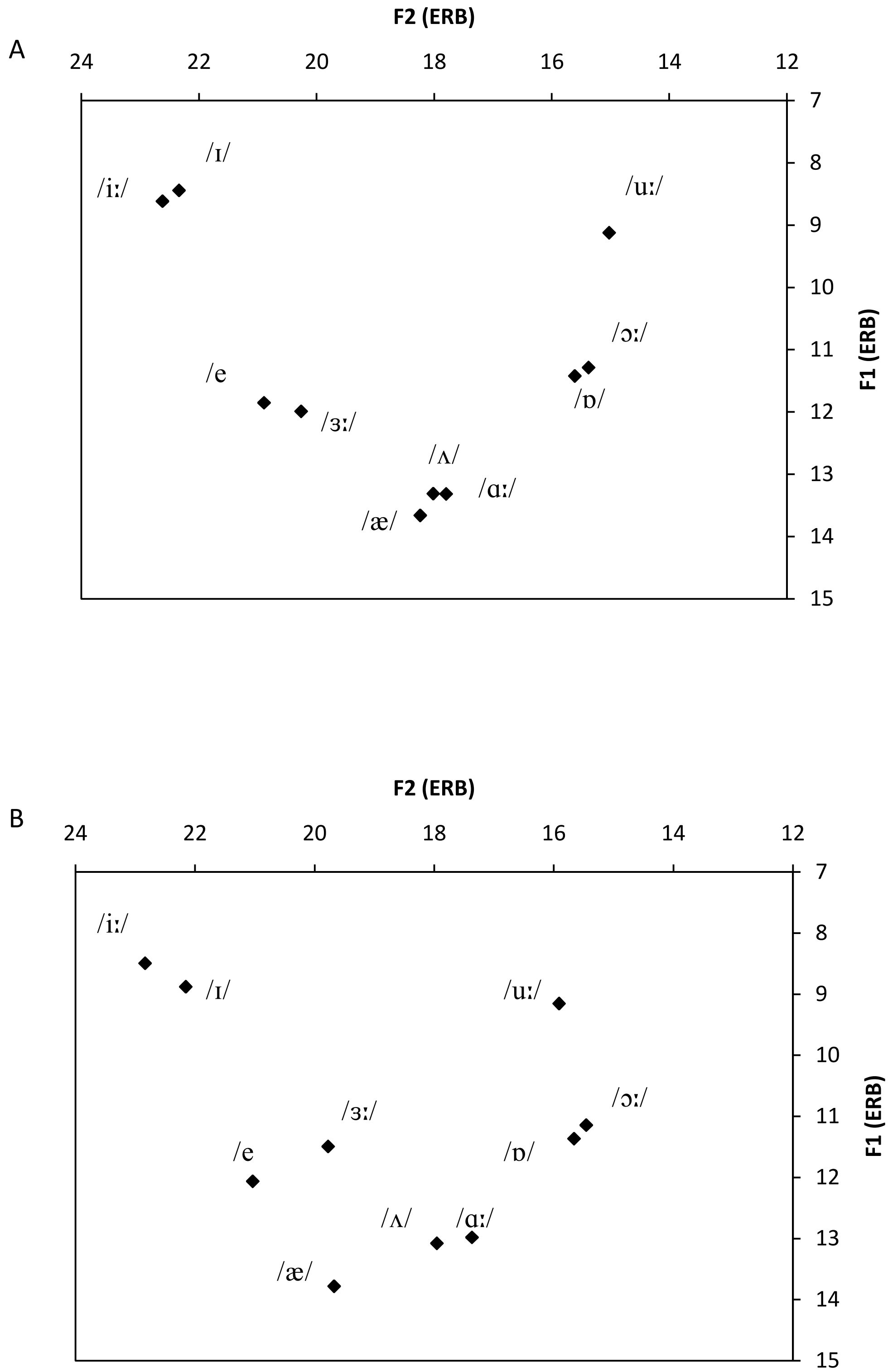


Figure 6.9: F1 and F2 frequencies (ERB) for English vowels produced by the trained group before auditory training (A) and after auditory training (B).

Table 6.5: Duration and F1 and F2 frequencies (ERB) for 10 English vowels produced by the trained group of Greek speakers before and after auditory training. Duration for 10 English vowels produced in /bVp/ context by three Southern British English speakers from Study 1 is also given. Standard deviations for all measures are given in parentheses.

	English speakers	Greek speakers pre-test			Greek speakers post-test		
Vowel	**Duration**	Duration	F1 (ERB)	F2 (ERB)	Duration	F1 (ERB)	F2 (ERB)
iː	**145**	143	8.51	22.54	184	8.37	22.72
	(9)	(24)	(1.1)	(1.1)	(30)	(1.0)	(1.1)
ɪ	**98**	113	8.43	22.38	111	8.76	22.04
	(12)	(25)	(1.0)	(1.2)	(22)	(1.2)	(1.3)
e	**112**	129	11.87	20.87	130	12.00	21.05
	(9)	(26)	(1.1)	(1.0)	(26)	(0.5)	(1.1)
ɜː	**188**	163	12.04	20.33	186	11.45	19.59
	(8)	(26)	(0.6)	(0.9)	(36)	(1.2)	(0.9)
æ	**140**	147	13.62	18.13	187	13.73	19.79
	(7)	(29)	(0.6)	(1.0)	(43)	(0.6)	(1.0)
ʌ	**111**	151	13.29	18.00	134	13.05	17.83
	(13)	(28)	(0.7)	(0.9)	(22)	(1.1)	(0.9)
ɑː	**191**	159	13.20	17.81	205	12.92	17.28
	(10)	(30)	(0.9)	(0.9)	(42)	(0.6)	(1.1)
ɒ	**110**	142	11.44	15.55	135	11.35	15.72
	(6)	(29)	(0.6)	(0.9)	(31)	(1.1)	(0.7)
ɔː	**178**	160	11.33	15.34	196	11.14	15.51
	(13)	(31)	(0.7)	(0.9)	(38)	(1.0)	(0.9)
uː	**166**	143	9.12	15.02	173	9.16	15.97
	(21)	(22)	(.07)	(0.8)	(27)	(.08)	(0.9)

mean duration for 10 English vowels (ms) for the trained group before and after auditory training (standard deviations in parentheses). Further, mean duration for 10 English vowels produced by three English speakers in /bVp/ context from Study 1 are also given for comparison (standard deviations in parentheses). As can be seen, the change in duration for most English vowels produced by Greek speakers after training is in the correct direction even though some English vowels were produced with even longer duration than those produced by English speakers. This may be related to the training stimuli which contained vowels in both voiced and voiceless contexts; the trainees may thus learned that some vowels should be pronounced with long durations but failed to learn the English rule that vowels are shorter before a voiceless context than before a voiced context and to produce the /bVt/ words accordingly. This kind of allophonic variation is likely more difficult to learn, especially after just 5 sessions of auditory training where the learner must infer the particular allophonic rule.

6.4 Correlations across experimental measurements

Given that all previous analyses showed no changes in performance from pre-test to post-test across tasks for the control group, this section presents correlations only for the trained group (n = 18). It should be mentioned though that all pre-test correlations reported for 28 Greek speakers hold when looking only at the trained group. A first set of analyses examined post-test correlations across tasks that were expected to tap into similar processing abilities. As shown in Table 6.6, identification of natural English vowels in quiet was significantly correlated with identification of natural English vowels in noise (r = .594, p <.01), suggesting that L2 vowel perception in quiet and in noise were aligned for individuals. As expected,

Table 6.6: Correlations (*r)* among tasks with natural vowels in the post-test for the trained group of Greek speakers.

	L1 identification in noise	L2 identification in quiet	L2 identification in noise
L1 identification in noise	1		
L2 identification in quiet	.080	1	
L2 identification in noise	.042	.594**	1

***p*<0.01

Table 6.7: Correlations (*r*) among identification slopes for 5 synthetic pairs (two Greek and three English) in the post-test for the trained group of Greek speakers.

	/i/-/e/	/iː/-/ɪ/ *natural*	/iː/-/ɪ/ *neutralized*	/a/-/o/	/æ/-/ʌ/
/i/-/e/	1				
/iː/-/ɪ/ *natural*	-.104	1			
/iː/-/ɪ/ *neutralized*	-.058	.516*	1		
/a/-/o/	.116	.161	.408	1	
/æ/-/ʌ/	.141	.504*	.629**	.263	1

*p<0.05
**p<0.01

identification of natural Greek vowels in noise was not correlated with identification of natural English vowels in quiet (r = -.080, p >.05). Further, identification of natural Greek vowels in noise was not correlated with identification of natural English vowel in noise (r = -.042, p >.05); however, as mentioned in the previous chapter, this may be due to different levels of noise in Greek and English. On the whole, these results are very similar to those obtained in the pre-test.

With respect to identification boundaries, none of the correlations run either within or between L1 and L2 reached significance which was also the case in the pre-test. With respect to identification slopes, the picture emerging was different from that in the pre-test (where no correlations were found within or between L1 and L2); as shown in Table 6.7, after auditory training there were significant correlations between /iː/-/ɪ/ *natural duration* and /iː/-/ɪ/ *neutralized duration* (r = .52, p <0.05), between /iː/-/ɪ/ *natural duration* and /æ/-/ʌ/ (r = .50, p <0.05), and between /iː/-/ɪ/ *neutralized duration* and /æ/-/ʌ/ (r = .63, p <0.01). These correlations show that, after receiving perceptual training, individuals showed consistently strong or poor identification abilities (steep or shallow identification slopes respectively) across English vowel continua demonstrating a symmetrical learning of L2 vowels.

