Counteracting Age Related Effects in L2 Acquisition: Training to distinguish between French vowels

Rachel M. M. Macdonald

PhD
The University of Edinburgh
2012

Declaration

I hereby declare that this thesis is of my own composition and that it contains no material previously submitted for the award of any other degree. The work reported in this thesis has been executed by myself, except where due acknowledgement is made in the text.

Rachel Macdonald

August 2012

Abstract

Two key methods of perceptually training difficult L2 contrasts are the perceptual fading (PF) technique and the high variability phonetic training (HVPT) technique, and perceptual benefits from using both of these methods have also been found to transfer to pronunciation. However, these techniques have not been compared in their classic forms (PF with one speaker vs. HVPT with multiple speakers) with regard to perceptual gains, nor have they been compared with regard to gains in pronunciation accuracy or how any improvement is retained in the long term. Furthermore, whilst a number of studies suggest that motivation, the concern for L2 pronunciation accuracy aspect in particular, along with perception and/or pronunciation training may contribute to more nativelike pronunciation in late L2 learners, this has not been examined with specific reference to these training techniques. The present work compares these techniques for training native English speaking learners of French on difficult L2 French contrasts (/u/ vs. /y/ and /ɑ̃/ vs. /ɔ̃/), and assesses participant concern for pronunciation accuracy in order to ascertain an optimal training technique to improve the perception and pronunciation of less able learners.

Experiment 1 of this thesis compares HVPT and PF using multiple and single speakers and found that the single speaker HVPT technique was significantly less effective than the others immediately after training. Testing again after at least one month suggested that training was best retained either through using PF with one speaker or HVPT with multiple speakers, that is, the techniques in their classic forms. Experiment 2 examines the benefits of these perceptual training techniques vs. pronunciation training vs. perception AND pronunciation training for both perceptual and pronunciation improvement. Undergoing multiple speaker HVPT + pronunciation training (over the same timescale as training in a single modality) appeared to be most beneficial for perception and pronunciation. Experiment 3 examines the relationship between average pronunciation improvement and participant concern for pronunciation accuracy as measured Elliott's (1995) Pronunciation Attitude Inventory and found that a high concern for pronunciation accuracy is only related to greater improvements when specific, perhaps more monotonous, training techniques (using only one modality and speaker) are used.

Overall, the present results provided no evidence of transfer of perceptual training benefits to pronunciation, and only slight evidence of transfer of pronunciation training benefits to perception, although there was a clear link between participant perception and pronunciation ability before training commenced. This is likely to be at least partly why some training in both modalities emerged as most successful in terms of improvements in both domains. It was therefore suggested that it may be prudent to consider the relationship between perceptual and production learning as distinct from any links between perception and production in general.

Acknowledgements

This work would not have been completed without the input of a number of people.

For academic support my first thanks goes to my principal supervisor Dr. Mits Ota for his invaluable suggestions, prompt and patient responses to problems or panicked questions, apparently limitless knowledge, and friendly understanding of a fellow new(ish) parent. My thanks also go to my second supervisor Prof. Alice Turk for her helpful comments and encouragement. Thanks must also be made to Prof. Bob Ladd for his inspiring introduction to phonetics and phonology, now some time ago.

For genial and invaluable technical and administrative support my gratitude goes to Mike Bennett, Ziggy Campbell, Eddie Dubourg, Katie Keltie and Toni Noble. Financially, I acknowledge the ESRC without whose backing I could never started this PhD process; and Margaret Begbie without whose backing I could never have finished it.

My gratitude also goes to the French department staff at the Universities of Edinburgh, Glasgow and Strathclyde and the Erasmus office staff at the University of Edinburgh in permitting and aiding the recruitment of their students; and to those students who agreed to give up their time to participate.

Finally I'd like to thank all of my family and friends for their encouragement along the way. In particular, I could not have seen this process through without the support, patience and understanding of my husband Iain and the joy brought to me by my daughter Laurie.

Contents

List of Tables

List of Figures

1 Introduction

1.1 Background

It is well documented that whilst first language acquisition is universally successful in normal children, this is not true in the case of adult second language acquisition (e.g. Johnson & Newport, 1989). Production errors by adult second language (L2) learners can occur at any linguistic level, for example, phonology, morphology, syntax or semantics (Major, 2001). Such errors are often cited in support for the critical period hypothesis (CPH) (Lenneberg, 1967; Flege, 1987). This states that nativelike L2 attainment is only possible if L2 acquisition begins before a certain age (Hyltenstam & Abrahamsson, 2003). Lenneberg (1967), who provided the original conceptualisation for the CPH, suggested puberty as the cutoff point (Hyltenstam & Abrahamsson, 2003). Whilst there is much debate surrounding the CPH, Hyltenstam and Abrahamsson (2003) note that in terms of L2 acquisition: "…few researchers today would deny long-term advantages for the child starters…" (p. 539).

A number of studies suggest that nativelike pronunciation in the L2 is one of the aspects of language which is particularly difficult to achieve for adult learners, and this idea was introduced by, for example, Scovel (1969, see also e.g., 1988, 2000; Long, 1990). The Polish author Josef Conrad is often given as evidence that this is the case, as he wrote English fluently but spoke English with a strong Polish accent (see, e.g., Abu-Rabia & Kehat, 2004). Furthermore, Newport (2002) also states that age of exposure does not affect all aspects of language learning in the same way, with the acquisition of vocabulary and semantic processing occurring relatively normally in late learners. The author states that instead, age-related effects appear to focus on phonology, morphology and syntax (Newport, 2002). However, even within these latter formal properties of language, Newport (2002) notes that various aspects may be more or less dependent on age of language exposure.

Specific evidence of the sensitivity of L2 pronunciation to age-related effects comes from, for example, Flege, Yeni-Komshian and Liu (1999) who examined the effect of age of arrival (AOA) in the United States on the L2 English foreign accent and morphosyntactic knowledge of native speakers of Korean. The authors found that as

AOA increased, the subjects' degree of foreign accent and errors in grammaticality judgement tasks also increased. However, the effect of AOA on the grammaticality judgement errors became non-significant when confounding variables (such as total years of education in the United States and amount of English use) were controlled for, which was not the case with the foreign accent ratings. In addition, more individual participants, as well as more participant groups defined by AOA, differed from the native English speaking controls in foreign accent ratings than in the grammaticality judgement scores (Flege et al, 1999). The authors concluded that these findings can be taken to support the view that AOA has a greater effect on degree of foreign accent and acquisition of L2 phonology than it does on acquisition of L2 morphosyntax (Flege et al, 1999).

A large body of work also attests to the problems later learners have in perceiving L2 contrasts (see e.g., Flege, Bohn & Jang, 1997; Bohn & Flege, 1997, Flege & MacKay, 2004, Flege, MacKay & Meador (1999)). When examining age-related effects on L2 pronunciation, consideration of the sensitivity of L2 perception to age related effects is equally important as there is evidence to suggest that differences between native and non-native perception may limit how accurately the L2 can be produced. For example, Rochet (1995) investigated the phenomenon that when learners attempt to speak a second language which makes use of the three high vowels /i/, /y/ and /u/, those whose language contains only two high vowels /i/ and /u/ find it difficult to produce three distinct vowels. The author found that speakers of some languages (e.g. English) will produce an /u/-like vowel for /y/ whereas speakers of other languages (e.g. Portuguese) will produce an /i/-like vowel for /y/. Rochet (1995) notes that this problem is unlikely to be at an articulatory level (i.e. production led) as his speakers of both English and Portuguese produce both /i/ and /u/. Similarly, through an imitation task, the author found that both the English and Portuguese speakers could reproduce French /y/ in approximately 50% of cases. This demonstrated that speakers of both languages can produce the high front rounded vowel /y/ and that the faulty reproductions (the /u/-like vowel by English speakers and the /i/-like vowel for Portuguese speakers) of this vowel could not be solely at an articulatory level also.

Evidence for a perceptual basis to this production problem (and therefore for the importance of considering L2 perception when examining accented L2 production) comes from the finding that in a perceptual task the crossover from /i/ to /u/ was located significantly higher on an F2 scale for these English speakers than these Portuguese speakers. Rochet (1995) noted that the average F2 value of the French /y/ the participants were given to imitate fell within the bounds of the /i/ category for the Portuguese speakers and the /u/ category for English speakers as demonstrated by the perceptual task. The author concluded that "[t]he parallelism between the results of the imitation task and those of the perceptual task appear to support the hypothesis that accented pronunciations of L2 sounds may be perceptually motivated" (p.385).

Despite such evidence that L2 perception and pronunciation are particularly sensitive to age related effects, a number of studies have found evidence of nativelike L2 pronunciation in late learners, across a number of L1-L2 pairings. These findings are discussed below.

1.2 Nativelike L2 Pronunciation by Late Learners

Whilst nativelike pronunciation by L2 learners is well demonstrated (see below), a number of researchers have noted that nativelike L2 pronunciation by late learners could, at least in part, be dependent upon the L1-L2 pairing (see, e.g., Birdsong & Molis, 2001; Piske, MacKay & Flege, 2001), with Piske et al (2001) noting that "Smaller typological L1-L2 differences may also account for ...greater [pronunciation] success." (p. 202). The basis for this assertion is mainly the L2 English pronunciation success of L1 Dutch or German speakers (and vice versa) documented in a series of well known studies by Bongaerts and his colleagues (e.g. Bongaerts, Planken & Schils, 1995; Bongaerts, van Summeren, Planken & Schils, 1997; Bongaerts, Mennen & van der Slik, 2000).

Bongaerts et al (1995 and 1997) compared native English speaker ratings of the productions of a group of native English speakers, a group of Dutch highly successful learners of English and a group of Dutch students of English with varying degrees of proficiency. Bongaerts et al (1995) found that the highly successful learners were indistinguishable from the native speakers. The authors concluded that as the successful learners had all began learning English after the age of 12, late learners can acquire nativelike L2 pronunciation (Bongaerts et al, 1995). Methodological changes by Bongaerts et al (1997) (rectifying a mismatch between variety of English spoken by the successful Dutch group, the native English group and the native English judges) resulted in the similar finding that 5 out of the 11 highly successful learners used were rated as having a nativelike accent for at least some of their productions. Of these 5 participants, 3 were rated as nativelike across all productions (6 sentences). Similarly, looking at the pronunciation of L2 Dutch, Bongaerts et al (2000) examined the accent ratings of advanced learners of Dutch living in the Netherlands who were speakers of 11 different languages. The authors found that one participant with L1 English and one participant with L1 German were rated as nativelike across most of their productions (7 out of 10 sentences; see also Moyer (1999) for nativelike pronunciation of L2 German by L1 English speakers). Speakers of the other languages such as Armenian, Berber, Czech, Greek and Swedish did not perform to this standard. Summing up all of their previous work the authors conclude that it is not impossible for late learners to achieve a native accent in a second language. At the same time, Bongaerts et al (2000), like Birdsong and Molis (2001) and Piske et al (2001) also noted that typological distance between the L1 and the L2 could be related to ultimate nativelike pronunciation.

However, although, as Piske et al (2001) note, it is not possible to gauge the overall typological distance between various language pairings, nativelike L2 pronunciation (and perception) in late learners has also been attested in a number of other, perhaps less related, L1-L2 pairings. For example, an often cited study regarding nativelike L2 pronunciation and perception of Arabic by native English speakers is that of Ioup, Boustagui, El Tigi and Moselle (1994) who carried out a case study on two highly proficient late learners of Arabic. In a test of spontaneous speech, the participants were rated as native speakers by 8 out of 13 native speaker judges, with 6 judges believing that they were both native, and 2 rating one but not the other as native (Ioup et al, 1994). In terms of their perception, in an accent discrimination task, both participants were 100% accurate in discriminating Egyptian and non-Egyptian accents in Arabic, performing better than 2 of the 11 native speaker judges. Furthermore, one participant was also able to discriminate to a certain extent between Egyptian Arabic accents that were and were not from Cairo (Ioup et al, 2004). Further evidence for nativelike pronunciation from those with perhaps less related L1-L2 pairings comes from Abu-Rabia and Kehat (2004). The authors examined native speaker ratings of 10 highly proficient speakers of L2 Hebrew with varying L1s (Russian, English, Bulgarian, Romanian, Polish and Afrikaans). The authors found that three L2 Hebrew speakers (two with L1 English and one with L1 Romanian) performed within the native speaker range on at least one of their three tests with one speaker (L1 Romanian) performing at a native level in all tests. Furthermore, two L1 speakers of Russian were also found to sound native by at least two of the five native judges in one of the tests.

Returning to a perhaps more closely related L1-L2 pair, of particular relevance to the present work is whether nativelike production of L2 French by L1 speakers of English has been attested, as it is intended to use this pairing. Birdsong (2003, 2007), examined the accentedness of 22 highly successful native English (American) learners of French who had been living in Paris for a minimum of 5 years. The author examined the degree of accent at two levels, the segment, using acoustic analysis, and at a global level, using native speaker judgements. It was found that two of the 22 subjects could pass for native speakers on all measures (Birdsong 2003, 2007; see also Palmen, Bongaerts & Schils, 1997, and Bongaerts, 1999, who found nativelike pronunciation by extremely proficient speakers L2 French who had Dutch as their L1).

As can be seen from the findings of the above studies, whilst the proportion of participants who are indistinguishable from native speakers can be low, none of the studies cited above found zero participants able to perform at a native level, and this is the case with a variety of language pairings. Nativelike L2 pronunciation by some late learners therefore certainly seems possible. Attempts to explain what makes these exceptional participants such successful L2 learners are detailed below.

1.3 Attempts to Explain Nativelike Pronunciation

From the knowledge of their participant learning histories, Bongaerts et al (1997) suggested that certain learning situations and learner characteristics could combine in a favourable way to override age-related effects. Many of the other authors of the studies cited above also agree that this combination makes their exceptional participants so successful. In particular, it appears that the learning situation common to the most successful participants across the studies is having participated in perceptual and/or pronunciation training of some nature, and this combines successfully with the learner characteristic common to these participants which is a high degree of motivation, as detailed below.

1.3.1 Training

The importance of training for accurate and even nativelike pronunciation has been noted by many of the authors who found such proficiency in their late L2 learning participants. Bongaerts et al (1997) noted that input enhancement in the form of perceptual training, along with production training, could be important in achieving nativelike attainment, as the participants used in their study had experienced such interventions. Although the participants in the Bongaerts et al (2000) study had in contrast had received very little formal pronunciation instruction, they had received much more intensive exposure to natural target language input because these participants were acquiring the native language of the country they had moved to (Bongaerts et al, 2000). Overall, because the participants in their papers prior to Bongaerts et al (2000) had performed to a slightly higher level, the authors concluded that extensive exposure to target language input, motivation (see 1.3.2. below) and intensive L2 perception and production training may all be important for ultimate attainment.

Moyer (1999) found that the type of phonological feedback the participants had received as learners was one of two variables (the other being motivation, see 1.3.2. below) which accounted for the most variance in accent ratings overall. Those participants who had received suprasegmental as well as segmental feedback tended to receive lower (and therefore more nativelike) foreign accent ratings. Similarly, Abu-Rabia and Kehat (2004), highlight the importance of training for the pronunciation accuracy in their results concluding that: "…formal instruction, attempting to increase the learner's awareness and motivation on the one hand, and providing the appropriate exposure and extensive practice on the other, may enhance a native-like accent." (p. 97).

Birdsong (2003, 2007) further examined the language backgrounds of the two participants who performed at a nativelike level in his study in order to identify common characteristics for pronunciation success. To this end, the author asked whether the participants had received any pronunciation training. One participant had indeed taken such a course whilst at university 15 years previously, whilst the other, being an actor, asked her friends to correct the pronunciation of her lines before performing in a play. However, Birdsong (2003/2007) does point out that three of his less successful participants had also taken formal courses in phonetics. The author concludes that whilst training is important in ultimate attainment in pronunciation, it does not guarantee nativelike pronunciation (Birdsong 2003/2007).

It can therefore be concluded that those authors cited above agree that having received some training is an important characteristic of those participants who have accurate or nativelike L2 pronunciation. The second important characteristic, motivation, is dealt with below.

1.3.2 Motivation

The importance of motivation for pronunciation accuracy has also been noted by many of the authors who identified nativelike speakers in their studies. Bongaerts et al (1997) note that a very high level of motivation could be one learner characteristic which could help override critical period effects as their participants stated that it was very important for them to be able to speak their L2 without a foreign accent. Bongaerts et al (2000) also note that their nativelike participants were highly motivated to achieve excellent pronunciation, in particular because they were now living in the country where their L2 was spoken.

Moyer (1999) found that professional motivation was the other of the two aspects (along with training) which accounted for the most variance in her participant accent ratings overall. Those participants who had high levels of professional motivation tended to receive lower (and therefore more nativelike) foreign accent ratings. The author found only one participant in her study who performed to a nativelike level and this participant did report some professional motivation. In addition, the major difference the author found was that this participant had a strong wish to acculturate and sound German, which was true of very few other participants. However, as the author notes: "…such integrative motivation is difficult to quantify, much less to influence, and its relationship to ultimate attainment has yet to be determined." (p. 98). Abu-Rabia and Kehat (2004) similarly highlight the importance of motivation in their pronunciation accuracy results and conclude from their findings that: "Outside the classroom…it is largely the individual's way of life (how important it is to his/her prestige or profession, his/her awareness and motivation, the amount of practice he/she gets) that may influence the level of L2 proficiency." (p. 97).

Birdsong, (2003, 2007), in his attempt to identify common characteristics in his two participants with nativelike pronunciation, also administered a questionnaire with some items measuring motivation. Firstly, on a scale of 1 (not at all important/motivated) to 10 (very important/motivated), the two participants stated that they had a very high motivation to learn French, both in a formal context at school and university, as well as in an immersion setting in France. In addition both participants stated that they found authenticity and accuracy in pronunciation to be very important and one stated that it was very important to them to be taken as a native speaker by native listeners (the other nativelike participant did not answer this question). However, Birdsong (2003/2007) does point out that many other participants (the author did not state how many) reported high levels of motivation. The author concludes, as with training, that whilst motivation is important in ultimate attainment in pronunciation, it does not guarantee nativelike pronunciation (Birdsong 2003/2007).

As with training, it can therefore be seen that the authors cited above are also in agreement that being highly motivated is another important characteristic of those participants who have accurate or nativelike L2 pronunciation. How this evidence will be used is described below.

1.4 Training and Concern for Pronunciation Accuracy

Evidence from the studies cited above suggests that perception and/or production training of some nature along with an individual's motivation (a high concern for pronunciation accuracy in particular, the relationship between the two is examined further in Chapter 3) appear to be key factors in obtaining nativelike pronunciation in late learners. It is also acknowledged, however, that these factors will not in themselves guarantee nativelike pronunciation.

The present work will therefore ask whether the general consensus that training and motivation are important (or even necessary) in obtaining a nativelike accent can be used to help those who are currently less proficient in L2 pronunciation. In other words, attempts will be made to ascertain whether L2 pronunciation and perception can be improved towards a native level through perception and pronunciation training, which training techniques are most successful (a comparison which has rarely been made, see Chapter 2) and whether differing training techniques are more or less successful with varying levels of concern for pronunciation accuracy (again, a topic which has rarely been examined, if at all).

It should be noted that using the evidence that motivation and training are important in obtaining nativelike pronunciation is not intended to imply that obtaining nativelike pronunciation is or should be the aim of the pronunciation and perception training in this study. Indeed, as can be seen in Section 1.1 above, nativelike pronunciation by L2 learners is possible, but by no means the norm, and such an aim is therefore likely to be unrealistic (Derwing & Munro, 2005). Derwing and Munro (e.g. Derwing & Munro, 2005, 2009; Derwing, Munro & Wiebe, 1998) suggest a distinction between accentedness, intelligibility and comprehensibility. Accentedness concerns how an individual's pattern of speech sounds is different to that found in the local community, comprehensibility concerns how easy or difficult a listener finds it to understand speech, and intelligibility concerns how much of an utterance a listener actually understands. The authors note that it is possible for heavily accented L2 speech to be perfectly intelligible and easily comprehensible. However, unintelligible and incomprehensible L2 speech is always heavily accented. Derwing and Munro (2005, 2009) therefore suggest that increasing intelligibility and/or comprehensibility is a more realistic and more important aim for pronunciation instruction. The conclusions from the research described in Section 1.3 above will therefore be used to attempt to assist L1 English speaking L2 French learners to move towards a native standard through improving their perception of L2 French vowel contrasts which do not exist in English (or improving how intelligible French vowels are for these L2 listeners) and increasing the intelligibility of their productions of French words containing these vowels.

In the following chapter (Chapter 2) I provide a review of the language learning training literature, before moving on to examine motivation and its link to concern for pronunciation accuracy in Chapter 3. This examination of the literature provides the motivation and grounding for the research questions at the end of Chapter 3.

2 Training

2.1 Early Training Studies

The earliest language training studies did not focus on non-native language contrasts but instead tended to use and train synthetic Voice Onset Time (VOT) continua in a non language specific manner to investigate the limits of participant psychophysical sensitivities (Bradlow, 2008, see, e.g., Carney, Widin & Viemeister, 1977). One of the first studies to take techniques from studies of this nature and apply them to L2 learning was the seminal paper by Strange and Dittman (1984). The authors noted that their paper merged research on L2 perception with the psychophysical training studies. Strange and Dittman (1984) based their training technique on Carney et al's (1977) VOT study and attempted to eight train native speakers of Japanese on the English /r/-/l/ contrast. The experimental procedure took the form of a pre-test, training then a post-test, a model to be followed by most subsequent language training studies.

Strange and Dittman (1984) found that participant ability to discriminate synthetic *rock-lock* stimuli improved during training, and this improvement was further demonstrated by improvement from pre-test to post-test testing discrimination of the synthetic *rock-lock* stimuli. In addition the authors found some evidence of transfer to a new *rake-lake* continuum. However, the participants did not perform to a native standard post-training, and performance on the untrained continuum was significantly less accurate. Furthermore, there was no improvement from pre-test to post-test in identification of naturally produced minimal pairs. With these mixed results, the authors therefore concluded that training of at least some L2 contrasts is likely to require much time and effort but that an improvement of the techniques used in their study may be helpful (Strange & Dittman, 1984).

One of the few early training studies to use natural training stimuli and again look at L2 learning is that of Tees and Werker (1984). The authors also made use of a delayed re-test after testing in order to examine how well any training effects were retained, another procedure to become common in training studies. Native English speaking Americans were trained on either a Hindi place of articulation contrast, the unvoiced unaspirated retroflex /ʈa/ versus the dental stop /ta/ (15 participants) or a Hindi VOT contrast, the unvoiced aspirated dental stop/t^ha/ versus the breathy voiced dental stop/d^ha/ (15 participants). Tees and Werker (1984) found that all participants reached their pre-determined (high) criterion level of discrimination ability during pre-testing and training of the VOT contrast, and 14/15 of the participants retained this ability in the delayed re-test 30-40 days after training. In contrast only 7/15 participants reached the criterion level for the place of articulation contrast by the end of training and only 3/15 retained this ability 30-40 days after training. Further analysis revealed that training had a significant impact on the voicing contrast but not on the place of articulation contrast. The authors noted that further work was necessary to develop a most effective procedure for producing changes in discriminability of particularly difficult contrasts (Tees & Werker, 1984).

2.2 Introduction to Current Training Studies

Many subsequent studies have sought to build upon the early work using synthetic and natural training stimuli by Strange and Dittman (1984) and Tees and Werker (1984) by examining the effectiveness of various training techniques on learning to perceive and produce L2 sounds. These studies have varied along several dimensions. The most commonly used overarching perceptual training paradigms used today are the perceptual fading (PF) technique (e.g. Jamieson & Morosan, 1986) and the High Variability Phonetic Training (HVPT) technique (e.g. Logan, Lively & Pisoni, 1991). The authors of both these papers noted in particular that the techniques used in Strange and Dittman (1984) described above did not result in transfer of training to natural speech, and therefore sought to improve training techniques as indeed was suggested by Strange and Dittman (1984) in their paper. Training can focus on perception or pronunciation or both, the use of segmental or suprasegmental contrasts, can make use of audio, visual or audiovisual stimuli and feedback and the stimuli themselves can be natural or synthetic. Training is generally deemed to be successful if it at least results in significant improvement in identification and/or discrimination and/or pronunciation of the contrast with familiar stimuli, but it is also preferable that a) the training transfers to novel (new talker and/or new contrast position) stimuli and b) the effects of the training are retained in the long term (see, e.g. Bradlow, 2008).

The perceptual fading (PF) technique trains unfamiliar contrasts by beginning with stimulus contrasts (simple isolated words) which have the normal perceptual differences between them so exaggerated that the participant can consistently identify which of the stimuli is being presented. The differences are then gradually reduced at such a pace that identification errors remain low. Stimuli, in particular those used for the most and least exaggerated contrasts, tend to be synthesised. The least exaggerated contrasts are usually less salient than those found in everyday speech (e.g. McClelland, 2001; Iverson, Hazan & Bannister, 2005). According to Jamieson and Morosan (1986): "Using this technique, a high level of identification and discrimination performance can be attained in a short interval of time, without frustrating the subject." (p. 208). In contrast, the HVPT technique uses natural speech tokens (again, simple isolated words) produced by a number of native speakers, with the contrasts in a number of different phonetic positions in the words used. According to Logan et al (1991) presenting stimuli containing the contrast to be trained from various phonetic contexts exposes listeners to the full range of acoustic-phonetic cues that characterise the contrast to be trained across different environments. (p. 876). The authors further note that using different speakers results in additional stimulus variability due to the fact that different talkers produce varying acoustic output, and that this allows participants to overcome such variability.

Although most studies have focused on perceptual training to improve perceptual accuracy (see below), some studies have examined the effect of pronunciation training on pronunciation accuracy (e.g. Derwing, Munro & Wiebe, 1998), many have looked at the effects of perceptual training on pronunciation (e.g. Rochet, 1995; Bradlow, Pisoni, Yamada & Tohkura, 1997), and some have examined the effects of pronunciation training on perception (e.g. Leather, 1997; Gómez Lacabex & García Lecumberri, 2010). The argument for training perception in order to improve pronunciation has been made for many years. For example in an early study by Pimsleur (1963), the author noted that language laboratories for teaching pronunciation were of little use if the students could not hear how accurate their productions were in comparison to the native speaker they were listening to and trying to imitate. The author found evidence to suggest that those participants who had received discriminatory training in French sounds before using the language lab to practice listening to and repeating native speaker productions then performed better in a pronunciation test. In his more recent work, Rochet (1995) sums up the importance of perceptual training for pronunciation, stating:

"There are good practical and theoretical reasons for wanting to teach L2 phonetic contrasts by means of perceptual training: 1) learners must be able to identify L2 phones in order to understand the target language; 2) most agree that a learner cannot master the production of L2 contrasts without being able to label the sounds in question correctly; and 3) if it works, perceptual training is highly desirable, because it is easier to administer than production training." (pp 395-396)

Therefore, whilst some training studies may not examine the effects of perceptual training on pronunciation, the methods used and conclusions drawn remain important for pronunciation accuracy given that it has been demonstrated by a number of studies that perceptual training will aid pronunciation (e.g. Bradlow, Pisoni, Yamada & Tohkura, 1997, Lambacher, Martens, Kakehi, Marasinghe and Moholt, 2005, Rochet, 1995 - see also 1.1 for further detail on how the author demonstrates how differences between native and non-native perception may limit how accurately the L2 can be produced). For the same reasons, it is intended to draw conclusions in the present work about optimal training techniques for perception as well as pronunciation. The contrasting features of training techniques and training studies and how well they meet the previously stated criteria for success are now examined in more detail below. As suprasegmental training (e.g. Wayland & Li, 2008) and multimodal training (e.g. Hardison, 2003) are beyond the scope of the current work they will not be dealt with in the review below, although the benefits of using audiovisual as opposed to audio training will be discussed in Chapter 8.

2.3 The Perceptual Fading (PF) Technique

2.3.1 Improvement from Pre-Test to Post-Test

Early evidence that use of the Perceptual Fading (PF) technique can result in perceptual improvement from pre-test to post-test comes from Jamieson and Morosan (1986). Jamieson and Morosan (1986) were among the first to use the term 'perceptual fading' and to use the technique to train non-native language contrasts. The authors noted that the technique was introduced by Terrace (1963) who trained pigeon sensitivity to colour, but did not use the term 'perceptual fading' to describe his approach. Jamieson and Morosan (1986) investigated the acquisition of the /ð/-/θ/ contrast by ten

Canadian francophone late learners of English and noted that these sounds are particularly difficult for adult learners, often being confused with /d/ and /t/ respectively. The authors used a mixture of naturally produced CV syllable stimuli by one talker and eight synthesised stimuli on a continuum of 1-8 from extremely voiceless to extremely voiced, such that tokens 1 and 8 were exaggerated exemplars of /θ/ and /ð/ respectively. The experimental procedure took the form of a pre-test, training, then post-test.

Looking firstly at identification of the synthesised tokens, the authors found that the post-test scores were significantly higher than the pre-test scores, with significant improvement in performance with all stimuli on the continuum, which was not the case for a control group who received no training. Moving on to examine identification of the natural tokens, it was again found that there was a general improvement in identification accuracy from pre-test to post-test. Finally, looking at the discrimination of the synthesised tokens, it was found that the participants' sensitivity to cross-category differences significantly improved after training. The authors concluded that this training technique is effective for training contrasts and that training with synthetic stimuli transfers onto natural tokens (Jamieson & Morosan, 1986).

Further evidence that use of the PF technique can result in perceptual improvement from pre-test to post-test comes from McClelland (2001). The author arrived at a perceptual fading training paradigm from examination of Hebbian theories of learning and proposed that failures to acquire nonnative speech contrasts such as the /r/-/l/ distinction can be explained by such accounts. McClelland (2001) notes that Hebb's rule of learning suggests that given an input, synaptic modification mechanisms will establish whatever response pattern this input elicited, and subsequent similar inputs will result in the same pattern. After this, a given inappropriate input will result in synaptic adjustment such that both subsequent appropriate and subsequent inappropriate inputs will result in the same activation. McClelland (2001) therefore suggests that failure to learn the /r/-/l/ contrast results from "…undesirable strengthening of inappropriate preexisting activations." (p.102). The author notes that Japanese has a single alveolar liquid that approximately spans both /r/ and /l/ in English. Therefore, presentation of a /r/ or /l/ would result in a pattern of neural activation corresponding to the Japanese alveolar liquid, such that input from native English speakers would reinforce the perception of this one Japanese sound rather than the two English sounds.

In order to solve this problem, McClelland (2001) suggests that it is necessary to find a way of having the English /r/ and /l/ inputs activate different representations. The author proposes a solution which essentially makes use of the perceptual fading technique, stating that use of inputs that exaggerate the difference between the two sounds may solve this problem, and that if sufficiently exaggerated, the stimuli will activate separate representations. Exaggerated stimuli can then be used to reinforce the two representations, before the difference is gradually reduced so that less exaggerated (and therefore more similar to real-life) differences will continue to be assigned to two perceptual categories. A key difference in this approach, however, is that learning is expected to occur without feedback. McClelland (2001) found that all of the participants made significant gains in their identification and discrimination of /r/ and /l/ stimuli. However, training on one of the continua did not transfer to the other (*rock-lock* and *road-load*), meaning that learning in this instance was very specific (McClelland, 2001).

Having achieved some success with this technique in improving perception from pre-test to post test, the authors moved on to ascertain whether the improvement could transfer to new words and/or word positions as described below.

2.3.2 Generalisation to New Words/Word Positions

Morosan and Jamieson (1989) used the same technique and a similar procedure as Jamieson and Morosan (1986) to examine whether or not the training would generalise to novel stimuli. As before, the authors found that identification of CV stimuli significantly improved from pre-test to post-test. In addition the authors found that the training with synthetic stimuli also transferred to novel natural CV stimuli produced by two male and two female speakers. However, the study demonstrated some limitations to the training in that identification of /ð/ and /Θ/ did not improve when they were presented in word medial and word final positions. In addition, the authors attempted to train participants on the /ð/-/d/ contrast and found that training did not improve identification of synthetic or natural exemplars (Morosan & Jamieson, 1989).

Jamieson and Moore (1991) examined how the perceptual fading technique used in previous studies to train the /ð/-/θ/ contrast could be modified in order to increase its generalisation. In order to do this they used a VCV rather than CV synthetic continuum in the training procedure. The authors hypothesised that training with synthetic VCV tokens may generalise to VC and CV natural speech tokens as well as VCV tokens due to the VCV phonetic environment providing the acoustic information for all three situations. In the tests of generalisation, the authors used natural VC, CV and VCV /ð/-/θ/ nonsense stimuli, spoken by various talkers and various vowel environments, and additionally tested whether training would generalise to /d/ versus /t/. The authors found that overall identification significantly improved following training both for the synthetic VCV tokens and natural VCV tokens produced by both male and female talkers. The greatest improvement was in the VCV stimuli used in training, next for those where the vowel was altered, and least improvement was found for the VC, CV and /d/-/t/ conditions. The authors also found that training generalised to new voices, but suggested that learning may be limited to the phonetic environment in which the training sounds occurred (Jamieson & Moore, 1991).

McCandliss, Fiez, Protopatas, Conway and McClelland (2002), provided further detail regarding the perceptual fading/Hebbian technique used in McClelland (2001) described in 2.2.1 above, and further investigated the role of feedback in the results along with transfer of training to novel stimuli. As noted by McClelland (2001), the training with no feedback resulted in significant improvements from pre-test to post-test in identification of members of the *rock-lock* continuum used in training. However, with the addition of feedback in training as to whether participant responses were correct or incorrect, the authors found that PF training resulted in substantial gains in identification ability and that this PF training with feedback resulted in significantly greater gains than that without it. In addition PF training with feedback transferred from the trained *rock-lock* to the untrained *road-load* continuum (McCandliss et al, 2002).

2.3.3 Transfer of Perceptual Training to Pronunciation Accuracy

Having demonstrated that PF training can result in pre-post test improvements and that it can generalise to new words or word positions, the next criterion is to ascertain whether or not the PF training perceptual improvement can transfer to pronunciation accuracy. Using PF techniques, Rochet and Chen (1992) sought to establish how valuable auditory training is to the teaching and learning of L2 pronunciation. Until this study, little work had been carried out on whether perceptual fading training carried over to pronunciation. The authors sought to train native speakers of Mandarin Chinese living in Canada on the voiced/voiceless contrast on labial, dental or velar consonants.

The authors firstly found that perceptual training lead to modification of the perception of the /pu/-/bu/ continuum, with the mean VOT boundary moving significantly closer to the native French norm. Secondly, training only on the /pu/-/bu/ continuum also transferred in a similar way to /p/-/b/continua using other vowels, and voiceless-voiced continua using /u/ preceded by dental and velar stops. Furthermore, the training transferred to perception of voiceless natural stimuli in the word initial position, although there was no significant change in the identification of voiced initial or all intervocalic stops.

Moving on to examine pronunciation, a significantly smaller number of items were mispronounced in the post-test than in the pre-test, as judged by three native speakers of French, although the change for voiced stops narrowly failed to reach significance. There was no improvement in the pronunciation of intervocalic stops. However, the mean VOT durations of token initial voiced and voiceless stops significantly increased towards the native French norm. The authors concluded that there is evidence that perceptual fading training can lead to improvement in perception and pronunciation performance and that training may transfer to some degree to other environments, but at the same time acknowledged that participants learned in a very context dependent fashion (Rochet, 1995; Rochet & Chen, 1992).

2.4 High Variability Phonetic training (HVPT)

2.4.1 Improvement from Pre-Test to Post-Test and Generalisation Tests

Perhaps the best known work on the HVPT technique is a series of studies carried out by Pisoni and his colleagues on training native speakers of Japanese on the /r/-/l/ contrast (Logan, Lively & Pisoni, 1991; Lively, Logan & Pisoni, 1993; Lively, Pisoni, Yamada, Tohkura & Yamada, 1994; Bradlow, Pisoni, Yamada & Tohkura, 1997; Bradlow, Yamada, Pisoni & Tohkura, 1999). In their seminal paper, Logan et al (1991) pointed out that studies prior to the work carried out in their paper had reported little success in training the /r/-/l/ contrast to Japanese listeners using synthetic stimuli, particularly when testing transfer to natural stimuli (e.g. Strange & Dittman, 1984). The authors therefore wished to ascertain whether changes to previously used techniques could allow acquisition of this contrast in naturally occurring stimuli and in novel stimuli which had not been used in the training procedure. As the first to formally conceptualise the technique, Logan et al (1991) provided early evidence that use of the HVPT technique can result in perceptual improvement from pre-test to post-test. The experiment took the form of a pre-test, training, post-test and then two tests of generalisation, the first with new words and a familiar speaker and the second with new words and a new speaker. This pattern of testing generalisation continued to be followed by many of those examining the success of training with natural stimuli.

Overall, the authors found that there was a significant increase from the pre-test mean percentage of correct responses (78.1%) to the post-test mean (85.9%), with all participants demonstrating some improvement. More specifically, it was found that performance was better when the /r/-/l/ contrast was found in final and intervocalic positions rather than in initial singleton and initial cluster positions, across all participants. Training resulted in a marked improvement for the contrasts in the initial cluster and intervocalic positions but only a slight improvement for those contrasts in the initial and final positions. The tests of generalisation showed a slight effect, with 83.7% mean correct responses for novel words and 79.5% mean correct responses for a novel talker and novel words, suggesting that identification is easier with a familiar talker and therefore that learning, to an extent, may still be talker specific in this study (Logan et al, 1991).

2.4.2 Talker Variability vs. Variation in Phonetic Environment

Although not one of the criteria to demonstrate training success, Lively et al (1993) attempted to clarify the findings of Logan et al (1991) by investigating whether it was the variability from using multiple talkers or the variability from using the contrast to be trained in a number of phonetic environments which contributed most to the success of the HVPT technique. This has been included in the present discussion as it would be important in determining whether future work using this technique should concentrate upon multiple word positions or multiple speakers and multiple word positions. In a first experiment Japanese listeners were trained with /r/-/l/ minimal pairs with the contrasts in initial singleton, initial consonant cluster and intervocalic positions produced by five talkers. In a second experiment another group of listeners was trained with /r/-/l/ minimal pairs with the contrasts in the initial singleton, initial consonant cluster, intervocalic, final consonant cluster and final singleton positions produced by only one talker.

The results of the first study using five speakers in training replicated the finding of Logan et al (1991) in that mean identification accuracy significantly improved from 79.96% in the pre-test to 85.57% in the post-test, with accuracy being poorest with initial consonant clusters. Of particular interest is the improvement in identification of members of a /r/-/l/ contrast that was not trained (only three phonetic environments were trained, but four were tested). The authors note that this means that less variability in terms of phonetic position can still lead to improvements in identification. The tests of generalisation demonstrated a more convincing generalisation to that found previously, with accuracy comparable to that found in the final week of training (Lively et al, 1993). The authors state that this means that the categories have certainly been acquired.

In the second experiment Lively et al (1993) examined the effect of reduced variability by only using one talker in training. The authors hypothesised that although there may still be improvements from pre- to post-test, generalisation may not be so strong due to this reduction in stimulus variability. The same procedures and stimuli as previously were followed, although with one speaker only. Overall, the participants' ability to identify /r/ and /l/ significantly improved only for some phonetic environments (initial consonant clusters), although the other environments had slight improvements. Furthermore, the tests of generalisation revealed that the participants failed to generalise both to the new tokens produced by the new talker and also new tokens produced by the familiar talker, highlighting the importance of the variability arising from the use of multiple talkers in this training paradigm (Lively et al, 1993).

2.4.3 Transfer of Perceptual Training to Pronunciation Accuracy

Having established that the particular strength of their technique lay with talker variability, Bradlow et al (1997) examined whether HVPT training would meet the next criterion for success and transfer to pronunciation accuracy by using a similar procedure as in previous HVPT studies and adding a production task to the pre- and post-tests.

In terms of perception, the authors found, as before, significant improvements from pre- to post- test and also significant evidence of generalisation to new stimuli and a new talker. In terms of pronunciation, the distribution of the native speaker preference ratings of the participant productions was skewed in favour of the higher ratings in the post-test, which demonstrated a preference for the post-test readings over those of the pre-test. This gave an initial indication that the perceptual training had had a beneficial impact on pronunciation. Furthermore, the authors found that the native speakers were able to correctly identify significantly more participant post-test than pre-test productions. In addition this held true for both words that were used in the training phase and novel words that were produced in the post-training test but not used in the training phase (Bradlow et al, 1997).

Further evidence of the success of the HVPT perceptual training technique and its transfer to pronunciation accuracy comes from Lambacher, Martens, Kakehi, Marasinghe and Moholt (2005). The authors further examined the benefits of this training technique by extending the training from the binary /r/-/l/ contrast to training native speakers of Japanese on the five American English low vowels /æ/, /ɑ/, /ʌ/, /ɔ/ and /ɝ/. Overall, improvement in vowel identification significantly rose from 54% pre-test to 70% post-test, and the improvement in identification was also significant for each vowel individually. There were no such improvements for a control group who were not trained.

In order to ascertain whether or not there were any improvements in pronunciation, the participants were recorded producing words containing the vowels within a varied CVC context before and after the training period. These productions were then analysed in two ways, with native speakers identifying which of the five vowels they thought was being produced and by acoustic analysis comparing native and participant formant and temporal characteristics. It was firstly found that overall identifiability of the productions significantly improved from the pre-test (a mean 39%) to post-test (a mean 47%), which was not the case for an untrained control group. Although improvement did not occur on one of the vowels (/ɔ/), this was the most accurately produced prior to training, and the most poorly produced vowel (/ɝ/) pronunciation accuracy improved the most with training. Acoustic analyses demonstrated that the productions had moved towards native norms in terms of formant frequencies and vowel duration (Lambacher et al, 2005).

2.4.4 Retention of Improvement and Generalisation

Unlike many of the early PF studies (however, see, e.g., Wang & Munro, 2004, described in 2.5 below), Lively et al (1994) examined whether the perceptual benefits of HVPT training on the /r/-/l/ contrast would be retained. Overall the authors found that the participants' ability to identify /r/ and /l/ rose from an average of 65% to an average of 77% from pre-test to post-test. Again, improvement was dependent upon phonetic environment. The tests of generalisation showed a better performance with a familiar talker from training (82%) than an unfamiliar talker (77%), with these mean percentages showing some evidence of generalisation. After three months, from those who returned, there was no significant difference between the re-test scores and the post-test scores from three months previously, nor between the scores from the tests of generalisation carried out at these two times. After six months, accuracy was still on average 4.5% greater than on the pre-test, although this was not significant. However, six of the eight subjects who returned after six months were still significantly more accurate after six months than on the pre-test, and there were no significant differences in the scores between the original tests of generalisation and those carried out six months later. The authors concluded that their high variability training procedure (participants took part in 15 40 minute training sessions over three weeks) quickly modifies the perception of non-native listeners, generalises to novel speakers and phonetic locations and is retained for at least six months (Lively et al 1994).

Bradlow et al (1999) examined the retention of improvements in pronunciation as well as perception of the /r/-/l/ contrast. From a native English speaker preference rating task, the authors found that the performance of the Japanese participants did not significantly differ after three months from the test carried out immediately post-training, with the post-test and three-month productions being preferred over the pre-test productions 44.4% and 43.9% of the time respectively (the reverse was true 33.1% and 32.2% respectively, with the remainder showing no preference). In a native English speaker minimal pair identification task, again no significant difference was found between mean accuracy post-test (73.3%) and at three months (77.15%). For the transcription task, the authors noted that overall accuracy was much lower than for the other tasks due to this being a very stringent measure. Although a significant difference was found between mean accuracy post-test and at 3 months, this was actually in a positive direction, 41.85% and 47.01% respectively. The authors conclude that their HVPT technique improves not only perceptual skills, but also pronunciation skills without specific pronunciation training, and that this improvement is long-term (Bradlow et al, 1999).

Having examined how each of the training techniques meets the success criteria noted in Section 2.2, Section 2.5 below compares the two techniques.

2.5 Perceptual Fading vs. HVPT

Very few studies have attempted to directly compare the perceptual fading and HVPT techniques. Wang and Munro (2004) used a combination of both techniques to train the /i/-/I/, /u/-/ʊ/ and /ɛ/-/æ/ English contrasts to native speakers of Cantonese and Mandarin. The authors found that the participants improved significantly in perceptual performance, that this transferred to novel stimuli and that the improvement was also retained three months after the experiment. However the authors did not compare the relative efficacy of the two techniques.

One study which has sought to compare the techniques is that of Iverson et al (2005), which compared the effectiveness of these techniques (among others) in training the English /r/-/l/ contrast to adult native speakers of Japanese. As usual, the HVPT technique involved using natural words from multiple talkers. The identification training words were spoken by 10 native speakers of English and consisted of 100 /r/-/l/ initial position minimal pair words (e.g. *road* and *load*). The test words were recorded by two additional speakers and consisted of 40 words from the training set, 40 other /r/-/l/ initial minimal pairs, 40 medial position /r/-/l/ minimal pairs and 40 consonant cluster /r/-/l/ minimal pairs. In this case the perceptual fading technique also used these natural recordings which were altered by signal processing. On the first day the stimuli were fully enhanced such that the F3s were set to extreme values during the closure (which enhances the difference between /r/ and /l/ (Iverson et al, 2005)) and the duration of the closure was increased to 100ms in order that it was long enough to be audible. This enhancement was gradually reduced such that by the final day of training the difference between /r/ and /l/ was less than normal. For both training types the talker was changed daily (Iverson et al, 2005). Training and testing took the same identification format, with feedback being provided in training. All subjects were pre- and post- tested with trained talkers and words, new talkers and trained words and new words with /r/ and /l/ in initial, medial and consonant cluster positions (Iverson et al, 2005).

Analysis of the natural stimuli demonstrated that /r/-/l/ identification performance improved after both types of training by an average of 18%, and that accuracy was higher for the trained talkers and words. In addition, the training generalised to new words and talkers, but there was no significant difference in improvement between the two training methods. Analysis of the interim test data suggested that there were some differences in the rate of learning. The authors concluded that training with natural speech may be the best method, simply because it is less labour intensive but additionally note that there is no particular advantage to having fully natural variability (Iverson et al, 2005).

One potential problem with the conclusions of this paper is that it appears that the only difference between the perceptual fading training and the HVPT training is that the perceptual fading stimuli underwent signal processing in order to grade the enhancement or exaggeration of the stimuli. According to the procedure section of this paper "Except for the stimulus differences between conditions, the training procedures were identical. The training comprised 10 sessions…There was a different talker each day…" (p. 3271). It is therefore unclear as to whether the perceptual fading used here is true perceptual fading, as although the stimuli were manipulated, the perceptual fading training group also had the benefits of multiple talkers. In the perceptual fading studies described above, the training materials are based on productions from one talker, and the strength of the technique comes from exaggerating the differences from between a contrast. Whilst use of multiple talkers may be an improvement to the perceptual fading technique, this is arguably borrowed from the HVPT technique and therefore comparing the two in this manner may not allow for the conclusion than one is no better than the other. This may particularly be the case because, as noted above, Lively et al (1993) found that it was use of various talkers (rather than the contrast in variable positions during training) which resulted in the greatest success in their HVPT technique, especially in terms of generalisation.

McClelland (2001) also argues against the conclusion of the Iverson et al (2005) paper that the HVPT technique may be better to use simply because it is less labour intensive. This may be so for the experimenter, but not so for the trainees. Using the Pisoni studies as examples of the HVPT technique, McClelland (2001) notes that progress of around 20% is at the expense of lengthy training (Bradlow et al (1999) state 15-22.5 hours over 2-3 weeks) whereas the PF training in the McClelland papers took 1-3 hours over 3 days (see McCandliss et al, 2002). Arguably, then, the relative strengths of the perceptual fading and HVPT techniques remain unclear and they have rarely been compared. Further work is therefore necessary in this area.

2.6 The Present Work

As can be seen from the range of studies presented above, it is now widely accepted that perceptual training improves both L2 perception and L2 pronunciation to some degree, and there does appear to be some benefit of training with multiple talkers with the

HVPT technique in particular. What is lacking, however, is a volume of work comparing the various methods and approaches, with the few comparisons that have been carried out proving inconclusive. In particular, as has been noted above, the effectiveness of the HVPT vs. the perceptual fading techniques has not been well established. Furthermore, whether perceptual training alone actually improves pronunciation more than specific pronunciation training has rarely been addressed. Whilst Pimsleur (1963) found that discriminatory training was beneficial to subsequent pronunciation training, Catford and Pisoni (1970) found that those who received articulatory training perceived and produced exotic sounds more accurately than those who had only received perceptual-like training. Little work appears to have been carried out on this issue since then (but see Leather, 1997, and the combined results of Gómez Lacabex, García Lecumberri & Cooke (2008) and Gómez Lacabex & García Lecumberri (2010) which suggest that training in either mode is equally beneficial to either skill, see also Chapter 8). It is therefore the aim of this work to carry out these comparisons in order to contribute to and refine the knowledge regarding optimal second language pronunciation and perceptual training techniques.