Table 6.8: Correlations (*r*) among discrimination tasks for 6 synthetic pairs (5 vowel and 1 non-speech pair) in the post-test for the trained group of Greek speakers.

	/i/-/e/	/iː/-/ɪ/ *natural*	/iː/-/ɪ/ *neutralized*	/a/-/o/	/æ/-/ʌ/	F2 only
/i/-/e/	1					
/iː/-/ɪ/ *natural*	.510*	1				
/iː/-/ɪ/ *neutralized*	.325	.801**	1			
/a/-/o/	.476*	.650**	.735**	1		
/æ/-/ʌ/	.614**	.760**	.656**	.637**	1	
F2 only	.464*	.710**	.464*	.646**	.677**	1

*p<0.05
**p<0.01

Given that discrimination did not change as a result of training in any of the five vowel continua or the non-speech continuum, it was expected that, as in the pre-test, post-test discrimination accuracy would correlate across continua. As can be seen in Table 6.8, all correlations except one (between /iː/-/ɪ/ *neutralized* and /i/-/e/ where correlation failed to reach significance but was in the ‘correct’ direction) were significant. These results confirmed that individuals were consistently successful or unsuccessful discriminators across L1, L2 and non-speech.

Next, it was examined whether performance in the L2 vowel tasks (perception of natural vowels in quiet and in noise, perception of synthetic vowels and vowel production) correlated with performance in the L1 vowel tasks (perception of natural vowels in noise and perception of synthetic vowels) and the non-speech discrimination task. Before doing so, composite scores for synthetic speech perception tasks were calculated as described in previous chapter, giving an L1 ID BOUNDARY, an L1 ID SLOPE, an L1 DISCRIMINATION, an L2 ID BOUNDARY, an L2 ID SLOPE and an L2 DISCRIMINATION score. As shown in Table 6.9, post-test identification of natural English vowels in quiet correlated with post-test L1 vowel discrimination ($r = .627, p <.01$) and non-speech frequency

Table 6.9: Correlations (*r*) among tasks tapping into different processing abilities in the post-test for the trained group of Greek speakers.

L2 tasks	L1 natural id noise	L1 ID BOUNDARY	L1 ID SLOPE	L1 DISCRIMINATION	F2 discrimination
L2 natural id quiet	.080	-.177	.105	-.627**	-.497*
L2 natural id noise	.042	-.267	-.073	-.478*	-.426*
L2 ID BOUNDARY	-.362	.224	.227	.017	-.009
L2 ID SLOPE	.151	-.087	.177	-.303	-.225
L2 DISCRIMINATION	-.077	-.062	-.278	.759**	.739**
L2 Production	-.013	.368	.110	-.444*	-.652**

*p<0.05
**p<0.01

discrimination (r = .497, p <.05); likewise, post-test identification of natural English vowels in noise correlated with post-test L1 DISCRIMINATION (r = .478, p <.05) and non-speech (F2 only) discrimination (r = .426, p <.05). Further, L2 DISCRIMINATION correlated with L1 DISCRIMINATION (r = .759, p <.01) and non-speech discrimination (r = .739, p <.01) which is consistent with previous analyses where vowel pairs were analyzed separately (see Table 6.8). One important difference between pre-test and post-test is that only in the latter was L2 vowel production correlated with both L1 DISCRIMINATION (r = .444, p <.05) and non-speech discrimination (r = .652, p <.01), i.e. the most successful discriminators were judged to produce more native-like English vowels than the less successful discriminators. To confirm the link between L2 vowel perception and production after perceptual training, contrary to what was found in the pre-test, L2 production was correlated with both identification of English vowels in quiet (r = .563, p <.01) and in noise (r = .594, p <.01).

To take a closer look at the effect of training for individuals, the relation between pre-test identification of natural English vowels across noise conditions and degree of improvement relative to pre-test was examined. As shown in Figure 6.10, there

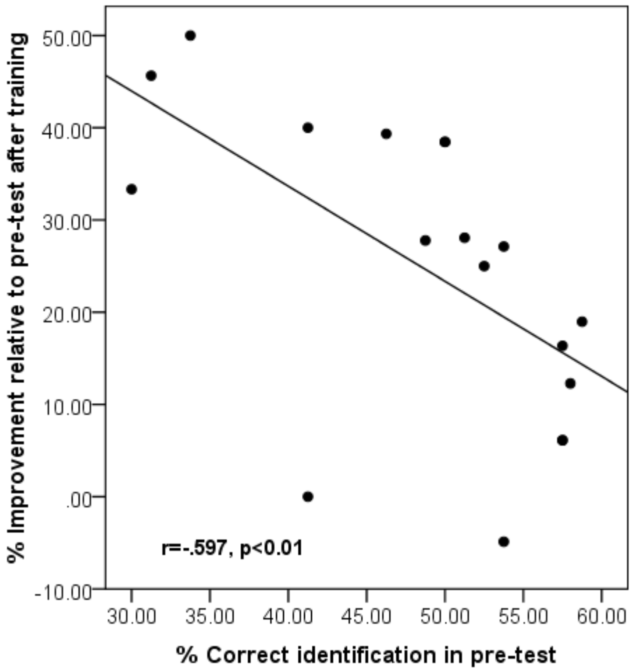


Figure 6.10: Scatterplot showing the relation between percent correct identification of natural English vowels (across noise conditions) in pre-test and degree of improvement relative to pre-test after auditory training.

was a negative correlation (r = -.597, p <.01) between the two measures; those who performed poorly in the pre-test improved more than those who performed well in the pre-test, a finding that cannot be attributed to a ceiling effect given that the highest score obtained was 81.25% correct. At the same time, when examining the relation between pre-test and post-training identification of natural English vowels across noise conditions, a positive correlation was found (r = .517, p <.05). This means that those trainees who were the most accurate before perceptual training were also the most accurate after training despite showing less improvement than those who performed poorly in pre-test.

Similarly, the relation between pre-training performance on English vowel production and degree of improvement relative to pre-training performance was examined. Again, as shown in Figure 6.11, a negative correlation (r = -.515, p <.05)

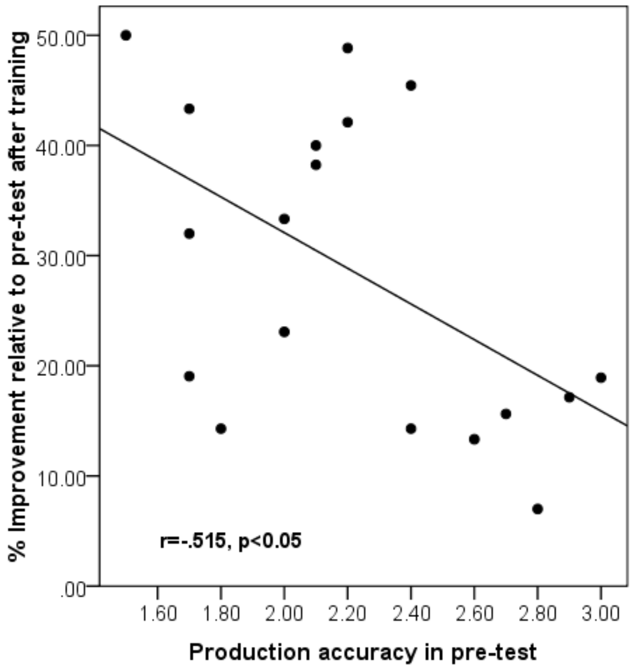


Figure 6.11: Scatterplot showing the relation between English vowel production accuracy in pre-test and degree of improvement relative to pre-test after auditory training.

was found between the two measures; those who performed poorly in the pre-test improved more than those who performed well in the pre-test. However, this time there was no correlation between pre-test and post-test English vowel production accuracy (r = .278, p >.05), suggesting that those who produced more accurate vowels in the pre-test were not the same individuals who produced more accurate vowels in the post-test. This can be attributed to the fact that while pre-test natural English vowel identification was related to L1, L2 and non-speech discrimination accuracy, pre-test English vowel production was rather random in the sense that it did not relate to subjects' performance on any of L1, L2 or non-speech discrimination tasks; although all subjects imposed their 5-vowel system to L2 vowel production, the vowels produced by some participants were judged by English listeners as closer to the target vowels than the vowels produced by others.

Table 6.10: Individual *z*-scores for trained (T) participants on eight tasks after English vowel training. Individuals are ranked based on their perception of natural English vowels scores. 'Good' performers are shown in bold and 'poor' performers are shown in italics. Shadowed cells indicate individuals who performed consistently well or poorly across L1, L2 and non-speech discrimination.

	Perception of natural vowels		Perception of synthetic vowels				F2 discr.	L2 Production
Case	L1	L2	L1 SLOPE	L2 SLOPE	L1 DISCR	L2 DISCR		
T5	0.08	**1.36**	-0.03	-0.02	**0.65**	**1.00**	**1.39**	**1.86**
T12	**0.82**	**1.36**	-0.44	*-0.68*	**0.96**	0.42	*-0.82*	0.13
T17	**0.82**	**0.91**	**2.07**	**1.32**	**0.57**	0.47	**0.79**	**0.85**
T1	0.45	**0.69**	*-0.53*	0.07	**1.80**	**1.20**	**1.27**	0.42
T10	*-2.13*	**0.58**	**2.07**	-0.35	**1.19**	**0.59**	**0.79**	**0.56**
T18	*-2.13*	0.47	-0.44	*-0.68*	-0.20	0.06	-0.01	-0.16
T3	0.08	0.47	*-0.60*	**0.66**	0.49	**0.97**	0.07	**1.00**
T2	**1.19**	0.36	*-1.34*	**0.72**	0.03	0.21	**1.15**	**1.72**
T4	*-1.02*	0.25	-0.38	0.08	0.34	**0.64**	**0.99**	-0.45
T6	**0.82**	0.25	**0.59**	**1.08**	**0.57**	0.21	*-0.58*	-0.16
T7	**0.82**	0.14	**1.17**	**2.58**	0.03	0.47	**0.87**	**0.56**
T13	0.08	0.14	*-1.01*	*-0.51*	0.03	*-0.70*	0.07	-0.16
T14	*-1.02*	-0.42	*-0.60*	-0.35	-0.04	0.47	-0.05	-0.16
T8	*-0.66*	-0.42	-0.07	*-0.85*	*-0.97*	*-1.87*	*-0.74*	*-1.46*
T16	**0.82**	*-0.75*	0.14	0.00	*-0.51*	*-0.73*	*-0.90*	*-0.74*
T9	0.45	-1.31	*-0.57*	*-1.39*	*-1.81*	*-1.69*	*-1.70*	*-0.88*
T15	0.08	-1.86	**0.98**	*-1.54*	*-2.09*	*-2.10*	*-1.78*	*-1.46*
T11	0.45	*-2.20*	*-1.02*	-0.13	*-1.04*	0.37	*-0.82*	*-1.46*

Finally, post-test identification of natural English vowels across noise conditions was found to correlate with pre-test accuracy in L1 (r = .549, p <.05), L2 (r = .563, p <.05) and non-speech discrimination (r = .553, p <.05). Similarly, post-test English vowel production accuracy correlated with pre-test accuracy in L1 (r = .524, p<.05), L2 (r = .680, p <.01) and non-speech discrimination (r = .654, p <.01).

Table 6.10 presents *z*-scores for the 18 Greek speakers that received English vowel training on eight tasks: identification of natural Greek vowels in noise, identification of natural English vowels averaged across quiet and noise, L1 SLOPE, L2 SLOPE, L1 DISCRIMINATION, L2 DISCRIMINATION, non-speech discrimination and L2 vowel production. As in the pre-test, individuals who performed above 0.5 standard deviation of the mean performance in each task (i.e. approximately top 20th percentile), were considered as ‘good’ performers and are shown in bold; individuals who performed below 0.5 standard deviation of the mean performance in each task (approximately bottom 20th percentile), were considered as ‘poor’ performers and are shown in italics. Participants are ranked according to their accuracy in perceiving natural English vowels. Individual data confirm that participants were generally consistent across L1, L2 and non-speech discrimination; in particular, 3 participants can be described as ‘good’ performers across all 3 discrimination tasks (shown in shadowed cells) and 4 participants can be described as ‘poor’ performers across all 3 discrimination tasks (again shown in shadowed cells). Additionally, another 2 participants can be described as ‘good’ performers in two out of three discrimination tasks and 1 participant can be described as ‘poor’ performer in two out of three discrimination tasks. Further, as in the pre-test, ‘good discriminators’ were highly ranked on natural English vowel perception. Finally, contrary to what happened in the pre-test, those individuals who were most accurate in producing English vowels were also in the upper part of the table.

6.5 Summary of results

This chapter examined Greek speakers’ post-training performance on the same battery of perceptual tasks used in the pre-test with natural and synthetic Greek and English vowels, a non-speech (F2 only) discrimination task and an English vowel production task. The relationships between tasks were explored that would provide some explanations for individual differences in the trainees’ post-test performance. Since all pre/post test comparisons for the control group showed no learning from test repetition, the following apply only to the group of Greek speakers who received auditory training.

First, the effect of training on Greek speakers' identification of natural English vowels in quiet and in noise was examined. Since participants were trained in quiet, any improvement in English vowel perception in noise would indicate generalization of learning to more naturalistic settings. A new speaker that the participants had not heard before was also included in the post-test to examine generalization of learning both in quiet and in noise. It is important here to remember that the pre/post tests used different speakers and different words to those used in the training materials so there is a definite degree of generalization even looking at the post-test in quiet and without looking at the results for the new speaker. Training significantly improved Greek speakers' identification performance about 20 percentage points in quiet (from 56.9% to 76.67% correct) and 15 percentage points in noise (from 40.3% to 55.3% correct) and learning generalized to a new speaker both in quiet (78.6% correct) and in noise (60.9% correct). Subjects performed better with the new speaker, especially in noise. One plausible explanation could be that the particular speaker was more intelligible than the other two speakers. Regarding the effect of English vowel training on Greek speakers' identification of Greek vowels in noise, it comes as no surprise that no change from pre-test to post-test was found. Further, no change from pre-test to post-test was found in the identification scores for Greek vowels in noise for the group of Greek speakers who received no training which confirmed that no learning would come from test repetition.

Next, the effect of training on Greek speakers' identification of synthetic Greek and English vowels was examined. Regarding the location of phoneme boundary, no changes from pre-test to post-test were found; the trainees placed the boundary at around the same position across languages and tests. As expected, training did not change Greek speakers' consistency in labeling the Greek /i/-/e/ and /a/-/o/ continua. Training improved Greek speakers' consistency in labeling the English /iː/-/ɪ/ *natural duration* and /iː/-/ɪ/ *neutralized duration* continua but not the /æ/-/ʌ/ continuum. In fact, after perceptual training, identification consistency for /iː/-/ɪ/ *natural duration* and /iː/-/ɪ/ *neutralized duration* reached that for /æ/-/ʌ/, suggesting that participants may reached a limit in their ability to identify synthetic English vowels, at least after five sessions of training. Finally, perceptual training did not change subjects' discrimination of the two Greek continua, the three English continua or the non-speech continuum. The lack of any improvement in English vowel discrimination suggests that listeners may learn to better label categories but their discrimination does not change, at least after this short period of training.

Perceptual training also improved English vowel production by Greek speakers as shown by both the success with which English listeners identified English vowels produced in the post-test compared to the same vowels produced in the pre-test and an acoustic analysis of those vowels. Overall identification scores improved from 61.9% correct in the pre-test to 75.8% correct in the post-test. Correct identification improved for all vowels although probably due to large variability in scores improvement was significant only for /iː/, /æ/, /ʌ/, /ɔː/ and /uː/. Acoustic analyses confirmed that English vowels produced by Greek speakers after perceptual training were more differentiated than before training. The trainees also learned to make the duration distinctions between English vowels more clear although in some cases they produced some vowels even longer than native English speakers did. Importantly, there was one vowel, namely English /ʌ/ which was produced with a shorter duration in the post-test than in the pre-test. This demonstrates that improvement is not limited to learning to produce longer vowels in an L2 than in L1 as commonly reported in the literature.

Correlation analyses showed that, as in the pre-test, individual patterns of discrimination accuracy extended across L1, L2 and non-speech so that participants appeared to be consistently 'good' or 'poor' discriminators. Further, English vowel production was correlated with the L1 DISCRIMINATION composite score and non-speech discrimination as well as with natural English vowel perception in quiet and in noise, supporting the existence of a perception-production link after perceptual training. Finally, post-test natural English vowel perception and English vowel production were correlated with pre-test L1 vowel discrimination, L2 vowel discrimination and non-speech discrimination, suggesting that those individuals who were successful discriminators were those who performed well after perceptual training, a finding that favours the *auditory processing* hypothesis.

Chapter 7

General discussion and conclusion

This thesis examined the acquisition of English vowels by native speakers of Greek with the ultimate goal of shedding some light on the issue of individual variability in L2 vowel learning. Current theoretical accounts offer several explanations for individual differences in learners' success in acquiring an L2. However, the vast majority of these explanations concern experience-related factors such as the relationship between the segmental inventory of the L1 and the L2, the age of L2 learning, the length of residence in L2 country and the degree of ongoing L1 use thus providing no explanation for differences in performance within groups of learners with similar profiles in terms of those background factors. This thesis aimed mainly at investigating how vowel processing in L2 is related to individual variability in vowel processing in L1 and frequency discrimination acuity in a relatively (given the difficulty in controlling all of the factors that have been found to influence L2 learning well) homogenous L2 group. The effects of auditory training on the perception of natural and synthetic L2 (and L1) vowels, the production of L2 vowels and the perception of a non-speech continuum were examined focusing on natural English vowel perception not only in quiet but also in noise conditions and on English vowel production. Another issue addressed concerned the availability or not of durational cues in L2 vowel perception for speakers with no such L1 experience. Given the nature of the instruction the participants in this work had received, they are different to those considered in many L2 studies where L2 learners are immersed in the L2 country.

7.1 Perceptual salience of duration for L2 learners?

Study 1 tested the perceptual assimilation (Experiment 1) and discrimination (Experiment 2) of Southern British English vowels by Greek speakers in two consonantal contexts, /bVb/ and /bVp/. The use of these contexts allowed measuring the effect of vowel duration differentiations introduced by the voicing vs. voicelessness of the stop consonant following the vowel on Greek speakers' perception of English vowels. Experiment 1 showed that Greek speakers were sensitive to such differentiations which seems against McAllister *et al.*'s (2002) feature hypothesis and, on the surface, in line with Bohn's (1995) desensitization hypothesis. However, separate analyses conducted for each vowel testing the effect of context on assimilation patterns showed that English vowels fitted better to Greek categories when their duration was closer to the duration of the spectrally closest Greek vowel. For instance, English /iː/ was found to be perceptually closer to Greek /i/ (which has a duration of around 100 ms) when placed in a /bVb/ context than when placed in a /bVp/ context the reason being that in the former case English /iː/ had a duration of 195 ms whereas in the latter case English /iː/ had a duration of 145 ms. It therefore seems that Greek speakers assimilate L2 vowels to their L1 categories on the basis of durational cues in the same way they assimilate L2 vowels to their L1 categories on the basis of spectral cues which is different from saying that Greek speakers are able to use duration because it is a salient cue compared to spectral cues. Experiment 2 showed that, in general, perceptual assimilation patterns predicted discrimination accuracy as hypothesized by the Perceptual Assimilation Model (Best, 1995; Best *et al.*, 2001; Best & Tyler, 2007). The results concerning the use of duration confirmed that Greek speakers were sensitive to durational cues when discrimination English vowels. However, separate analyses for each vowel testing the effect of context on English vowel discrimination showed that Greek speakers where not simply comparing the durations of the two vowels in each English pair as the desensitization hypothesis would predict. Instead, Greek speakers' discrimination accuracy depended on the cross-language relationships in terms of both durational and spectral cues.