The need for this work has been attested by a number of researchers, for example Jamieson (1995) who noted that further work was necessary to optimise training techniques, that few studies had compared alternative training techniques, and that it was unlikely that an optimal technique could be created from a single one of the techniques which had been examined at that time. Whilst Jamieson's (1995) views are now over 15 years old, they remain relevant. The Iverson et al (2005) study is one of the few which attempts to compare many of the techniques discussed by Jamieson (1995) and the authors note that there is still further scope for extension to their findings in that differences between techniques may emerge if pronunciation and long term retention of training benefits were measured. In addition, although Iverson et al (2005) have begun to address Jamieson's (1995) problems by comparing techniques, Bradlow et al's (1999) statement that "…it is still an open question whether the high-variability approach is more effective in promoting long term improvement in production than other 'low variability' approaches" (p. 983) continues to remain relevant, as the findings of one study alone cannot be taken to be definitive.

The present work will therefore firstly amend the comparison carried out by Iverson et al (2005) by comparing the PF and HVPT techniques using single and multiple speakers to ascertain whether adding multiple speakers to the PF technique is a beneficial borrowing from the HVPT technique. Secondly, this amendment has the additional benefit of carrying out a further comparison of the benefits of high vs. low variability training as suggested by Bradlow et al (1999). Thirdly, the present work will extend the work carried out by Iverson et al (2005) as the authors suggested, by including examination of long term retention and pronunciation. Finally, as can be seen from the above studies, most training work at the segmental level has been carried out on training the /r/-/l/ distinction in Japanese learners of English. The present work will deal with training French contrasts which are difficult for native speakers of English who are learning French, thereby contributing to the knowledge about these pairings.

It is anticipated that the relative success of the varying training techniques and approaches to be examined may be more or less dependent upon the varying levels of concern for pronunciation accuracy a participant has. The next chapter provides a detailed review of the work carried out on motivation, and its concern for pronunciation accuracy aspect, in second language learning.

3 Motivation and Pronunciation Accuracy

The best known theory regarding motivation and achievement in language learning is that of Gardner and his colleagues (e.g. Gardner, 1985; Tremblay & Gardner, 1995) who essentially founded the field of research in this area (Dörnyei, 2001). Whilst other theories of motivation in language learning have now been developed (see, e.g. Csizér & Dörnyei, 2005) and the work of Gardner and colleagues has been criticised for being too dependent on identification with the target language community and the Canadian language learning context (e.g. Dörnyei, 1994; Oxford & Shearin, 1994), it is the Gardnerian model which will be addressed here because a) many of the other theories have a basis in that of Gardner and his colleagues (e.g. Csizér & Dörnyei, 2005) and b) it is motivation as formulated by Gardner's theories that is of most relevance to the present work.

3.1 The Socio-Educational Model

3.1.1 Background

Initial studies on motivation carried out by Gardner and his colleagues (e.g. Gardner & Lambert, 1959) developed into the socio-educational model of language learning (Gardner, 1985) in which motivation plays a key role. Gardner's (1985) definition of motivation in this model has three components, the effort given to achieve a goal, desire to learn a language and satisfaction with the language learning task. These three components are assessed with corresponding scales (Motivational Intensity, Desire to Learn the Language, Attitudes Towards Learning the Language) in the Attitude/Motivation Test Battery (AMTB) (e.g. Gardner, 1985) designed by Gardner and his colleagues in an attempt to quantify motivation. According to Gardner (1985) all three components are necessary to describe motivation as, for example, there may be considerable effort made in order to please someone else, without any real desire to learn and/or satisfaction with the learning task, therefore meaning that no real motivation is present.

In this motivation component of the socio-educational model there are two classes of variables which are said to influence motivation. The first is called Integrativeness, which is a positive view of those who speak the language, and this is assessed by the

Attitudes Towards the Target Language Group, Interest in Foreign Languages and Integrative Orientation scales from the AMTB. The second class of variables seen to influence motivation is Attitudes Towards the Learning Situation which is measured by the Attitudes Towards the Language Course and Attitudes Towards the Language Teacher AMTB scales. It was initially hypothesised that these two classes separately caused motivation (Gardner, 1985), but subsequent analysis demonstrated that it was more powerful to link them together into one construct (e.g. Tremblay & Gardner, 1995). Therefore Integrativeness and Attitudes Towards the Learning Situation are now together referred to as Language Attitudes.

Of key importance in the socio-educational model is the relationship between orientation and motivation. Orientations are not part of motivation per se, but are motivational precedents and contribute to the cause of motivation in some way by explaining why the individual has a certain goal. For example the Integrative orientation contributes to Integrativeness/Language Attitudes, which in turn influence motivation. As Gardner (1985) notes: "The [motive/orientation] distinction can be clarified by considering the difference between an integrative orientation and an integrative motive. An integrative orientation refers to that class of reasons that suggest an individual is learning a second language in order to learn about, interact with, or become closer to, the second language community…The concept of the integrative motive includes not only the orientation but also the motivation (i.e. attitudes towards learning the language, plus desire, plus motivational intensity) and a number of other attitude variables…" (p. 54). The counterpart of the Integrative orientation is the Instrumental orientation which assesses the extent to which people learn a language for pragmatic reasons such as gaining employment or a higher salary (Gardner, Tremblay & Masgoret, 1997). It should be noted that Gardner (1985) did not consider these orientations to be mutually exclusive or fixed across an individual's lifespan.

In 1995 Tremblay and Gardner extended Gardner's (1985) view of L2 motivation to include aspects from mainstream psychological literature, rather than L2 approaches to motivation alone. The authors were responding to calls that research examining motivation in L2 acquisition would benefit from this expansion (e.g. Oxford & Shearin, 1994). In addition, it was felt that the motivational impact of the learning environment should be of more importance in the model (e.g. Dörnyei, 1994). The authors found that goal salience, valence and self-efficacy mediated between attitudes and motivational behaviour. Goal salience assesses the degree to which individuals had specific goals associated with language study and the strategies used to aid in achieving the goals. Self-efficacy assesses the level of anxiety when called upon to use the L2 and the level of belief on the part of a participant that they would have reached a certain standard by the end of the course, and valence assesses attitudes towards learning the language and the language teacher (Tremblay & Gardner, 1995). This extension helped reassure some critics about the flexibility of the Gardnerian model (e.g. Dörnyei, 2001), although it was also argued that it reduced the impact of some of the model's important social components (e.g. Smit & Dalton, 2000).

3.1.2 Language Achievement and the Socio-Educational Model

Over the years Gardner and Colleagues tested their model in relation to language achievement (e.g. Gardner, 1985, Gardner & MacIntyre, 1991, Tremblay & Gardner, 1995). For example, Tremblay and Gardner (1995), in extending the model as described above, administered motivational and attitudinal questionnaires to 75 secondary school pupils. The participants were also asked to write an essay, and their final grades in their French course were obtained at the end of the year. The authors hypothesised that motivational behaviour would have a direct influence on achievement and that language attitudes would indirectly influence motivational behaviour. The results indicated that this was the case. In addition, Gardner et al (1997) administered three questionnaires to 102 university students enrolled in an introductory French course. The first assessed attitudes, motivation, achievement, and self-ratings of French proficiency. The second assessed anxiety, learning strategies, aptitude and field dependence/independence. The third concerned the participants' language history. In addition, the participants' final grades from the course were obtained. The authors found that the final grade correlated slightly higher with the measures of motivation than with the other variables noted above. Further analysis with causal modelling indicated that language attitudes caused motivation and that motivation provided the greatest contribution to achievement, implying that language attitudes such as integrativeness are related to achievement but only indirectly, by acting through motivation (Gardner et al, 1997).

Having examined how motivation, as measured by the Socio-Educational model AMTB, relates to language achievement in general, how motivation in general applies to pronunciation success is described below. The links between the specific concern for pronunciation accuracy aspect of motivation and pronunciation success are then addressed.

3.2 Motivation (Concern for Pronunciation Accuracy) and Pronunciation Success

3.2.1 General Motivation and Pronunciation Success

Few studies have examined the link between motivation and pronunciation success in particular. Among the studies which have sought to do so with specific reference to the motivation described in the socio-educational model of language learning is Smit (2002) who examined the links between motivation and L2 pronunciation through studying the students taking a compulsory English pronunciation training/practical phonetics course at Vienna University. Smit (2002) examined the students' achievement on the pronunciation module and how this interacted with the motivational factors attested by the students through the administration of a language attitude test, an identity scale test and a general motivation test (see Smit & Dalton, 2000). The author used a questionnaire developed from her previous work (Smit & Dalton, 2000) in order to test for motivation, and added a number of questions regarding previous achievement. In addition, the participants' final grade in the pronunciation module was recorded. The questionnaire had three parts, two of which were completed at the beginning of term and two at the end such that one part was completed twice. The author acknowledges that a single grade for a pronunciation course is not sufficient to describe pronunciation accuracy, but notes that there is no standardised test for pronunciation. It was found that the only motivational construct to play a role in final grade was intrinsic motives (those motives driven by internal goals such as the satisfaction of learning something new or enjoying a challenge; seen by Smit, 2002, as akin to an integrative motive), and this was only of slight importance.

Further work which linked integrativeness or an integrative motive to pronunciation accuracy was carried out by Moyer (1999). The author recorded word list, sentence list, paragraph and free speech productions of 24 native English speakers of German, and asked recently arrived (in the US) native speakers of German to rate the productions for authenticity. The author found professional (arguably an instrumental-type) motivation to be one of the variables accounting most for variance in accent ratings, with those who had high levels tending to receive lower (and therefore more nativelike) foreign accent ratings. However, Moyer also found one exceptional participant who was generally perceived to be a native speaker, and the major difference with this participant was a strong wish to acculturate and sound German. The author interprets this as demonstrating integrative motivation (Moyer, 1999). A similar link was made by Polat (2011) who examined how motivation (along with other variables) contributed to the acquisition of a native Turkish accent by 13-18 year old Kurdish school pupils living in Turkey. Some participants had little exposure to Turkish until entering formal education at the age of 6 or 7. Participants filled in language background and motivation questionnaires and native speakers of Turkish rated the participants' readings of a paragraph in Turkish. The ratings of the participants' pronunciations ranged from very strong foreign accent to no foreign accent and Polat (2011) found that accents became more nativelike when levels of integrated orientation increased.

3.2.2 Concern for Pronunciation Accuracy and Pronunciation Success

From this work by Polat (2011), Moyer (1999), and (less strongly) Smit (2002) it can therefore be argued that the socio-educational model's integrativeness or integrative motive is closely linked to pronunciation accuracy and that the desire to sound like a native speaker is a key aspect of this integrative motive. Moyer (2007) carried out an additional study linking the desire to sound like a native speaker and pronunciation success. The author recorded 48 non-native English speaking students attending an American university carrying out a number of speech elicitation tasks and had native speakers of English rate these recording for accentedness. In addition, the participants completed a survey on their language background and attitude towards their target language (English) and (US) culture. The author found that the participants' desire to improve accent correlated significantly to their accent ratings in that the greater the desire to improve their accent, the better their English speaking accent was likely to be. In addition, in a multiple regression analysis, the desire to improve accent contributed significantly to the variance in accent ratings (Moyer, 2007).

A number of studies have described this desire to sound like a native speaker as a concern for pronunciation accuracy and have found links between this concern and pronunciation success. One of the earlier studies is that of Purcel and Suter (1980) who reanalysed data collected by Suter (1976). Suter's (1976) participants were 61 nonnative speakers of English attending University in California. Biographic detail pertaining to 20 hypothesised predictor variables was collected through one-to-one and small group interviews, tests and mimicry recordings. Of particular relevance to the present work are the motivational variables addressed: economic motivation, social prestige motivation, integration orientation and strength of concern for pronunciation accuracy. For rating, participants were also recorded carrying out a speech elicitation task. Purcel and Suter (1980) found that only four of their 20 proposed predictors combined significantly to explain the variation in pronunciation ratings and these were first language, aptitude for oral mimicry, residency in an English speaking country and with an English native speaker, and strength of concern for pronunciation accuracy (although the authors, in contrast to Moyer, 1999, felt this was a separate aspect to an integration orientation and noted that this orientation was less relevant in determining pronunciation accuracy) (Purcel & Suter, 1980).

Further evidence linking a concern for pronunciation accuracy and pronunciation success comes from Birdsong (2003, 2007). As noted in Chapter 1, the author asked English speaking learners of French to read aloud a word list and a list of paragraphs, and had the productions rated by native French speakers. He found that the two highest performing participants reported a very high motivation to learn French on a scale of 1 (not at all important/motivated) to 10 (very important/motivated). Furthermore, similarly to findings by Purcel and Suter (1980) and Moyer (1999) both participants stated that they found authenticity and accuracy in pronunciation to be very important and one stated that it was very important to them to be taken as a native speaker by native listeners (the other nativelike participant did not answer this question). However, Birdsong (2003, 2007) does add the caveat that many other of his participants reported high levels of motivation, although he does not examine whether they had a tendency to outperform those who did not.

An attempt to measure strength of concern of pronunciation accuracy was reported by Elliott (1995), who also found strong links between this concern and pronunciation success. The study looked at the pronunciation accuracy of American students enrolled in an intermediate Spanish programme. The author examined the relationship between 12 variables he believed might predict pronunciation accuracy and the scores that participants were awarded on pronunciation tests scored by native and near native speakers of Spanish. The author firstly found that strength of concern for pronunciation accuracy, as measured by his Pronunciation Attitude Inventory (PAI), correlated most with scores received on the pronunciation test, with the PAI significantly related to all test sections save word mimicry. Perhaps more importantly, however, multiple regression analysis demonstrated that the PAI score was also the most significant predictor of pronunciation accuracy (Elliott, 1995).

As can be seen from the studies described above, motivation, generally integrative motivation of some nature, and the strength of concern for pronunciation accuracy aspect in particular, has been found to play a role of varying importance in pronunciation success. At the same time, the paucity of studies carried out on the topic demonstrates a need for further work in this area. However, many of the studies linking motivation in general and pronunciation success have examined the ultimate attainment of L2 learners in the long term and not those still actively learning their L2. In the present work it is intended to carry out a relatively short term training study on L2 students for whom long term ultimate attainment is less relevant (M. Ota, personal communication). The present work will therefore examine how the specific aspect described above, strength of concern for pronunciation accuracy, may contribute to greater success in improving pronunciation through training. It is believed that a greater or lesser concern for pronunciation accuracy is more likely to have a direct effect on achievement after training in the shorter term. In particular, this concern has been linked to pronunciation success in students who are unlikely to have reached their final level of achievement (see Elliott, 1995), and it is intended to use a similar population in the present work.

3.3 Research Questions

The review of literature in the previous three chapters has demonstrated that there are some late L2 learners whose accents are indistinguishable from native speakers (e.g. Bongaerts, 1999), and that such learners have generally been found to have received some degree of pronunciation and/or perceptual training and also have a strong desire to sound like a native speaker, that is, have a high concern for pronunciation accuracy (e.g., Moyer, 1999). However, the perceptual training techniques commonly used have rarely been compared. The present work aims to use and contribute to this knowledge by training L2 learners on difficult L2 contrasts using a variety of techniques, by assessing their concern for pronunciation accuracy and in so doing, attempting to answer the following research questions:

Q1 Is the high variability phonetic training (HVPT) technique more successful than the perceptual fading (PF) technique (or vice versa) in terms of producing a generalisable, long-term improvement in perception?

Q2 Is the most successful perceptual training technique(s) suggested through answering research question 1 more successful than pronunciation training in terms of producing a generalisable, long-term improvement in pronunciation and perception?

Q3 With regard to using HVPT and/or PF and perception and/or production, does an optimal training technique emerge from those examined?

Q4 Do those with a stronger concern for pronunciation accuracy perform better or improve more in terms of pronunciation with training and with which techniques?

4 The Training Task

The purpose of this chapter is to describe three small scale studies which were conducted in order to inform key methodological aspects of the subsequent studies described in Chapters 5 and 6. These large scale studies could not be carried out without firstly establishing which contrasts (in this case French vowel contrasts) were difficult for second language learners to perceive, which task (identification or discrimination) to use for training and testing; and for the perceptual fading conditions, where to locate the 'fading' points on the vowel continua. The sections below describe how these questions were answered.

4.1 Which Contrasts?

4.1.1 Introduction

A majority of the previous research on second language (L2) perception and/or production (and their training) has focused on difficult L2 consonant contrasts, in particular the difficulty that native Japanese speakers have in perceiving the /r/-/l/ contrast in their L2 English (see e.g. Logan, Lively & Pisoni, 1991; Lively, Logan & Pisoni, 1993; Lively, Pisoni, Yamada, Tohkura & Yamada, 1994; Bradlow, Pisoni, Yamada & Tohkura, 1997; Bradlow, Yamada, Pisoni & Tohkura, 1999; McClelland, 2001; McCandliss, Fiez, Protopatas, Conway and McClelland 2002; Zhang, Kuhl, Imada, Iverson, Pruitt, Stevens, Kawakatsu, Tohkura & Nemoto, 2009). In comparison, a far smaller body of work exists on the perception and/or production of L2 vowels, although there has been great recent increased interest with regard to this topic. For example, Lambacher Martens, Kakehi, Marasinghe and Moholt (2005) examined the effectiveness of a high variability training technique on improving native Japanese speaker perception and production of American English mid and low vowels. Wang and Munro (2004) perceptually trained Mandarin and Cantonese speakers on three English vowel contrasts using synthetic and natural stimuli, and Jacewicz (2002) examined the ability of American English speakers beginning to learn German on their ability to accurately perceive and produce four lax German vowels. Lengeris and Hazan (2010) perceptually trained native speakers of Greek on up to 14 English vowels, again using synthetic and natural stimuli. However, this recent work on difficult vowel contrasts mostly (but not always, as can be seen above) relates to L2 English.

Although not related to L2 speech, evidence for the particular importance of vowels in determining comprehensibility of speech, in this case sentences, comes from Cole, Yan, Mak, Fanty and Bailey (1996) who carried out a series of experiments to test the importance of vowels vs. consonants to speech recognition. Participants listened to either unaltered sentences, sentences where the consonants were replaced by noise or sentences where the vowels were replaced by noise. The authors found that recognition of words within the sentences was more dependent upon vowels than consonants with significantly more words being recognised when the vowel information was retained in the sentences (vowels and consonants were equally represented). A hypothesis that the formant transitions at the vowel boundaries provide more information about adjacent consonants than the formant transitions at the consonant boundaries provide about adjacent vowels was partially supported (Cole et al, 1996). Kewley-Port, Burkle and Lee (2007) carried out a similar experiment using young normal-hearing participants and typical elderly hearing-impaired participants. The vowel only sentences were significantly more comprehensible than the consonant-only sentences for both groups.

Examination of the perception, production and training of difficult L2 vowel contrasts is at least as important as researching difficult L2 consonant contrasts for a number of similar reasons. In general terms, vowel perception and production has been found to pose more difficulties for second language learners in terms of perception, and production accuracy in particular (MacKay, 1997). More specifically, in terms of the importance of accurate vowel production, Ingram and Pittam (1987) examined accent changes in Vietnamese children learning English in Australia over the period of one year. The authors recorded participants naming pictures, reciting the days of the week and reciting the numerals 1-10 and re-recorded these participants approximately one year later. Acoustic analysis and perceptual judgements from native Australian English listeners determined that vowel features were more important than consonantal features in determining perceived accent change towards a native norm occurring over the year between recordings.

Similarly, Schairer (1992) examined the role of several phonetic vowel and consonant features on the accentedness and comprehensibility of the pronunciation of Spanish words by non-native Spanish speakers. She found that native speaker ratings of the non-native speaker productions were most influenced by the vowel features, and that it was the vowel features that posed most problems for the learners of Spanish (i.e., the vowel productions were less comprehensible and more accented than the consonants). The author concluded that:

> "Emphasis should be on native-like production of both stressed and unstressed vowels to enhance the communicative potential of the learner's speech. Specific attention to consonants and encouragement of rapid speech could be deferred until the vowel sounds are being produced more adequately." (p. 318)

Examination of French vowel contrasts, as the area of interest in the present study, reveals a number of difficulties for native speakers of English. As Dowd, Smith and Wolfe (1997) note, the /u/-/y/ distinction does not exist in English and many English speakers therefore experience difficulty in hearing and reproducing the difference between the two phonemes. This difficulty has been demonstrated in a number of studies. For example, Levy and Strange (2008) and Levy (2009) examined the performance of native American English speakers in an AXB discrimination task featuring Parisian French vowels, and found the /u/-/y/ contrast to be particularly difficult and among the most problematic of all the contrasts examined (nasal vowels were not tested), irrespective of participant experience with the French language. It was only for this contrast that inexperienced and experienced learners did not significantly differ in discrimination accuracy, showing it to be particularly resistant to perceptual learning (Levy 2009). In both studies the magnitude of the difficulty was dependent upon experience and the consonantal context of the vowels in the nonsense words used (bilabial rabVp vs. alveolar radVt). Similar work by Gottfried (1984) and Rochet (1995) also highlighted the difficulty of perceiving and producing the /u/-/y/ contrast for non-native speakers of French.

In terms of pronunciation of this contrast, Flege (1987) and Macdonald (2006) found that learners of French either made little distinction between the two vowels, or experienced particular difficulties in pronouncing French /u/ accurately. Levy and Law (2010) also found that their participants experienced difficulties in the pronunciation of this contrast, however where this difficulty lay was dependent on consonantal context with /u/ being pronounced less accurately than /y/ in an alveolar /radVta/ context and /y/ being pronounced less accurately than /u/ in a biliabial /rabVpa/ context.

The contrast between French nasal vowels also presents a difficulty for native speakers of English as English makes no use of the nasal distinction. Calbris (1978) and Tranel (1987) both state that the /ɑ̃/-/ɔ̃/ distinction is particularly difficult to perceive and produce, with Calbris (1978) further noting that the /ɑ̃/-/ɛ̃/ distinction is also often a source of confusion. One of the few studies to empirically demonstrate this difficulty is that of Garrott (2006). In order to investigate nasal pronunciation problems for native English speaking learners of French, the author asked participants to answer a series of elementary level French questions on general topics such as colours and foods. Participants also read a short passage regarding a typical student day. Garrott (2006) found that participants had particular problems in distinguishing /ɔ̃/ from/ɑ̃/ when carrying out the informal task of answering the questions and had a particular problem pronouncing /ɛ̃/ accurately in the formal reading task.

Overall evidence for the contribution of poor pronunciation of the nasal and oral contrasts described above to a noticeable foreign accent comes from Vieru-Dimalescu and Boula de Mareüil (1996). The authors asked native speakers of French to identify which of six foreign accents they could hear in the productions of learners of French. Although the native French speakers did not speak any of the languages concerned, they could identify the accents at well above chance levels. Participant self report confirmed by acoustic analysis demonstrated that "/u/ instead of /y/or vice versa, and a bad realisation of nasals reveal[ed] a foreign accent in general rather than [of] a particular origin." (p.442).

The existing literature, as cited above, does therefore pinpoint several likely candidates for which French vowel contrasts do actually pose real difficulties for native English speakers. Most empirical evidence suggests that differentiating /u/ and /y/ is problematic for native speakers of English. Furthermore, as noted above, Tranel (1987), Calbris (1967) and Garrott (2006) suggest that perceptual and production differentiation between the French nasal vowels can prove problematic. Although French has four nasal vowels: /ɛ̃/, /ɑ̃/, /ɔ̃/ and /œ̃/ only the first three listed are in common use in standard French today, with /œ̃/ often merged with /ɛ̃/ (Tranel, 1987). It was therefore also decided to investigate the difficulties in perceptually differentiating each of these first three vowels with the other, as well as each of these vowels and their oral counterparts.

The purpose of this preliminary study is therefore to identify which of these seven contrasts are the most problematic in terms of perceptual discrimination for native English speaking learners of French. This will ensure that the contrasts used in subsequent perception and pronunciation training studies merit attention. As well as investigating the relative difficulty of these French vowel contrasts, it was also of interest to investigate how this difficulty related to experience with French, i.e. for experienced vs. inexperienced learners. The study therefore compares experienced and inexperienced participant discrimination accuracy scores for each of the potentially difficult contrasts. An AXB task was used whereby the participant heard three words and had to decide whether the second word was the same as the first word or the third word. It was necessary to use this task as non-French speaking participants would not be able to carry out an identification task which would require them to listen to a French word and decide which of two words presented onscreen they had heard (see also Section 4.2). However, as each word in the AXB trial came from a different speaker, participants were required to make a vowel category match rather than an exact acoustic match making the task more comparable to real life situations (e.g. Strange, 1995, see also Section 4.2). The experimental methodology is detailed below.

4.1.2 Participants

A total of ten native English speakers with varying levels of experience with the French language participated in the study. The 'experienced' group consisted of 5 participants (3 female, 2 male), two of whom had just completed their second year of study at the University of Edinburgh and two of whom had just completed their second year of study at the University of Glasgow. The fifth participant (male) had graduated with a degree in French from the University of Glasgow in 2000. Although the latter participant performed the most accurately, at least one of the 2nd year students performed comparably, and the addition/removal of this participant made no difference to the final results. It was therefore decided to retain this participant's results for analysis within the experienced group, despite the participant being more experienced than the other participants in the group. Two members of this group were speakers of Southern English and the other three members were speakers of Standard Scottish English.

The 'inexperienced' group also consisted of 5 participants (3 female, 2 male), who were known to the experimenter. None had studied French to any more than Scottish Standard Grade level, and three out of the five spoke no French at all. Four members of the group were speakers of Standard Scottish English and the remaining participant was a speaker of Southern English. None of the ten participants reported any hearing difficulties.

4.1.3 Stimuli

Nineteen or twenty minimal pairs for each contrast (/ɑ̃/-/ɔ̃/; /ɑ̃/-/ɛ̃/; /ɛ̃/-/ɔ̃/; /ɑ̃/-/ɑ/; /ɔ̃/-/ɔ/; /ɛ̃/-/ɛ/ and /u/-/y/) were identified (see Appendix A for the full list). Due to the difficulty in finding sufficient minimal pairs to test, it was impossible to control for such factors as word frequency and vowel position (however, see Chapter 5 for considerations made once words were identified for the large scale training studies). Two native speakers of French (both female, with standard French accents) were recorded reading each word in isolation three times consecutively (the words were presented in a random order) in a soundproofed recording studio isolation booth at the University of Edinburgh. The sound was captured using an AKG CK98 hypercardoid microphone with and encoded on a custom designed PC based on a Shuttle XPC chassis with a Core2Duo processor running Sonar 4 Studio Edition software via a MOTU 828 audio interface. The sound was digitised at a sampling rate of 48 kHz with a resolution of 32 bits. A further native speaker (male, with a standard accent) was recorded reading each word three times using an M-Audio 24/96 digital recorder with a Sony ECM MS907 Electret Condenser desktop microphone in a quiet room in his home. The sound was again digitised at a sampling rate of 48 kHz with a resolution of 32 bits.

The second reading of each word was generally used for testing (as the second reading tended to be clearer than the first and third), excepting when there was some problem with its clarity (then the clearer of the other two readings was used). The readings of the words from each native speaker were put together for use in an auditory AXB task. Within each contrast tested, all possible permutations of the AXB task (ABB, AAB, BAA, BBA) and all possible permutations of speaker order were used in approximately equal numbers, as far as the number of stimuli allowed. Each minimal pair was used twice in order that sounds A and B in the pair each occupied the X position once. Half of the minimal pairs within each contrast were used in ABB and AAB trials and half of the minimal pairs within each contrast were used in BAA and BBA trials. Five hundred msec of silence was inserted between each word in a trial.

Due to inconsistencies between the native speakers in pronunciation, a number of minimal pairs were discarded from the /ɔ̃/-/ɔ/ contrast (in these discarded minimal pairs, one or more of the native speakers used /o/ for the oral vowel instead of /ɔ/). In addition, a number of minimal pairs from each contrast were chosen to be used as practice trials resulting in 12 practice trials with either one or two examples of each contrast. This resulted in 238 experimental items (18 minimal pairs for each contrast used twice, excepting the /ɔ̃/-/ɔ/ contrast with 11 minimal pairs used twice).

4.1.4 Procedure

For all participants, the experiment was run using E-Prime software on a Dell Inspiron 6400 laptop with headphones in a quiet room. The onscreen instructions explained that the participant would be played three French words and that their task was to decide whether the second word they heard was the same as the first word, or the same as the third word. The participants were instructed to press ‘1’ if the second word was the same as the first word and ‘3’ if the second word was the same as the third word (see Appendix B for the full instructions given). The 12 practice trials then occurred before the experimental block of 280 trials. The practice block and the experimental block trials were presented in a random order and no feedback was given.

4.1.5 Results

It was firstly hypothesised that the contrast tested would make some difference to the results, in other words, it was expected that some contrasts would result in a lower percentage of trials with a correct response than others. It was secondly hypothesised that the more experienced group was likely to perform more accurately. Overall, the participants performed reasonably accurately, with 86% of trials being answered correctly. The overall results can be found in Table 4.1.

Table 4.1: Mean Accuracy (%) and Standard Deviation in the AXB Task

Contrast	Accuracy	SD
/u/-/y/	84.17	13.42
/ɑ̃/-/ɔ̃/	61.67	9.15
/ɑ̃/-/ɛ̃/	85.55	12.55
/ɔ̃/-/ɛ̃/	90.83	7.18
/ɛ/-/ɛ̃/	96.94	3.33
/ɑ/-/ɑ̃/	94.17	6.98
/ɔ/-/ɔ̃/	90.00	10.00

The /ɑ̃/-/ɔ̃/ contrast has a particularly low score (although a one sample t-test reveals that this score is significantly different from chance levels $t(9) = 4.033$; p (two-tailed) = 0.003). In addition the scores for the /u/-/y/ and /ɑ̃/-/ɛ̃/ contrasts are also relatively low, at well under 90%.

Subsequent analyses suggested that the /ɑ̃/-/ɔ̃/contrast is difficult for both groups to discriminate and that the /u/-/y/ and /ɑ̃/-/ɛ̃/ contrasts are difficult for the inexperienced group to discriminate. A two-way mixed ANOVA with the seven French vowel contrasts as a within subjects factor and experience with French as a between subjects factor revealed a significant effect of Contrast [$F(6, 48) = 36.213$; $p < 0.001$], a significant effect of Experience [$F(1,8) = 28.776$, $p = 0.001$] and a significant Contrast*Experience interaction [$F(6,48) = 4.421$, $p = 0.01$]. The significant effect of French language Experience indicated that the experienced group obtained a significantly higher average percentage correct (Mean = 91.86%) than the inexperienced group (Mean = 80.52%).

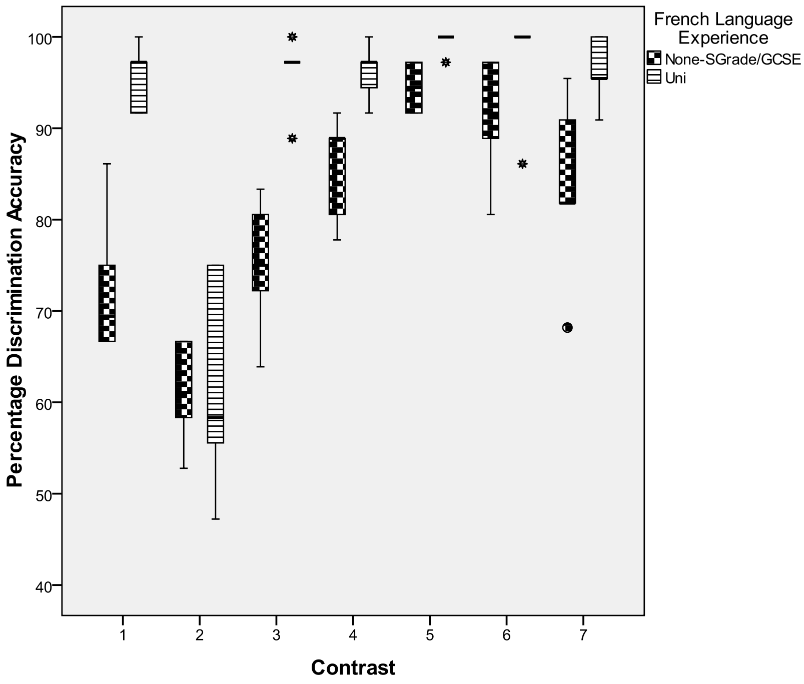
Post-hoc comparisons (Bonferroni adjusted) firstly revealed that the discrimination accuracy score for the /ɑ̃/-/ɔ̃/ contrast (Mean = 61.67%) is significantly lower than those for all other contrasts (all p < .005). Furthermore, the score for the /u/-/y/ contrast (Mean = 84.17%) is significantly lower than the /ɛ̃/-/ɛ/ contrast (Mean = 96.94%, p = .001) and /ɑ̃/-/ɑ/ contrast (Mean = 94.17%, p = .045). Finally, the score for the /ɑ̃/-/ɛ̃/ contrast (Mean = 85.55%) is also significantly lower than the /ɛ̃/-/ɛ/ contrast and the/ɑ̃/-/ɑ/ contrast (p = .004, p = .016). None of the other contrasts result in significantly lower scores in comparison with any other contrast.

Finally, Table 4.2 and Figure 4.1 illustrate the significant Contrast*Experience interaction.

Table 4.2: Mean Accuracy (%) and Standard Deviation in the AXB Task According to Experience

Contrast	Experienced Accuracy (%)	SD	Inexperienced Accuracy (%)	SD
/u/-/y/	95.56	8.19	72.78	3.73
/ɑ̃/-/ɔ̃/	62.22	5.89	61.11	12.36
/ɑ̃/-/ɛ̃/	96.11	7.61	75.00	4.21
/ɔ̃/-/ɛ̃/	96.11	6.02	85.56	3.17
/ɛ/-/ɛ̃/	99.44	2.77	94.44	1.24
/ɑ/-/ɑ̃/	97.22	6.91	91.11	6.21
/ɔ/-/ɔ̃/	96.36	10.46	83.64	3.80

Figure 4.1: The Interaction Between Contrast Tested and French Language Experience



(1 = /u/-/y/; 2 =/ɑ̃/-/ɔ̃/; 3 = /ɑ̃/-/ɛ̃/; 4 = /ɔ̃/-/ɛ̃/; 5 = /ɛ/-/ɛ̃/; 6 =/ɑ/-/ɑ̃/; 7= /ɔ/-/ɔ̃/)

The most striking effect here is that there appears to be little difference in the performance between the experienced and inexperienced groups on the /ɑ̃/-/ɔ̃/ contrast, whereas the groups seem more divided for the other contrasts. A one-way ANOVA confirms that there is no significant difference between the experienced and inexperienced groups in the /ɑ̃/-/ɔ̃/ contrast [M_{exp} = 62.22; M_{inexp} = 61.11; $F(1,8)$ = 0.033, p = .861], and the /ɑ̃/-/ɑ/contrast [M_{exp} = 97.22; M_{inexp} = 91.11; $F(1,8)$ = 2.161, p = .180], although in the latter case this is due to high discrimination accuracy by both groups. All other contrasts were more sensitive to experience with the discrimination accuracy of the experienced group being significantly more accurate than the inexperienced group (all $p < .05$).

4.1.6 Discussion and Conclusions

The purpose of this study was to identify the most problematic contrasts to use for the subsequent perception and production training studies. As the results above indicate, a clear problem area is the perception of the /ɑ̃/-/ɔ̃/ contrast, with experienced and inexperienced participants alike finding this far more of a problem than the other contrasts. In addition the /u/-/y/ and /ɑ̃/-/ɛ̃/ contrasts appear to be particularly sensitive to experience, the inexperienced participants scoring well below 80%, which also resulted in these contrasts having lower discrimination accuracy scores than two other contrasts overall.

Due to time constraints (in terms of the anticipated lengthy stimulus preparation time) it was likely that only two contrasts could be examined. It is therefore concluded that the /ɑ̃/-/ɔ̃/ contrast, as posing the clearest difficulties in the present study, and the /u/-/y/ contrast as being the next most problematic for the inexperienced group and having existing literature to support its difficulty, are the best candidates for further training through the proposed perception and production training studies.

Having established difficult French vowel contrasts for training, the next consideration was the way in which the training stimuli were to be presented. This is examined in Section 4.2 below.

4.2 Training and Testing Type: Identification vs. Discrimination

4.2.1 Introduction

Many training studies (see Chapter 2) use identification training and/or testing whereby participants are played a single stimulus on each trial and have to identify which among number of options they believe they have heard. This usually takes the form of a two-alternative forced-choice (2AFC) task, where both members of a minimal pair are provided and the participant identifies which member has been presented (e.g. Logan et al, 1991). In training, feedback as to whether the participant has responded correctly is usually provided.

An alternative task used is discrimination training and/or testing whereby at least two stimuli are presented on each trial, and the participant has to decide whether or not the stimuli are examples of the same category or different categories. Discrimination tasks usually take one of three forms. The first is AX (or same-different) discrimination where two stimuli are presented and the participant has to decide whether they are the same or different (Logan & Pruitt, 1995). The second form is ABX or AXB (or oddity) discrimination whereby three stimuli are presented on each trial and the participant has to decide whether the stimulus in position X is the same as the stimulus in position A or B; an alternative task is to decide which stimulus is different. The stimulus in position A or B which is not the same as X is the other member of the minimal pair of interest (Logan & Pruitt, 1995). Finally, in a category change task a stimulus is repeatedly presented and is then followed by repeated presentations of a different stimulus. The participant should indicate awareness of this change as soon as it occurs (Logan & Pruitt, 1995). Again, in discrimination training feedback as to whether the participant has responded correctly is usually provided.

The purpose of the present preliminary study was to ascertain the optimal task to use in the main training studies. Despite the clear differences between these methods, it appears that only Flege (1995), Wayland and Li (2008), Handley, Sharples and More (2009), and Shinohara (2012) to date have investigated whether or not one is more or less effective than the other. However, these studies compared the methods in terms of training L2 contrasts, whereas the purpose of the present study was to compare the methods in terms of testing, and the tasks used in the studies were not precisely same as those it was intended to use in the present study. Nonetheless, the discussion re the relative merits of these tasks for training is still relevant here as it was intended to use the same task for training as for testing in the main training studies. Choosing a particular testing method would therefore also have implications for training.

Flege (1995a) and Wayland and Li (2008) cite similar evidence to suggest that most researchers believe that using 2AFC identification training is preferable to using AX/AXB/ABX training for a number of reasons. Firstly, Logan et al (1991) suggested that 2AFC identification training results in participants developing and using long term memory phonetic codes instead of relying upon sensory information fading from short-term memory. Secondly, Lively et al (1994) suggested that 2AFC identification training plus immediate feedback can form more robust phonetic categories which are not sensitive to irrelevant stimulus properties such as speech rate or particular characteristics of individual speakers. Similarly, Jamieson and Morosan (1986) stated that discrimination training may result in participants attending to irrelevant within-category acoustic differences rather than the important between-category cues.

Doubt about the superiority of identification training is raised by Polka (1992), who notes that when performance in an identification task is not consistently accurate or inaccurate it is difficult to ascertain whether or not the problem is with recognising the category difference, or with incorrect assignment of the category label to category member, or both. In addition the author raises concerns that participants may learn to respond correctly in identification trials using different properties to differentiate categories than those used by native speakers (Polka, 1992). Furthermore, Logan and Pruitt (1995) note that if an identification task is used, participants must be able to match the sound being heard to one of the 'labels' (usually words) provided, which is difficult for inexperienced listeners or if the L2 uses a different orthographic system. Further confusion about the relative merits of each procedure comes from differences of approach within the discrimination method. For example Strange (e.g., 1995) makes the distinction between a traditional discrimination method where the two 'same' stimuli are physically identical and 'categorical' discrimination whereby the two 'same' stimuli are physically different but belong to the same phonetic category. Polka (1992) notes that using this categorical discrimination task requires the participant to attend to the important between-category cues and ignore irrelevant within category differences, and Wayland and Li (2008) suggest that the task encourages the participant to rely less on upon sensory information fading from short-term memory, thereby theoretically addressing some of the concerns noted above. Most recent discrimination studies, therefore, make use of this categorical discrimination task.

Flege (1995a) appears to be the first to directly compare the identification and categorical discrimination training tasks, in attempting to train L2 English speakers who were native speakers of Mandarin to distinguish between word final /t/ vs. /d/. The author found that both methods were equally successful in terms of improvement from pre-test to post-test, generalisation, and retention of improvement two months after completion of training. Wayland and Li (2008) compared the two training procedures with regard to increasing native English and native Chinese listeners' ability to perceive the difference between the mid and low tones in Thai. Similarly to Flege (1995a), the authors found no significant difference between the two training techniques. Handley et al (2009) compared the tasks with reference to training native speakers of Mandarin on the English /r/-/l/ contrast and Shinohara (2012) carried out this comparison on native speakers of Japanese. Both studies again found no significant difference between the two techniques.

Flege (1995a) concluded that his results cast doubt on the generally held view that identification training is the superior method, and suggested that using the categorical discrimination instead of the traditional discrimination task may be the reason for this, for the same reasons as outlined above. Wayland and Li (2008) similarly emphasise the importance of using this categorical discrimination task in order to obtain comparable results. However, at the end of the training, Flege (1995a) asked the participants about their enjoyment in participating in the study, about how beneficial they felt it had been, and their willingness to participate in more training. The author found that the participants who had received identification training responded more positively to these questions, and therefore tentatively suggested that, in the absence of any significant differences between training methods, identification training may be the method to use.

In the present work, as previously noted, it was necessary to use an AXB task to investigate the relative difficulty of the contrasts to be trained in the study described in 4.1 above, as the inexperienced participants would have been unable to carry out an identification task involving matching the sound being heard to one of two words presented onscreen. However, in the main training studies it was intended to use first and second year university students of French, which matched the participant profile of the experienced group in the 4.1 study, and these participants appeared to find little difficulty in discriminating the /u/-/y/ contrast in an AXB task (Mean accuracy = 95.56%; again the 'same' stimuli were nonidentical). At the same time, there is evidence to suggest that this contrast is difficult for learners of French at all levels of ability (e.g. Levy & Strange, 2008, Levy, 2009), which indicates that the AXB task may not be sufficiently difficult for participants at this level.

Subsequent results from carrying out early training of first and second year students of French provided further evidence that AXB training and testing may not be suitable for the /u/-/y/ contrast, as testing scores again demonstrated high accuracy (Pre-test Mean = 96.20%; Post-test Mean = 98.88%). The AXB task in this instance was slightly different as it made use of two native speakers of French (one female, one male; one speaker per trial) instead of three native speakers. The reason for this was that it was intended to compare training with multiple voices vs. training with only one voice in the first main study, and the initial concern was that exposing participants in the single talker conditions to three voices in an AXB pre-test would have an effect on the final results. However, as can be seen from the results above, the change from three to two speakers did not appear to make a difference to task difficulty, although it is difficult to compare these results as different words and participants were used.

Due to these early ceiling level test results it was therefore decided to consider whether an identification test would be more difficult in order that any training effects be clearly demonstrated. One immediate advantage of this training and testing is that it is more comparable the task listeners undertake in real life situations. The present work therefore examines the two difficult contrasts identified in 4.1 above and compares participant performance in the two-speaker AXB task with their subsequent performance in a 2AFC identification task using four speakers in order to increase difficulty.

4.2.2 Participants

The participants were 16 (12 female, 4 male) native English speaking first and second year students of French attending Edinburgh University. None of the 16 participants had extensive experience of the French language outside their studies (i.e. close French relatives, lengthy stays in a French speaking country) or reported any hearing difficulties.

4.2.3 Stimuli

Seven minimal pairs for each contrast (/ɑ̃/-/ɔ̃/ and /u/-/y/) were identified for the purposes of the testing (see Appendix C for the full list and word frequencies). Due to the difficulty in finding sufficient minimal pairs to use overall (pairs were also to be required for perceptual training, pronunciation training and tests of generalisation in the main training studies), it was impossible to control for such factors as word frequency and vowel position in selecting the words (however, again, see Chapter 5 for considerations made once words were identified).

4.2.3.1 AXB Task Stimuli

Two native speakers of French (one male, one female, with standard French accents) were recorded reading each word three times in isolation (the words were presented in a random order) in a soundproofed recording studio isolation booth at Edinburgh University. The sound was captured using an AKG CK98 hypercardoid microphone and encoded on a custom designed PC based on a Shuttle XPC chassis with a Core2Duo processor running Sonar software via a MOTU 828 audio interface. The sound was digitised at a sampling rate of 48 kHz with a resolution of 32 bits.

The first two readings of each word were generally used for testing (the participants did not read each word three times in a row, each word was presented three different times), excepting when there was some problem with its clarity or pronunciation (then the third reading was used). Using Audacity software, the readings of the words from each native speaker were put together for use in an auditory AXB task such that all possible permutations of the AXB task (ABB, AAB, BAA, BBA) were shared equally by the two speakers, and that only one speaker was heard per trial. In other words, each minimal pair was used four times, twice by each speaker, with the permutations used being counterbalanced across the two speakers.

Five hundred msec of silence was inserted between each word in a trial. There were 56 experimental trials (7 minimal pairs for two contrasts used four times). As each of the three words in the AXB trial were different productions by the same speaker, participants were still required to make a vowel category match rather than an exact acoustic match.

4.2.3.2 Identification Stimuli

For the identification testing, a further two native speakers of French (one male, one female, with standard French accents) were recorded using the same procedures and equipment as for the discrimination condition described above. These two native speakers were used in addition to the two originally recorded native speakers.

The first reading of each word was generally used for testing (the participants did not read each word three times in a row, each word was presented three different times in isolation to avoid a list intonation reading effect), excepting when there was some problem with its clarity or pronunciation (then the second or third reading was used). Individual readings of the words from the four native speakers were isolated for use in an identification task. Each minimal pair was again used four times, each member word was heard twice, once by a male speaker and once by a female speaker. The male/female pairing and which member of the contrasts of interest they read were counterbalanced across all four speakers. This resulted in 56 experimental trials (7 minimal pairs for two contrasts used four times).

4.2.4 Procedure

For all participants, both parts of the experiment were run on Dell PC computers using E-Prime software in sound deadened booths in the perception experiment laboratory at Edinburgh University. All 16 participants carried out the discrimination test first, before carrying out the identification test some weeks later (after the initial discrimination results indicated ceiling level performance).

4.2.4.1 Part 1: AXB Task

The onscreen instructions explained that the participant would be played three French words and that their task was to decide whether the second word they heard was the same as the first word, or the same as the third word. The participants were instructed to press '1' if the second word was the same as the first word and '3' if the second word was the same as the third word. The trials were presented in a random order and no feedback was given. The full text of the instructions is in Appendix B.

4.2.4.2 Part 2: Identification

The onscreen instructions explained that the participant would be played one French word, and given two options regarding what the word could be. One option (i.e. one member of the minimal pair) was displayed on the bottom left of the screen and the other option was displayed on the bottom right of the screen. The participants were instructed to press '1' on the keyboard if they thought they had heard the word on the left, and '2' if they thought that they had heard the word on the right (see Appendix D for the full instructions). The trials were counterbalanced such that '1' and '2' were the correct answer an equal number of times and such that each member of each minimal pair was seen twice on the left and twice on the right.

In order to avoid confusions due to orthography or unfamiliarity with the words, participants were also provided with an information sheet explaining the sound to spelling mappings used in the trials (see Appendix E for the full text).

4.2.5 Results

It was firstly hypothesised that the testing methodology would make some difference to the results in that it was anticipated that the 2AFC identification task may result in a lower percentage of trials with a correct response than the AXB task. It was secondly hypothesised that the contrast tested would make some difference to the results, in other words, it was expected that one of the contrasts would result in a lower percentage of trials with a correct response than the other.

The results for the identification task are presented with those from the discrimination task carried out by the same participants in Table 4.3.

Table 4.3: Mean AXB Task* vs. Identification Task Accuracy (%) and Standard Deviation

Contrast	AXB* Accuracy	SD	Identification Accuracy	SD
/u/-/y/	96.43	4.12	71.88	13.92
/ɑ̃/-/ɔ̃/	70.98	10.26	65.17	12.81

* Two speaker (one per trial) task

An initial comparison of the mean accuracy for identification vs. discrimination demonstrates that the participants indeed found the identification task more difficult, at least for the /u/-/y/ contrast (Identification Mean = 71.88%, Discrimination Mean = 96.43%). Taking these percentages as a guide (performance on both tasks for the /ɑ̃/-/ɔ̃/ contrast remains lower), there certainly appears to be room for improvement through identification training and testing of these contrasts.

Subsequent analysis revealed this initial comparison to be correct. A one-way repeated measures ANOVA with the four tests as four levels of the repeated measures factor were carried out and revealed a significant effect of Test [$F(3,45) = 49.039$, $p<.000$]. Post-hoc (Bonferroni-adjusted) comparisons indicated that the /u/-/y/ discrimination test was significantly easier (higher scoring) than the /u/-/y/ identification test, as well as both /ɑ̃/-/ɔ̃/ tests (all $p<.000$). The comparisons indicated no other significant differences between the scores. Whilst the testing task has no significant effect on performance with the /ɑ̃/-/ɔ̃/ contrast, the /u/-/y/ contrast appears to require identification testing in order to avoid a ceiling effect.

4.2.6 Discussion and conclusions

The purpose of this study was to ascertain the optimum testing (and therefore training) task for the first main experiment. As the results above indicate, the identification test is sufficiently difficult to demonstrate improvement after training in both contrasts of interest. Although the testing task has no significant effect on performance with the /ɑ̃/-/ɔ̃/ contrast, the accuracy level was not at ceiling for either test, again demonstrating the perceptual difficulty of this contrast for native speakers of English. It is likely that the greater difficulty of the AXB task arises from not having both members of the minimal pair presented for comparison.

As previously noted, Flege (1995a), Wayland and Li (2008), Handley et al (2009) and Shinohara (2012) had a slightly different focus from the present experiment in that they were comparing training methods rather than the testing methods used before and after training. Interestingly, Flege (1995a) and Handley et al (2009) used identification testing before and after both types of training, and Flege (1995a) hypothesised that slightly greater gains demonstrated by the identification training group in his study may have been down to greater familiarity with the task. In contrast, Wayland and Li (2008) used AXB testing before and after the using the same training techniques as Flege (1995a), and obtained very similar results, whilst Shinohara (2012) used both tasks in training. This mismatch, in addition to the similarity of the results, casts doubt upon whether any combination of differences in training and testing paradigms within an experiment are important.

However, with reference to results in 4.2.5 above and the contrasts of interest in the present work (the /u/-/y/ contrast in particular), if participants experience no real difficulties in an AXB test, it is unlikely that AXB training will be beneficial either, as they would perform at a high level in all training trials. This would leave participants ill-equipped to deal with the more real to life task of identifying the sounds produced by a number of speakers, as demonstrated by the lower identification test scores.

It is therefore concluded that the participants in the present training studies would benefit most from identification testing and training, and that conclusions about the relative merits of the discrimination vs. identification training and testing techniques can only be drawn with reference to specific contrasts. Whilst Flege (1995a), Wayland and Li 2008), Handley et al (2009) and Shinohara (2012) found no significant differences between the techniques in their studies, there is a clear candidate here for training the /u/-/y/ contrast in particular with the identification paradigm. AXB discrimination training and testing is much less likely to be effective and meaningful in this instance. This conclusion is supported by Smith and Baker (2010) who noted when using both training tasks (although not comparing them in terms of effectiveness) that although discrimination and identification scores are generally highly correlated, discrimination tasks tend to prove easier for the listener than identification tasks.

Having established the nature of the training task to be used, the final consideration was where to locate the 'fading' points on the vowel continua for the perceptual fading conditions. This is dealt with in Section 4.3 below.

4.3 Perceptual Fading Conditions: Formant and Nasal Values

4.3.1 Introduction

The final item for consideration before beginning the first major experiment was the creation of the vowel continua for the perceptual fading training to be carried out. The participants in Iverson et al's (2005) study carried out ten training sessions and therefore the continuum in their perceptual fading condition had ten points for each member of the minimal pair, or 20 points in total. It was decided that six training sessions was a more realistic aim for the present work (as it was felt that asking students to commit to ten training sessions was too much and likely to result in recruitment difficulties), therefore needing 12 points on the /u/-/y/ and 12 points on the /ɑ̃/-/ɔ̃/ continua. The aim here was to establish the formant values for these points by testing likely values on native French speakers. In addition, for the nasal /ɑ̃/-/ɔ̃/ contrast the frequency of the nasal pole (FNP) and frequency of the nasal zero (FNZ) values, which are manipulated in order to provide more or less 'nasality' to the sound being synthesised, would also need to be established.

4.3.1.1 The /u/-/y/ Continuum

Likely points on the /u/-/y/ continuum were the easiest to establish as work has already been carried out on the point at which native speakers of French stop hearing /u/ and start hearing /y/ or vice versa. Rochet (1995) examined how the high vowel continuum /i/-/y/-/u/ is perceived by those who only have /i/ and /u/ in their vowel inventory (native speakers of English and Portuguese) and those who have all three vowels (the native speakers of French). In order to do this the author prepared a synthetic /i/-/y/-/u/ vowel continuum, presented the 21 sounds ten times each in a random order and asked the native English and Portuguese speaking participants to identify them as either /i/ or /y/. Native French speaking participants were asked to identify the sounds as /i/, /y/ or /u/.

Using Klatt's (1980) cascade/parallel speech synthesiser, Rochet (1995) prepared the synthetic vowel continuum by holding constant the first formant (F1) dimension at 250 Hz and the third formant (F3) dimension at 2212 Hz for stimuli with second formant (F2) values below 1800 Hz. For those F2 values above 1800 Hz, the author cited the following Nearey (1989) formula to calculate F3: F3 = 1.4 x (F2-220). The variation in the stimuli was created along the F2 dimension, from 500Hz to 2500Hz in 100Hz steps. The native French speakers consistently identified those stimuli with F2 values between 1300Hz and 1900Hz as /y/ and those stimuli with F2 values between 500 Hz and 900 Hz as /u/, with F2 values between 1000 Hz and 1200 Hz being ambiguous. An F2 value of 1200 Hz was particularly ambiguous for these participants, with the F2 values of 1000 Hz and 1100Hz still generally being identified as /u/.

Creating a 12 point continuum bearing these values in mind therefore seems the most prudent course of action for the present work as the extreme and ambiguous values are well demonstrated by Rochet's (1995) work. However, the difference between the stimuli on the last day of training, the smallest and therefore most difficult, should still be reasonably consistently distinguishable to native speakers of French, even if smaller than the usual difference made by native speakers (see, e.g. McClelland, 2001 and Iverson et al 2005). It was therefore decided to remove the most ambiguous F2 Hz value (1200 Hz) as established by Rochet, as well as the two extreme F2 values (500Hz and 1900Hz, as there were too many points on the Rochet (1995) continuum for the present work), and therefore hold F3 constant at 2212 Hz as well as holding F1 constant at 250Hz. The proposed formant values for the /u/-/y/ continuum can be found in Table 4.4.

Table 4.4: Proposed Formant Values for the 12-Point /u/-/y/ Continuum

	1 /u/	**2**	**3**	**4**	**5**	**6 (intended /u/)**	**7 (intended /y/)**	**8**	**9**	**10**	**11**	**12 /y/**
F1	250	250	250	250	250	250	250	250	250	250	250	250
F2	600	700	800	900	1000	1100	1300	1400	1500	1600	1700	1800
F3	2212	2212	2212	2212	2212	2212	2212	2212	2212	2212	2212	2212

4.3.1.2 <u>The /ɑ̃/-/ɔ̃/ Continuum</u>

The basis for establishing a continuum to test for the /ɑ̃/-/ɔ̃/ contrast was Bognar and Fujisaki (1986), who used their own 'Analysis-by-Synthesis' approach to analyse the spectral characteristics of the French nasal vowels and their oral counterparts produced by one native speaker of French. Their analysis was based upon 12 to 20 productions of the vowel in question in a carrier sentence. The average values established by the authors for /ɑ̃/ and /ɔ̃/ can be found in Table 4.5.

Table 4.5: Average French nasal format and nasal values (Bognar & Fujisaki (1986))

Contrast	F1	F2	F3	FNP	FNZ
/ɑ̃/	590	790	2840	190	380
/ɔ̃/	450	690	2900	170	390

For simplicity, it was decided to hold the F3, FNP and FNZ values constant at their average values of F3 = 2870, FNP = 180 and FNZ = 385. With regard to the other formant values, an exploratory 12 point continuum was created with the F1 and F2 values suggested by Bognar and Fujisaki as endpoints. It was anticipated that using the values suggested by Bognar and Fujisaki (1986) would render words easily identifiable to native speakers and that therefore this would also be true of more exaggerated values. It was therefore necessary to ascertain the points at which major ambiguities occurred for native French listeners in order to create the experimental continuum. The exploratory /ɑ̃/-/ɔ̃/ continuum values can be found in Table 4.6.

Table 4.6: Proposed Exploratory Formant and Nasal Values for the 12-Point /ɑ̃/-/ɔ̃/ Continuum

	1 /ɑ̃/	2	3	4	5	6 (intended /ɑ̃/)	7 intended /ɔ̃/)	8	9	10	11	12 /ɔ̃/
F1	590	580	570	560	550	540	500	490	480	470	460	450
F2	790	783	776	769	762	755	725	718	711	704	697	690
F3	2870	2870	2870	2870	2870	2870	2870	2870	2870	2870	2870	2870
FNP	180	180	180	180	180	180	180	180	180	180	180	180
FNZ	385	385	385	385	385	385	385	385	385	385	385	385

Having ascertained a potential 12 point continuum for both contrasts it was necessary to test these formant values with native speakers in order to establish whether identifying the member of the minimal pair presented was a realistic task. Three native speakers of French carried out identification tasks with minimal pairs from both contrasts and from multiple points on the proposed 12 point continua as described below.

4.3.2 Participants

The participants were three native speakers (NSs) of French studying or living in Edinburgh (1 speaker, female) or Glasgow (2 speakers, both male). All had standard French accents and reported no hearing difficulties.

4.3.3 Stimuli

Fourteen minimal pairs for each contrast (/ɑ̃/-/ɔ̃/ and /u/-/y/) were identified to be used in the training section of the study (see Appendix C) and therefore it was these 56 words that required initial testing on each point of the continuum. Due to the difficulty in finding sufficient minimal pairs to use (pairs were also to be required for perceptual pre- and post- testing, pronunciation training and tests of generalisation; see Chapter 5), it was impossible to control for such factors as word frequency and vowel position (however, again, see Chapter 5 for considerations made once words were identified).

In the first training study there were to be two conditions using the perceptual fading technique, one with the same speaker each day, and one with a different speaker each day (with the fading resulting in the vowels becoming less distinct). Those hearing a different speaker each day were to hear the same speaker (male) on day one as those in the single speaker condition heard every day. To this end, six native speakers of French (three male, three female, with standard French accents) were recorded reading each word three times (the words were presented in a random order) in a soundproofed recording studio isolation booth at Edinburgh University. The sound was captured using an AKG CK98 hypercardoid microphone and encoded on a custom designed PC based on a Shuttle XPC chassis with a Core2Duo processor running Sonar software via a MOTU 828 audio interface. The sound was digitised at a sampling rate of 48 kHz with a resolution of 32 bits.

The first reading of each word was generally used for testing (the participants did not read each word three times in a row, each word was presented three different times), excepting when there was some problem with its clarity or pronunciation (then the second or third reading was used).

A Klatt (1980) synthesiser was then used to synthesise raw versions of the vowels at each stage of the continuum, with formant and nasal values as described in section 4.3.1 above. These raw versions were 500msec long. Using Audacity software, the portion of each word containing the vowel of interest read by the native speakers of French was cut out, and exactly the same length of the appropriate synthesised vowel was inserted into the word. The pitch traces of this hybrid were then manipulated in the PRAAT program to exactly match the original production by the French speaker. Finally, each sound file was normalised using Audacity software. This resulted in 11 versions of each word, six versions of the word produced by one speaker on points 1-6 or 7-12 on the continuum and a further 5 versions produced by different speakers on points 2-6 or 8-12 on the continuum. In order to ascertain that these formant values and procedures were reasonable before continuing with this lengthy procedure for all of the words a small selection of the hybrid words were tested for intelligibility by a native speaker of French who heard everything as intended (see Appendix F for spectrogram examples of the final versions of the synthesised and natural stimuli used in training).

4.3.4 Procedure

The native speakers of French carried out an identification task. The onscreen instructions explained that the participant would be played one French word, and given two options regarding what the word could be. One option (i.e. one member of the minimal pair) was placed on the bottom left of the screen and the other option was places on the bottom right of the screen. The participants were instructed to press '1' on the keyboard if they thought they had heard the word on the left, and '2' if they thought that they had heard the word on the right. The differing versions of the minimal pairs were paired off as points 1&12, 2&11, 3&10, 4&9, 5&8 and 6&7 on the continua. The trials were counterbalanced such that '1' and '2' were the correct answer an equal number of times and such that each member of each minimal pair was seen once on the left and once on the right (i.e. each version of the word was played twice). This resulted in 672 trials per identification task.

There were two identification tasks. The first was using speaker 1 only, with the six fading versions of his productions of each word played twice. The second task was using the six speakers such that the words on points 1&12 were spoken by speaker 1, the words on points 2&11 were spoken by speaker 2, and so on. One male and one female native speaker completed both tasks, and the other male was only able to complete the six speaker task due to time limitations.

The purpose of this trial was to ascertain the points at which the differences between the stimuli became too ambiguous or whether the points on the (/ɑ̃/-/ɔ̃/ and /u/-/y/) continua, in particular the points which were closer together, were actually in the 'correct place', that is, perceptually swapping from one member from the minimal pair to the other as intended. It was therefore proposed to examine the percentage of trials correctly answered by the native speakers, and then examine any errors made more closely to ascertain which 'direction' they took. In addition, it was hypothesised that the single speaker condition may be slightly easier for the native speakers as they became accustomed to his voice throughout the task.

4.3.5 Results

Table 4.7 details the percentage correct identifications from the fading stimuli and Figures 4.2-4.5 illustrate the native speaker identification functions for each contrast (oral vs. nasal) and task (one voice vs. six voices). Performance accuracy is high (>90%) for the /u/-/y/ contrast suggesting that the points and switch from /u/ to /y/ seem to be placed at the correct points in the continuum. However, as anticipated, the /ɑ̃/-/ɔ̃/ contrast appears to be more problematic, and in the six speaker condition in particular.

Table 4.7: Correct Identifications (%) of Perceptual Fading Stimuli by Native French Speakers

	Task 1: Single Voice		Task 2: Six Voices	
Speaker	**/u/-/y/**	/ɑ̃/-/ɔ̃/	**/u/-/y/**	/ɑ̃/-/ɔ̃/
NS1	97.62 (328/336)	83.33 (280/336)	97.02 (326/336)	61.90 (208/336)
NS2	93.45 (314/336)	92.26 (310/336)	94.94 (319/336)	76.49 (257/336)
NS3			96.43 (324/336)	77.68 (261/336)

Figure 4.2: Oral Contrast Identification Function for the Single Voice Task

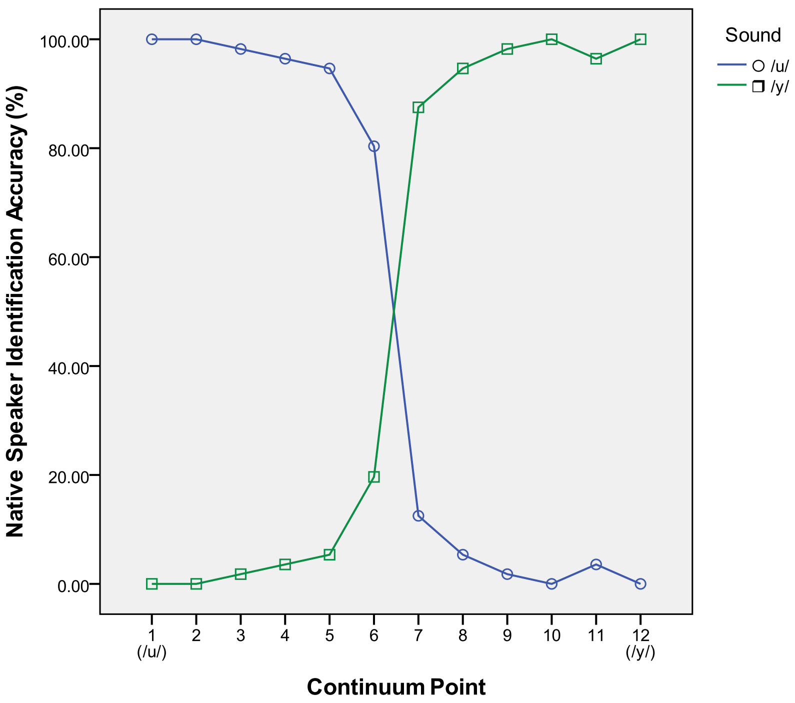


Figure 4.3: Nasal Contrast Identification Function for the Single Voice Task

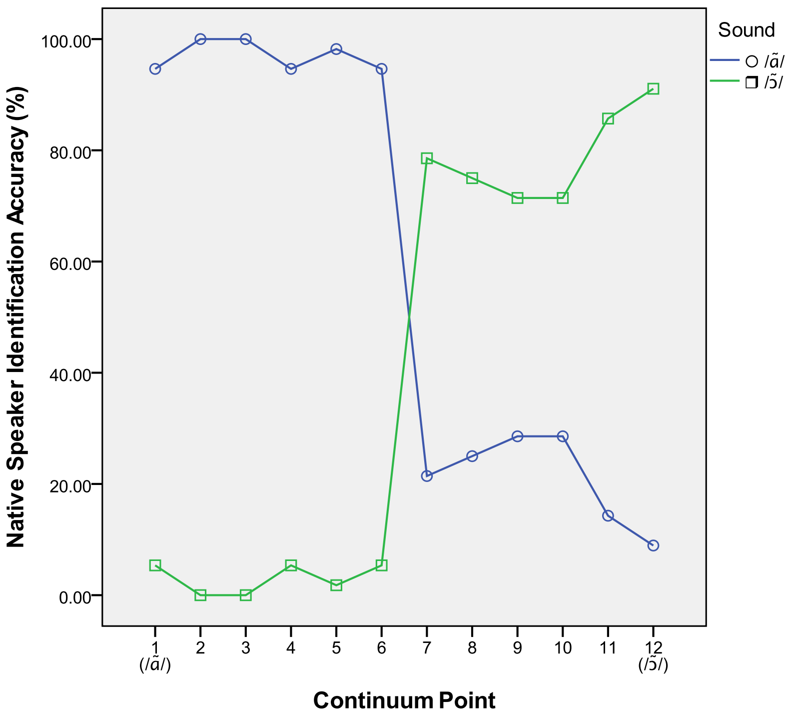


Figure 4.4: Oral Contrast Identification Function for the Six Voice Task

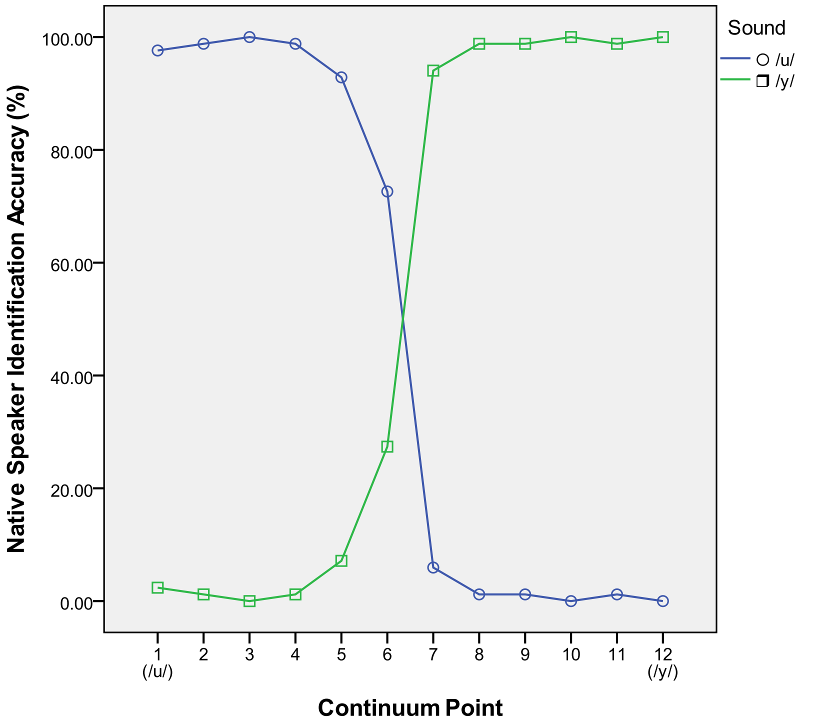
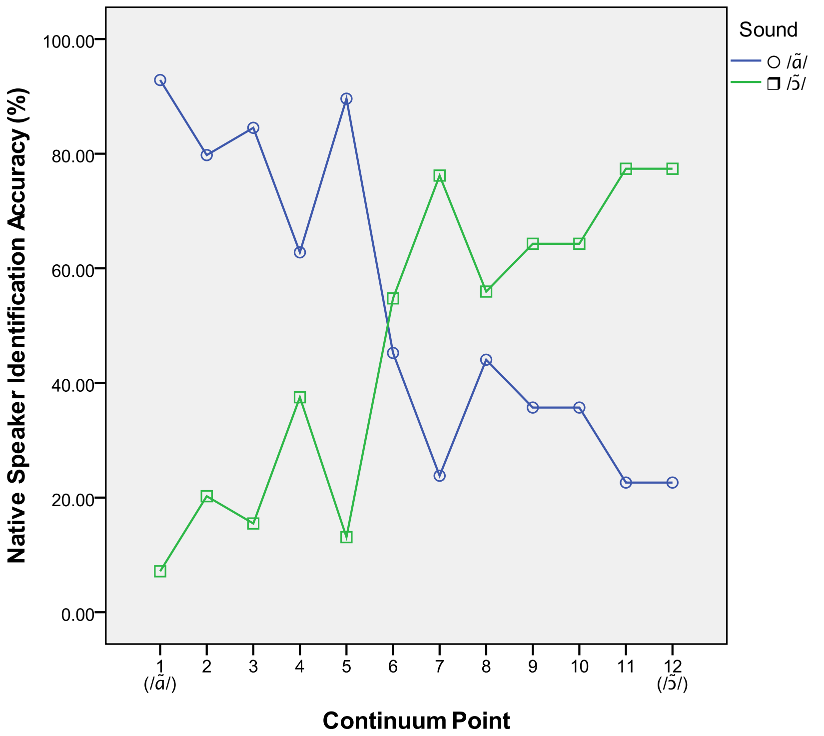


Figure 4.5: Nasal Contrast Identification Function for the Six Voice Task



A clue to where the problem lies is firstly to examine Figures 4.3 and 4.5 and along with how often /ɑ̃/ is being mistaken for /ɔ̃/and vice versa, these data are in Table 4.8. It appears that NS 1 has a tendency to hear /ɑ̃/ when /ɔ̃/ has been used and NSs 2&3 generally have the opposite tendency. This suggests that for some listeners some of the synthesised /ɔ̃/ sounds will sound too much like /ɑ̃/, and for others the synthesised /ɑ̃/ sounds will sound too much like/ɔ̃/. In other words, the same sound could be identified as either member of the contrast dependent on an individual listener's perception, and this was true even of fairly distant points on this continuum. However, whichever way the confusion lies for the NSs used here, it is unsurprisingly particularly marked at the closest points, 6, 7 and 8 on the continuum.

Table 4.8: Error Rates and Types from Identification of Perceptual Fading Stimuli by Native French Speakers

	Task 1: Single Voice			Task 2: Six Voices		
Speaker	**Total Error**	**/ɔ̃/ mistaken for /ɑ̃/**	**/ɑ̃/ mistaken for /ɔ̃/**	**Total Error**	**/ɔ̃/ mistaken for /ɑ̃/**	**/ɑ̃/ mistaken for /ɔ̃/**
NS1	56/336	55/56	1/56	128/336	122/128	6/128
NS2	26/336	18/26	8/26	79/336	22/79	57/79
NS3				75/336	10/75	65/75

4.3.6 Discussion and Conclusions

Examination of the /u/-/y/ continuum results shows that there are no problems with the chosen values and changeover points. It was therefore decided to take these values forward for use in the training study.

As there is a lack of consistency about the direction in which the confusion lies with the /ɑ̃/-/ɔ̃/ continuum, the only solution appears to be to move the sounds further apart in both directions, by creating a larger difference between the mid changeover points and also the extremes at either end. However, given that the native listeners had differing tendencies as to which member of the contrast they were hearing, the sounds cannot be moved apart too much in either direction to favour solving one problem as it would exacerbate the other. It was therefore decided that moving the continuum further apart by 2 'points' in both directions by removing the four middle values (positions 5-8 and adding two points with extrapolated F1 and F2 values to the top and bottom of the continuum) would achieve the best balance. This would theoretically change the /ɑ̃/-/ɔ̃/ error rate as the errors from positions 5, 6, 7 and 8 on the continuum could be discounted, as these stimuli would no longer exist (the new positions 5, 6, 7 and 8 being the former 3, 4, 9 and 10). The potential error and accuracy rate change is shown in Table 4.9, the potential new identification functions are shown in Figures 4.6 and 4.7 and the resultant proposed new values for the /ɑ̃/-/ɔ̃/ continuum to be used in the training study are shown in Table 4.10 (it should be noted that the error rates and identification functions have been calculated as if no errors were made at the new end points on the continuum for representative purposes only, whereas it is likely that some errors would still occur).

Table 4.9: Theoretical /ɑ̃/-/ɔ̃/ Contrast Error Rates and Types from Identification of Perceptual Fading Stimuli by Native French Speakers after Proposed Continuum Point Changes

	Task 1: Single Voice		Task 2: Six Voices	
Speaker	**Total Error (Old)**	**Total Error (New)**	**Total Error (Old)**	**Total Error (New)**
NS1	56/336	36/336 (89.28% correct)	128/336	82/336 (75.6% correct)
NS2	26/336	10/336 (97.02% correct)	79/336	44/336 (86.9% correct)
NS3			75/336	40/336 (88.1% correct)

Figure 4.6: Revised Nasal Contrast Identification Function for the Single Voice Task

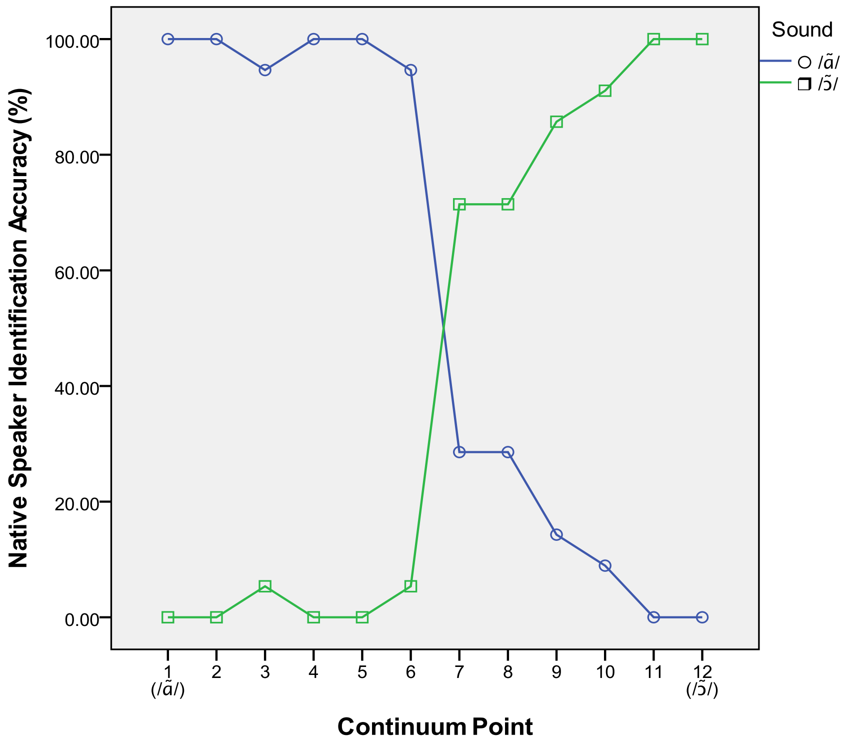


Figure 4.7: Revised Nasal Contrast Identification Function for the Six Voice Task

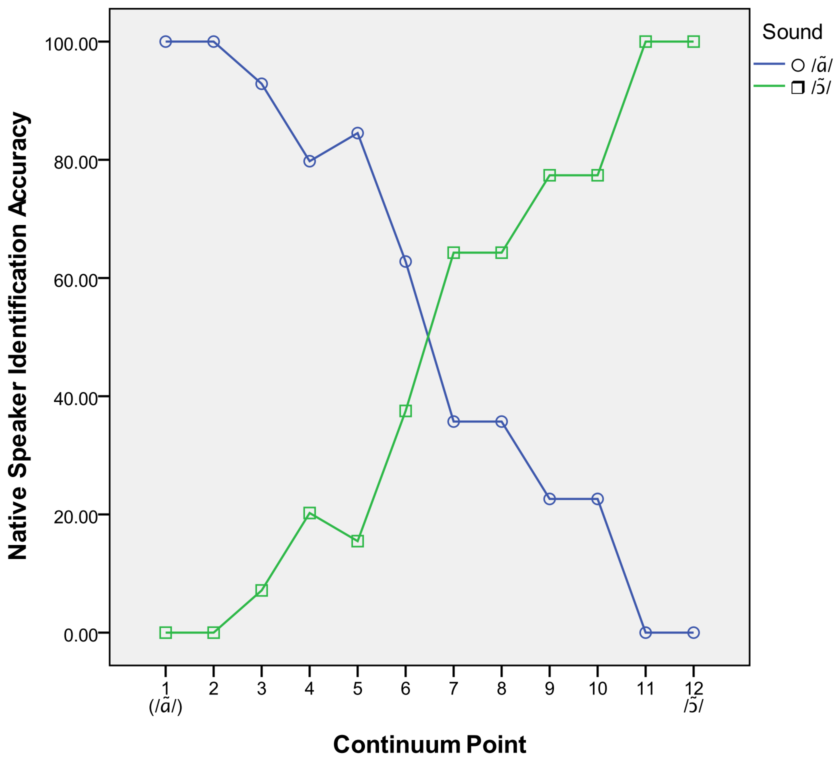


Table 4.10: Revised formant and nasal values for the /ɑ̃/-/ɔ̃/continuum

	1 /ɑ̃/	2	3	4	5	6 (intended /ɑ̃/)	7 intended /ɔ̃/)	8	9	10	11	12 /ɔ̃/
F1	610	600	590	580	570	560	480	470	460	450	440	430
F2	804	797	790	783	776	769	711	704	697	690	683	676
F3	2870	2870	2870	2870	2870	2870	2870	2870	2870	2870	2870	2870
FNP	180	180	180	180	180	180	180	180	180	180	180	180
FNZ	385	385	385	385	385	385	385	385	385	385	385	385

Whilst the six speaker nasal contrast continuum continues to show some ambiguities it was felt that the continuum points could not be moved further apart because, as noted above, there was a lack of consistency within the native speaker participants as to the direction of the confusion. In addition the proposed shift moves the values suggested by Bognar and Fujisaki (1986) towards the mid points for each sound and having more or less exaggerated versions at either side, which is more ideal for the perceptual fading task

Having established the difficult French vowel contrasts, training task and perceptual fading continuum points, the first main training study is described in the following chapter, Chapter 5.

5 Experiment 1: Perceptual Training and Effect on Perception Accuracy - Fading vs. HVPT

5.1 Introduction

This chapter describes the perceptual training study which compares the effects of Perceptual Fading (PF) training and High Variability Phonetic Training (HVPT) on perceptual accuracy. The vowel contrasts to be trained are the oral /u/-/y/ contrast and the nasal /ɑ̃/-/ɔ̃/contrast as these have been identified by the pilot study described in Chapter 4 (4.1) as problematic for native English speaking learners of French. These training techniques have rarely been compared and these difficult French vowel contrasts have rarely been used in training studies.

The main purpose of this chapter is to answer the first research question:

Q1 Is the HVPT technique more successful than the perceptual fading technique (or vice versa) in terms of producing a generalisable, long-term improvement in perception?

In answering this question further supplementary questions will also be answered:

1. Will perceptual training result in an improvement in perceptual accuracy in terms of improved identification scores after training, for both the oral and nasal vowel contrast to be trained?
2. Will this improvement generalise to new words and new voices, for both the oral and nasal vowel contrasts?
3. Will this improvement be retained at least one month after training, and for both the oral and nasal vowel contrasts?

As previously noted (see 2.2.5), one of the few studies which has sought to make this comparison of perceptual training techniques is Iverson et al (2005). The authors found that both methods were successful but neither was significantly more effective than the other. However, as the authors themselves noted, long term retention of training was not examined (nor was its effects on production, see Chapter 6) as it will be here.

Furthermore, again as previously noted, PF was used with multiple speakers and not in its classical single speaker form, possibly acting as an improvement to this technique. As a result there is a further supplementary question to be answered in carrying out this experiment:

4. Are each of these methods more or less successful in their 'classic' form (PF with one speaker; HVPT with multiple speakers) or their alternative form (PF with multiple speakers as used by Iverson et al (2005); HVPT with a single speaker for comparison purposes).

With the overall aim of attempting to identify an optimal training technique, the most successful perceptual training technique(s) identified by this study will then be compared to, and used along with, pronunciation training in the pronunciation training study described in Chapter 6.

5.2 Participants

A total of 55 (45 female, 10 male) first and second year students of French were recruited from the Universities of Edinburgh and Glasgow through emailing the French class lists. First and second year students of French were recruited as they had not yet completed their compulsory year in a French speaking country, and were therefore more likely to have more perceptual difficulties with the vowel contrasts to be examined due to this lack of experience. None of the participants had extensive experience with the French language outside of their studies and no participant reported any hearing difficulties. All participants were native speakers of English, and were either Scottish (34), English (15) or Irish (6). They had been learning French for 1-13 years with an average of 7.6 years of learning. The age range of the participants was 18-52 with a mean age of 20. However, there were two mature students over 40 and the rest of the participants were aged 18-26. The participants were paid for their time and were entered into a voucher prize draw, as it was unlikely that there would be a large enough number who would participate in a long term study without financial incentive.

The participants were randomly assigned to four groups: Group 1: Multiple Speaker HVPT (14 participants); Group 2: Multiple Speaker PF (14 participants); Group 3:

Single Speaker HVPT (14 participants); Group 4: Single Speaker PF (13 participants). It was decided not to have a control group for this study it is well established that these methods work. Instead, the purpose of this study was to establish which of the methods was most successful.

All 55 participants completed a language history questionnaire (Li, Sepanski & Zhao, 2006) and a Pronunciation Attitude Inventory (PAI) questionnaire (Elliott (1995), see Chapter 7 and Appendices G and H). The participants then undertook a perceptual identification pre-test and also a production (see Chapter 6) pre-test. Immediately after completing six training sessions (a maximum of twice per week), the participants took perception and production post-tests and generalisation tests. After a minimum of one month participants then returned to carry out retention testing. Twelve participants did not return due to lack of interest or scheduling difficulties, resulting in group numbers as follows: Group 1: Multiple Speaker HVPT (10/14 participants); Group 2: Multiple Speaker PF (12/14 participants); Group 3: Single Speaker HVPT (11/14 participants); Group 4: Single Speaker PF (10/13 participants).

5.3 Stimuli

This study made use of natural stimuli in training and testing and natural stimuli with synthesised vowels for training in the PF conditions only. A full list of words used can be found in Appendix I. All stimulus preparation was based upon the same set of recordings as detailed below.

Fourteen native speakers of French (7 male, 7 female) were used to provide the natural stimuli. All speakers were either born and raised in France or born in another French speaking country and moved to France when young. Although from differing areas of France, none self-reported, nor were heard by the French speaking experimenter, as having a strong regional accent. Recruitment was via email from the French student population attending Edinburgh University through the Erasmus scheme.

Fifty minimal pairs for each of the contrasts to be trained were identified (for example, *boule-bulle* and *angle-ongle)*. In addition, at the time of recording the stimuli, it was possible that a third contrast, the nasal /ɑ̃/-/ɛ̃/ contrast was going to be trained, therefore fifty minimal pairs for this contrast were identified also (for example *banc-bain)*. This resulted in a total of 300 words to be read, and as each was presented three times in case of reading error, each native speaker read 900 words in total. (It was later decided that the training the oral /u/-/y/ contrast and the nasal /ɑ̃/-/ɔ̃/contrast, the two more difficult contrasts identified by the pilot study (see 4.1), would be sufficient for this study, particularly given the lengthy time it was to take to prepare the synthesised stimuli). The speakers were recorded reading the words in a soundproofed recording studio isolation booth at Edinburgh University. The sound was captured using an AKG CK98 hypercardoid microphone and encoded on a custom designed PC based on a Shuttle XPC chassis with a Core2Duo processor running Sonar software via a MOTU 828 audio interface. The sound was digitised at a sampling rate of 48 kHz with a resolution of 32 bits. The words were presented one at a time on a monitor within the isolation booth using EPrime software, and were presented in a random order. The speakers were instructed to read in a neutral tone and moved to the next word to be read by using the spacebar on the keyboard attached to the monitor. Presenting the words individually on a monitor in this way prevented list intonation in the readings.

The first reading of the word was selected for use, unless there was a mistake in the reading, in which case the second reading was used. The third reading was used when the first and second readings were incorrect. Using Audacity software, the target words were separated from the recorded list, normalised, and saved as individual files for presentation during training and testing. Four native speakers of French (2 female, 2 male) screened the minimal pairs to be used in perceptual testing using a two-alternative forced choice identification task. Only one of the four native speakers made any mistakes in identification (3 errors from 56 trials) in the pre/post test. A check of these stimuli revealed no problems and they were therefore retained for use in testing. In the two tests of generalisation the native speakers made between 0 and 3 errors out of 56 trials, with no common confusions found.

Recordings from six of the speakers (3 female, 3 male) were randomly selected to be used in the multiple speaker training sessions. The first speaker (male) of this group was also used in the single speaker training sessions. This male speaker was also used for the first test of generalisation (new words, familiar speaker). The recordings of four further speakers (2 female, 2 male), were randomly selected to be used for the pre- and post-tests. Finally, the remaining four speakers (2 female, 2 male) were used for the second test of generalisation (new words, new speakers).

The stimuli for the PF conditions were the natural tokens as recorded above, with the natural vowel of interest removed and a synthetic vowel (matched for length and F0) added. The purpose of adding a synthetic vowel was to achieve the 'fading' effect by manipulation of the formant values. As there were to be six training sessions, 12 points on the /u/-/y/ and 12 points on the /ɑ̃/-/ɔ̃/ continua were required. Chapter 4 (4.3.2) describes how the pilot study with the synthesised data arrived at optimal formant values to create these 12 points and details the procedure for creation of the synthetic stimuli.

Due to the difficulty in finding sufficient minimal pairs to use for testing and training (pairs were to be required for perceptual pre- and post- testing, pronunciation pre- and post- testing and tests of generalisation), it was impossible to fully control for such factors as vowel position and word frequency. However, within each contrast, once the list of words to be used was produced, it was then attempted to distribute number of syllables and vowel position equally across the pre/post-test, training, generalisation test 1 and generalisation test 2 lists (see below for further details).

Word frequency was more difficult to distribute evenly once the length and vowel position considerations had been made and also because one member of a minimal pair often occurred more frequently than the other. However, individual word frequencies (where available) were noted (source: www.lexique.org) and can be found next to the stimuli used in Appendix I. Table 5.1 shows the averaged log transformed word frequencies for each test stimulus in occurrences per million. Log transformations were carried out for comparison purposes because 8 words (2 from the oral contrast Pre/Post pronunciation test list, 3 from the oral contrast training list, 2 from the nasal contrast training list and 1 from the oral contrast pronunciation generalisation test list) had particularly high frequencies of over 1000 occurrences per million and were falsely skewing the means. As can be seen, the frequencies happen to be reasonably well matched, however, the mean oral contrast frequency is clearly higher than the nasal contrast frequency for the pre/post pronunciation test, the mean nasal contrast frequency is clearly higher than the oral contrast for the perceptual generalisation test 2 and the mean nasal contrast frequency is clearly higher than the oral contrast for the pronunciation generalisation test.

Table 5.1: Log Transformed Word Frequency Descriptive Statistics (Occurrences per Million) for Each Test

Test	N	Min.	Max.	Mean	SD
NasalPrePostPerc	14	-.46	1.95	.98	.78
OralPrePostPerc	14	-.80	2.23	.83	.90
NasalPrePostPronun	14	-1.00	1.74	.60	.92
OralPrePostPronun	14	-.80	3.85	1.64	1.26
NasalTraining	28	-.05	4.03	1.37	.94
OralTraining	28	-.54	3.62	1.40	1.15
NasalGenPerc	14	-1.52	1.20	.44	.82
OralGenPerc	14	-1.00	1.47	.34	.69
NasalGen2Perc	14	-.54	1.79	1.00	.60
OralGen2Perc	14	-.59	2.00	.67	.80
NasalGenProd	14	.40	2.66	1.21	.67
OralGenProd	14	-.32	1.95	.87	1.07

It was furthermore attempted to make the words as orthographically transparent as possible, particularly in the pronunciation tests (see Chapter 6) where it was necessary to read words aloud. However, it was likely that some words would be unfamiliar to participants and a spelling guide was therefore provided which gave instructions on how to match up the spelling of any unfamiliar words with the sounds of interest (see Appendix E for the full text of this guide).

The majority of words used for training and testing the oral /u/-/y/ contrast were either monosyllabic (68/98), e.g. *bout-bu*, or disyllabic (28/98), e.g. *dessous-dessus,* (excepting one minimal pair which had words with three syllables (2/98), *écoulé-éculé*). The vowels of interest in the monosyllabic words were either in a CVC (38/98), e.g. *four,* CV (26/98), e.g. *bout,* or CCVC (4/98), e.g. *broute,* context. In the disyllabic words they were in a CVCV context with the vowel of interest as either the first (20/98), e.g. *bouter*, or second vowel (2/98), e.g. *dessous*, or a VCV (8/98), e.g. *écrou,* context with the vowel of interest as the second vowel.

The words used for training and testing the nasal /ɑ̃/-/ɔ̃/ contrast were either monosyllabic (16/98), e.g. *blanc-blond*, disyllabic (58/98), e.g. *allant-allons*, or trisyllabic (24/98), e.g. *achetant-achetons*. Minimal pairs for this contrast were easier to identify as French verbs have both a present participle ending in *-ant* (/ɑ̃/) and a first person plural form ending in *–ons* (/ɔ̃/). A majority of the minimal pairs did make exclusive use of these forms (28/49), and the vowel position of interest was therefore word final for these words. These stimuli were based upon the most frequently occurring French verbs as noted by Dudziak (2007). In addition, some minimal pairs had one member but not the other in this form (e.g. *massant-maçon*), or had one ambiguous member in that it could be representative of this form or be another word (e.g. *devant-devons, devant* is the present participle of the verb *devoir* (to have to) but is also a preposition meaning 'in front of') (9/49). Again, this resulted in a word final vowel position for the contrast of interest. However, it was also attempted to find as many minimal pairs as possible that did not follow this pattern (e.g. *angle-ongle*) in order that there would be vowel positions other than word final, and in order that the words did not resemble these conjugated verb forms (12/49). These twelve pairs had mono-, di- and trisyllabic words and featured the vowel contrast of interest in word initial, word medial and word final positions. Again, it was attempted to evenly distribute these minimal pairs which were different in form across all the training and testing lists.

5.4 Procedure

5.4.1 Pre-Testing

Prior to beginning the pre-tests, participants completed a language history questionnaire (Li et al, 2006), which elicited information about their native language(s) background and their L2 learning history (e.g. how long they had been learning their L2(s), in what environments and how they rated their abilities, see Appendix G). The participants then completed Elliott's (1995) Pronunciation Attitude Inventory which will be discussed in Chapter 7 (see Appendix H). The participants then undertook, in a random order, a production pre-test (to be discussed in Chapter 6) and the perceptual identification pre-test.

All testing made use of the naturally produced stimuli, as the training was intended to improve perception of the vowel contrasts as heard in everyday speech. The pre-test consisted of 56 experimental trials. The stimuli were produced by four native speakers of French (2 female, 2 male). Seven minimal pairs for each contrast were presented four times, and each member of the minimal pair was repeated twice, once by a male speaker and once by a female speaker. A list of the words used in the perceptual pre-test is given in Appendix I.

All testing and training took place in a sound-deadened booth with a Dell desktop PC in a perception laboratory at Edinburgh University or in a quiet room in Glasgow University library using a Macintosh MacBook laptop running boot camp Windows XP. The test task used the two alternative forced choice paradigm whereby the participant was presented with one member of the minimal pair on one side of a computer screen and one member on the other. At the same time, one of the native speaker word recordings was played and the participant was asked to press '1' on the computer keyboard if they believed that they had heard the word on the left of the screen and '2' if they believed that they had heard the word on the right of the screen. No feedback was given in the testing phase. The word presentations occurred in a random order and were counterbalanced in terms of whether the correct answer occurred on the left or right of the screen and the side of the screen on which each member of the minimal pair was presented. Participants were able to take as long as they needed to respond to each trial, the next word was played 500msec after they had responded.

Before testing commenced, instructions regarding the identification task were displayed on the computer screen. The experimenter then verified that the participant understood the task to be carried out. In order to avoid confusion as a result of unfamiliarity with a word, participants were provided with a 'Spelling Guide' sheet, explaining that despite other differences that may be seen onscreen, the words on the left and right of the screen would only differ by one vowel sound and then gave instructions on how to match up the spelling of the words with the sounds of interest (the full instructions and spelling guide text are in Appendix D and Appendix E respectively). This sheet was kept beside participants for reference throughout testing and training.

5.4.2 Training

Participants attended six training sessions, the first of which occurred immediately after the pre-tests had been completed. A maximum of two training sessions were carried out each week, and participants took between 4 and 6 weeks to complete the training. The training trials consisted of the same two alternative forced choice identification task as used in the pre-test, however in training the participant was provided with immediate feedback as to whether or not the response had been correct. Upon a correct response, three ascending tones were played and the word 'Correct!' was displayed on the computer screen. If the participant responded incorrectly, two descending tones were played, the words 'Incorrect – Listen Again! (The next trial will begin immediately afterwards)' were displayed, the misidentified word was replayed and then training moved on to the next word to be identified. Again, participants were able to take as long as they needed to respond to each trial, the next word was played (500msec) after they had responded.

Each training session had two blocks, one for each contrast, and the blocks were presented in a random order. Within each block there were 14 minimal pairs used, 28 words each repeated twice, resulting 56 training trials per block and 112 training trials in total for each session. At the end of each session there was a short identification test of ten items from each contrast to monitor improvement, using words randomly chosen from the training words. This test used natural speech from that day's speaker for all participants and resulted in a total of 132 trials per session. See Appendix I for the words used in training.

All participants heard the same male native speaker during their first training session. Participants being trained in the single speaker conditions heard this speaker on all subsequent training sessions. Those undertaking training in both multiple speaker conditions heard a different speaker for each training session (3 male and 3 female in total, alternating between male and female) and both multiple speaker conditions used the same speaker for each training session. Therefore the only difference between the multiple speaker conditions on any day of training was the treatment of the vowel contrast (natural vs. fading) to be trained. Participants undergoing PF training heard the vowels at the endpoints of the continuum on the first day of training and over the six days of training the differences between the vowels became less distinct as progressively closer points on the continuum were used.

Before training commenced, as with testing, instructions regarding the identification task were displayed on the computer screen. The experimenter then again verified that the participant understood the task to be carried out. Again, the participants were also provided with the 'Spelling Guide' sheet (see Appendix E) throughout training, again to avoid confusion due to unfamiliarity with a word.