Taken together, the results of Experiments 1 and 2 suggest that L2 vowels undergo both *temporal* and *spectral* perceptual assimilation to L1 category/categories and hence duration does not have a special status in L2 vowel perception compared to that of spectral cues. L2 learners who do not exploit duration in L1 may have access to temporal cues in an L2 provided that their 'single' L1 duration category does not temporally interfere with the perception of a given L2 contrast. The fact that the more-often tested English contrast in the literature, namely English /iː/-/ɪ/ does not suffer from L1 temporal interference seems to be the reason for the widespread view that listeners with a 'single' L1 vowel duration category have access to durational cues irrespective of the contrast to be perceived. Seen in this context, the results of this study are compatible with the perceptual interference account (Iverson *et al.*, 2003; Kuhl *et al.*, 2006; Kuhl *et al.*, 2008) and the current L2 speech perception models (PAM: Best, 1995; SLM: Flege, 1995a) that emphasize the role of L1 transfer. It seems that what is transferred is not an increased or decreased temporal acuity, depending on previous experience with duration in vowel distinctions, as the feature hypothesis would predict. Instead, the listeners transfer their L1 temporal pattern, which may impede or aid L2 perception depending on the cross-language temporal relationships. One explanation for the fact that the Latin American Spanish speakers in McAllister *et al.*'s (2002) study did not show any sensitivity to durational cues is given by the authors of the study in the discussion of their results. They draw attention to the fact that in their study L2 vowel perception was assessed by means of a word recognition task which does not exclude the possibility that instead of being unable to distinguish short from long vowels, some of the participants simply did not know whether a word contained a short or a long vowel. For listeners with no previous experience with duration in L1 vowel distinctions a more sensitive task such as a discrimination task may be therefore needed to capture their sensitivity to that acoustic cue.

The results concerning the identification and discrimination of synthetic English vowels in Study 3 seem to support the finding that L2 listeners perceive L2 contrasts via temporal perceptual assimilation to L1 duration category/categories rather than simply comparing the duration of the two vowels in a contrast. Greek speakers were equally consistent in their identification of the /iː/-/ɪ/ *natural duration* continuum and the /iː/-/ɪ/ *neutralized duration* continuum in the pre-test and showed similar degree of improvement in their labelling ability for the two continua in the post-test. Further, Greek speakers showed similar accuracy in discriminating the two duration versions of the English /iː/-/ɪ/ continuum which indicates that they did not benefit from the existence of differences in duration between the two members of the /iː/-/ɪ/ *natural duration* continuum. Both synthetic /iː/ and /ɪ/ endpoints had a duration that fell within the duration of Greek /i/ which resulted in Greek speakers' failure to use duration effectively in the perception of the /iː/-/ɪ/ *natural duration* synthetic continuum.

7.2 Effects of perceptual training

Relatively few training studies in the literature have examined vowels (Lambacher *et al.*, 2005; Iverson & Evans, 2007a; Nishi & Kewley-Port, 2007b, 2008; Iverson & Evans, 2009). With the exception of Iverson & Evans (2007a) and Iverson & Evans (2009) who trained German and Spanish native speakers' perception of English vowels, research on vowel training has examined Japanese (Lambacher *et al.*, 2005; Nishi & Kewley-Port, 2007b) or Korean speakers (Nishi & Kewley-Port, 2008) and only Lambacher *et al.* (2005) examined the impact of perceptual training on English vowel production. This work is the first training study in the literature to use a large pre/post battery of tests examining the trainees' perception of natural and synthetic L2 (English) and L1 (Greek) vowels, their English vowel production and their frequency discrimination ability. Further, natural English vowel perception was tested in quiet, as done in previous studies, and also in noise (English vowels were embedded in a multi-talker babble at an SNR of -4 dB). This tested whether learning was robust enough to translate to improvements in L2 vowel perception in degraded listening conditions.

First, English vowel training had no effect on the perception of natural or synthetic Greek vowels. There were no changes in natural Greek vowel identification in noise

(SNR = -10 dB) or in synthetic Greek vowel identification (boundary locations and identification slopes) and discrimination (thresholds). This is not surprising given the stability of the L1 categories after many years of experience with the ambient language.

With respect to L2 vowel learning, the results confirmed the effectiveness of high-variability phonetic training on L2 identification (+20 percentage points) for another L2 population. That is a larger degree of improvement than that reported in Iverson & Evans (2009) concerning Spanish speakers' learning of English vowels. This work used the same training materials and method as in Iverson & Evans (2009), however, there were 10 English vowels as response options instead of 14 English vowels in Iverson & Evans (2009) thus making the task less demanding. Although L2 identification was significantly lower in noise than in quiet (Mayo *et al.*, 1997; Cutler *et al.*, 2004), this work showed that learning is robust enough to transfer to vowel identification in noise conditions and to generalize to a new talker heard in noise. Given that even early bilinguals show decreased speech intelligibility in noise even when they show the same intelligibility rates as monolinguals in quiet (e.g. Mayo *et al.*, 1997) this transfer of learning supports further the effectiveness of high-variability auditory training.

Results also showed that learning generalized to synthetic speech reflected in steeper identification slopes which is consistent with the improvement seen for natural English vowels and further indicates successful learning. This improvement was found for the two duration versions of the English /iː/-/ɪ/ continuum but not for English /æ/-/ʌ/. Maybe only 5 sessions of training can improve up to a certain degree identification consistency for L2 vowel categories. English vowel training had no effect on English vowel discrimination. In a recent study, Heeren & Schouten (2008) successfully trained Dutch native speakers in identifying the Finish /t/-/tː/ contrast but without improving in their discrimination of the same contrast. The present work of course differs from Heeren & Schouten (2008) in terms of both the type of L2 contrasts tested (vowels vs. consonants), and of the type of discrimination task used (adaptive vs. non-adaptive) and therefore the results are not straightforwardly comparable. Iverson & Evans (2009) hypothesized that laboratory training improves the ability of the trainees to apply their already existing L1 and L2 category knowledge to L2 identification without changing the representation of L2 categories. Despite improving in English identification accuracy, Spanish and German speakers in Iverson & Evans (2009) did not improve in their English vowel space mapping after auditory training, i.e. their best exemplar locations for English vowels did not approach more the target vowels. The authors proposed that high-variability phonetic training may be more effective than low-variability training because stimulus variability trains the subjects in applying L2 categories to real speech. The results concerning improvement in labelling ability for at least one L2 contrast in this work suggest that some change in the representation of L2 categories is possible although the trainees' identification slopes for L2 vowels were still much shallower than their identification slopes for L1 vowels.

Results also demonstrated that perceptual training improved the production of English vowels by Greek speakers as judged by native English listeners and confirmed by an acoustic analysis of those English vowels. Instead of using their 5 Greek vowel qualities in English vowel production, the trainees learned to spectrally differentiate English vowels. With respect to length distinctions, the trainees learned not only to produce English /iː/, /ɜː/, /æ/, /ɑː/, /ɔː/ and /uː/ (all of the five long vowels in English and /æ/ whose duration approximates that of the long vowels) with longer durations than in the pre-test but also to produce English /ʌ/ with a shorter duration than in the pre-test thus approximating the target duration of that vowel. When comparing the durations of English /iː/, /ɜː/, /æ/, /ɑː/, /ɔː/ and /uː/ produced by Greek speakers after perceptual training with English vowels produced by English speakers in Study 1, it was found that these were somewhat longer than the target durations. However, it has to be noted that vowels in these comparisons were uttered in voiceless contexts (/bVt/ for Greek speakers and /bVp/ for English speakers) whereas during training trainees were exposed to both voiced and voiceless contexts. It therefore seems that during training subjects were able to pick up that some

English vowels were long; however, they failed to learn the allophonic variation in English according to which vowels are shorter before voiceless stops than before voiced stops. The improvement in English vowel production replicates the success of perceptual training on L2 production both for consonants (Bradlow *et al.*, 1997; Bradlow *et al.*, 1999) and vowels (Lambacher *et al*., 2005). Finally, as discussed in detail in the following section concerning individual differences in L2 learning, L2 vowel perception and production were aligned for individuals after perceptual training supporting a perception-production link (Flege *et al.*, 1997a; Flege, 1999; Flege *et al.*, 1999a).