5.4.3 Post-, Generalisation and Retention Testing

As soon as the sixth training session had been completed, the participants carried out the perception and production (see Chapter 6) post-tests, which were a repeat of the tests they carried out prior to training. In addition, they carried out a production generalisation test and two perceptual generalisation tests. The first perceptual generalisation test used new words read by a familiar speaker, the male speaker from the first training session. The second generalisation test used new words read by four new speakers (2 female, 2 male). As with the perceptual pre- and post-tests, both generalisation tests used the same two alternative forced choice procedure with no feedback. Again as with the pre-/post-tests, the generalisation tests consisted of 56 experimental trials; seven minimal pairs for each contrast presented twice (28 words) in a random order and counterbalanced in terms of correct answer location and contrast location. For the second test of generalisation each member of the minimal pair was spoken twice, once by one of the female speakers and once by one of the male speakers. Again, participants were able to take as long as they needed to respond to each trial, the next word was played as soon as they had responded (see Appendix I for a list of the words used in generalisation testing).

Participants then returned after a minimum of one month (and a maximum of two months) and carried out the three perception and two production (see Chapter 6) tests again to ascertain how well any training benefits were retained.

5.5 Results

The analysis was carried out upon the participant perceptual accuracy scores in the tests. It was firstly necessary to ensure that any differences between training groups after training were attributable to the training technique and not due to one group having a higher or lower pre-test score. To this end, two one-way ANOVAs were carried out with pre-test score as the dependent variable and training group as the between subjects factor, one ANOVA for each contrast of interest. There was no significant effect of training group for either contrast [Nasal: $F(3,51) = .377$, $p = .770$; Oral: $F(3,51) = .508$, $p = .678$] indicating that the pre-test scores were not significantly different for each group.

The results of the interim tests which were carried out at the end of each session were at ceiling on the first day, although the responses throughout training were not. That is, overall, participants performed very well on this test whilst performing far less accurately during training sessions. This meant that the data could not be analysed to monitor improvement over time in any meaningful way and the data were therefore not analysed.

The statistical tool R (R Development Core Team, 2011) with the R packages *lme4* (Bates, Maechler & Bolker, 2011) and *language*R (Baayen, 2011) was used to carry out binary logistic mixed effects analyses of the relationship between training groups, tests, contrast tested and time. This use of mixed effects models has a number of advantages. Firstly, it does not require averaging over responses made by a participant which could mask particular data trends (Drager, 2011). Secondly, it allows for inclusion of participants as random effects whereby each individual is assigned a coefficient and this coefficient is matched with all responses made by that participant. This reduces the likelihood that responses from one participant influence the results (Baayen, 2008). Finally, mixed effects models are less sensitive to missing data (as in the present work where some participants did not return to carry out the tests of retention) than other statistical methods such as MANOVA (Drager, 2011).

Analysis was carried out in three blocks. The first block examined the effects of training across the pre-test, post-test and delayed post-test. The second and third blocks examined the generalisation tests. In all cases, likelihood ratio tests comparing each model with fixed effects to a null model with only the random effects demonstrated that the fixed effects model differed significantly from the null model.

5.5.1 Block 1: Pre-Test, Post-test, Delay

The mean participant percentage perceptual identification accuracy and standard deviations for each test (pre-test, post-test, and delayed post-test according to participant training group and contrast are shown in Tables 5.2 and 5.3 and illustrated in Figures 5.1. and 5.2. These data suggest that the multi speaker HVPT and single speaker PF training may result in better results than other training types.

Table 5.2: Nasal Contrast: Mean Percentage Perceptual Identification Accuracy Scores According to Training Group

Test	Training Group	Mean	SD	N
Pre-Test	Multiple Speaker HVPT	59.69	8.80	14
	Multiple Speaker PF	58.68	10.46	14
	Single Speaker HVPT	62.76	9.57	14
	Single Speaker PF	60.44	13.16	13
Post-Test	Multiple Speaker HVPT	68.37	16.03	14
	Multiple Speaker PF	62.76	13.49	14
	Single Speaker HVPT	62.76	10.73	14
	Single Speaker PF	66.21	14.99	13
Delayed Post-Test	Multiple Speaker HVPT	67.50	14.33	10
	Multiple Speaker PF	58.12	8.91	11
	Single Speaker HVPT	58.63	8.27	12
	Single Speaker PF	68.93	15.07	10

Table 5.3: Oral Contrast: Mean Percentage Perceptual Identification Accuracy Scores According to Training Group

Test	Training Group	Mean	SD	N
Pre-Test	Multiple Speaker HVPT	68.39	15.68	14
	Multiple Speaker PF	64.03	13.25	14
	Single Speaker HVPT	63.78	6.40	14
	Single Speaker PF	67.03	9.91	13
Post-Test	Multiple Speaker HVPT	74.24	14.23	14
	Multiple Speaker PF	71.17	13.97	14
	Single Speaker HVPT	66.33	10.64	14
	Single Speaker PF	74.72	13.87	13
Delayed Post-Test	Multiple Speaker HVPT	76.43	16.51	10
	Multiple Speaker PF	63.31	16.30	11
	Single Speaker HVPT	61.91	5.56	12
	Single Speaker PF	79.64	14.29	10

Figure 5.1: Nasal Contrast: Mean Percentage Perceptual Identification Accuracy Scores According to Training Group

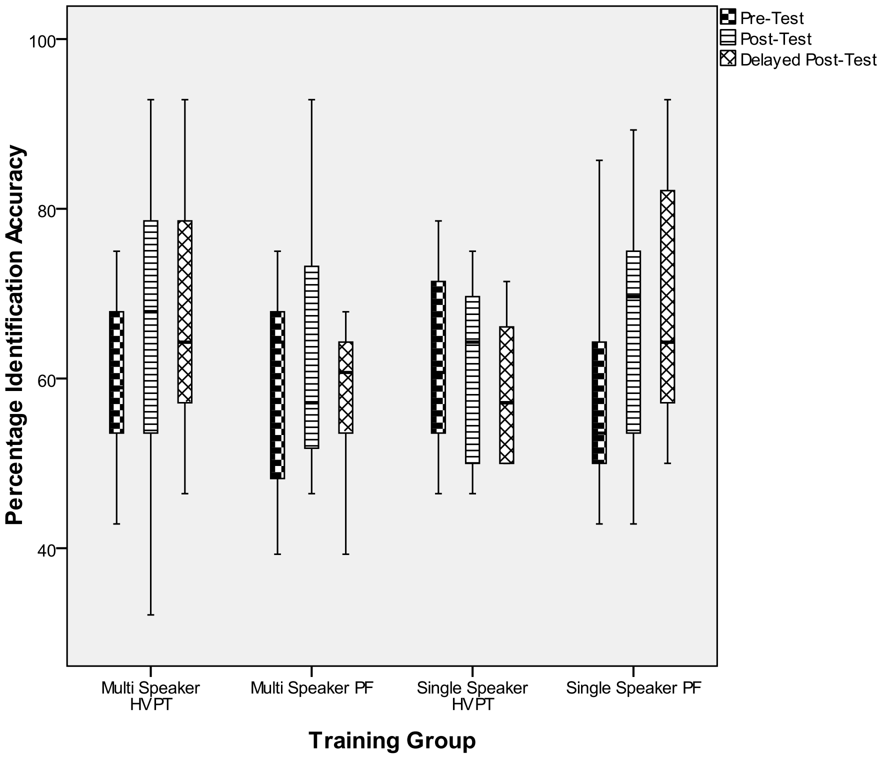
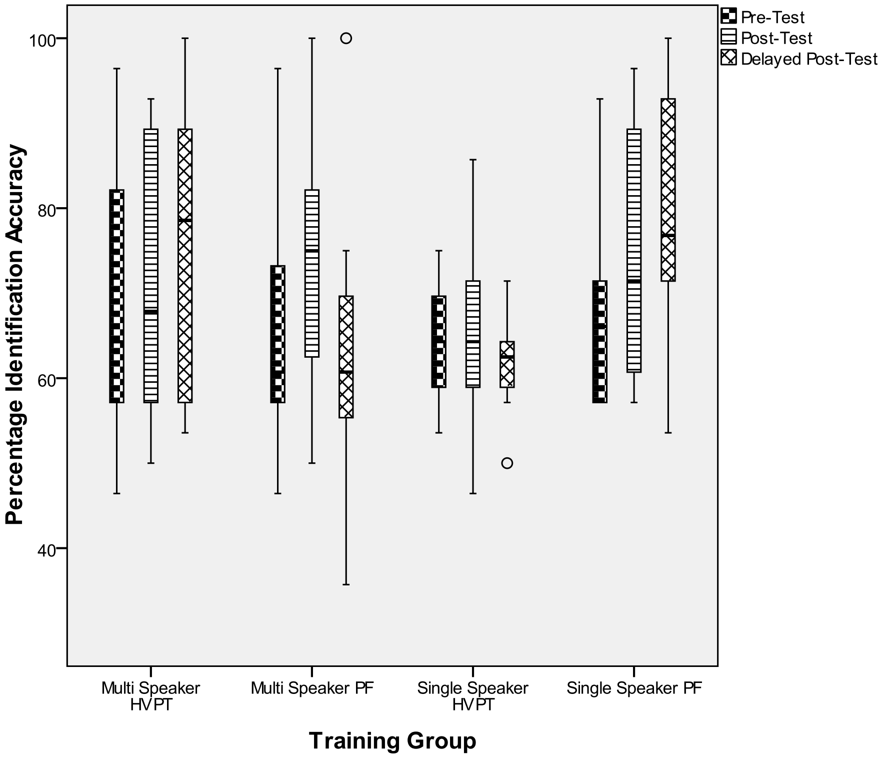


Figure 5.2: Oral Contrast: Mean Percentage Perceptual Identification Accuracy Scores According to Training Group



These data were then analysed using binary logistic mixed effects analysis in order to determine the significance of training group, test, contrast and time on whether or not participants answered correctly. In the present model the participant and test item were included as random effects. The dependent variable was whether the participant correctly identified the minimal pair member presented. The fixed effects or potential predictors tested were participant training group/method (SubGp), test (TestType, the three levels of this – pre, post, delay, also represented time) and Contrast (oral vs. nasal). The model is detailed in Table 5.4.

Table 5.4: Fixed Effects and Coefficients of the Pre-Test, Post-Test, Delay Model

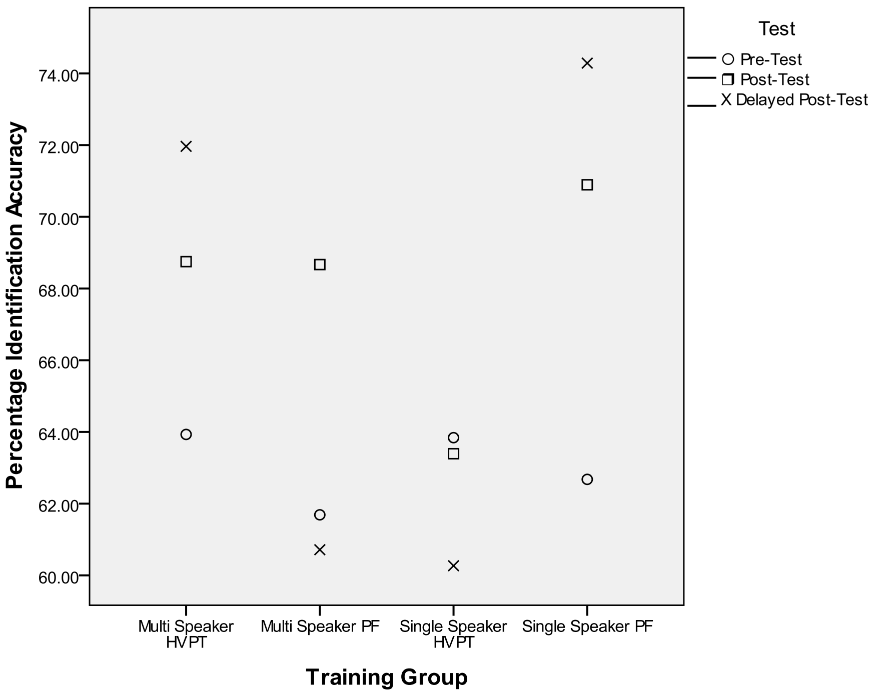
	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.464	0.165	3.851	0.0001****
SubGpSingle PF	-0.014	0.196	-0.070	0.9440
SubGpMulti PF	-0.130	0.192	-0.680	0.4964
SubGpSingle HVPT	-0.053	0.192	-0.277	0.7814
TestTypePost	0.371	0.113	3.281	0.0010***
TestTypeDelay	0.445	0.129	3.454	0.0005****
SubGpSingle PF:TestTypePost	-0.014	0.163	-0.085	0.9324
SubGpMulti PF :TestTypePost	-0.117	0.157	-0.745	0.4561
SubGpSingle HVPT :TestTypePost	-0.323	0.156	-2.066	0.0388**
SubGpSingle PF:TestTypeDelay	0.126	0.185	0.682	0.4954
SubGpMulti PF :TestTypeDelay	-0.502	0.175	-2.872	0.0041***
SubGpSingle HVPT :TestTypeDelay	-0.576	0.172	-3.361	0.0008****

***p<.1; **p<.05; ***p<.01, ****p<.001**

The figures above represent the models predictions, in log odds, of a factor's effect while holding other fixed effects in the model constant (Drager, 2011, p.111). The reference level for participant training group (SubGp) was the multiple speaker HVPT group and the reference level for test (TestType) was the pre-test. Contrast was not included in this final model as it only approached significance when included as a main effect (p = 0.08) and adding an interaction term resulted in no significant interactions. This means that the conclusions drawn from interpretation of the other interactions in this model hold for both contrasts and the data were collapsed across contrasts.

Looking firstly at the main effects, no group scored significantly differently from the multiple speaker HVPT group overall, however the significant interactions show that the groups do differ dependent upon test (see below). Examining the main effect of test, overall, post-test and delayed post-test scores were significantly greater than the pre-test scores. This implies that training worked and was retained over time. The post-test interactions suggest that the only group which scored differently from the multiple speaker HVPT group dependent upon whether the test was the pre-test or the post-test is the single speaker HVPT group. Similarly, looking at the delayed post-test scores, both the multiple speaker PF and single speaker HVPT groups scored significantly differently from the multiple speaker HVPT group dependent upon whether the test was the pre-test or the delayed post-test. An inspection of Figure 5.3 demonstrates the exact nature of the interactions.

Figure 5.3: The Interaction Between Participant Training Group and Test



Firstly, with reference to the post-test interactions it can be seen that whilst the single speaker HVPT group does not score significantly differently to the multiple speaker HVPT group at pre-test, the single speaker HVPT group scores less at post-test. With reference to the delayed post-test interactions, it can be seen that both the single speaker HVPT group and the multiple speaker PF group score less at delayed post-test than the multiple speaker HVPT group. At no time does the single speaker PF group score significantly differently than the multiple speaker HVPT group.

In sum, it can be seen that the participants in the multiple speaker HVPT group outperform the participants on the single speaker HVPT group at post-test and outperform the participants in both the single speaker HVPT and multiple speaker PF groups in the delayed post-test. The multiple speaker HVPT group and single speaker PF group do not perform differently at any time. Therefore, the single speaker PF and multiple speaker HVPT training methods appear to work better than the others, particularly in terms of retaining training effects.

5.5.2 Generalisation Test 1

The mean participant percentage perceptual identification accuracy and standard deviations for the first test of generalisation with new words but a familiar voice (from day 1 of training) are shown in Tables 5.5 and 5.6 and illustrated in Figures 5.4 and 5.5. The original pre-test scores are also included for reference. These data, as with the pre/post/delay data, suggest that the multi speaker HVPT and single speaker PF training may result in better results than other training types.

Table 5.5: Nasal Contrast Generalisation Test 1: Mean Percentage Perceptual Identification Accuracy Scores According to Training Group

Test	Training Group	Mean	SD	N
Pre-Test	Multiple Speaker HVPT	59.69	8.80	14
	Multiple Speaker PF	58.68	10.46	14
	Single Speaker HVPT	62.76	9.57	14
	Single Speaker PF	60.44	13.16	13
Gen Test 1 (Familiar voice, new words)	Multiple Speaker HVPT	67.60	15.31	14
	Multiple Speaker PF	63.01	13.48	14
	Single Speaker HVPT	64.03	13.69	14
	Single Speaker PF	68.96	14.24	13
Gen Test 1 (Delayed) (Familiar voice, new words)	Multiple Speaker HVPT	67.86	15.79	10
	Multiple Speaker PF	54.87	7.70	11
	Single Speaker HVPT	56.85	9.57	12
	Single Speaker PF	67.50	14.53	10

Table 5.6: Oral Contrast Generalisation Test 1: Mean Percentage Perceptual Identification Accuracy Scores According to Training Group

Test	Training Group	Mean	SD	N
Pre-Test	Multiple Speaker HVPT	68.39	15.68	14
	Multiple Speaker PF	64.03	13.25	14
	Single Speaker HVPT	63.78	6.40	14
	Single Speaker PF	67.03	9.91	13
Gen Test 1 (Familiar voice, new words)	Multiple Speaker HVPT	82.14	17.55	14
	Multiple Speaker PF	73.22	18.20	14
	Single Speaker HVPT	67.86	11.97	14
	Single Speaker PF	83.52	14.55	13
Gen Test 1 (Delayed) (Familiar voice, new words)	Multiple Speaker HVPT	77.50	18.60	10
	Multiple Speaker PF	60.71	16.83	11
	Single Speaker HVPT	61.01	6.36	12
	Single Speaker PF	80.71	9.70	10

Figure 5.4: Nasal Contrast Generalisation Test 1: Mean Percentage Perceptual Identification Accuracy Scores According to Training Group

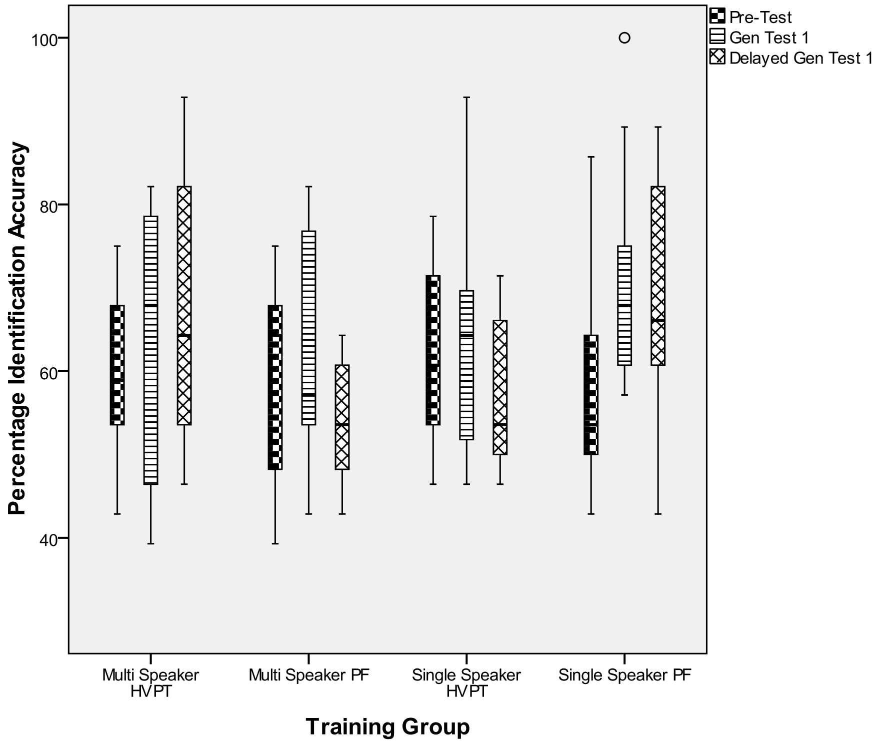
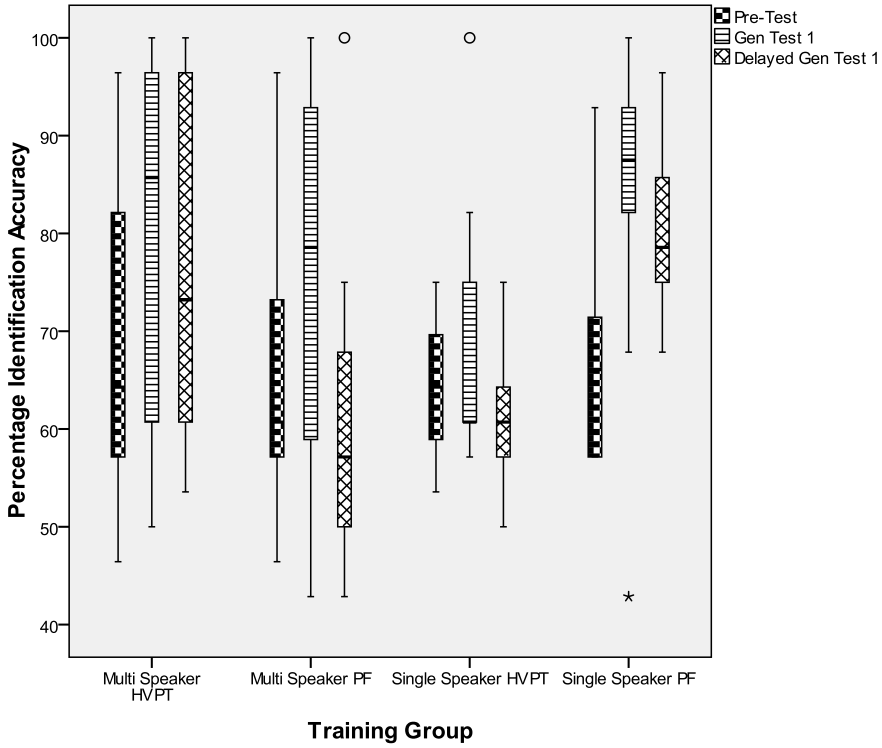


Figure 5.5: Oral Contrast Generalisation Test 1: Mean Percentage Perceptual Identification Accuracy Scores According to Training Group



As with the pre/post/delay data these data were then analysed using binary logistic mixed effects analysis in order to determine the significance of training group, test, contrast and time on whether or not participants answered correctly. The participant and test item were again included as random effects. The dependent variable was whether the participant correctly identified the minimal pair member presented. The fixed effects or potential predictors tested were participant training group/method (SubGp), test (TestType, the three levels of this – pre, gen1, gen1T2, also represented time) and Contrast (oral vs. nasal). The model is detailed in Table 5.7.

Table 5.7: Fixed Effects and Coefficients of the Generalisation Test 1 Model

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.486	0.192	2.535	0.0112**
SubGpSingle PF Only	-0.025	0.212	-0.118	0.9058
SubGpMulti PF Only	-0.141	0.208	-0.679	0.4973
SubGpSingle HVPT Only	-0.062	0.208	-0.300	0.7640
TestTypeGen1	0.547	0.182	3.007	0.0026***
TestTypeGen1T2	0.497	0.192	2.597	0.0094***
ContrastOral	0.372	0.148	2.514	0.0119**
SubGpSingle PF:TestTypeGen1	0.024	0.168	0.140	0.8884
SubGpMulti PF:TestTypeGen1	-0.257	0.161	-1.599	0.1098
SubGpSingle HVPT:TestTypeGen1	-0.494	0.160	-3.097	0.0020***
SubGpSingle PF:TestTypeGen1T2	-0.004	0.186	-0.024	0.9810
SubGpMulti PF:TestTypeGen1T2	-0.713	0.177	-4.039	<0.00001****
SubGpSingle HVPT:TestTypeGen1T2	-0.724	0.174	-4.170	<0.00001****

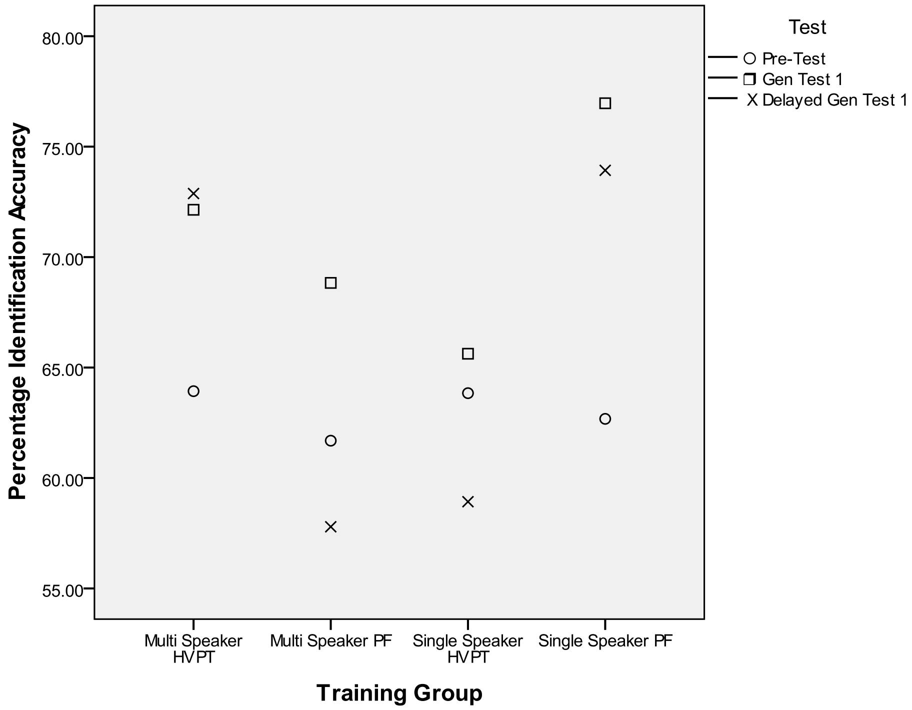
***p<.1; **p<.05; ***p<.01, ****p<.001**

As previously, the reference level for participant training group (SubGp) was the multiple speaker HVPT group, the reference level for Contrast was the nasal contrast, and the reference level for test (TestType) was the pre-test. Looking firstly at the main effects, no group scored significantly differently from the multiple speaker HVPT group overall, however the significant interactions show that the groups do differ dependent upon test (see below). The main effect of contrast shows that oral contrast test stimuli were answered correctly significantly more than nasal contrast test stimuli overall. However, there are no interaction terms with Contrast in this model as adding the interaction term resulted in no significant interactions. This means that the conclusions drawn from interpretation of the other interactions in this model hold for both contrasts, and the data were thus collapsed across contrasts. Finally, the main effect of test suggests that the overall generalisation test 1 and delayed generalisation test 1 scores were significantly greater than the pre-test scores. This implies that training generalised and this generalisation was retained over time.

This interpretation of the main effect of test should be made with caution as it is conceivable that had this generalisation test been administered before training it may have resulted in a significantly larger score than the actual pre-test (perhaps due to being easier) at that point and it is therefore difficult to directly compare the post-training generalisation scores to the original pre-test score and conclude that training has generalised if the scores are greater. However, if the training groups behave differently in these tests than in the pre-test as shown by significant interactions then this is likely to capture some generalisation of training for particular training groups. For example if some groups score greater than and some groups score less than the pre-test in the generalisation test after training, the random assignment of participants to groups mean this is likely to be capturing generalisation, as it is unlikely that those who may have scored higher or lower in any pre-training generalisation test than pre-test are all then assigned to the same training group.

The post training interactions suggest that the only group which scored differently from the multiple speaker HVPT group dependent upon whether the test was the pre-test or generalisation test 1 is the single speaker HVPT group. Similarly, looking at the delayed generalisation Test 1 scores, both the multiple speaker PF and single speaker HVPT groups scored significantly differently from the multiple speaker HVPT group dependent upon whether the test was the pre-test or the delayed generalisation test 1. Examination of Figure 5.6 demonstrates the exact nature of the interactions.

Figure 5.6: The Interaction Between Participant Training Group and Test



Firstly, with reference to the generalisation test 1 interactions, it can be seen that whilst the single speaker HVPT group does not score significantly differently from the multiple speaker HVPT group at pre-test, the single speaker HVPT group scores less in generalisation test 1. With reference to the delayed generalisation test 1 interactions, it can be seen that both the single speaker HVPT group and the multiple speaker PF group score less in the delayed generalisation test 1 than the multiple speaker HVPT group. At no time does the single speaker PF group score significantly differently than the multiple speaker HVPT group.

In sum, it can be seen that the participants in the multiple speaker HVPT group outperform the participants on the single speaker HVPT group in generalisation test 1 and outperform the participants in both the single speaker HVPT and multiple speaker PF groups in the delayed generalisation test 1. The multiple speaker HVPT group and single speaker PF group do not perform differently at any time. Therefore, the single speaker PF and multiple speaker HVPT training methods appear to work better than the others, particularly in terms of retaining generalisation effects.

5.5.3 Generalisation Test 2

The mean participant percentage perceptual identification accuracy and standard deviations for the second test of generalisation with new words and new voices are shown in Tables 5.8 and 5.9 and illustrated in Figures 5.7 and 5.8. The original pre-test scores are again included for reference. These data, as with the previous data analysed, suggest that the multi speaker HVPT and single speaker PF training may result in better results than other training types.

Table 5.8: Nasal Contrast Generalisation Test 2: Mean Percentage Perceptual Identification Accuracy Scores According to Training Group

Test	Training Group	Mean	SD	N
Pre-Test	Multiple Speaker HVPT	59.69	8.80	14
	Multiple Speaker PF	58.68	10.46	14
	Single Speaker HVPT	62.76	9.57	14
	Single Speaker PF	60.44	13.16	13
Gen Test 2 (New voices, new words)	Multiple Speaker HVPT	66.58	12.42	14
	Multiple Speaker PF	61.23	13.64	14
	Single Speaker HVPT	60.71	12.91	14
	Single Speaker PF	67.03	14.66	13
Gen Test 2 (Delayed) (New voices, new words)	Multiple Speaker HVPT	66.07	12.17	10
	Multiple Speaker PF	54.87	7.87	11
	Single Speaker HVPT	51.79	11.75	12
	Single Speaker PF	64.64	12.98	10

Table 5.9: Oral Contrast Generalisation Test 2: Mean Percentage Perceptual Identification Accuracy Scores According to Training Group

Test	Training Group	Mean	SD	N
Pre-Test	Multiple Speaker HVPT	68.39	15.68	14
	Multiple Speaker PF	64.03	13.25	14
	Single Speaker HVPT	63.78	6.40	14
	Single Speaker PF	67.03	9.91	13
Gen Test 2 (New voices, new words)	Multiple Speaker HVPT	78.57	15.22	14
	Multiple Speaker PF	71.68	20.85	14
	Single Speaker HVPT	71.17	13.40	14
	Single Speaker PF	76.99	16.17	13
Gen Test 2 (Delayed) (New voices, new words)	Multiple Speaker HVPT	77.50	17.98	10
	Multiple Speaker PF	60.71	17.57	11
	Single Speaker HVPT	61.31	6.61	12
	Single Speaker PF	81.79	8.66	10

Figure 5.7: Nasal Contrast Generalisation Test 2: Mean Percentage Perceptual Identification Accuracy Scores According to Training Group

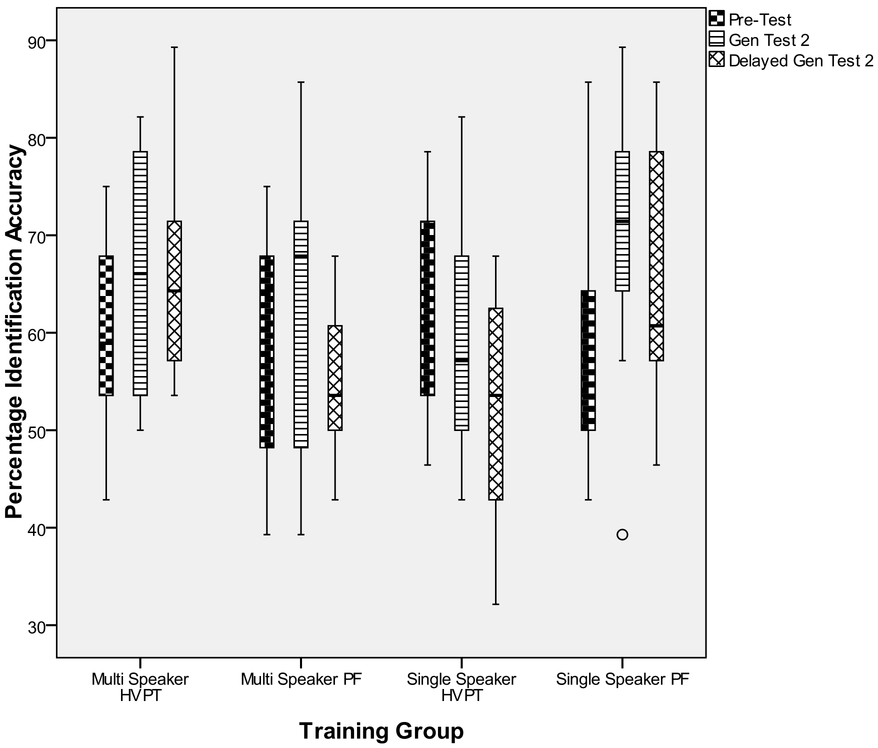
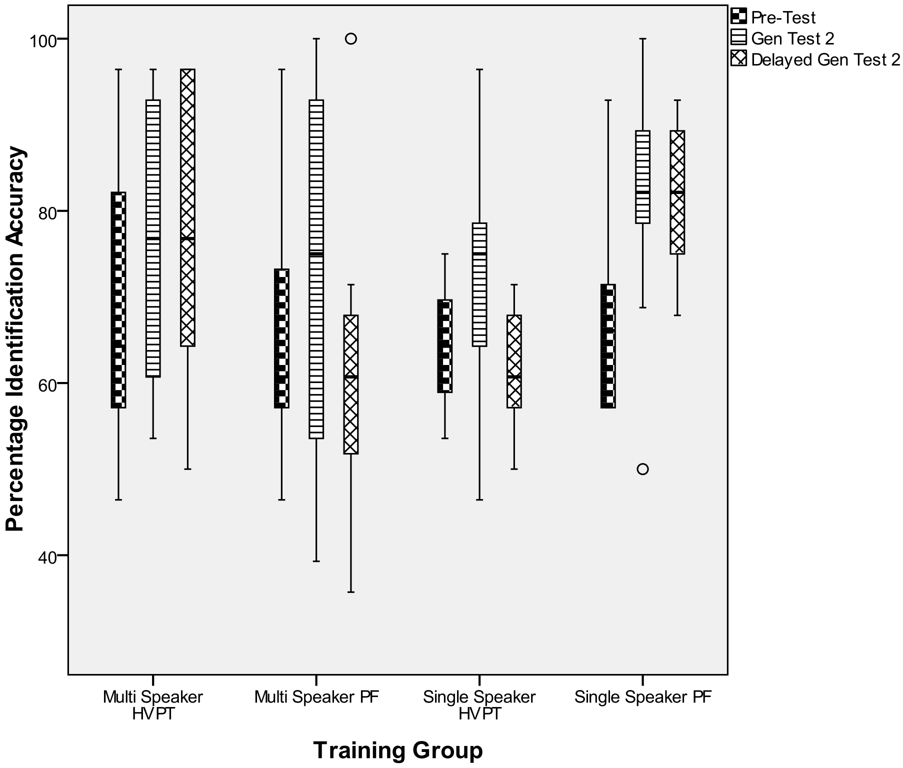


Figure 5.8: Oral Contrast Generalisation Test 2: Mean Percentage Perceptual Identification Accuracy Scores According to Training Group



As with the previous data, the generalisation test 2 data were then analysed using binary logistic mixed effects analysis in order to determine the significance of training group, test, contrast and time on whether or not participants answered correctly. The participant and test item were again included as random effects. The dependent variable was whether the participant correctly identified the minimal pair member presented. The fixed effects or potential predictors tested were participant training group/method (SubGp), test (TestType, the three levels of this – pre, gen2, gen2T2, also represented time) and Contrast (oral vs. nasal). The model is detailed in Table 5.10.

Table 5.10: Fixed Effects and Coefficients of the Generalisation Test 2 Model

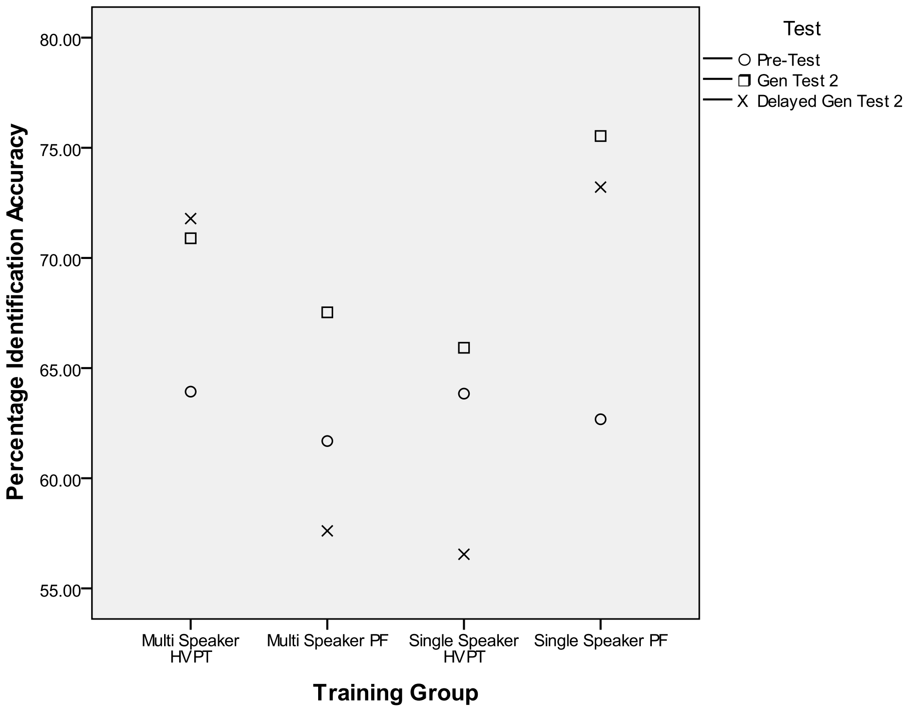
	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.455	0.178	2.564	0.0104**
SubGpSingle PF Only	-0.019	0.193	-0.096	0.9232
SubGpMulti PF Only	-0.126	0.190	-0.663	0.5076
SubGpSingle HVPT Only	-0.049	0.189	-0.259	0.7954
TestTypeGen2	0.391	0.174	2.250	0.0245**
TestTypeGen2T2	0.362	0.184	1.971	0.0487**
ContrastOral u-y	0.399	0.141	2.836	0.0046***
SubGpSingle PF:TestTypeGen2	-0.018	0.163	-0.109	0.9131
SubGpMulti PF:TestTypeGen2	-0.160	0.158	-1.014	0.3105
SubGpSingle HVPT:TestTypeGen2	-0.293	0.157	-1.861	0.0627*
SubGpSingle PF:TestTypeGen2T2	-0.001	0.183	-0.003	0.9978
SubGpMulti PF:TestTypeGen2T2	-0.587	0.174	-3.384	0.0007****
SubGpSingle HVPT:TestTypeGen2T2	-0.660	0.171	-3.867	0.0001****

***p<.1; **p<.05; ***p<.01, ****p<.001**

As with both previous models, the reference level for participant training group (SubGp) was the multiple speaker HVPT group, the reference level for Contrast was the nasal contrast, and the reference level for test (TestType) was the pre-test. Looking firstly at the main effects, no group scored significantly differently from the multiple speaker HVPT group overall, however the significant interactions show that the groups do differ dependent upon test (see below). The main effect of contrast shows that oral contrast test stimuli were answered correctly significantly more than nasal contrast test stimuli overall. However, there are no interaction terms with Contrast in this model as adding the interaction term resulted in no significant interactions. This means that the conclusions drawn from interpretation of the other interactions in this model hold for both contrasts, the data were thus again collapsed across contrasts. Finally, the main effect of test suggests that the overall generalisation test 2 and delayed generalisation test 2 scores were significantly greater than the pre-test scores. This implies that training generalised and this generalisation was retained over time. The caveats to this interpretation in 5.5.2 with reference to generalisation test 1 also apply here.

The post-test interactions suggest that the only group which may have scored differently from the multiple speaker HVPT group dependent upon whether the test was the pre-test or generalisation test 2 is the single speaker HVPT group. However, this interaction only approaches significance. Similarly, looking at the delayed generalisation test 2 scores, both the multiple speaker PF and single speaker HVPT groups scored significantly differently from the multiple speaker HVPT group dependent upon whether the test was the pre-test or the delayed generalisation test 2. An inspection of Figure 5.9 demonstrates the exact nature of the interactions.

Figure 5.9: The Interaction Between Participant Training Group and Test



Before the delay, the magnitude of the difference between the single speaker HVPT group score and multiple speaker HVPT group score for generalisation test 2 does not appear great, which explains the marginal significance of this interaction. With reference to the delayed generalisation test 2 interactions, it can be seen that both the single speaker HVPT group and the multiple speaker PF group score less in the delayed generalisation test 1 than the Multiple Speaker HVPT group. At no time does the single speaker PF group score significantly differently than the Multiple Speaker HVPT group.

In sum, it can be seen that the participants in the multiple speaker HVPT group outperform the participants on the single speaker HVPT group in generalisation test 1 and outperform the participants in both the single speaker HVPT and multiple speaker PF groups in the delayed generalisation test 1. The multiple speaker HVPT group and single speaker PF group do not perform differently at any time. Therefore, the single speaker PF and multiple speaker HVPT training methods appear to work better than the others, again particularly in terms of retaining generalisation effects.

5.6 Discussion

The aim of this chapter was to ascertain whether or not perceptual training resulted in improvement in perceptual accuracy for both contrasts trained and whether any method(s) emerged as most successful. Overall, the perceptual training resulted in an improvement in perceptual accuracy for both contrasts and this generalised to new words and new voices. This is in support of already well-documented evidence from the proponents of these techniques (e.g., Jamieson & Morosan, 1986; Logan et al, 1991). Iverson et al (2005) compared multiple speaker HVPT and multiple speaker PF training techniques (along with two others) and found no significant difference between these techniques immediately after training, and the findings of the present study are in agreement with this. The authors did not examine a single speaker HVPT technique in their study, and the present study suggested that this was the only technique which resulted in poorer results than the multiple speaker HVPT technique immediately post training (except, perhaps, in generalisation test 2). This suggests that training with only one voice, even if the contrast to be trained is in multiple phonetic contexts, is relatively ineffective (see Lively et al, 1993) unless perceptual fading manipulations are performed on the stimuli.

Stronger group effects emerged in the delayed tests, as Iverson et al (2005) anticipated in their comparison. In the delayed tests across both contrasts, the multiple speaker HVPT technique emerged as more successful than the multiple speaker PF and single speaker HVPT techniques, whilst any difference between the multiple speaker HVPT technique and single speaker PF technique was not significant at any time. In terms of answering the first, second and third supplementary research questions (1. Will perceptual training result in an improvement in perceptual accuracy in terms of improved identification scores after training, for both the oral and nasal vowel contrast to be trained? 2. Will this improvement generalise to new words and new voices, for both the oral and nasal vowel contrasts? 3. Will this improvement be retained at least one month after training, and for both the oral and nasal vowel contrasts?), perceptual training does result in some improvements in perceptual accuracy for both contrasts from pre- to post-test and there is some generalisation to new words and new voices. However, for both vowel contrasts, this improvement is only retained for at least one month after training if participants had undertaken the single speaker PF technique or the multiple speaker HVPT technique. For the other two techniques, across all tests, accuracy fell to at least pre-training levels.

The first main research question ('Is the HVPT technique more successful than the perceptual fading technique (or vice versa) in terms of producing a generalisable, long-term improvement in perception?) and fourth supplementary research question ('Are each of these methods more or less successful in their 'classic' form (PF with one speaker; HVPT with multiple speakers) or their alternative form (PF with multiple speakers as used by Iverson et al (2005); HVPT with a single speaker for comparison purposes'), can therefore be answered by noting that the HVPT technique and the PF technique in their 'classic' forms are significantly more successful at training the French vowel contrasts of interest than the HVPT technique and PF technique in their alternative forms in terms of producing a generalisable, long-term improvement in perception. These techniques appeared to result in phonetic memory which remained stronger over time, perhaps due to these techniques having ideal levels of variability or difficulty as discussed below.

Perhaps one of the most surprising findings is that the multiple speaker PF technique was shown to be inferior to the multiple speaker HVPT technique and the single speaker PF technique in terms of retaining the training. It was expected that the increased variability from single speaker PF of using multiple speakers at the same time as fading would be beneficial. In this case, however, it is possible that there was too much variation within each training stimulus when adding multiple speakers to perceptual fading which confused the participants over time. In addition, as shown by the native speaker identification functions in Chapter 4 (Figures 4.2-4.7) some minimal pairs used in this training condition were more difficult to discriminate for native speakers than in the other conditions, therefore indicating a more difficult training task (as explained in Chapter 4, this could not be avoided). At the same time, however, multiple speaker PF does not appear to result in any more variability than found in the successful multiple speaker HVPT technique, and it is therefore unclear why PF with multiple voices is less helpful for the present participants. The single speaker HVPT technique also appeared to be relatively ineffective, for the opposite reason, that is, not enough variability which is in agreement with the findings of Lively et al (1993) (see also Chapter 8, Section 8.2.1). Using just one voice in training can lead to trainees to focus on irrelevant speaker-specific information which prevents generalisation of learning to other voices (see, e.g., Lively et al, 1993; Wang & Munro, 2004). However, the success of the single speaker PF technique suggests that emphasis of the important features which determine the contrasts through perceptual fading appears to prevent participants focusing upon this irrelevant speaker-specific information.

It therefore appears that the high variability or perceptual fading training techniques in their 'classic' form are the best to take forward to the pronunciation training study. There appears to be no significant differences between them in terms of effectiveness, and it is therefore suggested that both techniques should be taken forward.

Finding these techniques equally as effective is perhaps not surprising as several authors have made use of both techniques within one training study without 'mixing' them as in the present work and Iverson et al (2005). For example, Wang and Munro (2004) used both techniques in training (although the PF training was with fully synthetic stimuli) because the authors believed it would maximise the effectiveness of the training by increasing variability. They noted that using synthesised stimuli in training (alongside the high variability of natural stimuli from multiple speakers) allowed for not only a fading effect but also a manipulation of pitch and vowel duration in order to direct listener attention away from these irrelevant cues and onto vowel quality instead (Wang & Munro, 2004).

Having established in the present work that these techniques are most effective in their 'classic' form, an interesting possibility for future work could be to compare the effect of using both techniques together in this form, versus using only one of the techniques. It is possible that using PF and HVPT stimuli separately within each training session would prove more effective than, for example, the multiple speaker PF technique used in the present work. Equally, however, if one technique emerges as equally or more successful than two in improving perceptual and pronunciation accuracy, then using both techniques would be unnecessary.

The results of the research conducted in this chapter have established that, for the French vowel contrasts used, the single speaker PF and multiple speaker HVPT perceptual training are superior to multiple speaker PF and single speaker HVPT at retaining perceptual training improvements over time, which does not appear to have been noted in any previous work. This study also supports the well established finding that perceptual training generally has a positive effect on perceptual learning. The next chapter (Chapter 6) provides further contributions by examining the effects of perceptual training alone, pronunciation training alone and both modes together on both perceptual and pronunciation learning.

6 Experiment 2: Perceptual vs. Pronunciation Training and Effect on Perceptual and Pronunciation Accuracy

6.1 Introduction

This chapter examines the impact on participant perceptual and pronunciation ability of adding pronunciation training conditions to the experimental design. The study compares the effectiveness of five training methods in improving both perceptual and pronunciation accuracy: 1) single speaker perceptual fading (PF), 2 multiple speaker high variability phonetic training (HVPT) (the two most successful perceptual training techniques identified by experiment 1), 3) pronunciation only, 4) single speaker PF + pronunciation and 5) multiple speaker HVPT + pronunciation.

Iverson et al (2005) did suggest that adding pronunciation testing as well as carrying out delayed testing may highlight differences between the perceptual training techniques they compared (Chapter 5 described how adding a delay did create a significant difference between some of these techniques). Furthermore, this kind of perceptual training vs. pronunciation training vs. both comparison has rarely been carried out and yet is important in the attempt to find an optimal training technique for difficult language contrasts. For example, it may be that training in one modality is sufficient for optimal improvement in both. Alternatively, some training in each modality over the same timescale may achieve the best results.

The main purpose of this chapter is to answer the second and third research questions:

RQ2 Are the most successful perceptual training techniques suggested through answering Research Question 1 more successful than pronunciation training in terms of producing a generalisable, long-term improvement in pronunciation and perception?

RQ3 With regard to using the multi speaker HVPT technique and/or the single speaker PF technique and/or pronunciation training, does an optimal training technique emerge from those examined in terms of producing a generalisable, long-term improvement in pronunciation and perception?

6.2 Participants

This study involved the use of data from three sets of participants. The first set was those participants who were part of the more successful single speaker PF and multiple speaker HVPT training groups as described in Chapter 5. The numbers in these groups were 13 and 14 respectively immediately after training, and 10 in both for the tests of retention due to lack of interest or scheduling difficulties. These participants are described in more detail in Chapter 5.

The second set of participants was those who participated in this second study. A total of 57 (45 female, 12 male) first and second year students of French were recruited from the Universities of Edinburgh, Glasgow and Strathclyde through emailing the French class lists. As with the perceptual experiment, first and second year students of French were recruited as they had not yet completed their compulsory year in a French speaking country, and were therefore more likely to have increased perceptual and pronunciation difficulties with the vowel contrasts to be examined due to this lack of experience. Again, none of the participants had extensive experience with the French language outside of their studies and no participant reported any hearing difficulties.

All participants were native speakers of English, and were either Scottish (26), English (26), American (3), Canadian (1) or Irish (1). They had been learning French for 2-12 years with an average of 7.65 years of learning. The age range of the participants was 18-22 years with a mean age of 18.7. The participants were again paid for their time and were entered into a voucher prize draw, as it was unlikely that there would be a large enough number who would participate in a long term study without financial incentive.

The 57 participants were randomly assigned to four groups: Group 1: Multiple Speaker HVPT + Pronunciation (14 participants); Group 2: Single Speaker PF + Pronunciation (14 participants); Group 3: Pronunciation Only (14 participants); Group 4: Control (15 participants). The participants completed a language history questionnaire (Li, Sepanski & Zhao, 2006) and a Pronunciation Attitude Inventory (PAI) questionnaire (Elliott

(1995), see Chapter 7 and Appendices G and H). The participants then undertook a perceptual identification pre-test and also a pronunciation pre-test. Immediately after completing six training sessions (a maximum of twice per week, non-control participants only), the participants took perception and pronunciation post-tests, and generalisation tests. Control participants were necessary in this experiment as in order to demonstrate that undertaking training was more effective than no training. These control participants carried out the pre-tests and then returned after a minimum of three weeks to do the post-tests (that is, after at least the minimum amount of time it would take for an experimental participant to complete training).

After a minimum of one month all participants then returned to carry out retention testing. Three participants did not return due to lack of interest or scheduling difficulties, resulting in group numbers as follows: Group 1: Multiple Speaker HVPT + Pronunciation (13/14 participants); Group 2: Single Speaker PF + Pronunciation (13/14 participants); Group 3: Pronunciation Only (14/14 participants); Group 4: Control (14/15 participants).

The final set of participants were 26 (19 female, 7 male) native speakers of French who identified and rated the pronunciation data from the French learning participants. They were all attending the University of Edinburgh under the Erasmus scheme and were therefore all normally resident in France. They were recruited through emailing around the French Erasmus class lists. None reported any hearing difficulties and they were paid for their time.

6.3 Stimuli

The preparation of the stimuli used by those undergoing training in the present study is detailed in Chapter 5. As previously noted, recordings from six (3 female, 3 male) of the native speakers recorded to provide stimuli were randomly selected to be used in the HVPT multiple speaker perceptual training sessions. The first speaker (male) of this group was also used in the single speaker PF perceptual training sessions (with the natural vowels removed and fading synthesised vowels inserted). This male speaker was also used for the first perceptual test of generalisation (new words, familiar speaker). The recordings of four further speakers (2 female, 2 male), were randomly selected to be used for the perceptual pre- and post-tests. Finally, the remaining four speakers (2 female, 2 male) were used for the second perceptual test of generalisation (new words, new speakers). It should be noted that for the Control group, no voices are familiar, thus both generalisation tests test new words and new voices.

The pronunciation training stimuli were the same words as used in the perceptual training sessions so that all participants were trained using the same stimuli. In order to avoid over-complicating the experimental design, the pronunciation training made use of a single speaker (multiple speaker vs. single speaker pronunciation training conditions would make for interesting future work). This single speaker was the same male speaker used in the first multiple speaker HVPT perceptual training session, all the single speaker PF perceptual training sessions and the first perceptual test of generalisation. The pronunciation training stimuli were therefore exactly the same as the perceptual training stimuli used on the first day of perceptual training. The pronunciation pre-/post test and generalisation test each made use of 28 new words, that is, seven minimal pairs or 14 words for each contrast.

The final set of stimuli was the recordings of the training participants reading French words. These recordings (along with all training and testing) took place in a sound-deadened booth with a Dell desktop PC at Edinburgh University or in a quiet room in Glasgow and Strathclyde University libraries using a Macintosh MacBook laptop running boot camp Windows XP. Sound was captured using a Crown noise cancelling headset microphone and a Microtrack digital recorder. The sound was digitised at a sampling rate of 48 kHz with a resolution of 32 bits.

After completing these recordings it was realised that identifying and rating all 140 words (28 pre-test, 28 post-test, 28 generalisation test, 28 post-test(T2) and 28 generalisation test (T2)) produced by 27 participants from the perceptual training study and from 57 participants in this perceptual and/or pronunciation study was too great a task to be completed in a timely fashion. It was therefore decided to analyse three minimal pairs randomly chosen from each contrast for each test, resulting in 12 words per test and 60 words in total to be analysed for each participant. The pairs chosen for analysis are asterisked in Appendix I.

The participant recordings were copied from the Microtrack unit onto a computer. Using Audacity software, the target words were then separated from the recorded lists, and background noise was eliminated where necessary. The recordings were then normalised and saved as individual files for presentation during analysis.

6.4 Procedure

6.4.1 Pre-Testing

The perceptual pre-testing procedure for this study is identical to that for the perceptual study and is described in detail in Chapter 5 (5.4.1), using the two alternative forced choice paradigm with no feedback. Again, the words from both contrasts to be trained (14 words from each contrast repeated twice, resulting in 56 experimental trials) were presented in a random order, and participants were able to take as long as they needed to respond.

For pronunciation pre-testing, participants were recorded reading 28 test words (7 minimal pairs from each contrast) which were presented randomly one at a time on a monitor using EPrime software. Participants were instructed to read in a neutral tone and moved to the next word to be read by using the spacebar on the keyboard attached to the monitor. The full instructions given and pre-test/post-test words are listed in Appendix J and Appendix I respectively. Again, a spelling guide was provided to participants which gave instructions on how to match up the spelling of any unfamiliar words with the sounds of interest (see Appendix E for the full text of this guide).

6.4.2 Training

As with the perceptual training alone described in Chapter 5, participants attended six training sessions, the first of which occurred immediately after the pre-tests had been completed. A maximum of two training sessions were carried out each week, and participants took between 3 and 6 weeks to complete the training.

Those participants undertaking both perceptual and pronunciation training received 3 sessions of perceptual training and 3 sessions of pronunciation training. The training modality undertaken in the first session was randomly assigned and subsequent training sessions alternated between the two modalities. This meant that those undergoing training in both modalities received half of the training in each modality received by those undergoing training in one modality.

6.4.2.1 Perceptual Training

Those in the multiple speaker HVPT + pronunciation training group heard two native speakers in each perceptual training session such that they heard all six speakers used by those in the perception-only HVPT group across three sessions instead of six. Each perceptual session had four blocks. The first two blocks used a male speaker and the second two blocks used a female speaker, therefore ensuring participants heard the six speakers in the same order as those who underwent perceptual HVPT training only. One block per speaker trained the oral contrast and the other block trained the nasal contrast, and these two blocks were presented in a random order within each speaker.

Within each block of the perceptual-training-only conditions (one for each contrast) there were 14 minimal pairs used, 28 words each repeated twice, resulting 56 training trials per block and 112 training trials in total for each session. In order to match this 112 training trials per session for the perceptual + pronunciation conditions, the 14 minimal pairs or 28 words were only used once, resulting in 28 trials per block. Participants therefore still heard 28 words, each repeated twice for each contrast, but the word was spoken once by the male speaker and once by the female speaker.

The compression of perceptual training for those in the single speaker PF + pronunciation training group followed a similar logic. Participants heard stimuli pairs at two subsequent points on the fading continuum in one session, such that they heard the stimuli on all 12 of the points on the fading continua across three sessions instead of six. Again, each perceptual session had four blocks. The first two blocks used the two points which were furthest apart on the continuum and the next two blocks used the next two closest points, therefore ensuring participants heard the progressively more similar sounding stimuli in the same fashion as those who underwent perceptual training PF only. Within each pair of fading points on the continuum one block trained the oral contrast and the other trained the nasal contrast, and these two blocks were presented in a random order.

Again, in order to match the 112 training trials per session for perceptual training only conditions, the 14 minimal pairs or 28 words were only used once, resulting in 28 trials per block. Participants therefore still heard 28 words each repeated twice for each contrast but the word was spoken once as part of the further apart pair and once as part of the closer together pair. Table 6.1 summarises this perceptual training schedule and gives the schedule for all training groups.

Table 6.1: Training Schedule for Experimental Groups

	Single Speaker PF	**Multi Speaker HVPT**	**Single Speaker PF + Pronunciation**	**Multi Speaker HVPT + Pronunciation**	**Pronunciation Only**
Day 1	Points 1 and 12	Speaker 1	Points 1 & 12 AND Points 2 & 11	Speakers 1&2	Pronunciation
Day 2	Points 2 and 11	Speaker 2	Pronunciation	Pronunciation	Pronunciation
Day 3	Points 3 and 10	Speaker 3	Points 3 & 10 AND Points 4 & 9	Speakers 3&4	Pronunciation
Day 4	Points 4 and 9	Speaker 4	Pronunciation	Pronunciation	Pronunciation
Day 5	Points 5 and 8	Speaker 5	Points 5 & 8 AND Points 6 & 7	Speakers 5&6	Pronunciation
Day 6	Points 6 and 7	Speaker 6	Pronunciation	Pronunciation	Pronunciation

NB: For perception + pronunciation conditions the order may have been reversed due to random assignment of which training modality came first. The perceptual training for these conditions consisted of half the time being spent on each part of training as compared to perceptual training only. For PF conditions continuum points 1-6 represent one member of the minimal pair and 7-12 represent the other.

The perceptual training tasks followed the same procedure as detailed in Chapter 5 (5.4.2), using the two alternative forced choice paradigm with feedback as to whether or not the participant had responded correctly. Again, participants were able to take as long as they needed to respond to each trial, the next word was played 500msec after the response.

6.4.2.2 Pronunciation Training

The pronunciation training took the form of simple 'listen and repeat' training sessions on computers running EPrime software. Instruction sheets were provided to complement onscreen instructions. The participants were trained using the same words as in the perceptual training sessions, and these sessions were split into two blocks, one for each contrast. Each contrast block was separated into 2 sub-blocks, one for each sound in the minimal pair. Each word was presented once resulting in 14 words per sub-block, 28 words (14 minimal pairs) per block and 56 words (28 minimal pairs) in total trained. The blocks, and the sub-blocks within them, were presented in a random order.

The first screen in the training program directed participants to the first page of the instruction sheet which explained the meaning of the terms 'hard palate' and 'soft palate' which were to be used in the later pronunciation instructions. Pressing the spacebar to move on to the next screen began the first block, and participants were informed that the two sounds they were to learn next were often confused with each other and so to listen carefully. Moving on to the next screen began the first sub-block and this informed the participant how the first sound they were about to learn was written in the IPA and how it was written/spelled in French words. They were then asked to press the spacebar to hear the sound and then instructed to turn to the relevant page in the instruction sheets. The next screen provided pronunciation instructions based upon Gregg (1960) and Tranel (1987). These instructions covered mouth position, lip position, tongue position, palate position (if relevant) and an approximate comparison with English. The instruction sheet repeated this information to allow participants to have these instructions to refer to throughout training. Participants were then asked to firstly mouth the sound whilst listening to the native speaker pronouncing it and then repeat the sound. They were then informed they were going to learn some words using that sound. Each word was played three times.

The training screen told participants which word they were hearing, advised them to mouth it whilst it was being pronounced by the speaker and then instructed them to say it aloud three times once the speaker had finished, referring to the instruction sheet if necessary. Participants then pressed the spacebar to hear the next word. Once the participants had heard all the words featuring the sound being trained they were informed that they had finished the training for that sound. Pressing the spacebar began the second sub-block with how the sound was written in the IPA and how it was written in French words. The second sub-block of training then took the same form as the first. At the end of this second sub-block, participants heard both of the sounds they had just learned pronounced together. Next, pressing the spacebar began the second block, again with the explanation that the two sounds they were about to learn were often confused with each other. Training then progressed in the second block in the same manner as the first block described above. Having heard the second set of sounds together, participants were then informed that training was complete for the day. The full text for the training and pronunciation instructions is in Appendix K.

Whilst the pronunciation training does have a perceptual element in terms of participants hearing native speakers produce the words, it was felt that this was necessary in order that the participants were able to undergo pronunciation training autonomously. Due to the constraints of running a large number of participants alone, it was not to be possible for the experimenter to provide the one-to-one training and feedback about how accurate participant productions were from only following articulatory instructions alone in the manner of studies such as that of Pimsleur (1963) which sought to avoid any perceptual element at all (the implications of participants not receiving any feedback about the accuracy of their productions are discussed in 8.2.2.2). Given that articulatory instructions are provided, the minimal pair sounds are only presented together once after the training session (thus the focus is not on perceptually differentiating the members of the minimal pair), and the training sessions focus upon providing pronunciation practice it can be argued that the perceptual training element to this pronunciation training is minimal.

6.4.3 Post-, Generalisation and Retention Testing

As soon as the sixth training session had been completed, the participants carried out the perception and pronunciation post-tests, which were a repeat of the tests they carried out prior to training. In addition, they carried out a pronunciation generalisation test and two perceptual generalisation tests (new words; new voices and new words). Details regarding perceptual generalisation testing can be found in Chapter 5 (5.4.3).

The pronunciation generalisation test was carried out in a similar fashion to the pronunciation pre- and post-testing. The only difference was the stimuli, which were 28 new test words to be read. Participants then returned after a minimum of one month (and a maximum of two months) and carried out the perception and pronunciation post-tests and generalisation tests again to ascertain how well any training benefits were retained.

6.4.4 Native Speaker Pronunciation Analysis

Twenty six native speakers of French were used to analyse the reduced 60 word sample from the 84 participants. Analysis was carried out across 41 sessions whereby fourteen native speakers carried out one analysis session, 12 native speakers carried out two analysis sessions and 1 native speaker carried out 3 analysis sessions.

Each rating session consisted of the following words:

- Heard in all sessions, the same for each rater: The 60 word sample produced by one randomly chosen (female) speaker.
- Heard in all sessions, the same for each rater: All of the test words produced by two native speakers of French (1 female, 1 male). As there was no effect of time for the native speakers this sample consisted of 48 words in total.
- Unique to each session, different for each rater: Either the 60 word sample from two speakers who returned to do the retention tests, resulting 120 words in total; or a reduced 36 word sample (as a result of speakers not returning to do the retention test) from three speakers resulting in 108 words in total.

This resulted in a total of either 228 or 216 words analysed per session. All analysis took place in a sound-deadened booth with a Dell desktop PC at Edinburgh University.

The native speaker analysis procedure made use of EPrime software. The first screen was the instruction screen which explained, in French, the tasks to be carried out. The native speaker was informed that words in French would be played and then there would be two tasks. The instructions then explained that the first task was to decide which of two options the word played could be, with one option on the left and one option on the right of the screen (a two alternative forced choice identification task). Secondly the instructions noted that the native speaker would then be informed which of the two words the speaker was trying to produce and that the pronunciation was to be rated on a scale of '1' (very accurate/nativelike) to '7' (very inaccurate/clearly not native). The instruction screen finally informed the native speakers that the particular sounds of interest were those which differentiated, for example, *tous* and *tu* (the oral contrast) and *devant* and *devons* (the nasal contrast), and that they should pay particular attention to how accurately those sounds were produced when giving their ratings. These instructions were also given on paper so that the native speakers could refer to them throughout their analysis. The full text of the instructions can be found in English and French in Appendix L.

Pressing the spacebar began the analysis procedure. The individual word recordings were played in a random order in terms of contrast, participant, and the test which the participant had undertaken. Whilst each word was played the accompanying screen displayed the minimal pair of which the word was a member, one word on the left and one word on the right and above this the instruction (in French) to '[p]ress '1' if you think the word played sounds most like the word on the left. Press '2' if you think the word played sounds most like the word on the right.' The native speakers could take as long as they needed to respond, and the word presentations were counterbalanced in terms of which side of the screen on which the correct answer appeared. Once the native speaker had responded, the next screen informed them (in French) that the word they had just heard 'was supposed to be XXXX,' and then asked 'please rate the accuracy of the pronunciation of the word on a scale of 1 to 7 where 1 means 'very accurate/nativelike' and 7 means 'very inaccurate/clearly not a native.' After rating the pronunciation (again, participants had as long as they needed to respond), the next word was played, and this procedure was repeated for all 228 or 216 words in the testing session.

6.5 Perception Results

As with the previous perception data, the analysis was carried out upon the participant perceptual accuracy scores in the tests. It was firstly necessary to ensure that any differences between training groups after training were attributable to the training technique and not due to one group having a higher or lower pre-test score. To this end, two one-way ANOVAs were carried out with pre-test score as the dependent variable and training group as the between subjects factor, one ANOVA for each contrast of interest. There was no significant effect of training group for either contrast [Nasal: $F(5,83) = .433$, $p = .825$; Oral: $F(5,83) = .280$, $p = .923$] indicating that the pre-test scores were not significantly different for each group.

As detailed in Chapter 5, the statistical tool R (R Development Core Team, 2011) with the R packages *lme4* (Bates, Maechler & Bolker, 2011) and *language*R (Baayen, 2011) was used to carry out binary logistic mixed effects analyses of the relationship between training groups, tests, contrast tested and time. The perceptual analysis was again carried out in three blocks. The first block examined the effects of training across the pre-test, post-test and delayed post-test. The second and third blocks examined the generalisation tests. In all cases, likelihood ratio tests comparing each model with fixed effects to a null model with only the random effects demonstrated that the fixed effects model differed significantly from the null model.

6.5.1 Block 1: Pre-Test, Post-Test, Delay

The mean participant percentage perceptual identification accuracy and standard deviations for each test (pre-test, post-test, and delayed post-test according to participant training group and contrast are shown in Tables 6.2 and 6.3 and illustrated in Figures 6.1. and 6.2. These data suggest that the control group may perform worse than the other groups, and that the multiple speaker HVPT, single speaker PF and multiple speaker HVPT + pronunciation groups may perform well.

Table 6.2: Nasal Contrast: Mean Percentage Perceptual Accuracy Scores According to Training Group

Test	Training Group	Mean	SD	N
Pre-Test	Multi HVPT + Pronun	61.22	10.28	14
	Single PF + Pronun	56.12	10.62	14
	Pronunciation Only	61.23	9.79	14
	Multi Speaker HVPT	59.69	8.80	14
	Single Speaker PF	60.44	13.16	13
	Control	60.24	11.99	15
Post-Test	Multi HVPT + Pronun	68.37	15.21	14
	Single PF + Pronun	67.35	11.87	14
	Pronunciation Only	65.31	7.20	14
	Multi Speaker HVPT	68.37	16.03	14
	Single Speaker PF	66.21	14.99	13
	Control	60.24	14.78	15
Delayed Post-Test	Multi HVPT + Pronun.	68.08	15.49	13
	Single PF + Pronun.	65.11	12.81	13
	Pronunciation Only	63.27	12.95	14
	Multi Speaker HVPT	67.50	14.33	10
	Single Speaker PF	68.93	15.07	10
	Control	59.18	9.78	14

Table 6.3: Oral Contrast: Mean Percentage Perceptual Accuracy Scores According to Training Group

Test	Training Group	Mean	SD	N
Pre-Test	Multi HVPT + Pronun.	68.88	13.17	14
	Single PF + Pronun.	63.52	17.05	14
	Pronunciation Only	67.09	12.78	14
	Multi Speaker HVPT	68.40	15.68	14
	Single Speaker PF	67.03	9.91	13
	Control	65.00	16.65	15
Post-Test	Multi HVPT + Pronun.	76.78	10.17	14
	Single PF + Pronun.	71.17	15.37	14
	Pronunciation Only	68.62	13.95	14
	Multi Speaker HVPT	74.24	14.23	14
	Single Speaker PF	74.72	13.87	13
	Control	67.14	14.55	15
Delayed Post-Test	Multi HVPT + Pronun.	77.75	10.33	13
	Single PF + Pronun.	68.13	17.71	13
	Pronunciation Only	67.09	10.50	14
	Multi Speaker HVPT	76.43	16.51	10
	Single Speaker PF	79.64	14.29	10
	Control	62.50	13.20	14

Figure 6.1: Nasal Contrast: Mean Percentage Perceptual Accuracy Scores According to Training Group

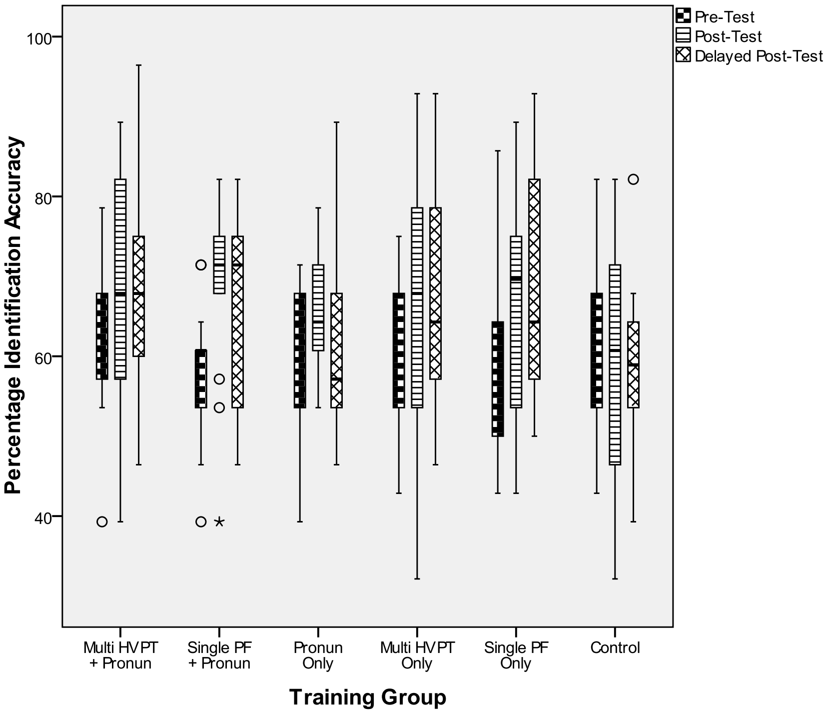
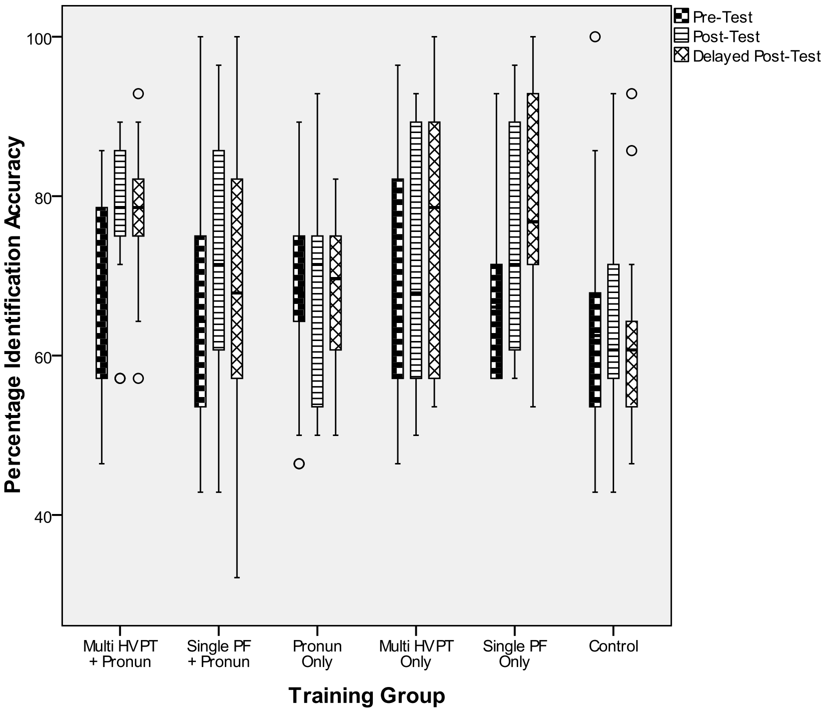


Figure 6.2: Oral Contrast: Mean Percentage Perceptual Accuracy Scores According to Training Group



These data were then analysed using binary logistic mixed effects analysis in order to determine the significance of training group, test, contrast and time on whether or not participants answered correctly. In the present model the participant and test item were included as random effects. The dependent variable was whether the participant correctly identified the minimal pair member presented. The fixed effects or potential predictors tested were participant training group/method (SubGp), test (Test, the three levels of this – pre, post, delay, also represented time) and Contrast (oral vs. nasal). The model is detailed in table 6.4.

Table 6.4: Fixed Effects and Coefficients of the Pre-Test, Post-Test, Delay Model

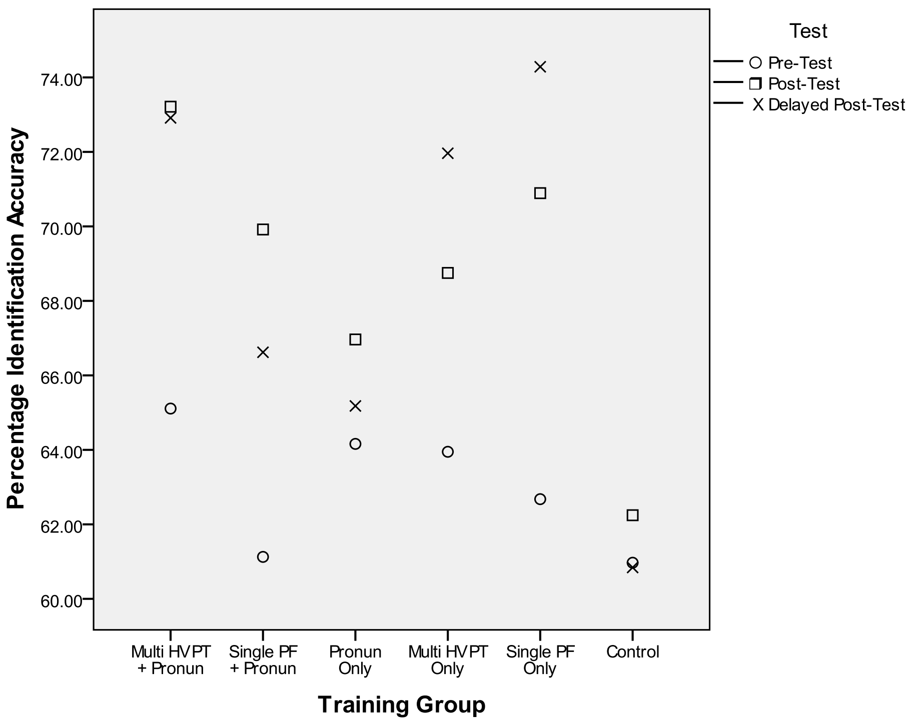
	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.421	0.183	2.306	0.0211**
SubGpSingle PF + Pronunciation	-0.133	0.202	-0.657	0.5109
SubGpPronunciation Only	0.045	0.202	0.221	0.8249
SubGpMulti HVPT Only	0.056	0.202	0.279	0.7805
SubGpSingle PF Only	0.043	0.206	0.211	0.8332
SubGpMulti HVPT + Pronunciation	0.079	0.202	0.390	0.6966
TestPost	0.050	0.105	0.477	0.6332
TestPostDelay	-0.034	0.107	-0.320	0.7490
SubGpSingle PF + Pronunciation:TestPost	0.405	0.153	2.646	0.0081***
SubGpPronunciation Only:TestPost	0.082	0.152	0.541	0.5883
SubGpMulti HVPT Only:TestPost	0.318	0.154	2.064	0.0390**
SubGpSingle PF Only:TestPost	0.305	0.157	1.938	0.0526*
SubGpMulti HVPT + Pronunciation:TestPost	0.347	0.155	2.242	0.0249**
SubGpSingle PF + Pronunciation:TestPostDelay	0.358	0.156	2.299	0.0215**
SubGpPronunciation Only:TestPostDelay	0.118	0.152	0.773	0.4396
SubGpMulti HVPT Only:TestPostDelay	0.479	0.167	2.864	0.0042***
SubGpSingle PF Only:TestPostDelay	0.603	0.170	3.54	0.0004****
SubGpMulti HVPT+ Pronunciation:TestPostDelay	0.334	0.157	2.128	0.0333**

***p<.1; **p<.05; ***p<.01, ****p<.001**

The figures above represent the models predictions, in log odds, of a factor's effect while holding other fixed effects in the model constant (Drager, 2011, p.111). The reference level for participant training group (SubGp) was the control group and the reference level for test (Test) was the pre-test. Contrast was not included in this final model as it only approached significance when included as a main effect (p = 0.06) and adding an interaction term resulted in no significant interactions. This means that the conclusions drawn from interpretation of the other interactions in this model hold for both contrasts and the data were collapsed across contrasts.

Looking firstly at the main effects, no group scored significantly differently from the control group overall, however the significant interactions show that the groups do differ dependent upon test (see below). In addition, examining the main effect of test, overall the post-test and delayed post-test scores were not significantly greater than the pre-test scores. The post-test and delayed post-test interactions suggest that the only group which did not score differently from the control group dependent upon whether the test was the pre-test vs. the post-test or vs. the delayed post-test is the pronunciation only group. An inspection of Figure 6.3 demonstrates the exact nature of the interactions.

Figure 6.3: The Interaction Between Participant Training Group and Test



With reference to both the post-test and the delayed post-test interactions it can be seen that whilst the control group does not score significantly differently to the other groups at pre-test, this group scores significantly less than all other groups except the pronunciation only group at post-test and at the delayed post-test.

In sum, it can be seen that the participants in all groups except the pronunciation only group outperform the participants in the control group at post-test and in the delayed post-test. Therefore, all the training methods including a perceptual element appear to work more effectively than pronunciation training only or no training in terms of producing and retaining perceptual improvement. In order to investigate whether any of the more successful groups outperformed the others the analysis was re-run with the control group and pronunciation only group omitted. The model is detailed in Table 6.5 below.

Table 6.5: Fixed Effects and Coefficients of the Second Pre-Test, Post-Test, Delay Model

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.453	0.182	2.490	0.0128**
SubGpSingle PF + Pronunciation	-0.193	0.196	-0.983	0.3257
SubGpSingle PF Only	0.013	0.201	0.065	0.9479
SubGpMulti HVPT + Pronunciation	-0.008	0.196	-0.038	0.9694
ContrastOral u-y	0.368	0.164	2.243	0.0249**
TestPost	0.395	0.057	6.972	<0.00001****
TestPostDelay	0.393	0.061	6.476	<0.00001****

***p<.1; **p<.05; ***p<.01, ****p<.001**

As the reference control group had been removed, the reference group used for this model was the multiple speaker HVPT group used as reference in the analysis in Chapter 5. Contrast is now included as a main effect, however, no interactions including training group and/or contrast were significant, therefore the remaining data are best explained by the main effects model. It therefore appears that in terms of the perceptual pre-test, post-test and delayed post-test that the control and pronunciation only techniques emerge as less successful. No other training groups score differ significantly differently from the multiple speaker HVPT group once the less successful techniques are removed.

6.5.1.1 Group Comparison With Native Speakers

After removal of the control and pronunciation only groups the model was also re-run with the results from native listeners as the reference group in order to ascertain how the trainee groups differed from a group of native speakers. The performance of the remaining groups along with the native group collapsed across contrast (there continued to be no significant interactions in terms of contrast, see below) is illustrated in Figure 6.4 and the model is detailed in Table 6.6.

Figure 6.4: Mean Overall Percentage Perceptual Identification Accuracy Scores According to Remaining Training Groups and Native Speaker Group

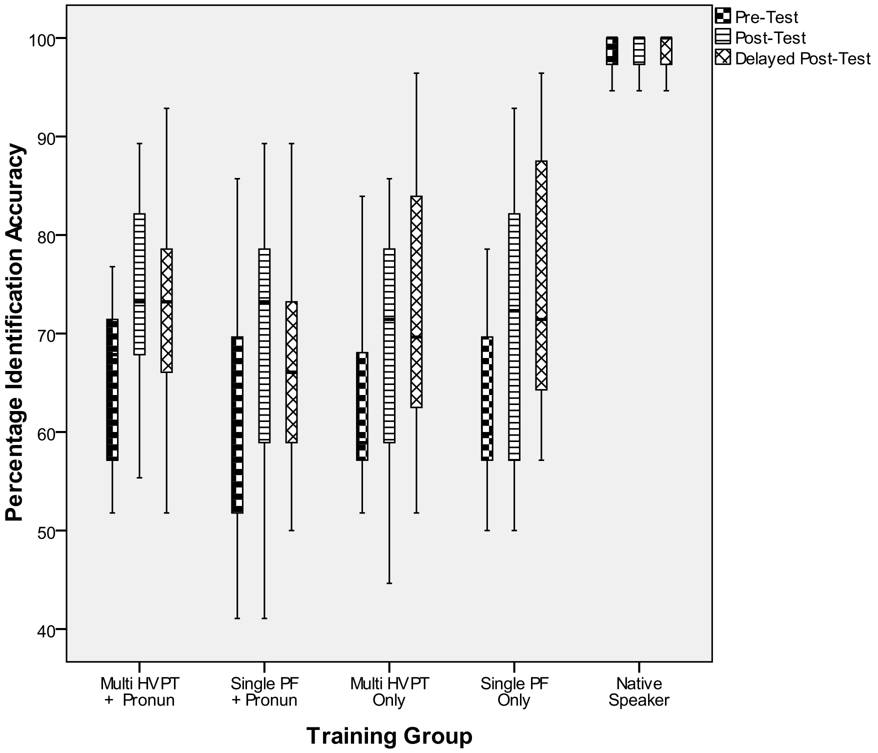


Table 6.6: Fixed Effects and Coefficients of the Native Speaker Pre-Test, Post-Test, Delay Model

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	4.247	0.468	9.079	<0.00001****
SubGpSingle PF + Pronun	-3.989	0.476	-8.388	<0.00001****
SubGpSingle PF Only	-3.782	0.478	-7.920	<0.00001****
SubGpMulti HVPT Only	-3.796	0.476	-7.976	<0.00001****
SubGpMulti HVPT + Pronun	-3.804	0.476	-7.997	<0.00001****
ContrastOral u-y	0.377	0.165	2.286	0.0223**
TestPost	0.394	0.057	6.961	<0.00001****
TestPostDelay	0.392	0.061	6.461	<0.00001****

With the native speaker group as the reference group the remaining data are best explained by the main effects model. Contrast is again included as a main effect, however, no interactions including training group and/or contrast were significant. This model suggests that the native speaker group scores significantly better than all other groups overall and the lack of significant interaction terms means that this is not dependent upon training group or time, in other words, no training technique results in a non-significant difference between native speaker group and any learner group.

6.5.2 Generalisation Test 1

The mean participant percentage perceptual identification accuracy and standard deviations for the first test of generalisation with new words but a familiar voice (from day 1 of training) are shown in Tables 6.7 and 6.8 and illustrated in Figures 6.5 and 5.6. The original pre-test scores are also included for reference. These data, as with the pre/post/delay data, suggest that the control group may perform worse than the other groups, and that the multiple speaker HVPT, single speaker PF and multiple speaker HVPT + pronunciation groups may perform well.

Table 6.7: Nasal Contrast Generalisation Test 1: Mean Percentage Perceptual Identification Accuracy Scores According to Training Group

Test	Training Group	Mean	SD	N
Pre-Test	Multi HVPT + Pronun	61.22	10.28	14
	Single PF + Pronun	56.12	10.62	14
	Pronunciation Only	61.23	9.79	14
	Multi Speaker HVPT	59.69	8.80	14
	Single Speaker PF	60.44	13.16	13
	Control	60.24	11.99	15
Gen Test 1 (Familiar voice, new words)	Multi HVPT + Pronun	69.64	12.51	14
	Single PF + Pronun	63.26	12.95	14
	Pronunciation Only	63.27	9.65	14
	Multi Speaker HVPT	67.60	15.31	14
	Single Speaker PF	68.96	14.24	13
	Control	52.38	11.74	15
Delayed Gen Test 1 (Familiar voice, new words)	Multi HVPT + Pronun.	66.76	22.37	13
	Single PF + Pronun.	63.19	17.03	13
	Pronunciation Only	66.33	13.42	14
	Multi Speaker HVPT	67.86	15.79	10
	Single Speaker PF	67.50	14.53	10
	Control	57.65	14.21	14

Table 6.8: Oral Contrast Generalisation Test 1: Mean Percentage Perceptual Identification Accuracy Scores According to Training Group

Test	Training Group	Mean	SD	N
Pre-Test	Multi HVPT + Pronun.	68.88	13.17	14
	Single PF + Pronun.	63.52	17.05	14
	Pronunciation Only	67.09	12.78	14
	Multi Speaker HVPT	68.40	15.68	14
	Single Speaker PF	67.03	9.91	13
	Control	65.00	16.65	15
Gen Test 1 (Familiar voice, new words)	Multi HVPT + Pronun.	81.63	20.00	14
	Single PF + Pronun.	72.45	19.78	14
	Pronunciation Only	71.17	13.76	14
	Multi Speaker HVPT	82.14	17.55	14
	Single Speaker PF	83.52	14.55	13
	Control	68.81	16.56	15
Delayed Gen Test 1 (Familiar voice, new words)	Multi HVPT + Pronun.	78.57	16.17	13
	Single PF + Pronun.	67.31	15.14	13
	Pronunciation Only	64.29	13.80	14
	Multi Speaker HVPT	77.50	18.60	10
	Single Speaker PF	80.71	9.70	10
	Control	61.22	13.86	14

Figure 6.5: Nasal Contrast Generalisation Test 1: Mean Percentage Perceptual Identification Accuracy Scores According to Training Group

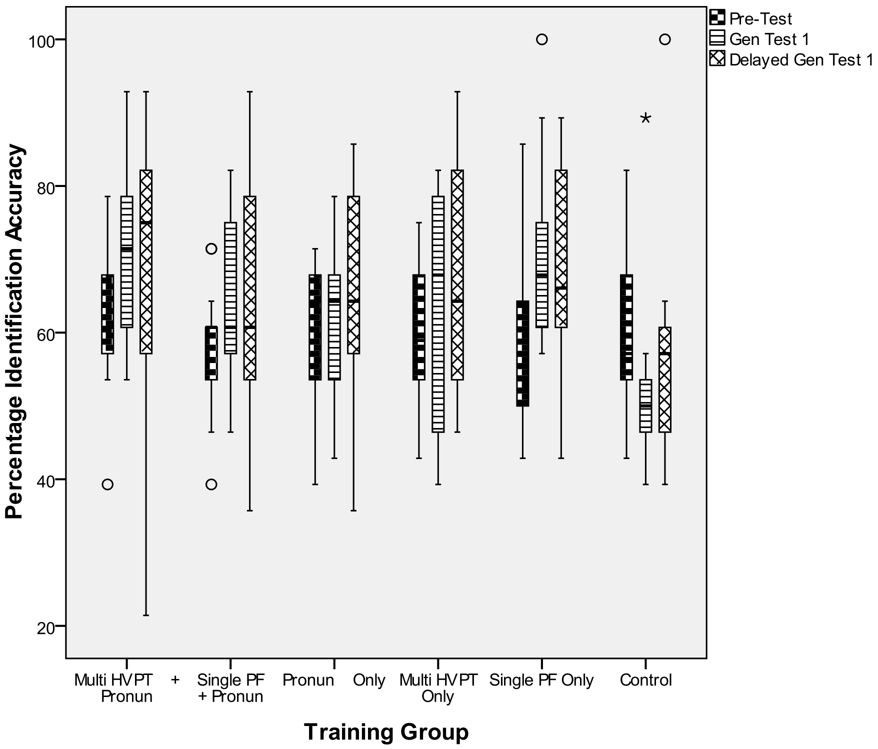
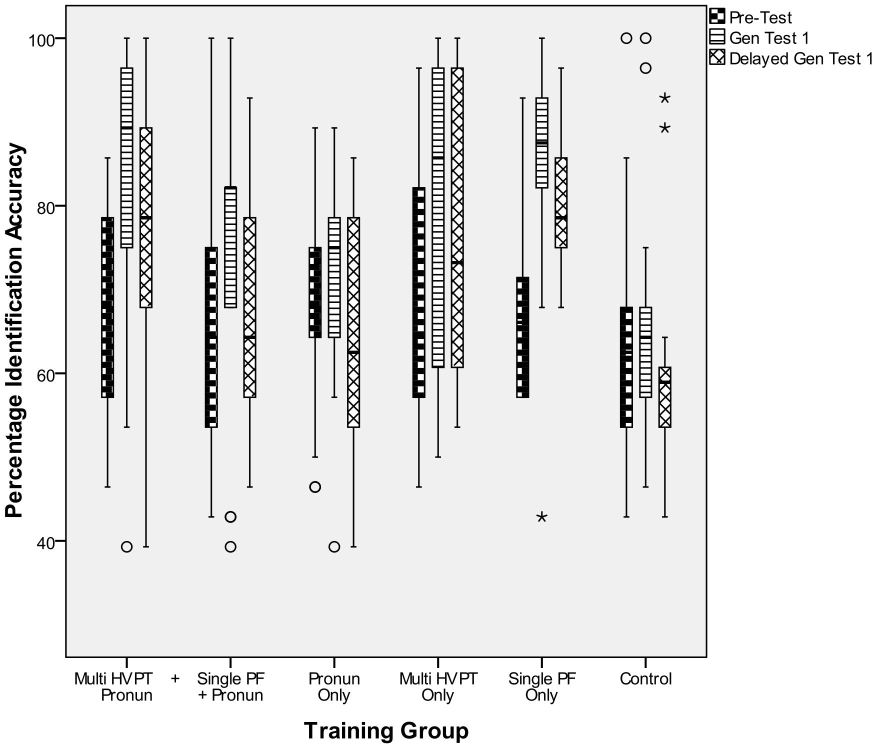


Figure 6.6: Oral Contrast Generalisation Test 1: Mean Percentage Perceptual Identification Accuracy Scores According to Training Group



As with the pre/post/delay data these data were then analysed using binary logistic mixed effects analysis in order to determine the significance of training group, test, contrast and time on whether or not participants answered correctly. The participant and test item were again included as random effects. The dependent variable was whether the participant correctly identified the minimal pair member presented. The fixed effects or potential predictors tested were participant training group/method (SubGp), test (Test, the three levels of this – pre, gen1, gen1T2, also represented time) and Contrast (oral vs. nasal). The model is detailed in Table 6.9.

Table 6.9: Fixed Effects and Coefficients of the Generalisation Test 1 Model

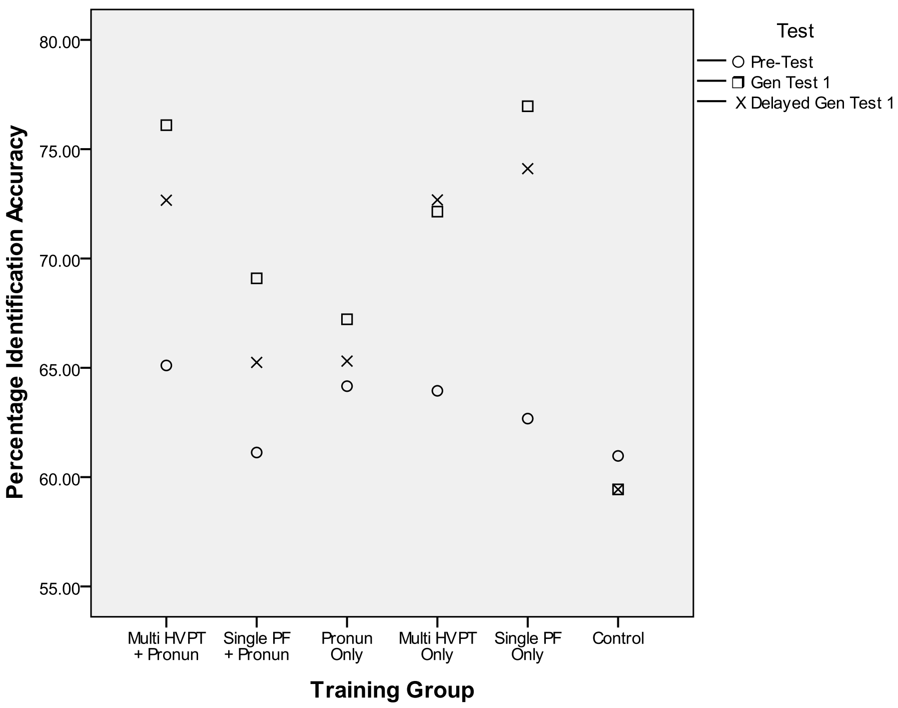
	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.389	0.192	2.031	0.0422**
SubGpMulti HVPT Only	0.051	0.229	0.224	0.8230
SubGpMulti HVPT + Pronun	0.073	0.229	0.318	0.7507
SubGpSingle PF + Pronun	-0.146	0.228	-0.640	0.5225
SubGpPronun Only	0.027	0.228	0.119	0.9054
SubGpSingle PF Only	0.026	0.233	0.113	0.9104
TestGen1	-0.107	0.161	-0.666	0.5055
TestGen1Delay	-0.144	0.163	-0.885	0.3764
ContrastOral u-y	0.446	0.129	3.467	0.0005****
SubGpMulti HVPT Only:TestGen1	0.673	0.157	4.283	<0.00001****
SubGpMulti HVPT + Pronun:TestGen1	0.670	0.157	4.257	<0.00001****
SubGpSingle PF + Pronun:TestGen1	0.485	0.153	3.161	0.0016***
SubGpPronun Only:TestGen1	0.371	0.152	1.594	0.0212**
SubGpSingle PF Only:TestGen1	0.696	0.160	4.352	<0.00001****
SubGpMulti HVPT Only:TestGen1Delay	0.661	0.170	3.900	<0.00001****
SubGpMulti HVPT + Pronun:TestGen1Delay	0.527	0.160	3.305	0.0009****
SubGpSingle PF + Pronun:TestGen1Delay	0.487	0.157	3.102	0.0019***
SubGpPronun Only:TestGen1Delay	0.361	0.154	2.340	0.0192**
SubGpSingle PF Only:TestGen1Delay	0.654	0.170	3.846	0.0001****

***p<.1; **p<.05; ***p<.01, ****p<.001**

As previously, the reference level for participant training group (SubGp) was the control group, the reference level for Contrast was the nasal contrast, and the reference level for Test was the pre-test. Looking firstly at the main effects, no group scored significantly differently from the control group overall, however the significant interactions show that the groups do differ dependent upon test (see below). The main effect of contrast shows that oral contrast test stimuli were answered correctly significantly more than nasal contrast test stimuli overall. However, there are no interaction terms with contrast in this model as adding the interaction term resulted in no significant interactions. This means that the conclusions drawn from interpretation of the other interactions in this model hold for both contrasts, and the data were thus collapsed across contrasts. Finally, the lack of a main effect of test suggests that the overall generalisation test 1 and delayed generalisation test 1 scores were not significantly greater than the pre-test scores.

As noted in Chapter 5, it is difficult to directly compare the post-training generalisation scores to the original Pre-Test score as it is conceivable that had this generalisation test been administered before training it may have resulted in a significantly larger score than the actual pre-test (perhaps due to being easier) at that point. However, again as noted in Chapter 5, if the training groups behave differently in these tests than in the pre-test as shown by significant interactions then this is likely to capture some generalisation of training for particular training groups. For example if some groups score greater than and some groups score less than the pre-test in the generalisation test after training, the random assignment of participants to groups mean this is likely to be capturing generalisation, as it is unlikely that those who may have scored higher or lower in any pre-training generalisation test than pre-test are all then assigned to the same training group. The post training interactions suggest that all groups scored significantly differently from the control group dependent upon whether the test was the pre-test or generalisation test 1 or the pre-test or the delayed generalisation test 1. Examination of Figure 6.7 demonstrates the exact nature of the interactions.

Figure 6.7: The Interaction Between Participant Training Group and Test



It can be seen that whilst the control group does not score significantly differently from the any of the other groups at pre-test, the group scores significantly less than the other groups in generalisation test 1 and in the delayed generalisation test 1. In other words, the participants in all groups outperform the participants in the control group at generalisation test 1 and in the delayed generalisation test 1. Therefore, all the training methods appear to work more effectively than no training in terms of producing and retaining generalisable perceptual improvement. In order to investigate whether any of the trained groups outperformed the others the analysis was re-run with the control group omitted. The model is detailed in Table 6.10 below.