7.3 Individual differences

In line with previous work in L1 (Kewley-Port & Watson, 1994; Kewley-Port, 2001; Gerrits & Schouten, 2004), L2 (Bradlow *et al.*, 1997; Bradlow *et al.*, 1999; Hazan *et al.*, 2005) and non-speech perception (Surprenant & Watson, 2001; Kidd *et al.*, 2007; Lee *et al.*, 2007), there were large individual differences in performance across tasks both before and after auditory training. Despite this variability, subjects were generally consistent across tasks, i.e. this variability was not random for most (but not all) tasks. First, natural English vowel identification in quiet correlated with natural English vowel identification in noise both before training (for all 28 participants) and after training (for the trained participants). There were also significant positive correlations between L1, L2 and non-speech frequency discrimination. That is, individuals who were successful in Greek vowel discrimination were also successful in English vowel discrimination and non-speech discrimination (across pre/post-tests). Importantly, L1 DISCRIMINATION score, L2 DISCRIMINATION score, and non-speech discrimination correlated with natural English vowel identification in quiet and in noise (across pre/post-tests). Before training, English vowel production did not correlate with natural English vowel identification in quiet or in noise but did so after training. Finally, natural English vowel identification and English vowel production in the post-test were correlated with L1, L2 and non-speech frequency discrimination accuracy in the pre-test, suggesting that those individuals who were more efficient discriminators when tested for the first time achieved better scores in English vowel perception and production after auditory training than those individuals who were less efficient discriminators in the pre-test.

In the Introduction it was hypothesized that individual differences in L2 vowel perception and production may be explained on an *L1 phonetic* and/or an *auditory processing* level. The *L1 phonetic* hypothesis was based on the well-attested effect of L1 experience on L2 learning and is compatible with current cross-language/L2 models. The SLM and the NLM (and NLM-e) are of most relevance here. According to SLM, age effects are due to changes in how the L1 and L2 systems interact (Flege *et al.*, 2003). As the L1 categories become more established with age (Hazan & Barrett, 2000), they become more likely to 'assimilate' L2 sounds. Similarly, according to the NLM, L1 experience sharpens L1 perception but unavoidably interferes with L2 learning (Iverson *et al.*, 2003; Kuhl *et al.*, 2006; Kuhl *et al.*, 2008). If this L1-L2 perception trade-off extents to adult learners, we would expect listeners with more robust L1 categories to find it harder to retune their system when learning an L2. Similarly, Maye (2007) offers an attentional-weighting explanation for individual differences in L2 learning based on Goldinger's (2007) complementary-systems model; it is hypothesized that there might be individual differences in the long-term acquisition of L1-appropriate attentional cue weighting, i.e. in the acquisition of a filter for exemplar encoding which would result in less efficient phonological processing but at the same time the ability to develop native-like L2 phonologies. Indirect support for the *auditory processing* hypothesis comes from studies by Wong and colleagues showing that auditory pitch ability, as measured using non-speech stimuli, can predict success in the use of pitch patterns in lexical identification by L2 learners (Lee *et al.*, 2007; Wong & Perrachione, 2007). The reader should bear in mind that these two hypotheses are not necessarily mutually exclusive, i.e. a non-speech auditory ability may underlie both L1 and L2 vowel processing.

The results showing that participants were consistent across L1, L2 and non-speech discrimination tasks and that discrimination accuracy related to post-test natural English vowel identification in quiet and in noise and English vowel production showcase an underlying auditory acuity component for L2 speech processing. Previous studies have failed to find such a connection between speech and non-speech tasks, however, as noted in Surprenant & Watson (2001), speech and non-speech are typically measured using tasks tapping into different processing abilities; speech ability is measured via recognition-in-noise tasks whereas non-speech ability is measured using discrimination tasks that require analytic listening. Surprenant & Watson (2001) propose that non-speech discrimination or identification tasks that require more global listening may be more appropriate for the prediction of individual variability in speech perception. Rather than using more global non-speech tasks, this work employed more analytical speech tasks and a connection was indeed found between non-speech processing and both L1 and L2 vowel processing.

These findings, of course, do not reject the importance of L1 interference when learning an L2 as acknowledged by current cross-language models. At a group level, there was a clear effect of L1 vowel experience on L2 vowel perception and production. Greek speakers had difficulty in perceiving and producing English vowels as shown in Study 1 examining L2 discrimination of natural English vowels and in Study 3 examining the identification slopes and discrimination performance for synthetic English vowels, the identification of natural English vowels in quiet and in noise and the production of English vowels. What is shown in this work is that while L1 experience affects L2 vowel processing, some people are better in using acoustic information to overcome L1 biases, and in that respect the hypothesis by Maye (2007) seems to be supported.

7.4 Limitations and future research

One limitation with regard to the training programme in this work is the lack of a task that would test long-term retention of learning as done is some but not all training studies in the literature. This was mainly due to practical reasons as retention testing would have required another trip to Greece to retest all participants. However, given that all studies showing generalization of training also show retention (Lively *et al.*, 1994; Bradlow *et al.*, 1997; Bradlow *et al.*, 1999; Iverson & Evans, 2009 with the same materials), there is no reason to expect that no retention would be found if tested especially given the robustness of training shown in this work (generalization of learning to a new talker, transfer to noise conditions, transfer to synthetic speech identification and production improvement).

Given that non-speech perception was tested using a formant-like stimulus (range = 1250 -1500 Hz) it might be claimed that a pure-tone task should be more appropriate to test non-speech discrimination accuracy. The rationale behind the use of the particular type of non-speech task was the following: the non-speech continuum should have a harmonic structure (thus sharing similar acoustic properties with the vowel continua) without resembling speech. To this end, this was the first task completed by subjects. Still, future research could include a broader range of non-speech tasks with pure-tones or formant-like stimuli at different frequencies. Future research could also use regression analyses to quantify the effect of factors related to language aptitude such as PSTM and musical ability as well as motivational and social factors on acquiring the sounds of an L2. Finally, a test of generalization of production improvement to sentence materials and, ultimately, to conversational speech could assess the effectiveness of training on more naturalistic materials that are close to everyday communication.

Appendix A

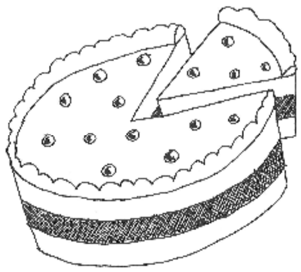
Table I: Most frequent and second most frequent percentage classification of English vowels in terms of Greek vowel categories with the relevant goodness ratings assigned.

	/bVb/						/bVp/					
	Modal response			2nd response			Modal response			2nd response		
SBE	Greek listeners											
iː	i	100	5.0	-	-	-	i	100	5.4	-	-	-
ɪ	i	100	5.4	-	-	-	i	100	5.0	-	-	-
e	e	97	5.0	-	-	-	e	92	5.2	i	4	2.0
ɜː	e	87	3.0	o	9	2.5	e	77	3.2	o	14	2.4
æ	a	95	4.7	o	5	1.5	a	95	5.2	o	3	1.0
ʌ	a	62	4.2	o	36	5.1	a	66	4.1	o	30	4.3
ɑː	o	57	4.0	a	43	3.6	o	54	4.1	a	46	3.9
ɒ	o	97	5.0	a	3	1.0	o	97	5.0	-	-	-
ɔː	o	55	4.1	u	45	3.8	o	52	4.1	u	48	3.8
ʊ	u	92	4.0	o	4	1.8	u	84	4.2	o	8	2.4
uː	u	82	3.8	i	14	2.5	u	92	3.8	i	8	1.9

SBE = Southern British English

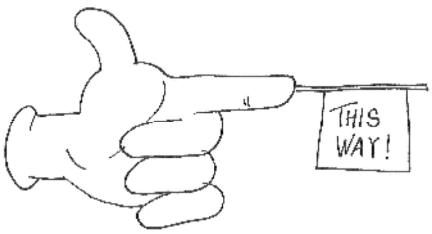
Appendix B

Pictures used in the synthetic speech perception experiment. The first four were used in the Greek vowel continua and the other four in the English vowel continua.



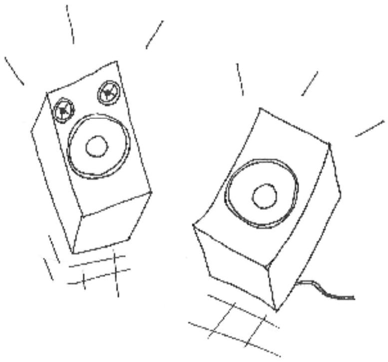
/pita/ "pie"

/peta/ "throw"



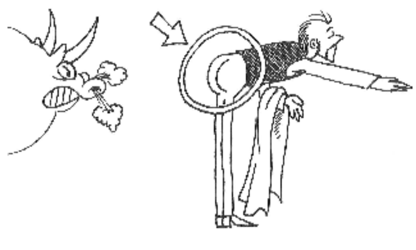
/pate/ "go"

/pote/ "when"



/biːt/ “beat”

/bɪt/ “bit”



/bæt/ “bat”

/bʌt/ “butt”

Bibliography

Alexander, J.A., Wong, P.C.M. & Bradlow, A.R. (2005) Lexical tone perception in musicians and nonmusicians. In: *9th European Conference on Speech Communication and Technology*, Lisboa, Portugal.

Aoyama, K., Flege, J.E., Guion, S.G., Akahane-Yamada, R. & Yamada, T. (2004) Perceived phonetic dissimilarity and L2 speech learning: The case of Japanese /r/ and English /l/ and /r/. *Journal of Phonetics*, **32**, 233-250.

Arvaniti, A. (1999) Standard Modern Greek. *Journal of the International Phonetic Association*, **29**, 167-172.

Arvaniti, A. (2001) Comparing the phonetics of single and geminate consonants in Cypriot and Standard Greek. In: *Fourth International Conference on Greek Linguistics*, pp. 37-44. University Studio Press, Thessaloniki, Greece.

Arvaniti, A. (2007) Greek phonetics: The state of the Art. *Journal of Greek Linguistics*, **8**, 97-208.