Table 6.10: Fixed Effects and Coefficients of the Second Generalisation Test 1 Model

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.422	0.189	2.229	0.0258**
SubGpMulti HVPT + Pronun	0.022	0.224	0.097	0.9225
SubGpSingle PF + Pronun	-0.196	0.224	-0.878	0.3799
SubGpPronun Only	-0.023	0.223	-0.104	0.9175
SubGpSingle PF Only	-0.025	0.228	-0.109	0.9134
TestGen1	0.568	0.165	3.433	0.0005****
TestGen1Delay	0.519	0.176	2.948	0.0032***
ContrastOral u-y	0.475	0.126	3.773	0.0002****
SubGpMulti HVPT + Pronun:TestGen1	-0.002	0.165	-0.013	0.9899
SubGpSingle PF + Pronun:TestGen1	-0.188	0.161	-1.173	0.2410
SubGpPronun Only:TestGen1	-0.429	0.160	-2.681	0.007***
SubGpSingle PF Only:TestGen1	0.023	0.167	0.138	0.8902
SubGpMulti HVPT + Pronun:TestGen1Delay	-0.132	0.176	-0.751	0.4524
SubGpSingle PF + Pronun:TestGen1Delay	-0.173	0.174	-0.995	0.3196
SubGpPronun Only:TestGen1Delay	-0.299	0.172	-1.741	0.0817*
SubGpSingle PF Only:TestGen1Delay	-0.006	0.185	-0.033	0.9733

***p<.1; **p<.05; ***p<.01, ****p<.001**

As the reference control group had been removed, the reference group used for this model was again the multiple speaker HVPT group used as reference in the analysis in Chapter 5. There are no interaction terms with contrast in this model as adding the interaction term resulted in no significant interactions. The interactions and marginally significant interactions suggest that the only group which scored differently from the multiple speaker HVPT group dependent upon whether the test was the pre-test or generalisation test 1 or whether the test was the pre-test or generalisation test 2 is the pronunciation only group. Examination of Figure 6.7 suggests that while the pronunciation only group and the multiple speaker HVPT group do not differ at pre-test, the multiple speaker HVPT group scores significantly more than the pronunciation only group in generalisation test 1 and the delayed generalisation test 1.

In sum, it can be seen that the participants in the Multiple Speaker HVPT group outperform the participants in the pronunciation only group in generalisation test 1 and the delayed generalisation test 1. The Multiple Speaker HVPT group does not perform differently to any other group at any time. Therefore, whilst it is more effective than no training at all, the pronunciation only training method appears to be less effective than the other training methods in terms of generalisation and retaining this generalisation.

6.5.2.1 Group Comparison With Native Speakers

After removal of the control group the model was also re-run with the results from native listeners as the reference group in order to ascertain how the trainee groups differed from a group of native speakers. The performance of the remaining groups along with the native group collapsed across contrast (there continued to be no significant interactions in terms of contrast, see below) is illustrated in Figure 6.8 and the model is detailed in Table 6.11.

Figure 6.8: Mean Overall Percentage Perceptual Identification Accuracy Scores According to Remaining Training Groups and Native Speaker Group

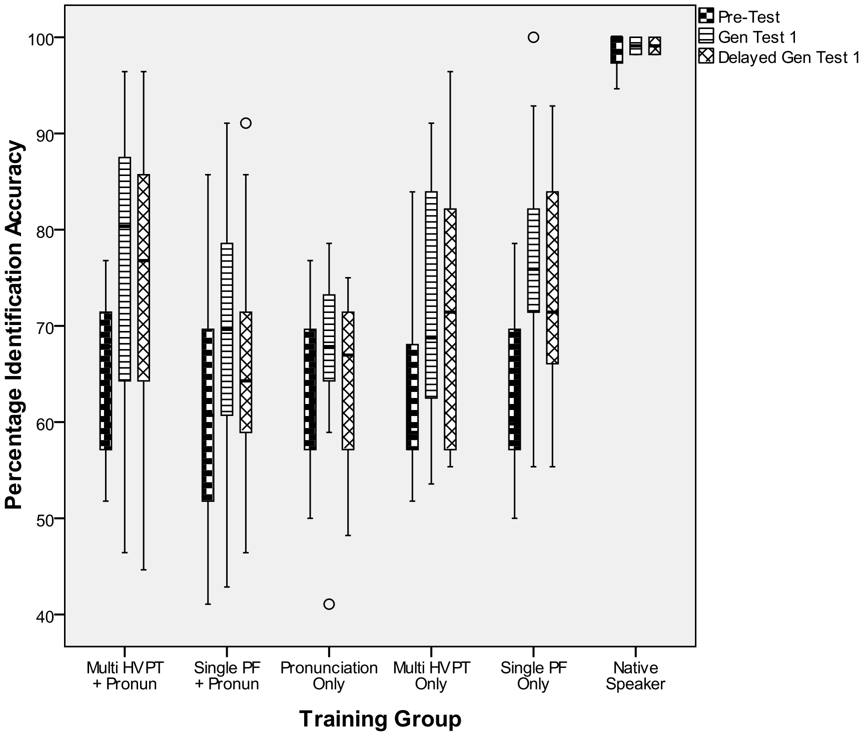


Table 6.11: Fixed Effects and Coefficients of the Native Speaker Generalisation Test 1 Model

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	4.279	0.482	8.878	<0.00001****
SubGpMulti HVPT Only	-3.782	0.492	-7.684	<0.00001****
SubGpMulti HVPT + Pronun	-3.795	0.492	-7.716	<0.00001****
SubGpSingle PF + Pronun	-4.086	0.492	-8.310	<0.00001****
SubGpPronunciation Only	-4.032	0.491	-8.203	<0.00001****
SubGpSingle PF Only	-3.799	0.494	-7.695	<0.00001****
ContrastOral u-y	0.476	0.125	3.792	0.0002****
TestGen1	0.439	0.128	3.423	0.0006***
TestGen1Delay	0.387	0.129	2.998	0.0027**

With the native speaker group as the reference group the remaining data are best explained by the main effects model. No interactions including training group and/or contrast were significant. This model suggests that the native speaker group scores significantly better than all other groups overall and the lack of significant interaction terms means that this is not dependent upon training group or time, in other words, no training technique results in a non-significant difference between native speaker group and any learner group.

6.5.3 Generalisation Test 2

The mean participant percentage perceptual identification accuracy and standard deviations for the second test of generalisation with new words and new voices are shown in Tables 6.12 and 6.13 and illustrated in Figures 6.9 and 6.10. The original pre-test scores are again included for reference. These data, as with the previous data analysed, suggest that the control group may perform worse than the other groups, and that the multiple speaker HVPT, single speaker PF and multiple speaker HVPT + pronunciation groups may perform well.

Table 6.12: Nasal Contrast Generalisation Test 2: Mean Percentage Perceptual Identification Accuracy Scores According to Training Group

Test	Training Group	Mean	SD	N
Pre-Test	Multi HVPT + Pronun	61.22	10.28	14
	Single PF + Pronun	56.12	10.62	14
	Pronunciation Only	61.23	9.79	14
	Multi Speaker HVPT	59.69	8.80	14
	Single Speaker PF	60.44	13.16	13
	Control	60.24	11.99	15
Gen Test 2 (New voices, new words)	Multi HVPT + Pronun	64.54	8.80	14
	Single PF + Pronun	55.87	8.70	14
	Pronunciation Only	59.18	11.18	14
	Multi Speaker HVPT	66.58	12.42	14
	Single Speaker PF	67.03	14.66	13
	Control	54.52	8.59	15
Delayed Gen Test 2 (New voices, new words)	Multi HVPT + Pronun.	64.56	13.24	13
	Single PF + Pronun.	54.95	12.59	13
	Pronunciation Only	60.72	10.76	14
	Multi Speaker HVPT	66.07	12.17	10
	Single Speaker PF	64.64	12.98	10
	Control	56.38	12.93	14

Table 6.13: Oral Contrast Generalisation Test 2: Mean Percentage Perceptual Identification Accuracy Scores According to Training Group

Test	Training Group	Mean	SD	N
Pre-Test	Multi HVPT + Pronun.	68.88	13.17	14
	Single PF + Pronun.	63.52	17.05	14
	Pronunciation Only	67.09	12.78	14
	Multi Speaker HVPT	68.40	15.68	14
	Single Speaker PF	67.03	9.91	13
	Control	65.00	16.65	15
Gen Test 2 (New voices, new words)	Multi HVPT + Pronun.	82.40	8.57	14
	Single PF + Pronun.	75.51	12.98	14
	Pronunciation Only	73.47	13.27	14
	Multi Speaker HVPT	78.57	15.22	14
	Single Speaker PF	76.99	16.17	13
	Control	71.19	13.53	15
Delayed Gen Test 2 (New voices, new words)	Multi HVPT + Pronun.	80.49	12.76	13
	Single PF + Pronun.	67.31	13.35	13
	Pronunciation Only	66.58	12.10	14
	Multi Speaker HVPT	77.50	17.98	10
	Single Speaker PF	81.79	8.66	10
	Control	61.48	13.81	14

Figure 6.9: Nasal Contrast Generalisation Test 2: Mean Percentage Perceptual Identification Accuracy Scores According to Training Group

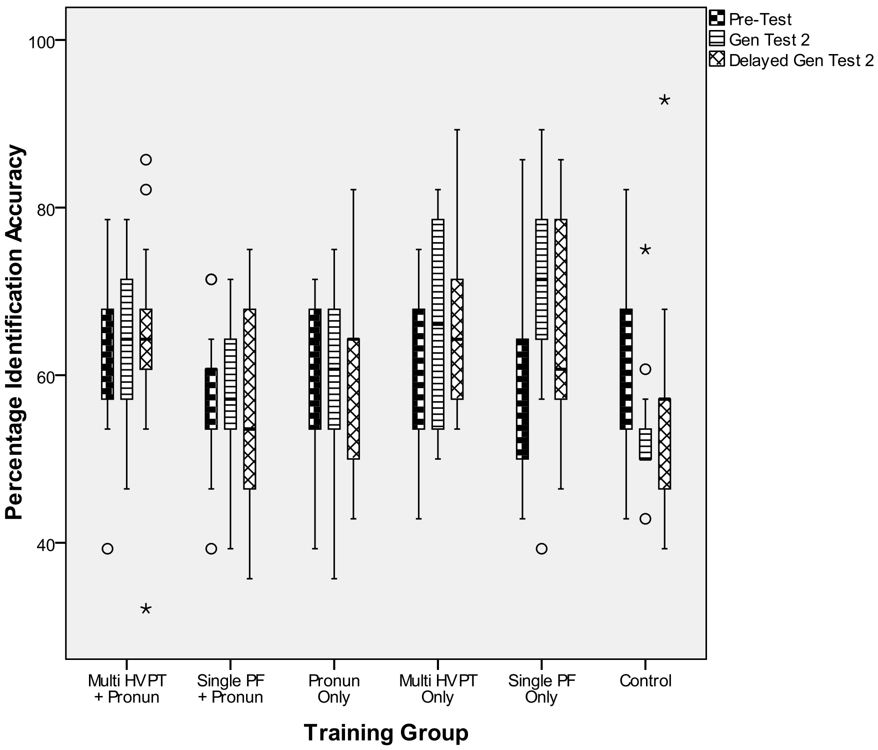
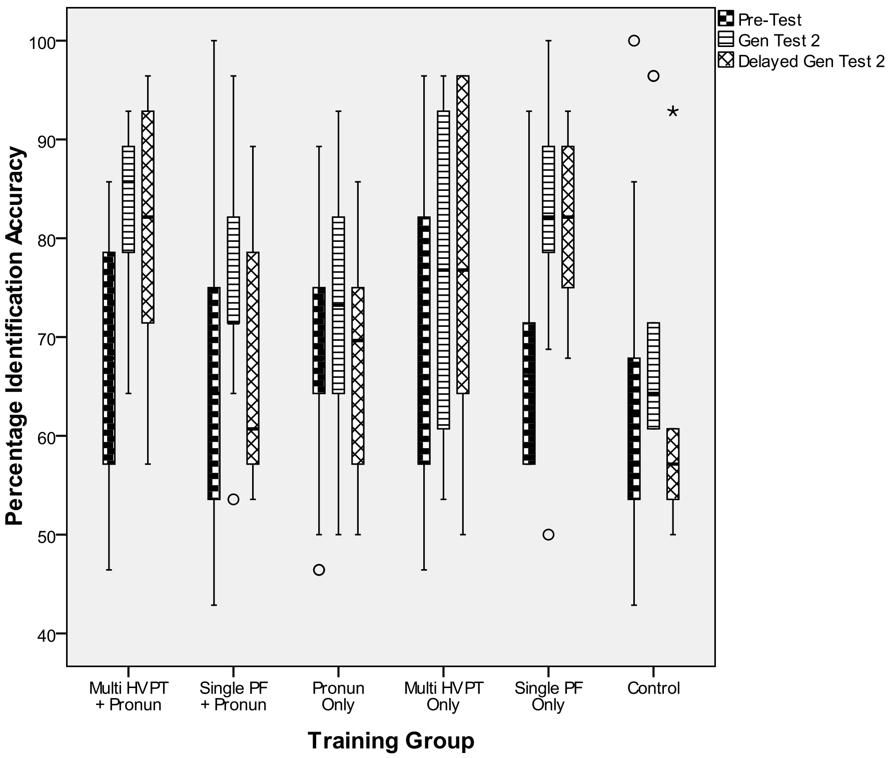


Figure 6.10: Oral Contrast Generalisation Test 2: Mean Percentage Perceptual Identification Accuracy Scores According to Training Group



As with the previous data, the generalisation test 2 data were then analysed using binary logistic mixed effects analysis in order to determine the significance of training group, test, contrast and time on whether or not participants answered correctly. The participant and test item were again included as random effects. The dependent variable was whether the participant correctly identified the minimal pair member presented. The fixed effects or potential predictors tested were participant training group/method (SubGp), test (Test, the three levels of this – pre, gen2, gen2T2, also represented time) and Contrast (oral vs. nasal). The model is detailed in Table 6.14

Table 6.14: Fixed Effects and Coefficients of the Generalisation Test 2 Model

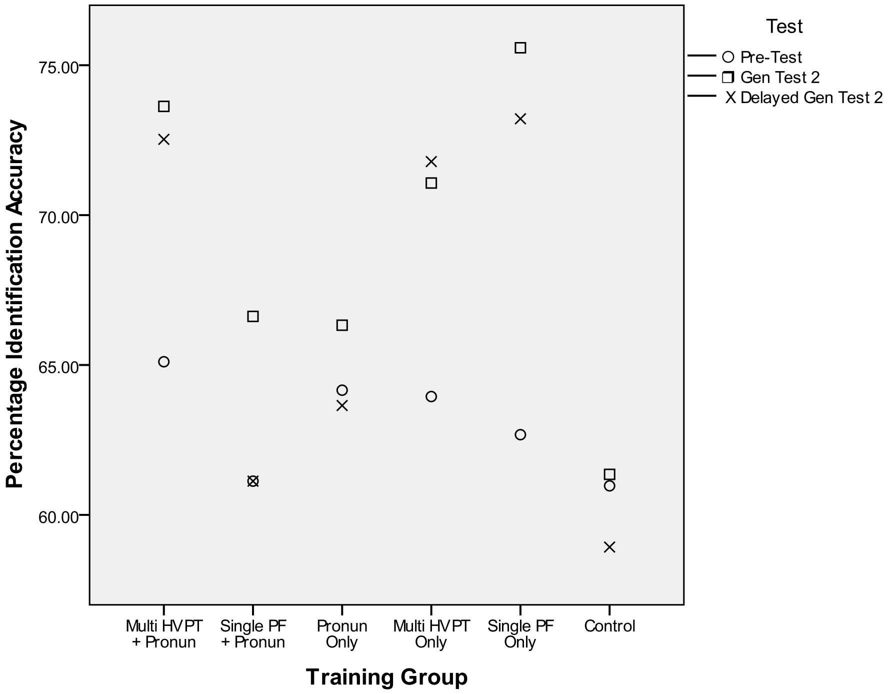
	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.315	0.164	1.918	0.0551*
SubGpMulti HVPT Only	0.053	0.183	0.293	0.7698
SubGpMulti HVPT + Pronun	0.071	0.183	0.388	0.6979
SubGpSingle PF + Pronun	-0.143	0.183	-0.786	0.4320
SubGpPronun Only	0.045	0.183	0.247	0.8047
SubGpSingle PF Only	0.035	0.187	0.190	0.8496
TestGen2	-0.006	0.158	-0.039	0.9687
TestGen2Delay	-0.117	0.159	-0.738	0.4604
ContrastOral u-y	0.547	0.124	4.397	<0.00001****
SubGpMulti HVPT Only:TestGen2	0.427	0.155	2.763	0.0057***
SubGpMulti HVPT + Pronun:TestGen2	0.439	0.155	2.839	0.0045***
SubGpSingle PF + Pronun:TestGen2	0.274	0.151	1.811	0.0701*
SubGpPronun Only:TestGen2	0.101	0.152	0.663	0.5073
SubGpSingle PF Only:TestGen2	0.409	0.157	2.596	0.0094***
SubGpMulti HVPT Only:TestGen2Delay	0.508	0.167	3.050	0.0023***
SubGpMulti HVPT + Pronun:TestGen2Delay	0.496	0.157	3.153	0.0016***
SubGpSingle PF + Pronun:TestGen2Delay	0.362	0.155	2.340	0.0193**
SubGpPronun Only:TestGen2Delay	0.342	0.153	2.228	0.0259**
SubGpSingle PF Only:TestGen2Delay	0.508	0.168	3.025	0.0025***

***p<.1; **p<.05; ***p<.01, ****p<.001**

As with previous models, the reference level for participant training group (SubGp) was the control group, the reference level for Contrast was the nasal contrast, and the reference level for Test was the pre-test. Looking firstly at the main effects, no group scored significantly differently from the control group overall, however the significant interactions show that the groups do differ dependent upon test (see below). The main effect of contrast shows that oral contrast test stimuli were answered correctly significantly more than nasal contrast test stimuli overall. However, there are no interaction terms with Contrast in this model as adding the interaction term resulted in no significant interactions. This means that the conclusions drawn from interpretation of the other interactions in this model hold for both contrasts, the data were thus again collapsed across contrasts. In addition, examining the main effect of test, overall the post-test and delayed post-test scores were not significantly greater than the pre-test scores.

The post-test interactions suggest that the only group which has not scored differently from the control group dependent upon whether the test was the pre-test or generalisation test 2 is the pronunciation only group. However, looking at the delayed generalisation test 2 scores, all groups scored significantly differently from the control group dependent upon whether the test was the pre-test or the delayed generalisation test 2. The caveats to this comparison of tests in 6.5.2 with reference to generalisation test 1 also apply here. An inspection of Figure 6.11 demonstrates the exact nature of the interactions.

Figure 6.11: The Interaction Between Participant Training Group and Test



It can be seen that whilst the control group does not score significantly differently from the any of the other groups at pre-test, the group scores significantly less than the other perceptually trained groups in generalisation test 2. In the delayed generalisation test 2 the control group scores significantly less then all trained groups. In other words, the participants in all groups but the pronunciation only group outperform the participants in the control group at generalisation test 2 and in all groups at the delayed generalisation test 2. Therefore, all the perceptual training methods appear to work more effectively than no training in terms of producing generalisable perceptual improvement and all the methods appear to work more effectively than no training in terms of retaining generalisable perceptual improvement. In order to investigate whether any of the trained groups outperformed the others the analysis was re-run with the control group omitted. The model is detailed in Table 6.15 below.

Table 6.15: Fixed Effects and Coefficients of the Second Generalisation Test 2 Model

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.344	0.160	2.155	0.0312**
SubGpMulti HVPT + Pronun	0.018	0.176	0.105	0.9165
SubGpSingle PF + Pronun	-0.196	0.175	-1.119	0.2630
SubGpPronun Only	-0.007	0.175	-0.041	0.9671
SubGpSingle PF Only	-0.018	0.179	-0.102	0.9184
TestGen2	0.426	0.161	2.654	0.0080***
TestGen2Delay	0.395	0.171	2.305	0.0211**
ContrastOral u-y	0.588	0.122	4.821	<0.00001****
SubGpMulti HVPT + Pronun:TestGen2	0.013	0.160	0.078	0.9376
SubGpSingle PF + Pronun:TestGen2	-0.154	0.157	-0.982	0.3262
SubGpPronun Only:TestGen2	-0.327	0.157	-2.076	0.0379**
SubGpSingle PF Only:TestGen2	-0.019	0.163	-0.117	0.9071
SubGpMultiHVPT+Pronun:TestGen2Delay	-0.010	0.172	-0.058	0.9541
SubGpSingle PF + Pronun:TestGen2Delay	-0.145	0.170	-0.852	0.3943
SubGpPronun Only:TestGen2Delay	-0.295	0.169	-0.975	0.0496**
SubGpSingle PF Only:TestGen2Delay	0.004	0.182	0.022	0.9823

***p<.1; **p<.05; ***p<.01, ****p<.001**

As the reference control group had been removed, the reference group used for this model was again the multiple speaker HVPT group used as reference in the analysis in Chapter 5. There are no interaction terms with contrast in this model as adding the interaction term resulted in no significant interactions. The interactions suggest that the only group which scored differently from the multiple speaker HVPT group dependent upon whether the test was the pre-test or generalisation test 2 or dependent upon whether the test was the pre-test or delayed generalisation test 2 is the pronunciation only group. Examination of Figure 6.11 suggests that while the pronunciation only group and the multiple speaker HVPT group do not differ at pre-test, the multiple speaker HVPT group may score significantly more than the pronunciation only group in generalisation test 2 and the delayed generalisation test 2.

In sum, it can be seen that the participants in the Multiple Speaker HVPT group outperform the participants in the pronunciation only group in generalisation test 2. The Multiple Speaker HVPT group does not perform differently to any other group at any time. Therefore, whilst it is more effective than no training at all, the pronunciation only training method appears to be less effective than the other training methods in terms of generalisation.

6.5.3.1 Group Comparison With Native Speakers

After removal of the control group the model was again also re-run with the results from native listeners as the reference group in order to ascertain how the trainee groups differed from a group of native speakers. The performance of the remaining groups along with the native group collapsed across contrast (there continued to be no significant interactions in terms of contrast, see below) is illustrated in Figure 6.12 and the model is detailed in Table 6.16.

Figure 6.12: Mean Overall Percentage Perceptual Identification Accuracy Scores According to Remaining Training Groups and Native Speaker Group

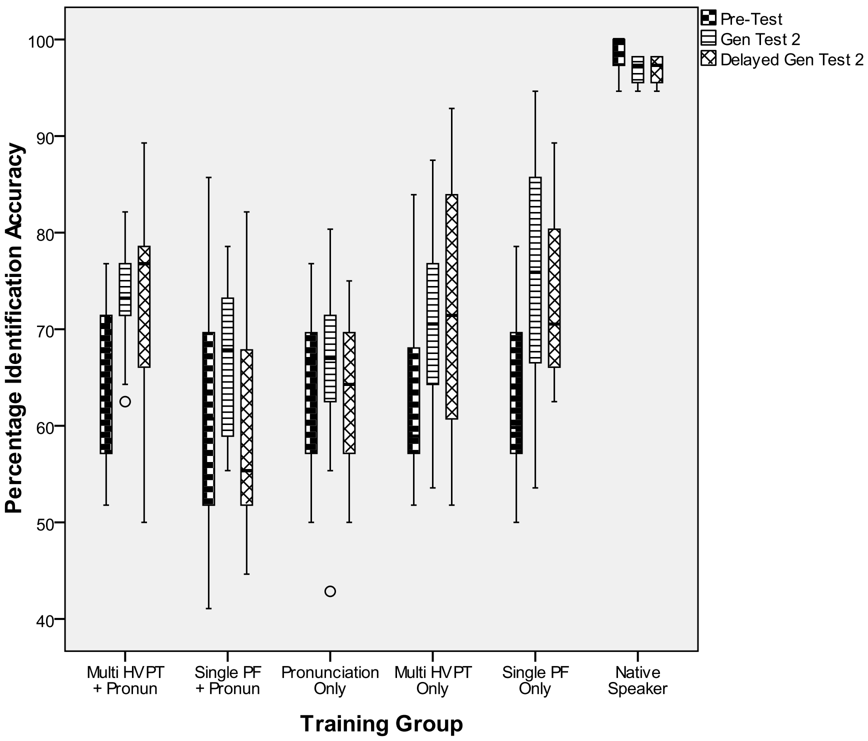


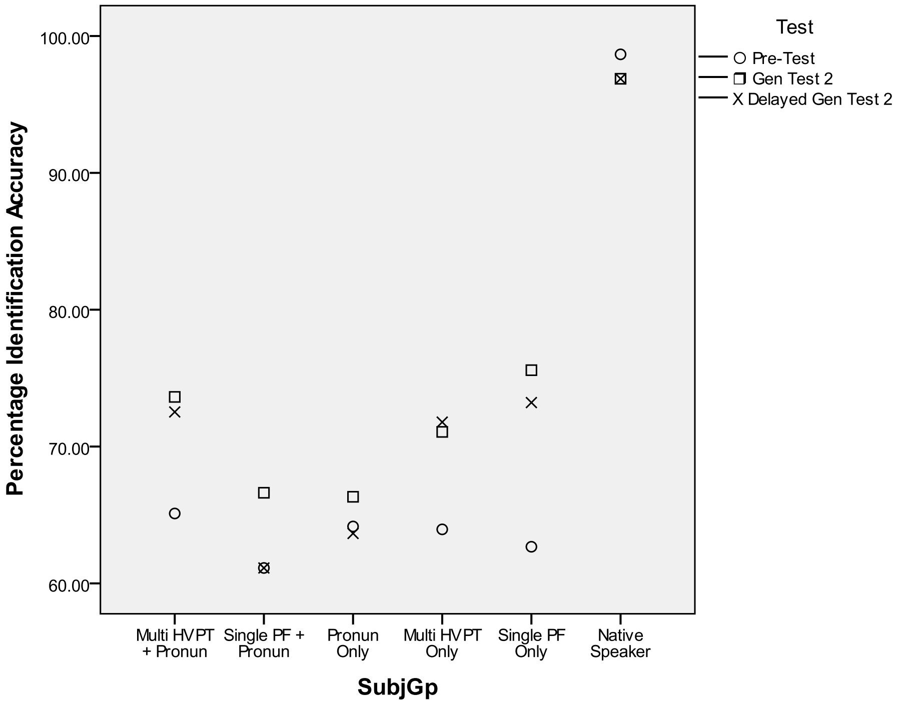
Table 6.16: Fixed Effects and Coefficients of the Native Speaker Generalisation Test 2 Model

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	4.179	0.628	6.656	<0.00001****
SubGpMulti HVPT Only	-3.839	0.632	-6.075	<0.00001****
SubGpMulti HVPT + Pronun	-3.821	0.632	-6.046	<0.00001****
SubGpSingle PF + Pronun	-4.035	0.632	-6.387	<0.00001****
SubGpPronunciation Only	-3.846	0.632	-6.087	<0.00001****
SubGpSingle PF Only	-3.857	0.633	-6.096	<0.00001****
TestGen2	-0.882	0.719	-1.227	0.2197
TestGen2Delay	-0.882	0.719	-1.227	0.2197
ContrastOral u-y	0.597	0.122	4.872	<0.00001****
SubGpMulti HVPT Only:TestGen2	1.310	0.719	1.822	0.0684*
SubGpMultiHVPT+ Pronun:TestGen2	1.322	0.719	1.840	0.0658*
SubGpSingle PF + Pronun:TestGen2	1.155	0.718	1.609	0.1077
SubGpPronunciation Only:TestGen2	0.983	0.718	1.368	0.1712
SubGpSingle PF Only:TestGen2	1.290	0.719	1.794	0.0728*
SubGpMultiHVPT Only:TestGen2Delay	1.278	0.721	1.772	0.0764*
SubGpMultiHVPT+ Pronun:TestGen2Delay	1.268	0.719	1.764	0.0778*
SubGpSinglePF+ Pronun:TestGen2Delay	1.133	0.718	1.577	0.1148
SubGpPronunciation Only:TestGen2Delay	1.113	0.718	1.550	0.1211
SubGpSinglePFOnly:TestGen2Delay	1.282	0.721	1.777	0.0755*

***p<.1; **p<.05; ***p<.01, ****p<.001**

When examining performance on generalisation test 2 with reference to a native speaker group there are marginal interactions worthy of investigation. These marginal interactions suggest that the multiple speaker HVPT, multiple speaker HVPT + pronunciation and single speaker PF groups approach not scoring significantly differently from the native speaker group dependent upon whether the test was the pre-test or generalisation test 2 and dependent upon whether the test was the pre-test or the delayed generalisation test 2 (bearing in mind the previous caveats). An inspection of Figure 6.13 demonstrates the exact nature of the interactions.

Figure 6.13: The Interaction Between Participant Training Group and Test with Native Speakers Included



It can be seen that whilst the multiple speaker HVPT, multiple speaker HVPT + pronunciation and single speaker PF groups do differ significantly from the native group in terms of pre-test scores, the difference between these groups and the native group are approaching non-significance in generalisation test 2 and delayed generalisation test 2, despite the difference being numerically large. In sum it can be very tentatively suggested that these training techniques obtain a move towards a native standard which the others do not in terms of generalisation to new words and new voices and retention of this generalisation.

6.5.4 Individual Comparison with Native Speakers

Further comparison with native speaker results at an individual level rather than a group level was also of interest, in order to ascertain whether anyone performed to a native standard either before or after training. For simplicity, a native standard was taken to be within the native range of scores. Perceptually, none of the 84 participants scored at a native level in all tests, however, there was evidence of nativelike performance in some cases. Whilst no participant performed within the native range for the nasal contrast pre-test two participants scored 100% (as all the native listeners did) in the oral contrast pre-test. In terms of post-training scores, across all 6 post training tests (post-test, generalisation test 1, generalisation test 2, post-test T2, generalisation test 1 T2, generalisation test 2 T2), no participants scored within the native range for all 12 tests (6 tests for each contrast). The 'best' participant scored within the native range in 3/6 nasal contrast tests and 4/6 oral contrast tests, with the numbers for the next best participant being 2/6 and 3/6 respectively. Of all 84 participants, 8 scored within the native range for at least one of the tests in both contrasts and a further 15 scored within the native range in at least one of the tests for one of the contrasts. Overall, whilst no participants scored within the native range across all tests, these results suggest that training did result in an improvement towards a native standard in some instances.

6.5.5 Perceptual Results Discussion

There is at best weak evidence of transfer of pronunciation training only to perception performance from the current results. At post-test and delayed post-test this group does not perform significantly differently to the untrained control group. In the first test of generalisation there is evidence to suggest that pronunciation training did result in some perceptual improvement, however the pronunciation group was the only group to perform worse than the multiple speaker HVPT group. Similarly in the second test of generalisation the pronunciation only group's performance was also significantly worse than the multiple speaker HVPT group. In addition, as found in Chapter 5, there were no interactions with contrast trained. While some main effects demonstrate that the oral contrast is easier (higher scoring perceptual accuracy) than the nasal contrast, the pattern of results is similar across both contrasts.

The removal of some perceptual training does not appear detrimental to perceptual performance, although the comparison with native speakers in the second test of generalisation suggests that this is marginally truer of the multiple speaker HVPT + pronunciation group than the single speaker PF + pronunciation group. The former technique results in a generalisation test 2 score closer to a native standard than the latter. The success of the training techniques featuring some perceptual training suggest that, unsurprisingly, some perceptual training is more effective than pronunciation training at improving perceptual performance.

With regard to answering the research questions posed at the start of this chapter (RQ2: are the most successful perceptual training techniques suggested through answering Research Question 1 more successful than pronunciation training in terms of producing a generalisable, long-term improvement in pronunciation and perception?; RQ3: With regard to using the multi speaker HVPT technique and/or the single speaker PF technique and/or pronunciation training, does an optimal training technique emerge from those examined in terms of producing a generalisable, long-term improvement in pronunciation and perception?) regarding improvement in perception, there is evidence to suggest that the single speaker PF, multiple speaker HVPT and the multiple speaker HVPT + pronunciation groups (and to a slightly lesser extent the single speaker PF + pronunciation group) are most successful at producing a generalisable improvement in perception. With regard to the effect of pronunciation training only on perception performance, the results do suggest that pronunciation training alone is not sufficient to achieve notable improvements in perception. Interestingly, replacing half of the single speaker PF training with pronunciation training appears to have a slightly more negative effect on perceptual performance than replacing half of the multiple speaker HVPT training with pronunciation training, suggesting that multiple speaker HVPT training may be more efficient.

The next section deals with how accurately the native speakers of French analysing the participant pronunciation could identify the word the participant was attempting to pronounce.

6.6 Pronunciation Results 1: Native Speaker Identification Accuracy

The identification accuracy data for each French learning participant (and French native speaker) for each pronunciation test undertaken was obtained. Next, the results from the 41 analysis sessions were averaged for the French learning participant whom everyone analysed and for both the native French speakers whom everyone also analysed. This resulted in a native speaker percentage identification accuracy score for each participant's pre-test, post-test, generalisation test, post-test (T2) and generalisation test (T2) word productions.

Before carrying out the analysis, it was firstly necessary to ensure that any differences between training groups after training in terms of native speaker identification accuracy were attributable to the training technique and not due to one group having a higher or lower pre-test native speaker identification accuracy score. To this end, two one-way ANOVAs were carried out with pre-test score as the dependent variable and training group as the between subjects factor, one ANOVA for each contrast of interest. There was no significant effect of training group for either contrast [Nasal: $F(5,82) = .700$, $p = .625$; Oral: $F(5,82) = .198$, $p = .962$] indicating that the pre-test scores were not significantly different for each group.

Binary logistic mixed effects analyses of the relationship between training groups, tests, contrast tested and time was then carried out in two blocks. The first block examined the effects of training across the pre-test, post-test and delayed post-test. The second block examined the generalisation test. Unless otherwise stated, likelihood ratio tests comparing each model with fixed effects to a null model with only the random effects demonstrated that the fixed effects model differed significantly from the null model.

6.6.1 Block 1: Pre-Test, Post-Test, Delay

The mean native speaker percentage perceptual identification accuracy and standard deviations of participant productions for each test (pre-test, post-test, and delayed post-test according to participant training group and contrast are shown in Tables 6.17 and 6.18 and illustrated in Figures 6.14. and 6.15. These data suggest that the control group may perform worse than the other groups, and that the pronunciation and multiple speaker HVPT + pronunciation groups may perform well.

Table 6.17: Nasal Contrast: Mean Percentage Identification Accuracy Scores by Native Speakers According to Participant Training Group

Test	Training Group	Mean	SD	N
Pre-Test	Multi HVPT + Pronun.	63.10	19.81	14
	Single PF + Pronun.	60.71	15.48	14
	Pronunciation Only	57.14	14.19	14
	Multi Speaker HVPT	69.23	23.42	13
	Single Speaker PF	58.54	14.96	13
	Control	62.22	21.33	15
Post-Test	Multi HVPT + Pronun.	67.86	23.08	14
	Single PF + Pronun.	65.48	21.15	14
	Pronunciation Only	70.24	19.81	14
	Multi Speaker HVPT	67.95	20.93	13
	Single Speaker PF	62.54	17.59	13
	Control	63.33	16.90	15
Delayed Post-Test	Multi HVPT + Pronun.	70.51	23.72	13
	Single PF + Pronun.	60.26	18.68	13
	Pronunciation Only	67.86	22.13	14
	Multi Speaker HVPT	65.00	22.84	10
	Single Speaker PF	61.14	21.32	10
	Control	63.10	18.70	14

Table 6.18: Oral Contrast: Mean Percentage Identification Accuracy Scores by Native Speakers According to Participant Training Group

Test	Training Group	Mean	SD	N
Pre-Test	Multi HVPT + Pronun.	63.10	20.86	14
	Single PF + Pronun.	60.71	18.03	14
	Pronunciation Only	65.48	15.28	14
	Multi Speaker HVPT	66.67	22.57	13
	Single Speaker PF	61.76	18.57	13
	Control	65.56	22.24	15
Post-Test	Multi HVPT + Pronun.	79.76	21.86	14
	Single PF + Pronun.	67.86	19.02	14
	Pronunciation Only	76.19	21.40	14
	Multi Speaker HVPT	73.08	22.10	13
	Single Speaker PF	76.74	17.14	13
	Control	67.78	20.38	15
Delayed Post-Test	Multi HVPT + Pronun.	80.77	23.42	13
	Single PF + Pronun.	71.80	17.19	13
	Pronunciation Only	77.38	20.26	14
	Multi Speaker HVPT	68.33	25.40	10
	Single Speaker PF	66.38	21.76	10
	Control	60.71	14.03	14

Figure 6.14: Nasal Contrast: Mean Percentage Identification Accuracy Scores by Native Speakers According to Participant Training Group

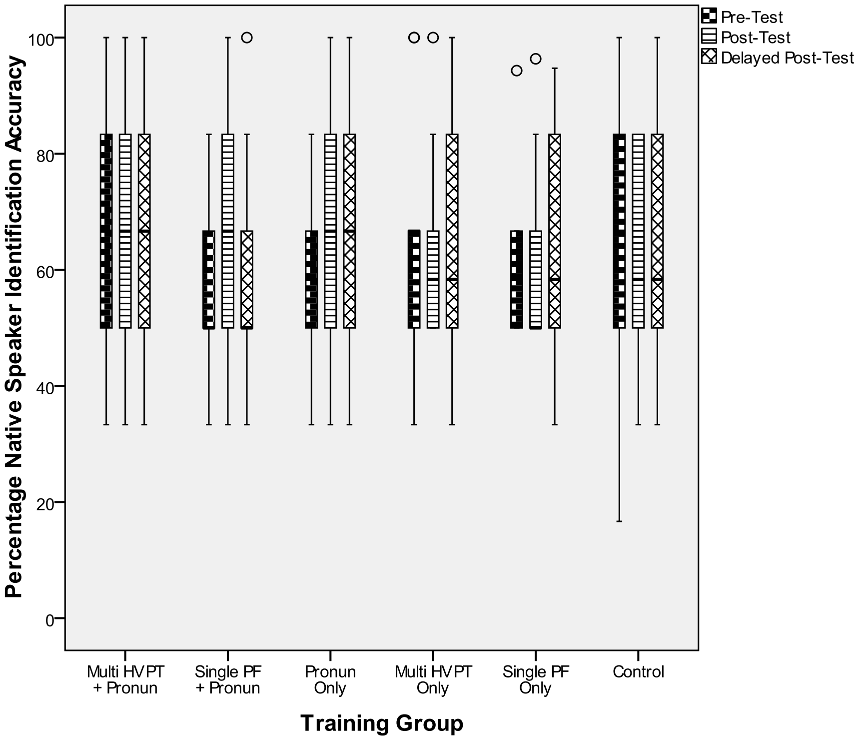
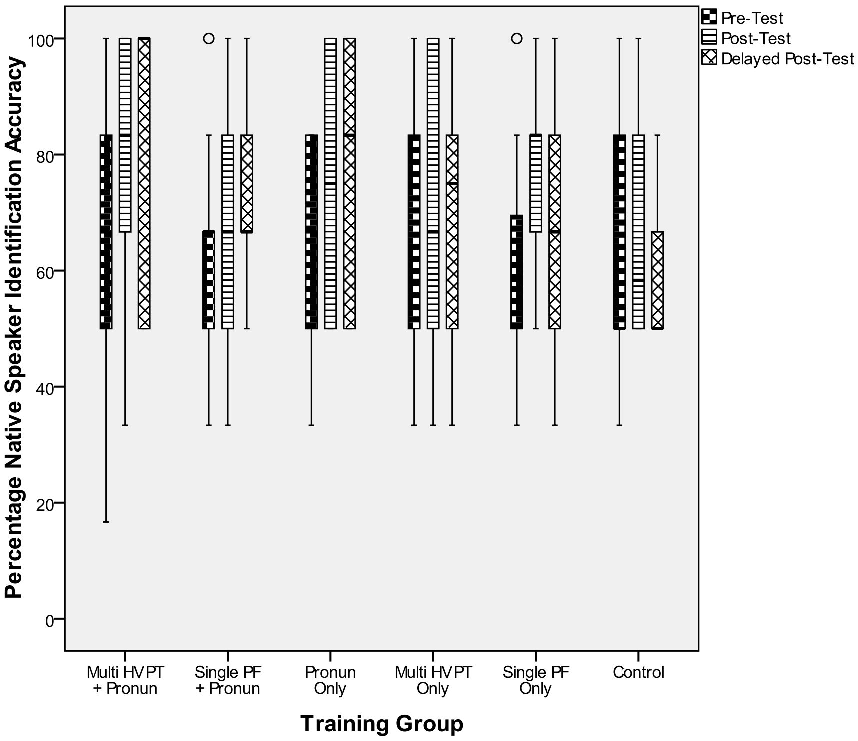


Figure 6.15: Oral Contrast: Mean Percentage Identification Accuracy Scores by Native Speakers According to Participant Training Group



As with the perceptual data, these pronunciation identification accuracy data were then analysed using binary logistic mixed effects analysis in order to determine the significance of training group, test, contrast and time on whether or not native speakers identified the correct member of the minimal pair from participant productions. In the present model the participant and test item were included as random effects. The dependent variable was whether the native speaker correctly identified the minimal pair member presented. The fixed effects tested were participant training group/method (SubjGp), test (Test, the three levels of this – pre, post, delay, also represented time) and Contrast (oral vs. nasal). The model is detailed in Table 5.19.

Table 6.19: Fixed Effects and Coefficients of the Pre-Test, Post-Test, Delay Model

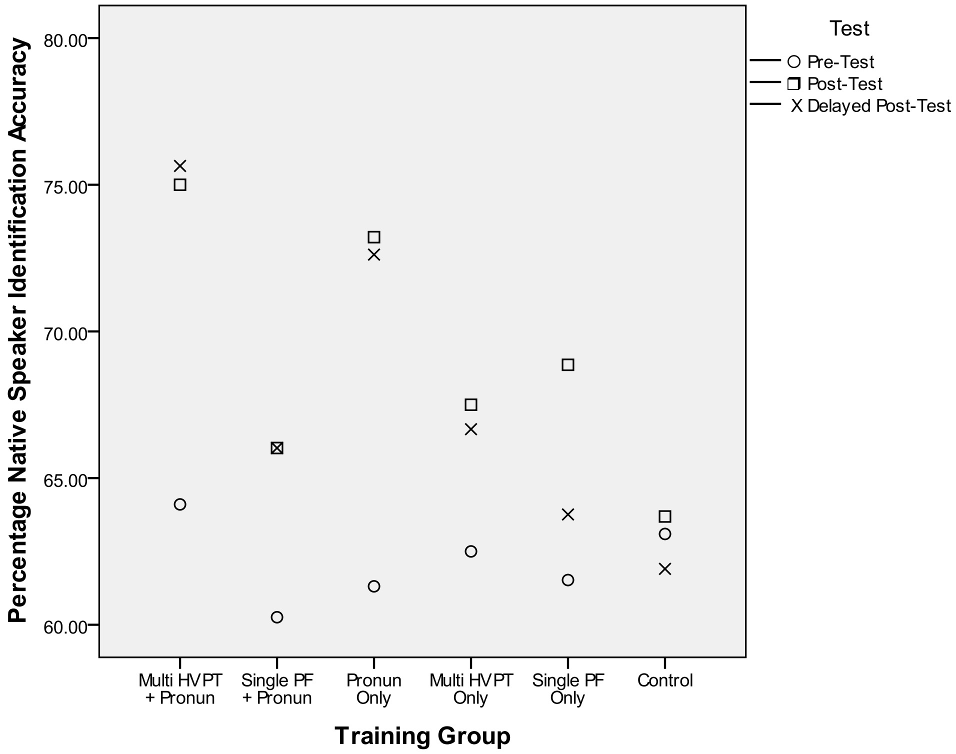
	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.524	0.414	1.265	0.2059
SubjGpMulti HVPT Only	0.250	0.348	0.717	0.4733
SubjGpMulti HVPT + Pronun	-0.033	0.338	-0.098	0.9223
SubjGpSingle PF + Pronun	-0.168	0.336	-0.502	0.6159
SubjGpPronun Only	-0.139	0.336	-0.413	0.6795
SubjGpSingle PF Only	-0.293	0.348	-0.842	0.3998
TestPost	0.088	0.240	0.369	0.7125
TestPostDelay	-0.058	0.243	-0.241	0.8098
SubjGpMulti HVPT Only:TestPost	0.060	0.360	0.166	0.8686
SubjGpMulti HVPT + Pronun:TestPost	0.525	0.353	1.485	0.1374
SubjGpSingle PF + Pronun:TestPost	0.223	0.344	0.648	0.5169
SubjGpPronun Only:TestPost	0.564	0.349	1.615	0.1064
SubjGpSingle PF Only:TestPost	0.376	0.357	1.054	0.2920
SubjGpMulti HVPT Only:TestPostDelay	0.135	0.376	0.358	0.7200
SubjGpMulti HVPT + Pronun:TestPostDelay	0.750	0.363	2.064	0.0390**
SubjGpSingle PF + Pronun:TestPostDelay	0.354	0.350	1.012	0.3115
SubjGpPronun Only:TestPostDelay	0.675	0.351	1.925	0.0542*
SubjGpSingle PF Only:TestPostDelay	0.162	0.376	0.431	0.6666

***p<.1; **p<.05; ***p<.01, ****p<.001**

As with the perceptual data in this chapter, the reference level for participant training group (SubjGp) was the control group, the reference level for Contrast was the nasal contrast, and the reference level for Test was the Pre-Test. The main effect of contrast was excluded from the model as it was not significant (p = 0.44) and addition of the interaction term resulted in no significant interactions. As with the perceptual data, this means that the conclusions drawn from interpretation of the other interactions in this model hold for both contrasts, and the data were thus collapsed across contrasts.

The significant interactions show that the groups do differ dependent upon test. Whilst no group performs significantly differently from the control group in the post-test, the interactions from the delayed post-test data suggest that the multiple HVPT + pronunciation group and, marginally, the pronunciation group score differently from the control dependent upon whether the test was the pre-test or the delayed post-test. Examination of Figure 6.16 demonstrates the exact nature of the interactions.

Figure 6.16: The Interaction Between Participant Training Group and Test



With reference to the delayed post-test interactions, it can be seen that whilst the control group does not score significantly differently from the multiple HVPT + pronunciation group and the pronunciation only group at pre-test (nor, apparently, differ enough at post-test), the control group scores significantly less than these two trained groups at the delayed post-test. In other words the multiple speaker HVPT + pronunciation technique and the pronunciation only technique appear to work better than no training at all in terms of retaining the training effects.

6.6.1.1 Group Comparison With Native Speakers

As with the perceptual data the model was also re-run with the results from natives listening to native speakers as the reference group in order to ascertain how the trainee groups' pronunciation differed from a group of native speakers. The performance of the groups along with the native group collapsed across contrast (there continued to be no significant interactions in terms of contrast, see below) is illustrated in Figure 6.17 and the model is detailed in Table 6.20.