Asher, J.J. & García, R. (1969) The optimal age to learn a second language. *The Modern Language Journal*, **53**, 334-341.

Aslin, R.N., Pisoni, D.B., Hennessy, B.L. & Perey, A.J. (1981) Discrimination of Voice Onset Time by Human Infants: New findings and implications for the effects of early experience. *Child Development*, **52**, 1135-1145.

Baltazani, M. (2007) Prosodic rhythm and the status of vowel reduction in Greek. In: *17th International Symposium on Theoretical and Applied Linguistics*, pp. 31-43, Thessaloniki, Greece.

Beddor, P.S. & Gottfried, T.L. (1995) Methodological issues in cross-language speech perception research with adults. *Speech perception and linguistic experience: Theoretical and methodological issues in cross-language speech research* (ed. by W. Strange), pp. 207-232. York Press, Timonium, MD.

Best, C.T. (1995) A direct realist view of cross-language speech perception. *Speech perception and linguistic experience* (ed. by W. Strange), pp. 171-204. York Press.

Best, C.T., Mcroberts, G.W. & Goodell, E. (2001) Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *Journal of the Acoustical Society of America*, **109**, 775-794.

Best, C.T., Mcroberts, G.W. & Sithole, N.M. (1988) Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance*, **14**, 345-360.

Best, C.T. & Tyler, M.D. (2007) Nonnative and second-language speech perception: Commonalities and complementarities. *Second language speech learning* (ed. by M. Munro and O.-S. Bohn), pp. 13-34. John Benjamins Publishing, Amsterdam.

Boersma, P. & Weenink, D. (2005) Praat: Doing phonetics by computer. In, Amsterdam, The Netherlands.

Bohn, O.-S. (1995) Cross-language speech perception in adults: First language transfer doesn't tell it all. *Speech perception and linguistic experience: Issues in cross-language research* (ed. by W. Strange), pp. 279-304. Baltimore: York Press.

Bongaerts, T., Planken, B. & Schils, E. (1995) Can late learners attain a native accent in a foreign language? A test of the critical period hypothesis. *The age factor in second language acquisition* (ed. by D. Singleton and Z. Lengyel), pp. 30-50. Multilingual Matters, Clevedon, UK.

Bongaerts, T., Van Summeren, C., Planken, B. & Schils, E. (1997) Age and ultimate attainment in the pronunciation of a foreign language. *Studies in Second Language Acquisition*, **19**, 447-466.

Bosch, L., Costa, A. & Sebastián-Gallés, N. (2000) First and second language vowel perception in early bilinguals. *European Journal of Cognitive Psychology*, **12**, 189-221.

Bosch, L. & Sebastian-Galles, N. (2003) Simultaneous bilingualism and the perception of a language-specific vowel contrast in the first year of life. *Language and Speech*, **46**, 217-243.

Botinis, A., Fourakis, M. & Prinou, I. (2000) Acoustic structure of the Greek stop consonants. *Glossologia*, **11-12**, 167-200.

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

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

Browman, C.P. & Goldstein, L. (1989) Articulatory gestures as phonological units. *Phonology*, **6**, 201-251.

Carney, A.E., Gregory, P.W., G. P. & Viemeister, N.F. (1977) Noncategorical perception of stop consonants differing in VOT. *The Journal of the Acoustical Society of America*, **62**, 961-970.

Carroll, J.B. (1981) Twenty-five years of research on foreign language aptitude. *Individual differences and universals in language learning aptitude* (ed. by K. Diller), pp. 83-118. MA: Newbury House, Rowley.

Carroll, J.B. (1993) *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge University Press, New York.

Carroll, J.B. & Sapon, S. (1959) *Modern language aptitude test (MLAT): Manual*. The Psychological Corporation, New York.

Cebrian, J. (2006) Experience and the use of duration in the categorization of L2 vowels. *Journal of Phonetics*, **34**, 372-387.

Crandell, C.C. (1991) Individual differences in speech recognition ability: Implications for hearing aid selection. *Ear and Hearing*, **12**, 100S-108S.

Cutler, A., Andrea, W., A., Smits, R. & Cooper, N. (2004) Patterns of English phoneme confusions by native and non-native listeners. *The Journal of the Acoustical Society of America*, **116**, 3668-3678.

Docherty, G. (1992) *The timing of voicing in British English obstruents*. New York: Foris, Berlin.

Dörnyei, Z. (2005) *The psychology of the language learner: Individual differences in second language acquisition*. Mahwah, NJ: Lawrence Erlbaum Associates.

Edman, T.R. (1980) Learning of intraphonemic discrimination for several synthetic speech continua. Unpublished PhD thesis, University of Minnesota.

Ehrman, M.E. & Oxford, R.L. (1995) Cognition plus: Correlates of language learning success. *The Modern Language Journal*, **79**, 67-89.

Eimas, P.D., Siqueland, E.R., Jusczyk, P. & Vigorito, J. (1971) Speech perception in infants. *Science*, **171**, 303-306.

Elliott, R.E. (1995) Field independence/dependence, hemispheric specialization, and attitude in relation to pronunciation accuracy in Spanish as a foreign language. *The Modern Language Journal*, **79**, 356-371.

Escudero, P. (2005) Linguistic perception and second language acquisition. Explaining the attainment of optimal phonological categorization. Unpublished PhD thesis, Utrecht University, Utrecht.

Escudero, P. & Boersma, P. (2004) Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition*, **26**, 551-585.

Evans, B.G. & Iverson, P. (2004) Vowel normalization for accent: An investigation of best exemplar locations in northern and southern British English sentences. *Journal of the Acoustical Society of America*, **115**, 352-361.

Evans, B.G. & Iverson, P. (2007) Plasticity in vowel perception and production: A study of accent in young adults. *Journal of the Acoustical Society of America*, **121**, 3814-3826.

Flege, J.E. (1987) A critical period for learning to pronounce foreign languages? *Applied Linguistics*, **8**, 162-177.

Flege, J.E. (1995a) Second language speech learning: Theory, findings, and problems. *Speech perception and linguistic experience: Issues in cross-language research* (ed. by W. Strange), pp. 233-277. York Press, Baltimore.

Flege, J.E. (1995b) Two procedures for training a novel second language phonetic contrast. *Applied Psycholinguistics*, **16**, 425-442.

Flege, J.E. (1999) The relation between L2 production and perception. In: *14th International Congress of Phonetic Sciences*, University of California, Berkeley.

Flege, J.E. (2002) Interactions between the native and second-language phonetic systems. *An integrated view of language development: Papers in honor of Henning Wode* (ed. by P. Burmeister, T. Piske and A. Rohde), pp. 217-244. Trier: Wissenschaftlicher Verlag.

Flege, J.E., Bohn, O.-S. & Jang, S. (1997a) Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics*, **25**, 437-470.

Flege, J.E. & Fletcher, K.L. (1992) Talker and listener effects on degree of perceived foreign accent. *Journal of the Acoustical Society of America*, **91**, 370-389.

Flege, J.E., Frieda, E.M. & Nozawa, T. (1997b) Amount of native-language (L1) use affects the pronunciation of an L2. *Journal of Phonetics*, **25**, 169-186.

Flege, J.E. & Liu, S. (2001) The effect of experience on adults' acquisition of a second language. *Studies in Second Language Acquisition*, **23**, 527-552.

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

Flege, J.E., Mackay, I.R.A. & Meador, D. (1999a) Native Italian speakers' production and perception of English vowels. *Journal of the Acoustical Society of America*, **106**, 2973-2987.

Flege, J.E., Munro, M. & Mackay, I.R.A. (1995) Factors affecting strength of perceived foreign accent in a second language. *Journal of the Acoustical Society of America*, **97**, 3125-3134.

Flege, J.E., Schirru, C. & Mackay, I.R.A. (2003) Interaction between the native and second language phonetic subsystems. *Speech Communication*, **40**, 467-491.

Flege, J.E. & Schmidt, A.M. (1995) Native speakers of Spanish show rate-dependent processing of English stop consonants. *Phonetica*, **52**, 90-111.

Flege, J.E., Yeni-Komshian, G. & Liu, H.-M. (1999b) Age constraints on second language acquisition. *Journal of Memory and Language*, **41**, 78-104.

Fourakis, M., Botinis, A. & Katsaiti, M. (1999) Acoustic characteristics of Greek vowels. *Phonetica*, **56**, 28-43.

Fowler, C.A. (1986) An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics*, **14**, 3-28.

Fry, D.B., Abramson, A.S., Eimas, P.D. & Liberman, A.M. (1962) The identification and discrimination of synthetic vowels. *Language and Speech*, **5**, 171-189.

Gagne, J.P. & Zurek, P.M. (1988) Resonance-frequency discrimination. *Journal of the Acoustical Society of America*, **83**, 2293-2299.

Gardner, R.C. (1980) On the validity of affective variables in second language acquisition: Conceptual, contextual and statistical consideration. *Language Learning*, **30**, 255-270.

Gerrits, E. & Schouten, M.E.H. (2004) Categorical perception depends on the discrimination task. *Perception and Psychophysics*, **6**, 363-378.

Giegerich, H. (1992) *English phonology: An introduction*. Cambridge University Press, Cambridge.

Glasberg, B.R. & Moore, B.C.J. (1990) Derivation of auditory filter shapes from notched-noise data. *Hearing Research*, **47**, 103-138.

Goldinger, S.D. (2007) A complementary-systems approach to abstract and episodic speech perception. In: *16th International Congress of Phonetic Sciences*, pp. 49-54, Saarbrucken, Germany.

Goto, H. (1971) Auditory perception by normal japanese adults of the sounds “l” and “r”. *Neuropsychologia*, **9**, 317-323.

Gottfried, T. (1984) Effects of consonant context on the perception of French vowels. *Journal of Phonetics*, **12**, 91-114.

Guion, S.G., Flege, J.E., Akahane-Yamada, R. & Pruitt, J.C. (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**, 2711-2724.

Halle, P.A., Best, C.T. & Levitt, A. (1999) Phonetic vs. Phonological influences on French listeners' perception of American English approximants. *Journal of Phonetics*, **27**, 281-306.

Harnsberger, J.D. (2001) On the relationship between identification and discrimination of non-native nasal consonants. *Journal of the Acoustical Society of America*, **110**, 489-503.

Hazan, V. & Barrett, S. (2000) The development of phonemic categorisation in children aged 6 to 12. *Journal of Phonetics*, **28**, 377-396.

Hazan, V. & Markham, D. (2004) Acoustic-phonetic correlates of talker intelligibility for adults and children. *Journal of the Acoustical Society of America*, **116**, 3108-3118.

Hazan, V. & Rosen, S. (1991) Individual variability in the perception of cues to place contrasts in initial stops. *Perception and Psychophysics*, **49**, 187-200.

Hazan, V., Sennema, A., Faulkner, A., Ortega-Llebaria, M., Iba, M. & Chung, H. (2006) The use of visual cues in the perception of non-native consonant contrasts. *The Journal of the Acoustical Society of America*, **119**, 1740-1751.

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.

Heeren, W.F.L. & Schouten, M.E.H. (2008) Perceptual development of phoneme contrasts: How sensitivity changes along acoustic dimensions that contrast phoneme categories. *Journal of the Acoustical Society of America*, **124**, 2291-2302.

Hirata, Y. (2004) Training native English speakers to perceive Japanese length contrasts in word versus sentence contexts. *The Journal of the Acoustical Society of America*, **116**, 2384-2394.

Hirata, Y., Whitehurst, E. & Cullings, E. (2007) Training native English speakers to identify Japanese vowel length contrast with sentences at varied speaking rates. *The Journal of the Acoustical Society of America*, **121**, 3837-3845.

Højen, A. & Flege, J.E. (2006) Early learners' discrimination of second-language vowels. *The Journal of the Acoustical Society of America*, **119**, 3072-3084.

Horwitz, E.K. (1987) Linguistic and communicative competence: Reassessing foreign language aptitude. *Foreign language learning: A research perspective* (ed. by B. Vanpatten, T.R. Dvorak and J.F. Lee), pp. 146-157. Newbury House, Cambridge.

House, A.S. (1961) On vowel duration in English. *Journal of the Acoustical Society of America*, **33**, 1174-1182.

Huckvale, M. (2008) Speech Filing System tools for speech research (SFS) University College London. URL http://www.phon.ucl.ac.uk/resource/sfs/

Hummel, K.M. (2009) Aptitude, phonological memory, and second language proficiency in nonnovice adult learners. *Applied Psycholinguistics*, **30**, 225-249.

Ingram, J.C.L. & Park, S.-G. (1997) Cross-language vowel perception and production by Japanese and Korean learners of English. *Journal of Phonetics*, **25**, 343-370.

Iverson, P. & Evans, B.G. (2003) A goodness optimization method for investigating phonetic categorization. In: *15th International Congress of Phonetic Sciences*, Barcelona, Spain.

Iverson, P. & Evans, B.G. (2007a) Auditory training of English vowels for first language speakers of Spanish and German. In: *16th International Congress of Phonetic Sciences*, pp. 1625-1628, Saarbrucken, Germany.

Iverson, P. & Evans, B.G. (2007b) Learning English vowels with different first-language vowel systems: Perception of formant targets, formant movement, and duration. *Journal of the Acoustical Society of America*, **122**, 2842-2854.

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**, 866-877.

Iverson, P., Kuhl, P.K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A. & C., S. (2003) A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition*, **87**, 47-57.

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

Johnson, D.M., Watson, C.S. & Jensen, J.K. (1987) Individual differences in auditory capabilities. *Journal of the Acoustical Society of America*, **81**, 427-438.

Johnson, J. & Newport, E. (1989) Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology*, **21**, 60-99.

Johnson, K., Flemming, E. & Wright, R. (1993) The hyperspace effect: Phonetic targets are hyperarticulated. *Language*, **69**, 505-528.

Kewley-Port, D. (2001) Vowel formant discrimination: Effects of stimulus uncertainty, consonantal context and training. *Journal of the Acoustical Society of America*, **110**, 2141-2155.

Kewley-Port, D. & Watson, C.S. (1994) Formant frequency discrimination for isolated English vowels. *Journal of the Acoustical Society of America*, **95**, 485-496.

Kewley-Port, D. & Zheng, Y. (1999) Vowel formant discrimination: Towards more ordinary listening conditions. **106**, 2945-2958.

Kidd, G.R., Watson, C.S. & Gygi, B. (2007) Individual differences in auditory abilities. *Journal of the Acoustical Society of America*, **122**, 418-435.

Klatt, D.H. & Klatt, L.C. (1990) Analysis, synthesis, and perception of voice quality variations among female and male talkers. *Journal of the Acoustical Society of America*, **87**, 820-857.

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

Kuhl, P.K., Conboy, B.T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M. & Nelson, T. (2008) Phonetic learning as a pathway to language: New data and native language magnet theory expanded (NLM-e). *Philosophical Transactions of the Royal Society*, **B, 363,** , 979-1000.

Kuhl, P.K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S. & Iverson, P. (2006) Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science*, **9**, 13-21.

Kuhl, P.K., Tsao, F.-M. & Liu, H.-M. (2003) Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning. *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 9096-9101.

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

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.

Lasky, R.E., Syrdal-Lasky, A. & Klein, R.E. (1975) VOT discrimination by four to six and a half month old infants from Spanish environments. *Journal of Experimental Child Psychology*, **20**, 215-225.

Lee, J., Perrachione, T., Dees, T. & Wong, P.C.M. (2007) Differential effects of stimulus variability and learners' pre-existing pitch perception ability in lexical tone learning by native English speakers. In: *16th International Congress of Phonetic Sciences*, Saarbrücken, Germany.

Lee, S., Potamianos, A. & Narayanan, S. (1999) Acoustics of children's speech: Developmental changes of temporal and spectral parameters. *Journal of the Acoustical Society of America*, **105**, 1455-1468.

Lengeris, A. (2009) Perceptual assimilation and L2 learning: Evidence from the perception of Southern British English vowels by native speakers of Greek and Japanese. *Phonetica*, **66**, 169-187.

Lengeris, A. & Hazan, V. (2007) Cross-language perceptual assimilation and discrimination of Southern British English vowels by native Greek and Japanese learners of English. In: *16th International Congress of Phonetic Sciences* (eds. J. Trouvain and W. Barry), pp. 1641-1644, Saarbruecken, Germany.

Lenneberg, E.H. (1967) *Biological foundations of language*.

Levitt, H. (1971) Transformed up-down methods in psychoacoustics. *Journal of the Acoustical Society of America*, **49**, 467-477.

Liberman, A.M., Harris, K.S., Eimas, P., Lisker, L. & Bastian, J. (1961) An effect of learning on speech perception: The discrimination of durations of silence with and without phonemic significance. *Language and Speech*, **4**, 175-195.

Liu, C. & Kewley-Port, D. (2004a) Formant discrimination in noise for isolated vowels. *Journal of the Acoustical Society of America*, **116**, 3119-3129.

Liu, C. & Kewley-Port, D. (2004b) Vowel formant discrimination for high-fidelity speech. *Journal of the Acoustical Society of America*, **116**, 1224-1233.

Liu, H.-M., Kuhl, P.K. & Tsao, F.-M. (2003) An association between mothers' speech clarity and infants' speech discrimination skills. *Developmental Science*, **6**, F1-F10.

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.

Lively, S.E., Pisoni, D.B., Yamada, R., 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.

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**, 874-886.

Mackay, I.R.A., Flege, J.E., Piske, T. & Schirru, C. (2001a) Category restructuring during second-language (L2) speech acquisition. *Journal of the Acoustical Society of America*, **110**, 516-528.

Mackay, I.R.A., Meador, D. & Flege, J.E. (2001b) The identification of English consonants by native speakers of Italian. *Phonetica*, **58**, 103-125.

Maye, J. (2007) Learning to overcome L1 phonological biases. In: *16th International Congress of Phonetic Sciences*, pp. 63-66, Saarbrucken, Germany.

Maye, J. & Weiss, D.J. (2003) Statistical cues facilitate infants' discrimination of difficult phonetic contrasts. In: *27th Annual Boston University Conference on Language Development.*

Maye, J., Werker, J.F. & Gerken, L. (2002) Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, **82**, B101-B111.

Mayo, H.L., Florentine, M. & Buus, S. (1997) Age of second-language acquisition and perception of speech in noise. *Journal of Speech and Hearing Research*, **40**, 686-693.

Mcallister, R., Flege, J.E. & Piske, T. (2002) The influence of L1 on the acquisition of Swedish quantity by native speakers of Spanish, English and Estonian. *Journal of Phonetics*, **30**, 229-258.

Meador, D., Flege, J.E. & Mackay, I.R.A. (2000) Factors affecting the recognition of words in a second language. *Bilingualism: Language & Cogniton*, **3**, 55-67.

Mermelstein, P. (1978) Difference limens for formant frequencies of steady-state and consonant-bound vowels. *Journal of the Acoustical Society of America*, **63**, 572-580.

Middelweerd, M.J., Festen, J.M. & Plomp, R. (1990) Difficulties with speech intelligibility in noise in spite of normal pure-tone audiogram. *Audiology*, **29**, 1-7.