Figure 6.17: Mean Overall Percentage Perceptual Identification Accuracy Scores According to Training Groups and Native Speaker Group

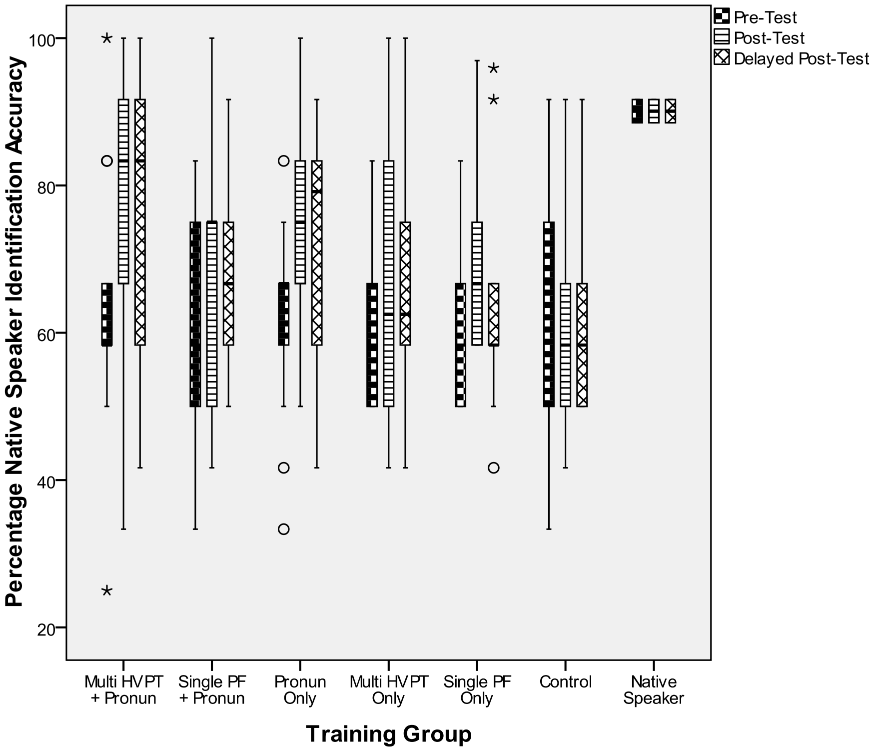


Table 6.20: Fixed Effects and Coefficients of the Native Speaker Pre-Test, Post-Test, Delay Model

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	2.338	0.712	3.287	0.0010***
SubjGp2Multi HVPT + Pronun	-1.660	0.654	-2.539	0.0111**
SubjGp2Single PF + Pronun	-2.010	0.653	-3.079	0.0021***
SubjGp2Pronunciation Only	-1.768	0.653	-2.708	0.0068***
SubjGp2Multi HVPT Only	-1.711	0.657	-2.606	0.0092***
SubjGp2Single PF Only	-2.132	0.658	-3.240	0.0012***
SubjGp2Control	-2.028	0.651	-3.116	0.0018***
Test2Post	0.371	0.104	3.586	0.0003****
Test2PostDelay	0.288	0.107	2.694	0.0071***

***p<.1; **p<.05; ***p<.01, ****p<.001**

With the native speaker group as the reference group the data are best explained by the main effects model. No interactions including training group and/or contrast were significant. As the main effect of contrast was not significant this was also not included in the model. This model suggests that the native speaker group scores significantly better than all other groups overall and the lack of significant interaction terms means that this is not dependent upon training group or time, in other words, no training technique results in a non-significant difference between native speaker group and any learner group.

6.6.2 Generalisation Test

The mean participant percentage perceptual identification accuracy and standard deviations for the generalisation test with new words are shown in Tables 6.21 and 6.22 and illustrated in Figures 6.18 and 6.19. The original pre-test scores are also included for reference.

Table 6.21: Nasal Contrast Generalisation Test : Mean Percentage Perceptual Identification Accuracy Scores According to Training Group

Test	Training Group	Mean	SD	N
Pre-Test	Multi HVPT + Pronun.	63.10	19.81	14
	Single PF + Pronun.	60.71	15.48	14
	Pronunciation Only	57.14	14.19	14
	Multi Speaker HVPT	69.23	23.42	13
	Single Speaker PF	58.54	14.96	13
	Control	62.22	21.33	15
Gen Test (New words)	Multi HVPT + Pronun.	67.86	21.15	14
	Single PF + Pronun.	60.71	23.21	14
	Pronunciation Only	65.48	16.62	14
	Multi Speaker HVPT	66.67	16.67	13
	Single Speaker PF	66.20	19.64	13
	Control	64.44	17.67	15
Delayed Gen Test (New words)	Multi HVPT + Pronun.	74.36	25.11	13
	Single PF + Pronun.	62.82	15.45	13
	Pronunciation Only	67.86	20.11	14
	Multi Speaker HVPT	61.67	17.66	10
	Single Speaker PF	59.09	23.59	10
	Control	64.29	18.32	14

Table 6.22: Oral Contrast Generalisation Test 1: Mean Percentage Perceptual Identification Accuracy Scores According to Training Group

Test	Training Group	Mean	SD	N
Pre-Test	Multi HVPT + Pronun.	63.10	20.86	14
	Single PF + Pronun.	60.71	18.03	14
	Pronunciation Only	65.48	15.28	14
	Multi Speaker HVPT	66.67	22.57	13
	Single Speaker PF	61.76	18.57	13
	Control	65.56	22.24	15
Gen Test (New words)	Multi HVPT + Pronun.	75.00	19.34	14
	Single PF + Pronun.	76.19	19.30	14
	Pronunciation Only	71.43	23.05	14
	Multi Speaker HVPT	74.36	19.97	13
	Single Speaker PF	64.92	22.32	13
	Control	65.56	20.38	15
Delayed Gen Test (New words)	Multi HVPT + Pronun.	76.92	18.68	13
	Single PF + Pronun.	71.80	20.84	13
	Pronunciation Only	76.19	19.30	14
	Multi Speaker HVPT	66.67	20.79	10
	Single Speaker PF	68.17	23.92	10
	Control	64.29	22.51	14

Figure 6.18: Nasal Contrast Generalisation Test : Mean Percentage Perceptual Identification Accuracy Scores According to Training Group

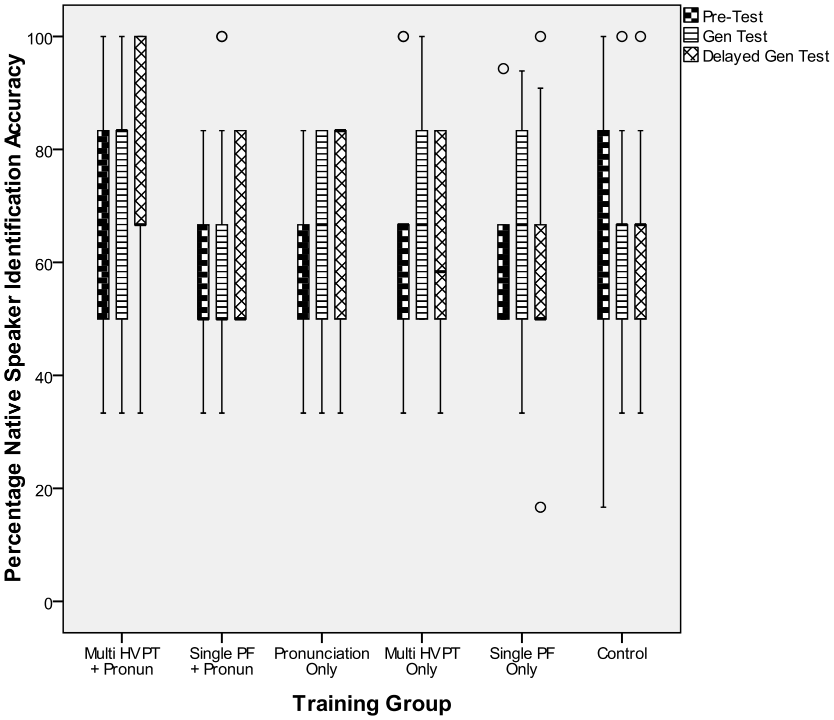
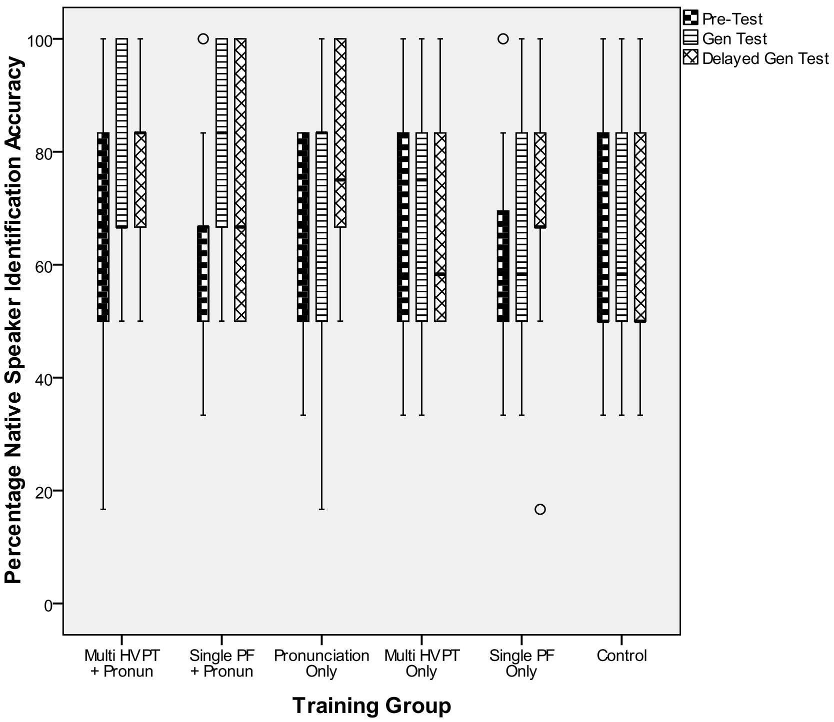


Figure 6.19: Oral Contrast Generalisation Test 1: Mean Percentage Perceptual Identification Accuracy Scores According to Training Group



As with the pre-test/post-test pronunciation data, these pronunciation identification accuracy generalisation data were then analysed using binary logistic mixed effects analysis in order to determine the significance of training group, test, contrast and time on whether or not native speakers identified the correct member of the minimal pair from participant productions. Again, as with the pre-test/post-test pronunciation data, the participant and test item were included as random effects. The dependent variable was whether the native speaker correctly identified the minimal pair member presented. The fixed effects tested were participant training group/method (SubjGp), test (Test, the three levels of this – pre, post, delay, also represented time) and Contrast (oral vs. nasal).

Here the best model which had any significant predictors only produced a marginally significant interaction (Pr(>|z|) = 0.08) between the control group and the multi speaker HVPT + pronunciation group and the pre-test and the delayed generalisation test. Inspection of values in Tables 6.21 and 6.22 suggests that whilst these groups did not differ at pre-test the multi speaker HVPT + pronunciation group may score higher than the control group in the delayed generalisation test. However, there was no significant difference between this model and the null model (Pr(>Chisq) = 0.839), and therefore this result cannot be interpreted. The fixed effects in this model have no more predictive power than the random effects alone. In other words training had no generalising effect on native speaker identification accuracy for any group and at any time.

6.6.2.1 Group Comparison with Native Speakers

As with the pre/post-test data, this non-significant model was re-run with the results from natives listening to native speakers as the reference group in order to investigate whether or not the addition of these data would result in a model with more predictive power and ascertain how the trainee groups' pronunciation differed from a group of native speakers. The performance of the groups along with the native group collapsed across contrast is illustrated in Figure 6.20 and (now significant) best model is detailed in Table 6.23.

Figure 6.20: Mean Overall Percentage Perceptual Identification Accuracy Scores According to Training Groups and Native Speaker Group

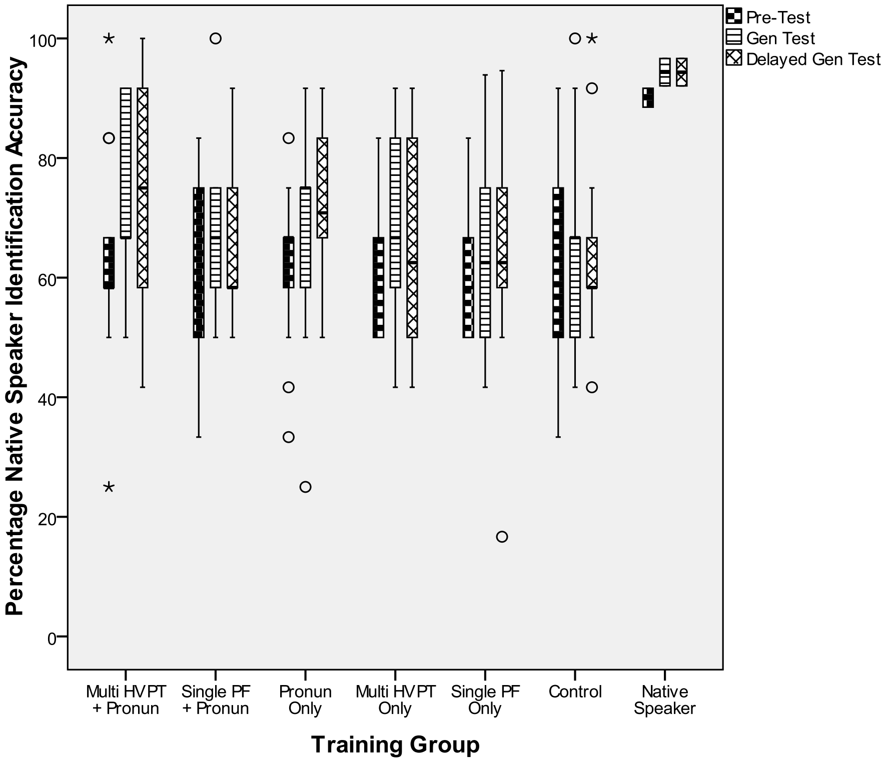


Table 6.23: Fixed Effects and Coefficients of the Native Speaker Generalisation Test Model

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	2.840	0.711	3.995	<0.00001****
SubjGp2Multi HVPT + Pronun	-2.157	0.687	-3.141	0.0017***
SubjGp2Single PF + Pronun	-2.408	0.686	-3.511	0.0005****
SubjGp2Pronunciation Only	-2.321	0.686	-3.385	0.0007****
SubjGp2Multi HVPT Only	-2.235	0.689	-3.246	0.0012***
SubjGp2Single PF Only	-2.633	0.690	-3.817	0.0001****
SubjGp2Control	-2.434	0.684	-3.556	0.0004****
Test2Gen	0.249	0.305	0.816	0.4145
Test2GenDelay	0.262	0.306	0.856	0.3922

***p<.1; **p<.05; ***p<.01, ****p<.001**

With the native speaker group as the reference group the data are best explained by the main effects model. No interactions including training group and/or contrast were significant. As the main effect of contrast was not significant this was also not included in the model. This model suggests that the native speaker group scores significantly better than all other groups overall and the lack of significant interaction terms means that this is not dependent upon training group or time, in other words, no training technique results in a non-significant difference between native speaker group and any learner group.

6.6.3 Identification Accuracy Results Discussion

The native speaker identification accuracy results have suggested that the multiple speaker HVPT + pronunciation group and (to a lesser extent) the pronunciation only group may be more effective than no training at all in improving pronunciation accuracy, but this was only true in the delayed post-test. In the post-test immediately after training and in all generalisation tests there was no evidence that any training method was more effective than no training at all.

Finding some evidence for the superiority of the multiple speaker HVPT + pronunciation group for pronunciation accuracy reinforces the effectiveness of this technique as suggested by the perceptual results. The lack of success of the perceptual training only groups suggests that there was no transfer of perceptual training to pronunciation. The evidence that the pronunciation only group may also be more successful than no training suggests, in parallel with the perceptual results, that some pronunciation training is more effective than perceptual training at improving pronunciation performance as measured by native speaker identification accuracy.

The removal of some pronunciation training does not appear detrimental to pronunciation performance in the multiple speaker HVPT + pronunciation group although why this is the case with this group and not the single speaker PF + pronunciation group is unclear. It may be that there is a small amount of pronunciation benefit from this perceptual training technique when used alongside pronunciation training that does not exist when using this technique alone or the single speaker PF technique with or without pronunciation training.

With regard to answering the research questions posed at the start of this chapter concerning improvement in pronunciation, there is some evidence to suggest that the multiple speaker HVPT + pronunciation group and (to a slightly lesser extent) the pronunciation only group are most successful at producing an improvement in pronunciation as determined by native speaker identification accuracy, however no technique emerges as better than no training at all in producing a generalisable improvement in pronunciation.

6.6.4 Individual Comparison with Native Speakers

As with the perceptual data these data were also compared at an individual level with native speaker results rather than at a group level in order to ascertain whether anyone performed to a native standard either before or after training. Again, for simplicity, a native standard was taken to be within the native range of scores. In terms of pronunciation as measured by native speaker identification accuracy, none of the 84 participants scored at a native level in all tests; however, there was evidence of nativelike performance in some cases. Seventeen of the 84 participants scored within the native range for the nasal contrast pre-test and 8 of the participants scored within the native range for the oral contrast pre-test. It should be noted that the native speaker accuracy range for native speaker nasal productions of the pre/post stimuli was relatively low at 77.84-84.96%.

Looking at post-training scores, across all 4 post training tests (post-test, generalisation test, post-test T2, generalisation test T2), no participants scored within the native range across all 8 tests. The strongest performing participant scored within the native range in 4/4 nasal contrast tests and 3/4 oral contrast tests, however, this participant was a control and not trained. The next best participants (6 participants, all trained) scored 4/4 and 2/4, or 3/4 and 3/4, or 2/4 and 4/4 respectively. Of all 84 participants, 18 scored within the native range in at least one of the tests for both contrasts and a further 31 scored within the native range in at least one of the tests for one of the contrasts. Again, overall, whilst no participants scored within the native range across all tests, these results suggest that training did result in an improvement towards a native standard in some instances.

The next section details further analysis of the participant productions. It deals with how the native speakers of French rated the participant productions and whether training resulted in any difference to these ratings.

6.7 Pronunciation Results 2 – Native Speaker Ratings

The raw rating accuracy data from each native speaker was sorted such that their rating data for each French learning participant and French native speaker for each pronunciation test undertaken could be obtained. The results from the 41 analysis sessions were averaged for a criterion speaker, the French learning participant whom everyone rated, and for both the native French speakers whom everyone also rated.

Krippendorff's Alpha revealed a very low level of agreement ($\alpha = .12$) on native speaker ratings of the criterion speaker due to the large number of rating sessions. In addition, the raters who returned more than once did not have a high level of agreement on their ratings of the criterion speaker. It was therefore firstly decided to treat the 41 rating sessions as if they were from 41 individual raters (as opposed to averaging out the data from the raters who returned more than once which would have resulted in 26 sets of rating data). It was then decided to convert the rating data to z-scores. For each of the 41 raters, the criterion speaker's rating data was transformed using the overall mean and standard deviation of these data and the data from the other two participants whom were also rated by the rater were also transformed using the criterion speaker's overall mean and standard deviation. The results of this manipulation are that inter-rater differences were adjusted for each rater's strictness and spread of use of the rating scale, and the ratings for the non-criterion speakers are normalised against the criterion speaker's scores. Comparison of these data are therefore less affected by inter-rater differences.

As previously, before carrying out the analysis, it was firstly necessary to ensure that any differences between training groups after training in terms of native speaker ratings were attributable to the training technique and not due to one group having a higher or lower pre-test native speaker identification accuracy score. To this end, two one-way ANOVAs were carried out with pre-test score as the dependent variable and training group as the between subjects factor, one ANOVA for each contrast of interest. There was no significant effect of training group for either contrast [Nasal: $F(5,82) = .968$, $p = .443$; Oral: $F(5,82) = .432$, $p = .825$] indicating that the pre-test ratings were not significantly different for each group.

For analysis of the rating data linear mixed effects analysis was used in this instance as the native speaker ratings dependent variable was continuous. The analyses of the relationship between training groups, tests, contrast tested and time was again carried out in two blocks. The first block examined the effects of training across the pre-test, post-test and delayed post-test. The second block examined the generalisation test. Unless otherwise stated, likelihood ratio tests comparing each model with fixed effects to a null model with only the random effects demonstrated that the fixed effects model differed significantly from the null model.

6.7.1 Block 1: Pre-Test, Post-Test, Delay

The mean participant percentage perceptual identification accuracy and standard deviations for each test (pre-test, post-test, and delayed post-test according to participant training group and contrast are shown in Tables 6.24 and 6.25 and illustrated in Figures 6.21. and 6.22. These data suggest that the multi speaker HVPT + pronunciation and pronunciation only training may result in better results than other training types (as lower ratings mean more nativelike pronunciation).

Table 6.24: Nasal Contrast: Mean Z-Transformed Ratings by Native Speakers According to Participant Training Group

Test	Training Group	Mean	SD	N
Pre-Test	Multi HVPT + Pronun.	.95	.75	14
	Single PF + Pronun.	1.20	.98	14
	Pronunciation Only	.77	.40	14
	Multi Speaker HVPT	1.11	.71	13
	Single Speaker PF	1.32	.77	13
	Control	1.12	.67	15
Post-Test	Multi HVPT + Pronun.	.78	.81	14
	Single PF + Pronun.	.89	.99	14
	Pronunciation Only	.59	.49	14
	Multi Speaker HVPT	1.00	.71	13
	Single Speaker PF	1.01	.63	13
	Control	1.23	.57	15
Delayed Post-Test	Multi HVPT + Pronun.	.91	.69	13
	Single PF + Pronun.	1.07	1.03	13
	Pronunciation Only	.62	.45	14
	Multi Speaker HVPT	1.07	.89	10
	Single Speaker PF	.87	.81	10
	Control	1.32	.59	14

Table 6.25: Oral Contrast: Mean Z-Transformed Ratings by Native Speakers According to Participant Training Group

Test	Training Group	Mean	SD	N
Pre-Test	Multi HVPT + Pronun.	1.11	.75	14
	Single PF + Pronun.	1.17	.93	14
	Pronunciation Only	1.18	.58	14
	Multi Speaker HVPT	1.37	.69	13
	Single Speaker PF	1.18	.97	13
	Control	1.46	.75	15
Post-Test	Multi HVPT + Pronun.	.82	.63	14
	Single PF + Pronun.	1.05	.84	14
	Pronunciation Only	.86	.71	14
	Multi Speaker HVPT	1.21	.70	13
	Single Speaker PF	1.11	.84	13
	Control	1.33	.82	15
Delayed Post-Test	Multi HVPT + Pronun.	.85	.60	13
	Single PF + Pronun.	.96	.90	13
	Pronunciation Only	.73	.73	14
	Multi Speaker HVPT	1.16	.93	10
	Single Speaker PF	1.00	.80	10
	Control	1.60	.72	14

Figure 6.21: Nasal Contrast: Mean Z-Transformed Ratings by Native Speakers According to Participant Training Group

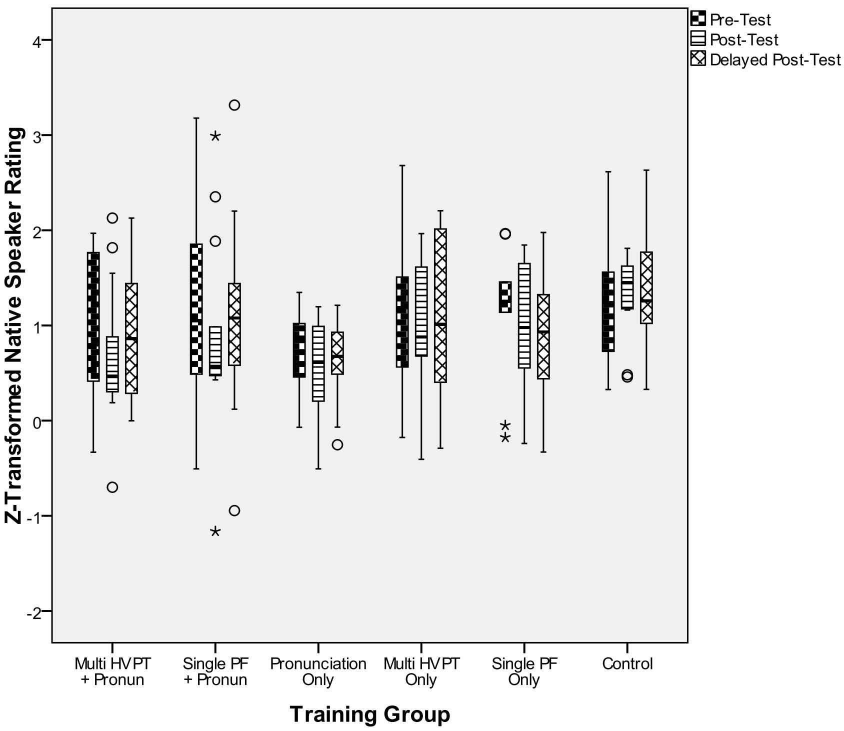
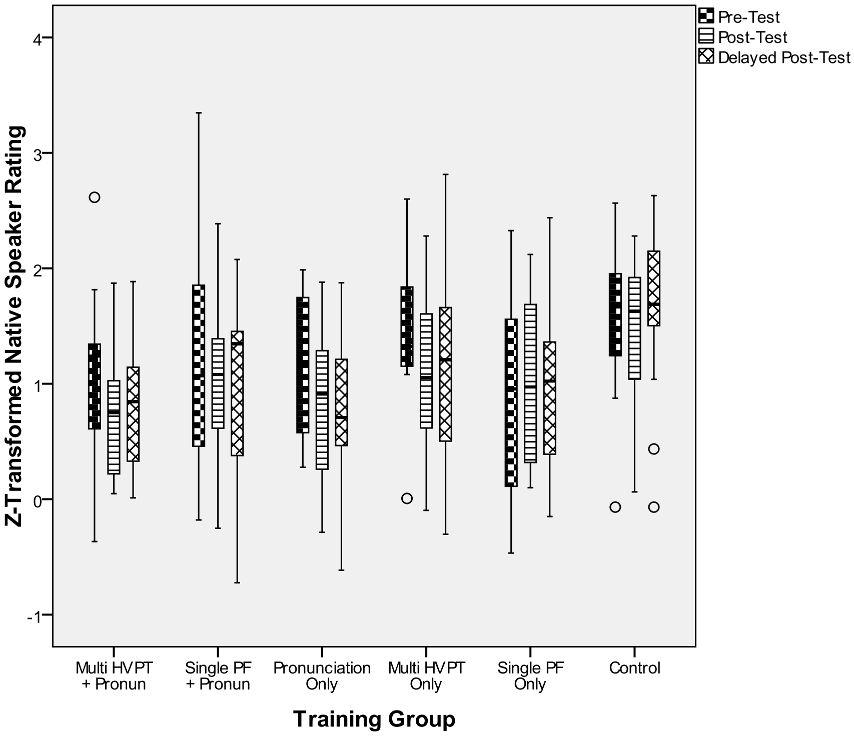


Figure 6.22: Oral Contrast: Mean Z-Transformed Ratings by Native Speakers According to Participant Training Group



These data were then analysed using linear mixed effects analysis in order to determine the significance of training group, test, contrast and time on how native speakers rated the participant productions. In the present model the participant and test item were included as random effects. The dependent variable was the z-transformed native speaker rating given for each word produced. The fixed effects or potential predictors tested were participant training group/method (SubGp), test (Test the three levels of this – pre, post, delay, also represented time) and Contrast (oral vs. nasal). The model is detailed in Table 6.26 below.

Table 6.26: Fixed Effects and Coefficients of the Pre-Test, Post-Test, Delay Model

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	1.342	0.231	5.807	<0.00001****
SubjGpMulti HVPT Only	-0.155	0.243	-0.639	0.5227
SubjGpMulti HVPT + Pronun	-0.388	0.237	-1.634	0.1024
SubjGpSingle PF + Pronun	-0.246	0.237	-1.038	0.2995
SubjGpPronun Only	-0.518	0.237	-2.184	0.0290**
SubjGpSingle PF Only	-0.179	0.243	-0.739	0.4600
TestPost	-0.170	0.054	-3.124	0.0018***
TestPostDelay	-0.126	0.057	-2.225	0.0262**

***p<.1; **p<.05; ***p<.01, ****p<.001**

As with the perception and pronunciation identification accuracy data, the reference level for participant training group (SubjGp) was the control group, the reference level for Contrast was the nasal contrast, and the reference level for Test was the pre-test. In this instance the best model is one with main effects only. There is no main effect of contrast as this was not significant when added to the model and addition of a contrast interaction term alone or with a group interaction term resulted in no interactions. Addition of a group interaction term initially resulted in an interaction which suggested that the control group and pronunciation only group differed dependent upon whether the test was the pre-test or the delayed post-test. Examination of Tables 6.24 and 6.25 together suggested that whilst the pronunciation group and the control group did not differ in z-ratings at pre-test, pronunciation group scored significantly lower (and therefore better) at the delayed post-test. However, the model with this interaction was not significantly different from the null model so this interaction cannot be recognised as valid.

The significant main effects of test in the present model cannot be interpreted as this is collapsed across groups which include the control group and the fact that the post-test and delayed post-test scores are greater than the pre-test scores is therefore meaningless. With regard to the main effect of group, it can be seen that the pronunciation only group scores significantly less (and therefore better) than the control group overall. As the ANOVAs demonstrated that the groups did not significantly differ at the pre-test, it can be tentatively concluded that this difference has some meaning and that pronunciation only training may be the only technique to result in ratings significantly lower than with no training at all.

6.7.1.1 Group Comparison with Native Speakers

As with the previous data in this chapter, this model was re-run with the results from natives rating native speakers as the reference group in order to ascertain how the trainee groups' pronunciation differed from a group of native speakers. The performance of the groups along with the native group collapsed across contrast is illustrated in Figure 6.23 and the best model is detailed in Table 6.27.

Figure 6.23: Mean Overall Z-Rating Scores According to Training Groups and Native Speaker Group

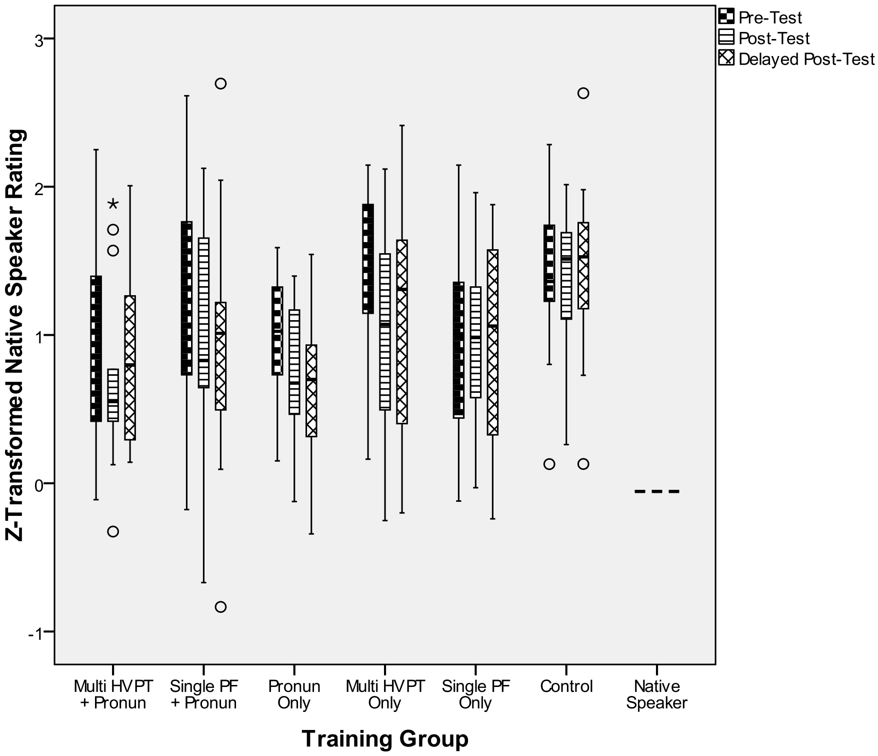


Table 6.27: Fixed Effects and Coefficients of the Native Speaker Pre-Test, Post-Test, Delay Model

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-0.008	0.475	-0.018	0.9860
SubjGpMulti HVPT + Pronun	0.976	0.479	2.038	0.0417**
SubjGpSingle PF + Pronun	1.118	0.479	2.333	0.0197***
SubjGpPronunciation Only	0.846	0.479	1.766	0.0774*
SubjGpMulti HVPT Only	1.209	0.482	2.510	0.0121**
SubjGpSingle PF Only	1.185	0.482	2.460	0.0139**
SubjGpControl	1.364	0.477	2.859	0.0043***
TestPost	-0.166	0.053	-3.110	0.0019***
TestPostDelay	-0.123	0.055	-2.215	0.0268**

***p<.1; **p<.05; ***p<.01, ****p<.001**

With the native speaker group as the reference group the data are best explained by the main effects model. No interactions including training group and/or contrast were significant. As the main effect of contrast was not significant this was also not included in the model. This model suggests that the native speaker group scores significantly better than all other groups overall and the lack of significant interaction terms means that this is not dependent upon training group or time, in other words, no training technique results in a non-significant difference between native speaker group and any learner group. However, the overall difference between the native group score and the pronunciation group score is marginally non significant. Again, as the ANOVAs demonstrated that the groups did not significantly differ at the pre-test it can perhaps be tentatively concluded that this is the only training group which results in ratings approaching a native standard.

6.7.2 Block 2: Generalisation Test

The mean participant percentage perceptual identification accuracy and standard deviations for each test (pre-test, post-test, and delayed post-test according to participant training group and contrast are shown in Tables 6.28 and 6.29 and illustrated in Figures 6.24 and 6.25. The original pre-test scores are also included for reference. These data suggest that the multi speaker HVPT + pronunciation and pronunciation only training may result in better results than other training types (as lower ratings mean more nativelike pronunciation).

Table 6.28: Nasal Contrast: Mean Z-Transformed Ratings by Native Speakers According to Participant Training Group

Test	Training Group	Mean	SD	N
Pre-Test	Multi HVPT + Pronun.	.95	.75	14
	Single PF + Pronun.	1.20	.98	14
	Pronunciation Only	.77	.40	14
	Multi Speaker HVPT	1.11	.71	13
	Single Speaker PF	1.32	.77	13
	Control	1.12	.67	15
Gen Test (New words)	Multi HVPT + Pronun.	1.04	.68	14
	Single PF + Pronun.	1.19	1.11	14
	Pronunciation Only	.64	.53	14
	Multi Speaker HVPT	1.33	.73	13
	Single Speaker PF	1.16	.73	13
	Control	1.23	.79	15
Delayed Gen Test (New words)	Multi HVPT + Pronun.	.86	.62	13
	Single PF + Pronun.	1.30	1.03	13
	Pronunciation Only	.73	.63	14
	Multi Speaker HVPT	1.21	.72	10
	Single Speaker PF	1.12	.94	10
	Control	1.36	.62	14

Table 6.29: Oral Contrast: Mean Z-Transformed Ratings by Native Speakers According to Participant Training Group

Test	Training Group	Mean	SD	N
Pre-Test	Multi HVPT + Pronun.	1.11	.75	14
	Single PF + Pronun.	1.17	.93	14
	Pronunciation Only	1.18	.58	14
	Multi Speaker HVPT	1.37	.69	13
	Single Speaker PF	1.18	.97	13
	Control	1.46	.75	15
Gen Test (New words)	Multi HVPT + Pronun.	.96	.69	14
	Single PF + Pronun.	.90	1.07	14
	Pronunciation Only	.94	.73	14
	Multi Speaker HVPT	.96	.81	13
	Single Speaker PF	1.06	.85	13
	Control	1.36	.67	15
Delayed Gen Test (New words)	Multi HVPT + Pronun.	.75	.87	13
	Single PF + Pronun.	1.00	.95	13
	Pronunciation Only	.70	.65	14
	Multi Speaker HVPT	.99	.87	10
	Single Speaker PF	1.04	1.14	10
	Control	1.45	.88	14

Figure 6.24: Nasal Contrast: Mean Z-Transformed Ratings by Native Speakers According to Participant Training Group

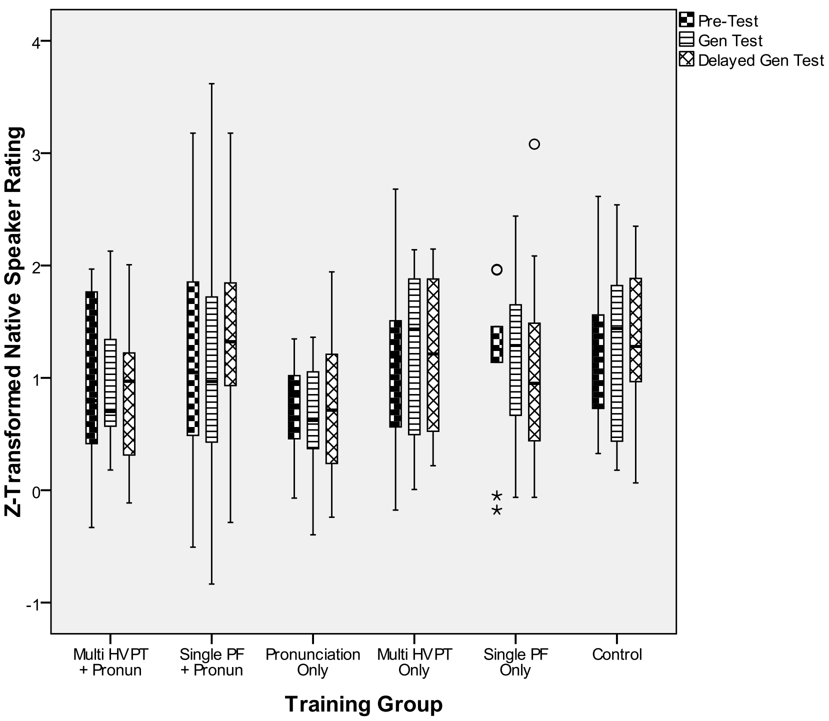
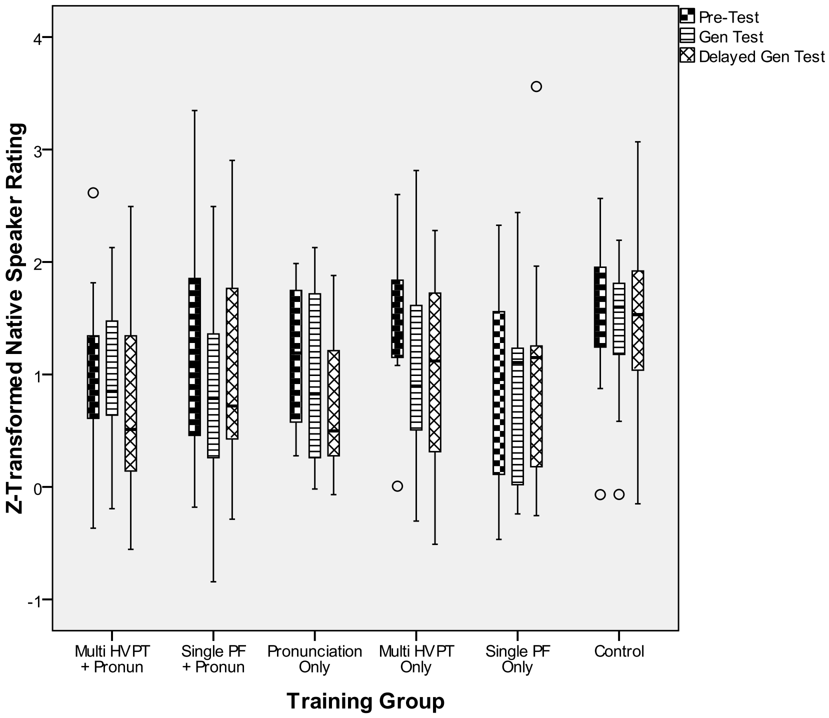


Figure 6.25: Oral Contrast: Mean Z-Transformed Ratings by Native Speakers According to Participant Training Group

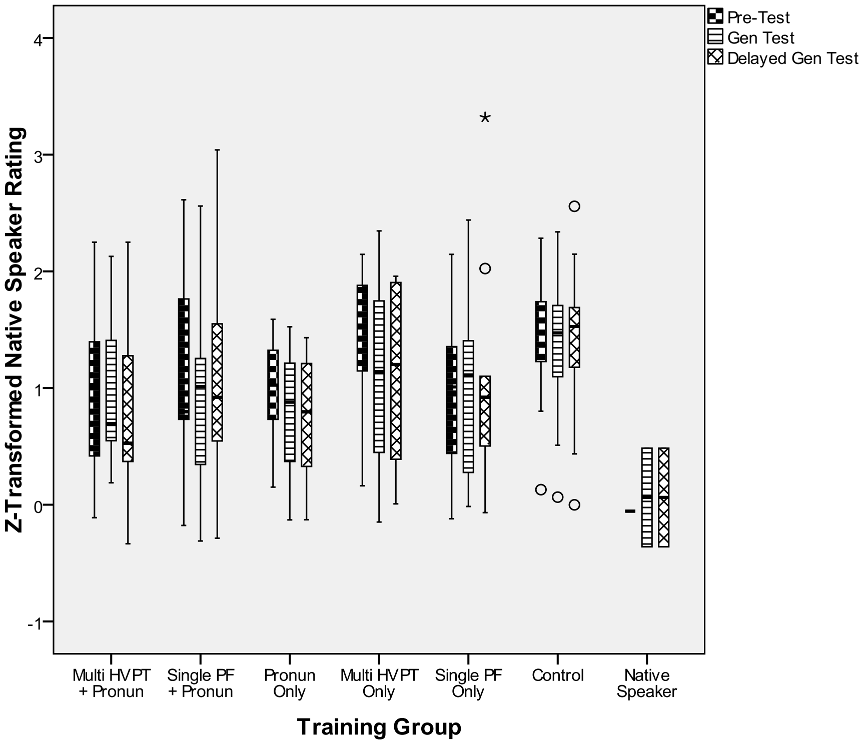


These data were then analysed using linear mixed effects analysis in order to determine the significance of training group, test, contrast and time on how native speakers rated the participant productions. In the present model the participant and test item were included as random effects. The dependent variable was the z-transformed native speaker rating given for each word produced. The fixed effects or potential predictors tested were participant training group/method (SubGp), test (TestType, the three levels of this – pre, post, delay, also represented time) and Contrast (oral vs. nasal). However, there was no significant difference between any model investigated and the null model. The fixed effects in this model have no more predictive power than the random effects alone. In other words training had no generalising effect on native speaker ratings for any group and at any time.

6.7.2.1 Group Comparison with Native Speakers

The non-significant models were re-run with the results from native speakers rating other native speakers as the reference group in order to investigate whether or not the addition of these data would result in a model with more predictive power and ascertain how the trainee groups' pronunciation differed from a group of native speakers. The performance of the groups along with the native group collapsed across contrast is illustrated in Figure 6.26.

Figure 6.26: Mean Overall Z-Rating Scores According to Training Groups and Native Speaker Group



Even with the native speaker group as the reference group there were no significant differences between any model and the null model. The fixed effects in this model continue to have no more predictive power than the random effects alone. In other words training had no generalising effect no matter which group and which time. Inspection of Figure 6.26 shows that the generalisation word ratings for natives were unexpectedly high with some degree of overlap with the other groups. This is therefore likely why the addition of native speaker group did not add any predictive power to the model.

6.7.3 Rating Results Discussion

The native speaker rating results have provided an indication that the pronunciation only group may be more effective than no training at all in improving pronunciation accuracy, and may result in accuracy (marginally) not significantly different from native speakers but this was only suggested by main effect data which did include pre-test scores. In all generalisation tests there was no evidence that any training method was more effective than no training at all.

In sum, the pronunciation only training technique for training pronunciation accuracy emerges as optimal from native speaker ratings of participant pronunciations. As with the pronunciation identification accuracy data, there is no significant evidence of transfer of perceptual training to pronunciation, with none of the groups with a perceptual training element performing better than no training at all. The slight evidence that the pronunciation only group may be more successful than no training suggests, in parallel with the perceptual and the previous pronunciation results, that some pronunciation training is more effective than perceptual training at improving pronunciation performance as measured by native speaker identification accuracy. In contrast to the native speaker identification data the removal of some pronunciation training is indeed detrimental to pronunciation performance in terms of native speaker ratings with no indication that the multiple speaker HVPT + pronunciation group or the single speaker PF + pronunciation group obtain results significantly different to those who had no training.

With regard to answering the research questions posed at the start of this chapter concerning improvement in pronunciation, there is some evidence to suggest that the pronunciation only group is most successful at producing an improvement in pronunciation as measured by native speaker ratings, however no technique emerges as better than no training at all in producing a generalisable improvement in pronunciation.

6.7.4 Individual Comparison with Native Speakers

These data were again compared at an individual level with native speaker results rather than at a group level in order to ascertain whether anyone performed to a native standard either before or after training. In terms of pronunciation as measured by native speaker z-transformed ratings, none of the 84 participants scored at a native level in all tests; however, there was evidence of nativelike performance in some cases. Twenty-one participants performed within the native range for the nasal contrast pre-test and none of the participants scored within this range for the oral contrast pre-test. The native z-transformed rating range was relatively high (inaccurate) for the nasal pre/post test words (1.59-.55), compared to the generalisation words (.55-.55) and all of the oral contrast words (-.55--.66) resulting in more participants falling within this range than anticipated. Again, no participants scored within the native range across all 8 post-training tests (post-test, generalisation test, post-test time 2, generalisation test time 2 for both contrasts), with the most successful participants scoring 4/4 for the nasal contrast tests and 1/4 in the oral contrast tests. Only six participants scored within the native range in any oral contrast test (one test each), therefore only these 6 participants scored within the native range for at least one of the tests in both contrasts. Due to the scoring of the nasal contrast, 57 participants scored within the native range for at least one test, a result that is likely to be less meaningful due to this anomaly. However, again, whilst no participants scored within the native range across all tests, these results suggest that training did result in an improvement towards a native standard in some instances.

6.8 Overall Discussion and Conclusions

The most successful perceptual training techniques suggested through answering Research Question 1 in Chapter 5 are indeed more successful than pronunciation training alone or (to a lesser extent) along with single speaker PF training in terms of producing a generalisable improvement in perception. However, the multiple speaker HVPT + pronunciation training technique also emerged as equally successful for perceptual training.

The multiple speaker HVPT + pronunciation technique and the pronunciation only technique emerged as better than no training at all for improvements in pronunciation as measured by native speaker identification and rating of the participant productions. The evidence for the superiority of the multiple speaker HVPT + pronunciation technique came from the native speaker identification of the participant productions and slight evidence for the superiority of the pronunciation only technique came from both native speaker identification accuracy and native speaker ratings. However, no technique emerged as beneficial for obtaining a generalisable improvement in pronunciation.

From the perceptual and pronunciation results overall, it therefore appears that despite evidence of transfer of perceptual training to pronunciation accuracy in previous work (see e.g. Bradlow et al, 1997), perceptual training is best for perceptual improvement and pronunciation training is best for pronunciation improvement in the present work. Halving the time spent on single speaker PF training has a slightly more detrimental effect on perceptual results than halving the time spent on multiple speaker HVPT training when comparing the performance of these groups to native speakers. It is possible that more time is needed at each point on the fading continuum in order for the contrast to be acquired, whereas in the multiple speaker HVPT condition the benefits of the variability from multiple speakers can be obtained in less time.

When compared with the control group, halving the time spent on pronunciation training had a slightly more detrimental effect on pronunciation identification accuracy when this was paired with the single speaker PF technique than with the multiple speaker HVPT technique, although it is unclear why this is the case, particularly as there is no clear trend for multiple speaker HVPT training alone being more successful than single speaker PF training alone in improving pronunciation.

In terms of whether an optimal training technique emerges, the multiple speaker HVPT + pronunciation training technique over the same timescale as perception training only or pronunciation training only appears to be the most effective. Overall, the perceptual and pronunciation results suggest that this technique is approximately as successful as the perception only techniques at improving perception and as the pronunciation only technique at improving pronunciation. As a very tentative conclusion, it therefore appears that this is the optimal technique from those examined for improving both perception and pronunciation over a set period of time. It is possible that other techniques may prove more successful if training is terminated only when a certain test performance level is reached in terms of time taken to reach that target.

7 Experiment 3: The Pronunciation Attitude Inventory

7.1 Introduction

This chapter concerns the administration of the Pronunciation Attitude Inventory (PAI) (Elliott, 1995). As noted in Chapter 3, motivation, generally integrative motivation of some nature, and the strength of concern for pronunciation accuracy aspect in particular, has been found to play some role in determining pronunciation accuracy (e.g. Moyer 1999; Birdsong 2003, 2007). With particular reference to the PAI, Elliott (1995) found that the more concerned the participant was about their pronunciation of their target language (in this instance Spanish) as measured by the PAI, the more accurate their pronunciation tended to be.

The purpose of the present experiment is to examine the relationship between any improvement in pronunciation through training and the motivational measure of strength of concern for pronunciation accuracy (from the PAI), and in so doing answer the fourth research question:

RQ4 Do those with a stronger concern for pronunciation accuracy perform better or improve more in terms of pronunciation with training and with which techniques?

Birdsong (2003, 2007), Bongaerts et al (e.g. 1997), Moyer (1999) and Abu-Rabia and Kehat (2004) found that a combination of motivation and undertaking perceptual and/or pronunciation training were common characteristics of those who were rated as having nativelike pronunciation. It is hoped that answering the fourth research question will provide further evidence to support these findings and therefore provide an indication of potential interventions to move learners towards nativelike (and therefore more comprehensible) pronunciation and perception.