Morrison, G.S. (2002) Effects of L1 duration experience on Japanese and Spanish listeners' perception of English high front vowels. In. Simon Fraser University, Burnaby, British Columbia, Canada.

Moyer, A. (1999) Ultimate attainment in L2 phonology. *Studies in Second Language Acquisition*, **21**, 81-108.

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

Nicolaidis, K. (2003) Acoustic variability of vowels in Greek spontaneous speech. In: *15th International Congress of Phonetic Sciences*, pp. 3221-3224, Barcelona, Spain.

Nishi, K. & Kewley-Port, D. (2007a) Second language vowel perception training: Effects of set size, training order, and native language. In: *16th International Congress of Phonetic Sciences*, pp. 1621-1624, Saarbrucken, Germany.

Nishi, K. & Kewley-Port, D. (2007b) Training Japanese listeners to perceive American English vowels: Influence of training sets. *Journal of Speech, Language, and Hearing Research*, **50**, 1496-1509.

Nishi, K. & Kewley-Port, D. (2008) Nonnative speech perception training using vowel subsets: Effects of vowels in sets and order of training. *Journal of Speech, Language, and Hearing Research*, **51**, 1480-1493.

O'brien, I., Segalowitz, N., Freed, B. & Collentine, J. (2007) Phonological memory predicts second language oral fluency gains in adults. *Studies in Second Language Acquisition*, **29**, 557-582.

Oyama, S. (1976) A sensitive period for the acquisition of a nonnative phonological system. *Journal of Psycholinguistic Research*, **5**, 261-283.

Pallier, C., Bosch, L. & Sebastián-Gallés, N. (1997) A limit on behavioral plasticity in speech perception. *Cognition*, **64**, B9-B17.

Pallier, C., Colomé, A. & Sebastián-Gallés, N. (2001) The influence of native-language phonology on lexical access: Exemplar-based versus abstract lexical entries. *Psychological Science*, **12**, 445-449.

Patkowski, M.S. (1980) The sensitive period for the acquisition of syntax in a second language. *Language Learning*, **30**, 449-472.

Patkowski, M.S. (1990) Age and accent in second language: A reply to James Emil Flege. **11**, 73-89.

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

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

Piper, T. & Cansin, D. (1988) Factors influencing the foreign accent. *The Canadian Modern Language Review*, **44**, 334-342.

Piske, T., Flege, J.E., Mackay, I.R.A. & Meador, D. (2002) The production of English vowels by fluent early and late Italian-English bilinguals. *Phonetica*, **59**, 49-71.

Piske, T., Mackay, I.R.A. & Flege, J.E. (2001) Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics*, **29**, 191-215.

Pisoni, D.B. (1973) Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception and Psychophysics*, **13**, 253-260.

Pisoni, D.B. (1975) Auditory short-term memory and vowel perception. *Memory and Cognition*, **3**, 7-18.

Pisoni, D.B., Aslin, R.N., Perey, A.J. & Hennessy, B.L. (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**, 297-314.

Polka, L. (1995) Linguistic influences in adult perception of non-native vowel contrasts. *Journal of the Acoustical Society of America*, **97**, 1286-1296.

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**, 577-592.

Polka, L. & Werker, J.F. (1994) Developmental changes in perception of non-native vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance*, **20**, 421-435.

Pruitt, J.S. (1993) Comments on "Training Japanese listeners to identify english /r/ and /l/: A first report" [J. S. Logan, S. E. Lively, and D. B. Pisoni, J. Acoust. Soc. Am. 89, 874-886 (1991)]. *Journal of the Acoustical Society of America*, **94**, 1146-1147.

Pruitt, J.S. (1995) Perceptual training on Hindi dental and retroflex consonants by native English and Japanese speakers. *Journal of the Acoustical Society of America*, **97**, 3417.

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

Repp, B.H. (1981) Two strategies in fricative discrimination. *Perception and Psychophysics*, **30**, 217-227.

Repp, B.H. (1984) Categorical perception: Issues, methods, findings. *Speech and language: Advances in basic research and practice* (ed. by N.J. Lass), pp. 244-335. FL: Academic Press, Orlando.

Rochet, B. (1995) Perception and production of second-language speech sounds by adults. *Speech perception and linguistic experience: Issues in cross-language speech research* (ed. by S. Strange), pp. 379-410. York Press, Timonium, MD.

Rosen, S.M. & Howell, P. (1987) Auditory, articulatory, and learning explanations of categorical perception in speech. *Categorical perception* (ed. by S. Harnad), pp. 113-160. New York, Cambridge.

Rupp, R.R. & Phillips, D. (1969) The effect of noise background on speech discrimination function in normal-hearing individuals. *Journal of Auditory Research*, **9**, 60-63.

Schmidt, A.M. (1996) Cross-language identification of consonants: Part 1. Korean perception of English. *Journal of the Acoustical Society of America*, **99**, 3201-3211.

Schmidt, A.M. & Flege, J.E. (1995) Effects of speaking rate changes on native and non-native production. *Phonetica*, **52**, 41-54.

Schouten, M.E.H. & Hessen, A.J.V. (1992) Modeling phoneme perception. I: Categorical perception. *The Journal of the Acoustical Society of America*, **92**, 1841-1855.

Sebastián-Gallés, N. & Soto-Faraco, S. (1999) On-line processing of native and non-native phonemic contrasts in early bilinguals. *Cognition*, **72**, 111-123.

Service, E. (1992) Phonology, working memory, and foreign-language learning. *The Quarterly Journal of Experimental Psychology*, **45A**, 21-50.

Sfakianaki, A. (2002) Acoustics characteristics of Greek vowels produced by adults and children. In: *14th International Symposium on Theoretical and Applied Linguistics*, pp. 383-394, Thessaloniki, Greece.

Skehan, P. (1998) *A cognitive approach to language learning*. Oxford University Press, Oxford.

Sleve, L.R. & Miyake, A. (2006) Individual differences in second-language proficiency: Does musical ability matter? *Psychological Science*, **17**, 675-681.

Snodgrass, J.G., Levy-Berger, G. & Haydon, M. (1985) *Human experimental psychology*. New York: Oxford University Press.

Sperbeck, M., Strange, W. & Ito, K. (2005) Training Japanese L2 learners to perceive difficult American vowel contrasts. *The Journal of the Acoustical Society of America*, **117**, 2400.

Strange, W. (1995) Cross-language studies of speech perception: A historical review. *Speech perception and linguistic experience: Issues in cross-language speech research* (ed. by W. Strange), pp. 3-45. Timonium, MD: York Press.

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

Streeter, L.A. (1976) Language perception of 2-month-old infants show both innate mechanisms and experience. *Nature*, **259**, 39-41.

Surprenant, A.M. & Watson, C.S. (2001) Individual differences in the processing of speech and nonspeech sounds by normal-hearing listeners. *Journal of the Acoustical Society of America*, **110**, 2085-2095.

Tahta, S., Wood, M. & Loewenthal, K. (1981) Foreign accents: Factors relating to transfer of accent from the first language to a second language. *Language and Speech*, **24**, 265-272.

Tajima, K., Kato, H., Rothwell, A., Akahane-Yamada, R. & Munhall, K.G. (2008) Training English listeners to perceive phonemic length contrasts in Japanese. *The Journal of the Acoustical Society of America*, **123**, 397-413.

Thompson, I. (1991) Foreign accents revisited: The English pronunciation of Russian immigrants. *Language and Learning*, **41**, 177-204.

Trehub, S.E. (1976) The discrimination of foreign speech contrasts by infants and adults. *Child Development*, **47**, 466-472.

Tsao, F.-M., Liu, H.-M. & Kuhl, P.K. (2004) Speech perception in infancy predicts language development in the second year of life: A longitudinal study. *Child Development*, **75**, 1067-1084.

Walley, A.C. & Flege, J. (1999) Effect of lexical status on childrens' and adults' perception of native and non-native sounds. *Journal of Phonetics*, **27**, 307-332.

Wang, Y., Jongman, A. & Sereno, J.A. (2003) Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training. *Journal of the Acoustical Society of America*, **113**, 1033-1043.

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

Watson, C.S., Jensen, J.K., Foyle, D.C., Leek, M.R. & Goldgar, D.E. (1982a) Performance of 146 normal adult listeners on a battery of auditory discrimination tasks. *Journal of the Acoustical Society of America*, **71**, S73.

Watson, C.S., Johnson, D.M., Lehman, J.R., Kelly, W.J. & Jensen, J.K. (1982b) An auditory discrimination test battery. *Journal of the Acoustical Society of America*, **71**, S73.

Werker, J.F., Gilbert, J.H.V., Humphrey, K. & Tees, R.C. (1981) Developmental aspects of cross-language speech perception. *Child Development*, **52**, 349-355.

Werker, J.F. & Lalonde, C.E. (1988) Cross-language speech perception: Initial capabilities and developmental change. *Developmental Psychology*, **24**, 672-683.

Werker, J.F. & Tees, R.C. (1983) Developmental change across childhood in the perception of non-native speech sounds. *Canadian Journal of Psychology*, **37**, 278-286.

Werker, J.F. & Tees, R.C. (1984) Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, **7**, 49-63.

Wong, P.C.M. & Perrachione, T.K. (2007) Learning pitch patterns in lexical identification by native English-speaking adults. *Applied Psycholinguistics*, **28**, 565-585.

Yeni-Komshian, G.H., Flege, J.E. & Liu, S. (2000) Pronunciation proficiency in the first and second languages of Korean-English bilinguals. *Bilingualism: Language and Cognition*, **3**, 131-149.