7.2 Participants

Two sets of participants, 84 participants in total, completed the PAI questionnaire and had their word productions analysed. The first set was those participants who were part of the more successful single speaker PF and multiple speaker HVPT training groups as described in Chapter 5. The numbers in these groups were 13 and 14 respectively immediately after training (who all completed the PAI questionnaire), and 10 in both for the tests of retention due to lack of interest or scheduling difficulties. These participants are described in more detail in Chapter 5.

The second set of participants was those who participated in the pronunciation training study described in Chapter 6. The participants were randomly assigned to a further four training groups: multiple speaker HVPT + pronunciation training (14 participants), single speaker PF + pronunciation training (14 participants), pronunciation only training (14 participants) and a control group (15 participants) who received no training. All participants completed the PAI questionnaire. After a minimum of one month the participants then returned to carry out retention testing. Three participants did not return due to lack of interest or scheduling difficulties, resulting in group numbers as follows: Multiple Speaker HVPT + Pronunciation (13/14 participants); Single Speaker PF + Pronunciation (13/14 participants); Pronunciation Only (14/14 participants); Control (14/15 participants). This second set of participants is described in more detail in Chapter 6.

7.3 Stimuli and Procedure

Participants completed the PAI questionnaire before undertaking the perception and production pre-tests described in Chapters 5 and 6. The questionnaire consisted of 12 experimental statements and 13 'filler' statements. Nine experimental statements were positively worded and three were negatively worded and were concerned with the acquisition of native or close to native pronunciation in French. As Elliott's (1995) study concerned the acquisition of Spanish pronunciation, the word 'Spanish' was substituted by the word 'French' for this study. The experimental questions were (Elliott 1995):

1. I'd like to sound as native as possible when speaking French.
2. Acquiring proper pronunciation in French is important to me.
3. I will never be able to speak French with a good accent.
4. I believe I can improve my pronunciation skills in French.
5. I believe more emphasis should be given to proper pronunciation in class.

6. One of my personal goals is to acquire proper pronunciation skills and preferably be able to pass as a near-native speaker of the French language.
7. I try to imitate French speakers as much as possible.
8. Communicating is much more important than sounding like a native speaker of French.
9. Good pronunciation skills in French are not as important as learning vocabulary and grammar.
10. I want to improve my accent when speaking French.
11. I'm concerned with my progress in my pronunciation of French.
12. Sounding like a native French speaker is very important to me.

The 13 'filler' statements (9 positive and 4 negative) were devised by the experimenter and concerned other aspects of French language learning. Thirteen filler statements were used in the present work to replicate the 13 used by Elliott (1995), however the author did not detail the statements used. These additional statements were used in order to avoid the purpose of the questionnaire being immediately obvious and consequently causing a reactivity effect (Elliott, 1995). The experimental and 'filler' statements appeared in a random order within the questionnaire and the full questionnaire can be found in Appendix H.

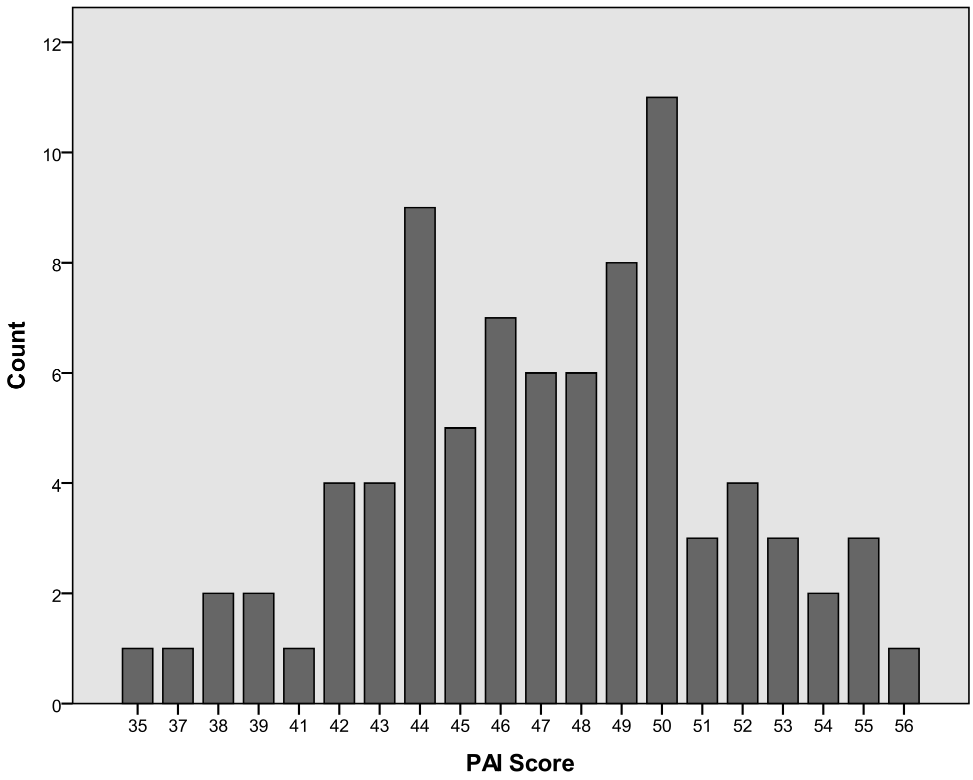
To complete the questionnaire, participants were asked to respond to the statements using the following response categories: 5 = Always or almost always true of me, 4 = Usually true of me, 3 = Somewhat true of me, 2 = Usually not true of me, 1 = Never or almost never true of me. This yielded a measure ranging from 12 (negative attitude) to 60 (positive attitude) (Elliott, 1995), with the negatively worded items having their scores reversed before being added to the total. The responses to the 'filler' questions were not used in the analysis. The PAI scores were then analysed alongside the native speaker identification accuracy data and native speaker rating (z-transformed) data obtained as described in Chapter 6 (6.4.4 and 6.7).

7.4 Results

7.4.1 PAI Scores

The distribution of PAI scores from the 84 participants is illustrated in Figure 7.1. The participant PAI scores ranged from 35-56 with a mean score of 47.11 (SD = 4.44) and a median score of 47 showing the data to be fairly symmetrical. However, virtually all of the participants scored in the upper half of the range of scores available, 36-60, with only one participant scoring (35) in the lower half of the range, 12-36.

Figure 7.1: PAI Score Distribution



7.4.2 Pre-training PAI Relationships

A first item of interest was to examine whether there was any relationship between participant PAI scores and participant pronunciation accuracy as measured by the percentage correct identification scores and native speaker ratings in the pre-test. This would replicate previous findings (e.g. Elliott, 1995) that PAI is correlated with pronunciation accuracy. Prior to training, any relevant relationship between PAI score and native speaker identification accuracy would be positive as native speaker identification accuracy was measured by percentage correct identification. However, any relevant relationship between PAI and native speaker rating should be negative as lower native speaker ratings reflected more accurate pronunciations. Figures 7.2 and 7.3 illustrate the relationship between PAI score and native speaker percentage identification accuracy for each contrast to be trained.

Figure 7.2: The Relationship Between PAI Score and Pre-test Native Speaker Identification Accuracy of the Nasal Contrast

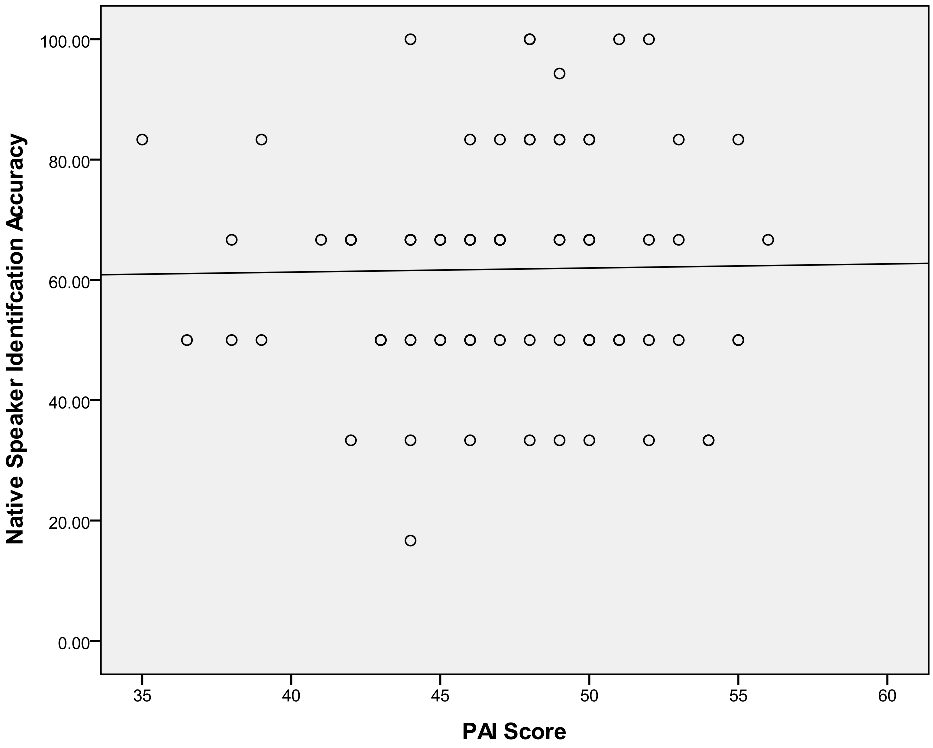
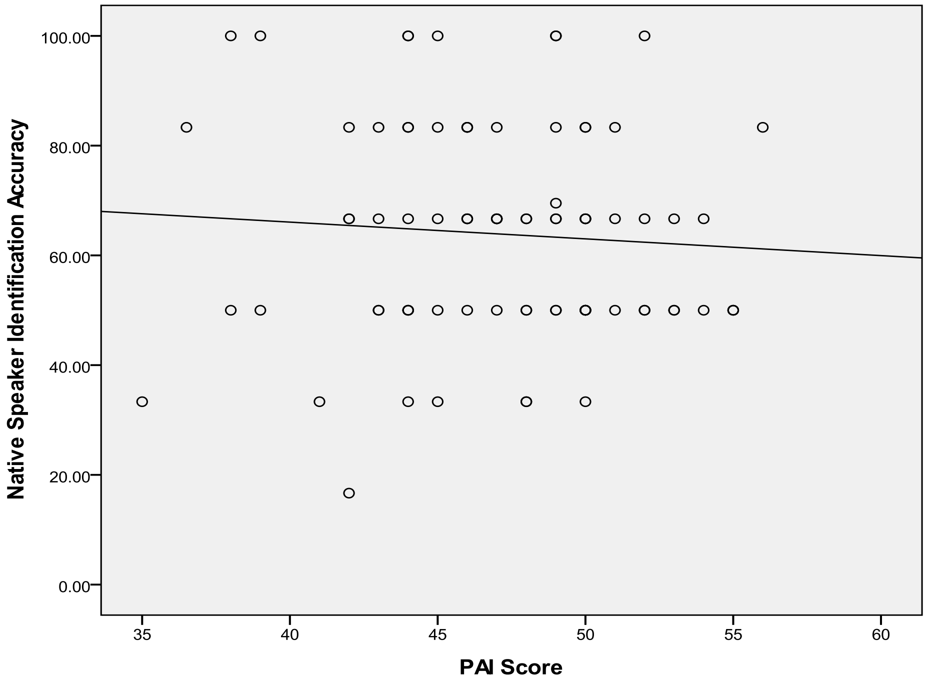


Figure 7.3: The Relationship Between PAI and Pre-Test Native Speaker Identification Accuracy of the Oral Contrast



Figures 7.4 and 7.5 illustrate the relationship between PAI score and native speaker z-ratings for each contrast to be trained.

Figure 7.4: The Relationship Between PAI and Pre-Test Native Speaker Z-Rating of the Nasal Contrast

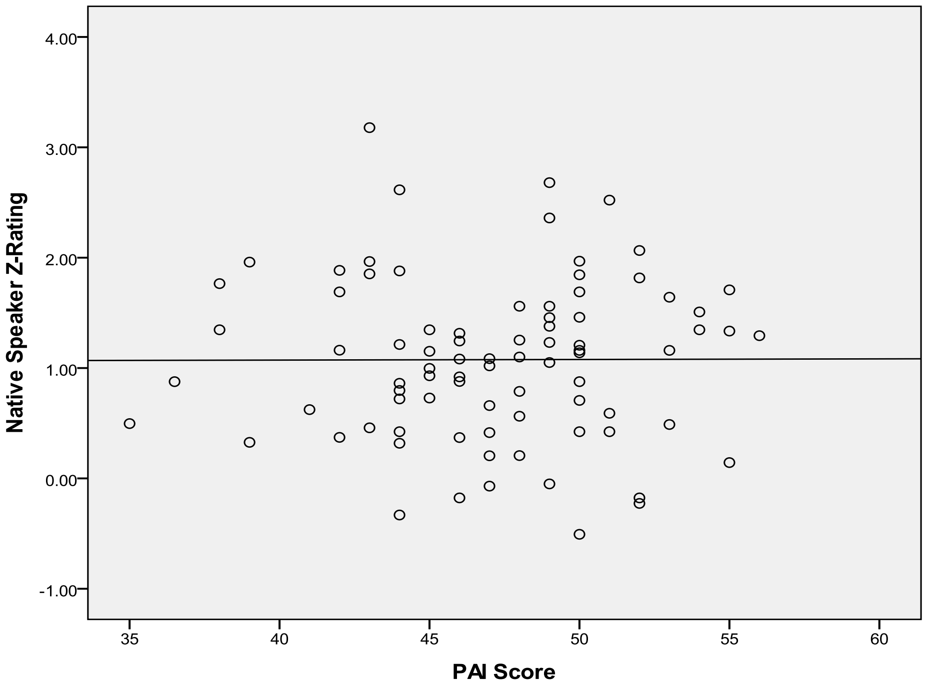
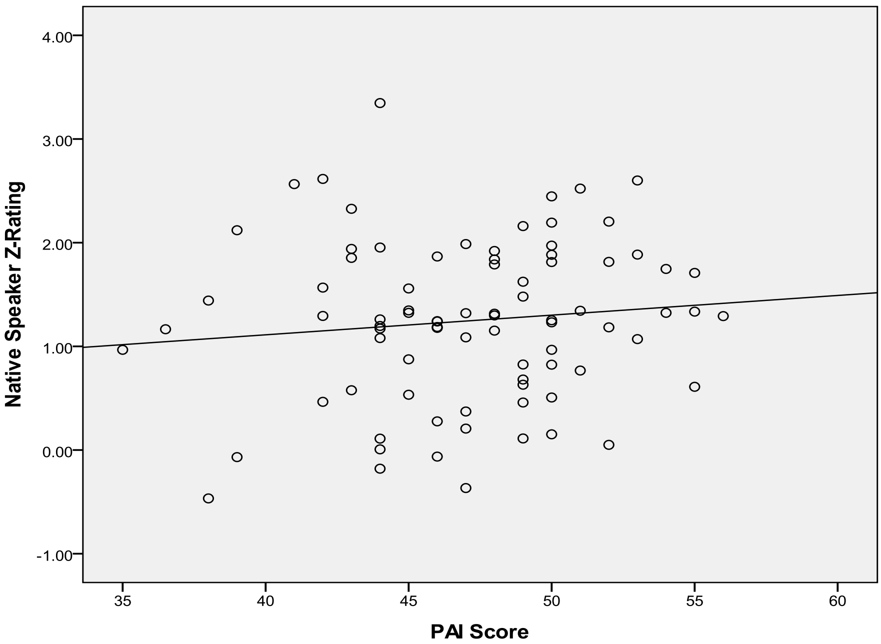


Figure 7.5: The Relationship Between PAI and Pre-test Native Speaker Z-Rating of the Oral Contrast



As suggested by the Figures, none of these correlations (Pearson's) are significant (Identification accuracy: Nasal r = .016, p = .883; Oral r = -.070, p = .528; Ratings: Nasal r = .003; p = .976; Oral r = .109, p = .328). This is in contrast to results with similar foreign language students found by Elliott (1995), who did not examine the relationship between PAI and training.

7.4.3 Post-Training PAI Relationships

Despite the lack of relationship between pre-test pronunciation ability and concern for pronunciation accuracy as measured by the PAI score, it was anticipated that any degree of improvement after training may be related to the PAI score, and that this relationship may differ dependent upon training group and/or the contrast trained. To this end, average improvement scores were calculated for all participants, for both native speaker identification accuracy and native speaker z-ratings. Visual examination of these data overall using scatterplots indicated no correlations, and the data were therefore examined on a training group basis and then broken down further according to contrast. The significant results are detailed below and all results according to training group can be found in Tables 7.1 and 7.2. Two-tailed tests are used as the direction of the relationships could not be anticipated, for example, those with low PAI scores may have responded well to a certain training technique.

Table 7.1: Correlation between PAI and Improvement in Pronunciation Identification and Pronunciation Z-Rating According to Training Group

	Correlation	
	PAI/Mean Identification Improvement	**PAI/Mean Z-Rating Improvement**
Multi Speaker HVPT + Pronun.	r = .24; p = .42	r = -.22; p = .46
Single Speaker PF + Pronun.	r = -.27; p = .34	r = .17; p = .56
Pronunciation Only	r = .23; p = .42	r = -.33; p = .25
Multiple Speaker HVPT	r = .31; p = .30	r = -.32; p = .28
Single Speaker PF	**r = .54; p = .06***	r = -.39; p = .19
Control	r = -.22; p = .43	r = .33; p = .23

*p < .1

Table 7.2: Correlation between PAI and Improvement in Pronunciation Identification and Pronunciation Z-Rating According to Training Group and Contrast Tested

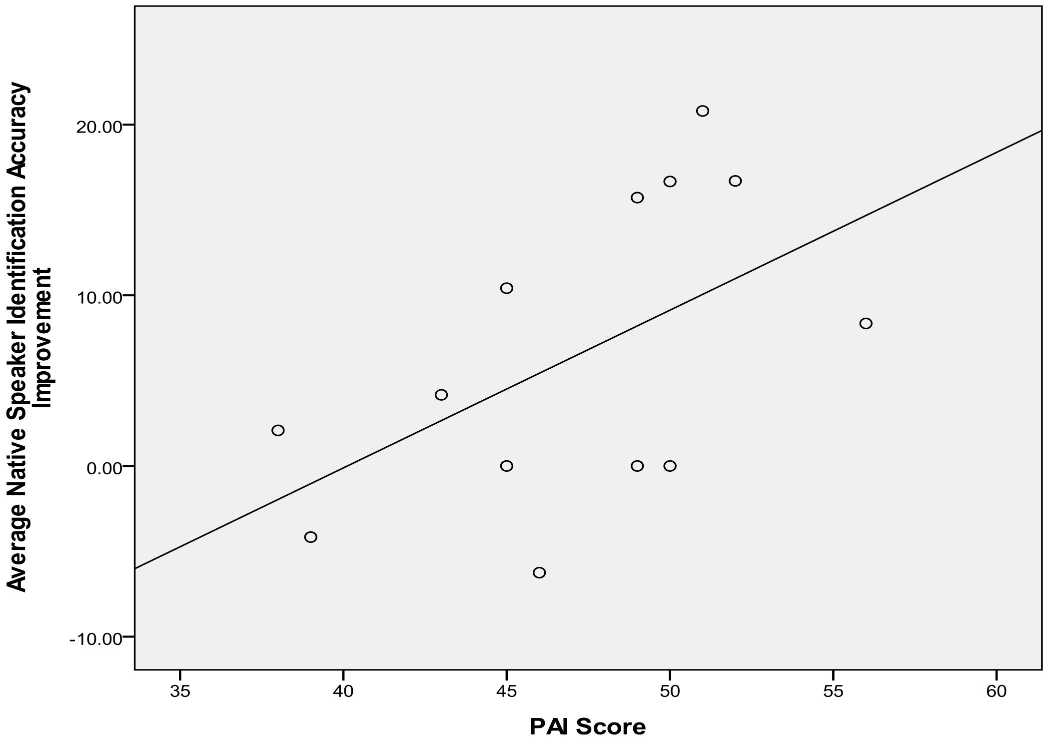
	Correlation			
	PAI/Mean Nasal Contrast Identification Improvement	**PAI/Mean Nasal Contrast Rating Improvement**	**PAI/Mean Oral Contrast Identification Improvement**	**PAI/Mean Oral Contrast Rating Improvement**
Multi Speaker HVPT + Pronun.	r = .19; p = .51	r = -.08; p = .79	r = .21; p = .48	r = -.29; p = .32
Single Speaker PF + Pronun.	r = -.06; p = .83	r = .07; p = .82	r = -.33; p = .25	r = .21; p = .46
Pronunciation Only	r = .44; p = .11	**r = -69; p = .01****	r = -.12; p = .69	r = 22; p = .45
Multiple Speaker HVPT	r = .09; p = .78	r = -.08; p = .84	r = 32; p = .28	r = -.38; p = .28
Single Speaker PF	r = 13; p = .68	r = -.08; p = .79	**r = .49; p = .09***	r = -.40; p = .20
Control	r = -.08; p = .78	r = .03; p = .92	r = -.25; p = .37	r = .37; p = .18

*p < .1 **p < .05

7.4.3.1 The PF Only Group

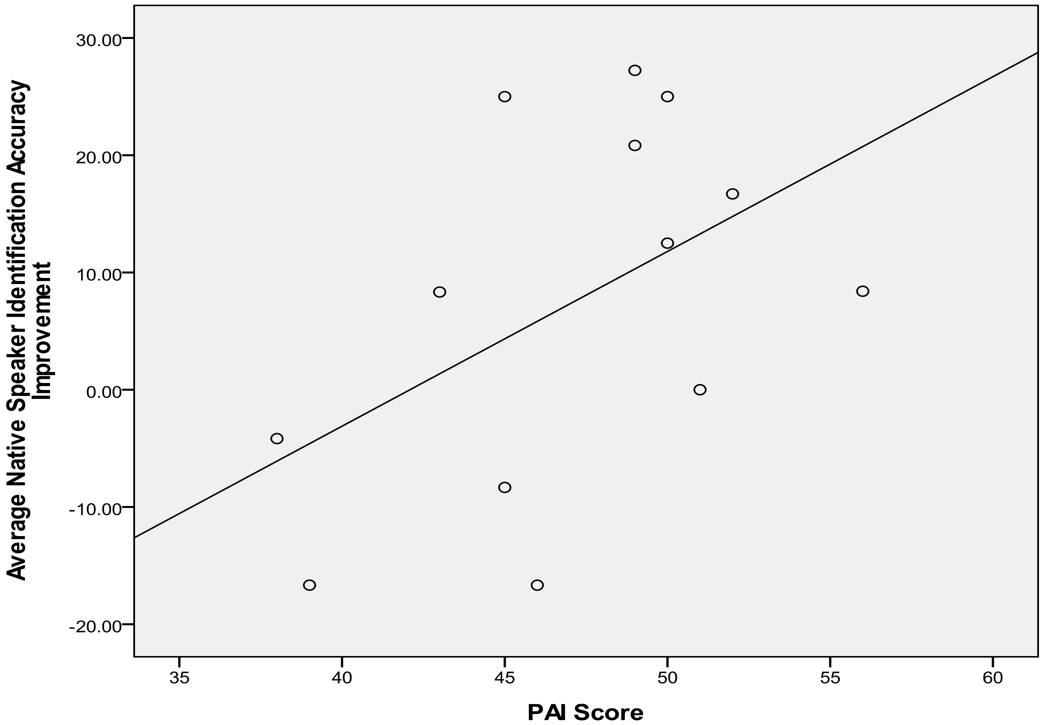
Visual examination of the data broken down by training group suggested a relationship between PAI and average improvement in native speaker identification accuracy for the PF only training group (see Figure 7.6). A Pearson's correlation confirmed that this relationship approached significance (r = .536, p = .059), suggesting that those participants with a higher PAI score who underwent PF training tended to improve more in pronunciation accuracy as determined by native speaker identifications. The R^2 value would therefore indicate that the PAI can account for 28.8% of the variation in this improvement in average native speaker identification accuracy.

Figure 7.6: PF Only Training Group: The Relationship Between PAI and Average Native Speaker Identification Accuracy Improvement



To investigate this relationship further, the data were broken down according to contrast trained. Pearson's correlations confirmed that the relationship between PAI score and improvement in native speaker identification accuracy of participant productions of the oral contrast approached significance ($r = .493$, $p = .087$, the correlation is illustrated in Figure 7.7). The relationship was not significant for the nasal contrast productions ($r = .126$, $p =.681$) or for native speaker ratings of either contrast (see Table 7.2). The R^2 value would therefore indicate that the PAI can account for 24.3% of the variation in this improvement in average native speaker identification accuracy of participant productions of the oral contrast.

Figure 7.7: PF Only Training Group: The Relationship Between PAI and Average Native Speaker Identification Accuracy Improvement for the Oral Contrast



Examination of these oral contrast data according to test indicated that the participants in this training group with higher PAI scores tended to improve more from pre-test to generalisation test time 2 in terms of native speaker identification accuracy of their oral contrast productions (r = .756, p = .011, see Figure 7.8), whereas no other relationships were significant (although moderate at r = .37, r = .4 and r = .45, see Table 7.3). The R^2 value would therefore indicate that the PAI can account for 57.2% of the variation in native speaker identification accuracy improvement from pre-test to generalisation test time 2.

Figure 7.8: PF Only Training Group: The Relationship Between PAI and Native Speaker Identification Accuracy Improvement for the Oral Contrast Pre-test to Generalisation Test Time 2

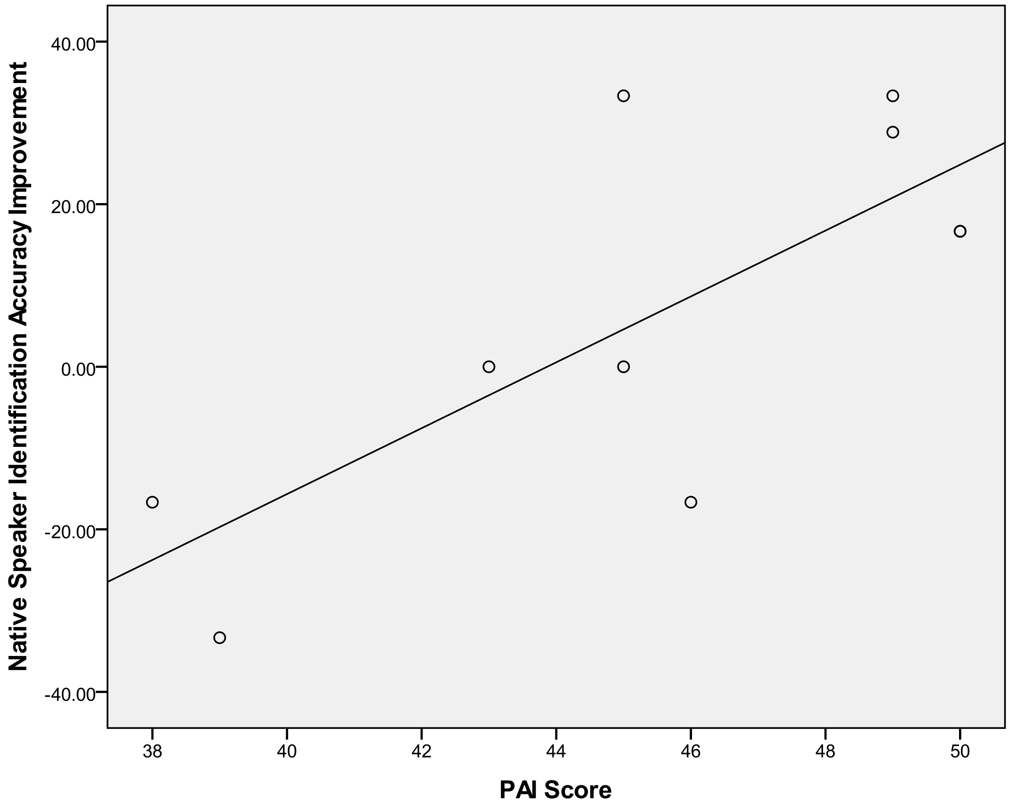


Table 7.3: Correlation Between PAI and the Native Speaker Identification Accuracy of the Single Speaker PF Training Group's Productions of the Oral Contrast

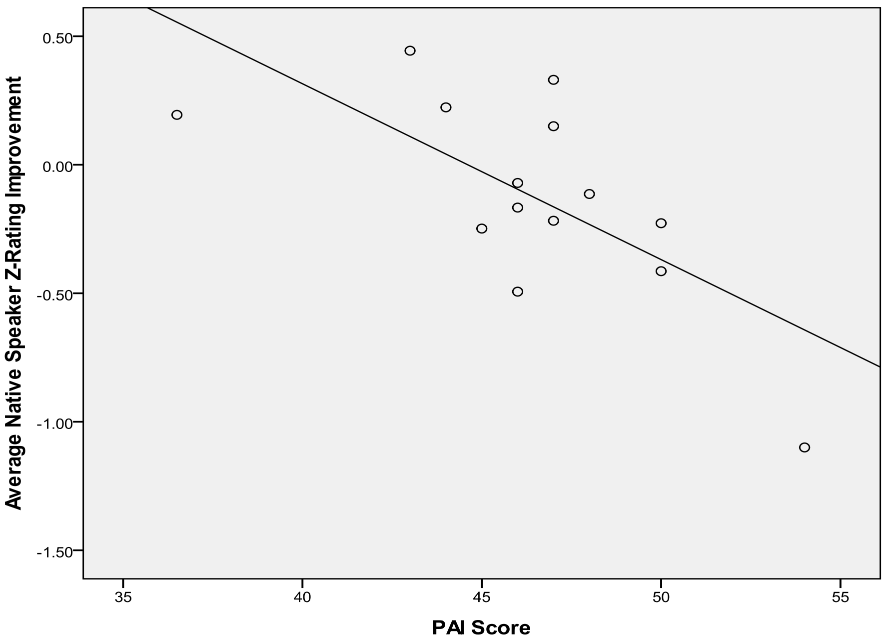
	Correlation			
	PAI/ Improvement Pre-Post	**PAI/Improvement Pre-Gen**	**PAI/Improvement Pre-Post2**	**PAI/Improvement Pre-Gen2**
Single Speaker PF Group	r = .37; p = .21	r = .40; p = .18	r = .45; p = .20	**r = .76; p = .01****

**p < .05

7.4.3.2 The Pronunciation Only Group

Examination of the pronunciation only group data according to contrast trained revealed a significant correlation between a higher PAI score and average improvement in native speaker ratings of participant nasal contrast productions (r = -.686, p = .007), this relationship is illustrated in Figure 7.9. No other relationships were significant (see Table 7.2). The R^2 value would therefore indicate that PAI score can account for 47.1% of the variation in average improvement in speaker ratings of participant nasal contrast productions.

Figure 7.9: Pronunciation Only Training Group: The Relationship Between PAI and Average Native Speaker Z-Rating Improvement for the Nasal Contrast



Moving on to look at these nasal contrast data according to test indicated that the participants in this training group with higher PAI scores tended to improve more from pre-test to generalisation test time 2 in terms of native speaker ratings of their nasal contrast productions ($r = -.844$, $p < .001$), no other relationships were significant (see Table 7.4). This relationship is illustrated in Figure 7.10. The R^2 value from the relationship would therefore indicate that PAI score can account for 71.2% of the variation in native speaker rating improvement.

Figure 7.10: Pronunciation Only Training Group: The Relationship Between PAI and Native Speaker Z-Rating Improvement for the Nasal Contrast Pre-test to Generalisation Test Time 2

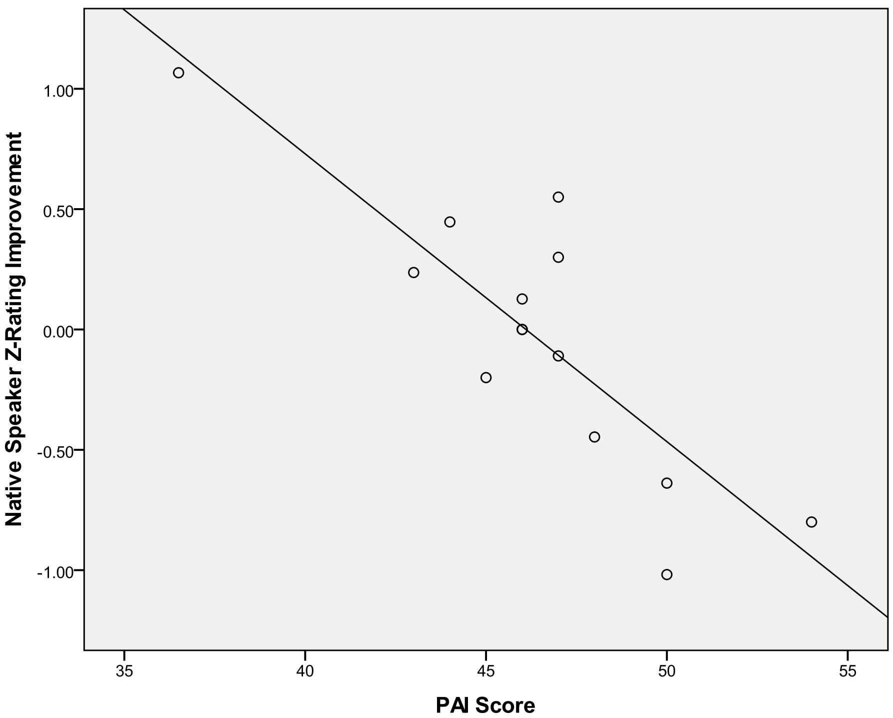


Table 7.4: Correlation Between PAI and the Pronunciation Only Training Group's Improvement in Pronunciation Accuracy (Z-Transformed Ratings) of the Nasal Contrast

	Correlation			
	PAI/ Improvement Pre-Post	**PAI/Improvement Pre-Gen**	**PAI/Improvement Pre-Post2**	**PAI/Improvement Pre-Gen2**
Pronun. Only Group	r = -.33; p = .24	r = -.07; p = .82	r = -.39; p = .17	**r = -.84; p < .00***

*** p < .00

7.5 Discussion and Conclusions

Very few significant relationships have emerged from the current data, of the five training groups and one control group, only a selection of results from two training groups have been significantly related to motivation or concern for pronunciation accuracy as measured by the PAI. In other words, it appears that motivation only plays a role in very specific instances in the current data. However, as PAI score had some role to play in the data from PF only and pronunciation only training groups, this arguably suggests that these techniques are sensitive to participant concern for pronunciation accuracy in a way that the other training techniques are not. Therefore, whilst this concern does not appear to be important if using the HVPT only, HVPT + pronunciation or PF + pronunciation techniques, those with high levels of motivation improve more with the PF only and pronunciation only techniques.

As can be seen from the data presented in Chapter 6, the untrained control group's results changed very little across tests and time and it is therefore unlikely that the results would be related to any other measure. It could also be speculated that the results from the single speaker PF + pronunciation, multiple speaker HVPT+ pronunciation and multiple speaker HVPT groups appear unrelated to PAI score due to there being more variety with these training techniques with new voices and/or new tasks (perception vs. pronunciation) being presented across the six training sessions. It is possible that those with lower levels of concern for pronunciation accuracy are still able to effectively engage with these training conditions. Conversely, the training conditions which are linked to PAI score involve only one task and one voice perhaps resulting in more monotony across the six training sessions meaning that only those with higher PAI scores are able to meaningfully engage with the training task. It is thus possible that the PF and pronunciation techniques should only be used if highly motivated participants (as measured by the PAI) are used.

The significant effects emerge for different contrasts and measures within each training group, with the PAI scores of the PF group being significantly related to native speaker identification of their oral contrast productions and the PAI scores of the pronunciation group being significantly related to native speaker rating of their nasal contrast productions. It is possible that PAI score is related to improvement in native speaker identification of the PF group's pronunciations and oral contrast pronunciation in particular due to the differing levels of difficulty in acquiring each contrast. Overall the results from Chapters 5 and 6 suggest the nasal contrast appeared to be the harder of the two to master both in pronunciation and perception, with scores from participants and native French speaking controls alike being lower for this contrast[1]. Due to the

---

[1] Whilst there is no significant difference in mean native speaker identification improvement between contrasts ($M_{NasalImp}$ = 4.47, $M_{OralImp}$ = 8.06, p = .214), the difference between mean participant

difficulty of the nasal contrast, a higher concern for pronunciation accuracy may not have been enough to have an effect when undergoing PF training. However, as the oral contrast appears easier to acquire, having a higher level of concern for pronunciation accuracy may have resulted in an additional benefit, although why this would only emerge from native speaker identification accuracy and not the native speaker ratings is unclear.

The significant results from the pronunciation only training group may have emerged for the opposite reasons. Results from Chapter 6 gave a slight indication that pronunciation training was more beneficial in terms of producing improvements in pronunciation than perceptual training. Significant relationships may have emerged for the difficult nasal contrast for this pronunciation group in particular because of the slight benefit of pronunciation training. However, due to the difficulty of the contrast, and as noted above, it is possible that improvements could only be made by those with a higher concern for pronunciation accuracy (as measured by PAI score) as those scoring highly would be more likely to engage fully with the training task. Again, why this effect would only emerge from native speaker ratings and not identification accuracy is unclear, as the latter is arguably the stricter measure of pronunciation. For the oral contrast, it is possible that no significant relationships emerged as it is easier to learn than the nasal contrast and therefore did not specifically require a higher level of concern for pronunciation accuracy (as measured by PAI score) to engage in the task sufficiently to result in greater pronunciation improvement.

A further area of interest was to investigate whether the significant relationships noted above specifically required a higher concern for pronunciation accuracy rather than a high motivation for the task of learning French in general. To this end scores were also calculated from 10 of the 13 PAI filler items for the members of the PF only and pronunciation only groups. These filler items should capture a general motivation to learn French, whilst the three omitted items (regarding when work on assignments is carried out, whether some French classes are preferred to others and whether French is the easiest subject the participant is studying, see Appendix H) appear unrelated to

improvement in identification is approaching significance ($M_{NasalImp}$ = 3.88, $M_{OralImp}$ = 6.91, p = .061) and the difference between mean native speaker Z-rating improvement is significant ($M_{NasalImp}$ = -.04, $M_{OralImp}$ = -.21, p = .034) motivation. Partial correlations controlling for the score from these filler items to separate concern for pronunciation accuracy from general motivation to learn French broadly indicated that the significant relationships detailed in 7.4.3 are specifically linked to the concern for pronunciation accuracy aspect of motivation in particular rather than motivation in general, although motivation in general appears to be related to overall improvement in native speaker identification of the PF only group productions. Examining the results for the PF only group, the relationship between PAI score and improvement in native speaker identification no longer approached significance ($r = .456$; $r^2 = .207$; $p = .218$). The relationship between PAI score and the improvement in native speaker identification for the oral contrast continued to approach significance ($r = .649$; $r^2 = .421$; $p = .059$), with a stronger relationship than identified by the bivariate correlation previously calculated. Finally, the relationship between PAI score and improvement in native speaker identification for the oral contrast from pre-test to generalisation test time 2 remained significant ($r = .742$; $r^2 = .550$; $p = .022$). For the pronunciation only group, the relationships between PAI score and improvement in native speaker ratings of participant nasal contrast productions both overall and from pre-test to generalisation test time 2 remained significant ($r = -.671$; $r^2 = .450$; $p = .012$ and $r = -.857$; $r^2 = .734$ $p <.000$).

Overall, the finding of a limited relationship between motivation/concern for pronunciation accuracy and actual pronunciation accuracy is in particular in contrast to Elliott (1995) who found that PAI score explained most of the variance in a model of the pronunciation success of university students of Spanish. However, Smit (2002) did find a limited relationship between the variables, and Birdsong (2003, 2007) also noted that a high level of motivation did not guarantee pronunciation success. Furthermore, this limited relationship is perhaps unsurprising as these participants volunteered to take part in the present perception and production training study. It is therefore probable that they will be more highly motivated/have a higher concern for pronunciation accuracy than those who did not volunteer. As noted in 7.4.1, the mean score in the present work was 47.11, which reflects a high concern for pronunciation accuracy. In comparison, the mean score from Elliott's (1995) participants was 32.79, lower than the lowest score recorded from the present group of participants, although it is unclear whether the participants were volunteers or whether an entire cohort participated. An increased number of significant correlations would be more likely from a group of participants with PAI scores across the entire possible range, a distribution that is not present in the current sample.

In sum, language learning motivation in general and/or the concern for pronunciation accuracy element of motivation as measured by PAI score plays only a limited role in the current data, in specific instances of the PF training only and pronunciation training only results. This provided some indication that these techniques may be more sensitive to participant motivation level and should therefore be avoided unless highly motivated participants are guaranteed.

8 General Discussion and Conclusions

8.1 Overview of the Present Work

8.1.1 Summary of Results

The main purpose of the present work was to compare the effectiveness of existing lab-based training techniques on improving the perception and pronunciation of two difficult French contrasts, and examine whether these results were related to concern for pronunciation accuracy as measured by the Pronunciation Attitude Inventory (Elliott, 1995). Chapter 5 compared the effects of PF and HVPT training with single and multiple speakers and the results suggest that PF training using one voice (as opposed to multiple voices) and HVPT training using multiple voices (as opposed to one voice) are the most successful techniques for training both the oral and nasal French contrasts in terms of improvement in participant identification accuracy. This effect only emerged when testing after a delay; only the single speaker HVPT technique was inferior immediately after training, but the single speaker PF and multiple speaker HVPT techniques were more effective in retaining improvements in the long term. These improvements generalised to identification of new words from familiar speakers and to identification of new words from new speakers.

Chapter 6 examined the benefits of using pronunciation training as well as perception training and analysed participant pronunciation as well as perception. In terms of improving participant perceptual identification accuracy, the perceptual training techniques were the most successful. However replacing half of the multiple speaker HVPT training with pronunciation training did not have a detrimental effect on perceptual identification accuracy. This effect of multiple speaker HVPT + pronunciation training producing comparable results to the perception only training methods was found in particular when comparing these training groups to native speakers in the second test of generalisation (aside from this the single speaker PF + pronunciation group was also successful). The multiple speaker HVPT + pronunciation technique also resulted in generalisation to identification of new words from a familiar speaker and new words from new speakers. The perceptual results only showed slight evidence of transfer from pronunciation only training in the generalisation tests.

In terms of participant pronunciation accuracy, the native speaker identification accuracy results suggested that both the multiple speaker + pronunciation technique and the pronunciation only technique produced better results than no training, however this was only true in the delayed post-test. All training types had little effect in terms of generalisation. Similarly, the native speaker rating results provided a slight suggestion that the pronunciation only technique produced better results than no training, however all training types again had little effect in terms of generalisation. In addition, the pronunciation results showed no evidence of transfer from perception-only training.

Chapter 7 examined the relationship between motivation, in particular its concern for pronunciation accuracy aspect, as measured by Elliott's (1995) Pronunciation Attitude Inventory, and pronunciation accuracy and average improvement in pronunciation accuracy as measured by native speaker ratings and native speaker identification accuracy. Participant concern for pronunciation accuracy did not appear to be strongly correlated to initial measures of pronunciation accuracy or improvement in pronunciation accuracy. However, some relationships between training and concern for pronunciation accuracy were found: For the single speaker PF group there was a significant (or approaching significance) relationship between concern for pronunciation accuracy and 1) improvement in native speaker identification accuracy of participant productions, 2) improvement in native speaker identification accuracy of participant oral contrast productions and 3) improvement in native speaker identification accuracy of participant oral contrast productions from pre-test to generalisation test time 2. For the pronunciation only group there was a significant relationship between concern for pronunciation accuracy and 1) improvement in native speaker ratings of participant nasal contrast productions and 2) improvement in native speaker ratings of participant nasal contrast productions from pre-test to generalisation test time 2.

How these results answer the research questions posed at the beginning of this thesis is considered in the next section.

8.1.2 Implication of Results for Research Questions

- RQ 1: Is the HVPT technique more successful than the perceptual fading technique (or vice versa) in terms of producing a generalisable, long-term improvement in perception?

The results from Chapter 5 suggest that the HVPT technique and PF technique in their classic forms are equally as successful in terms of producing a generalisable, long-term improvement in perception. Consistent with the findings of Iverson et al (2005), all techniques examined were equally as successful immediately after training (excepting single speaker HVPT training which was not examined by the authors). However, after a delay the single speaker PF technique and the multiple speaker HVPT technique emerged as more successful in retaining generalisable perceptual improvement.

- RQ 2: Is the most successful perceptual training technique(s) suggested through answering research question 1 more successful than pronunciation training in terms of producing a generalisable, long-term improvement in pronunciation and perception?

- RQ 3: With regard to using HVPT and/or perceptual fading and perception and/or production, does an optimal training technique emerge from those examined?

Taken together, the results from Chapter 6 suggest that the most successful perceptual training techniques identified in Chapter 5 are more successful than pronunciation training only in producing a generalisable improvement in perception. Evidence also suggests that pronunciation training only is more successful at improving pronunciation accuracy than the perceptual training only techniques in this instance but the pronunciation results showed no evidence of generalisation. Of the two groups which underwent perception and pronunciation training, only the multiple speaker HVPT + pronunciation group performed comparably to the perception training only groups in perception and the pronunciation only group in pronunciation. The tentative answer to the third research question is therefore that the multiple speaker HVPT + pronunciation technique is the optimal technique suggested by the present work.

- RQ 4: Do those with a stronger concern for pronunciation accuracy perform better or improve more in terms of pronunciation with training and with which techniques?

The results from Chapter 7 suggest that prior to training, in contrast to Elliott (1995), a stronger concern for pronunciation accuracy is unrelated to participant actual pronunciation accuracy as measured by native speaker identification accuracy and native speaker rating of participant pronunciations. After training, those with a stronger concern for pronunciation accuracy improve more than those with a weaker concern for pronunciation accuracy only if they have undertaken single speaker PF training or pronunciation training. However, this relationship was only present with specific measurements of pronunciation success: native speaker identification accuracy of the single speaker PF training group productions and native speaker rating of the pronunciation group productions. For all other techniques concern for pronunciation accuracy was not related to pronunciation success. This finding of a limited relationship between concern for pronunciation accuracy and pronunciation success was perhaps not surprising due to high levels of concern for pronunciation accuracy across all participants.

8.2 General Discussion of Present Results

8.2.1 Perceptual Fading vs. High Variability Phonetic Training

The greater success of the HVPT technique using multiple speakers in training as opposed to a single speaker as demonstrated in Chapter 5 is consistent with the original results of Lively et al (1993) and subsequent studies have generally continued to use the technique in this form (e.g., Bradlow et al, 1999, Lambacher et al, 2005, Lengeris & Hazan, 2010).

By the same logic of the benefit of increased variability, it would perhaps be expected that the multiple speaker PF technique would have proven to be more successful than the single speaker PF technique, which was not the case in the present work. It was hypothesised in Chapter 5 that the multiple speaker PF technique may have had too much variability as compared to the single speaker PF technique which confused participants over time. However, it was also noted that too much variability could not be the full reason as the multiple speaker PF technique does not appear to have more variability than the successful multiple HVPT technique. A possible explanation comes from the fact that evidence for the increased difficulty in identification of minimal pair members with stimulus variability from multiple speaker perceptual fading can also be found in the native speaker identification function data from Chapter 4 (Section 4.3). It is clear that even the native speakers found the task of differentiating the multiple speaker perceptual fading nasal contrast minimal pairs more difficult than differentiating the single speaker perceptual fading nasal contrast minimal pairs.

The native speaker data from Chapter 6 suggest that this difficulty arises with the nasals in particular because native speakers tend to be more variable in how they perceive and produce this contrast as compared to the oral contrast. The results from Chapter 4 indicate that this variability could arise from the fact that in ambiguous and also less ambiguous cases individual native speakers hear differing members of a minimal pair when presented with the same stimulus. The multiple speaker condition could further increase the difficulty in identification of the manipulated minimal pair members due to the fact that whilst the vowel formant and nasal values were held constant across single and multiple speaker training conditions, the vowel F0 and length were matched to the each native speaker's original production, thus creating more variability for the listener. Overall, the nasal multiple speaker PF training being too difficult could partly explain why participants in this condition performed less well.

8.2.2 Perception and/or Pronunciation Training

8.2.2.1 Discussion of Training Techniques

The two perceptual training techniques examined in the present work have a slightly different focus. Logan et al (1991) note that previous studies (e.g. Sheldon & Strange, 1982) had demonstrated that Japanese participants find the /r/-/l/ contrast more or less difficult dependent upon the phonetic context in which they occur. The HVPT technique therefore re-creates this real-life variability by making use of multiple speakers and multiple phonetic environments. In contrast, the PF technique makes use of only one voice, synthesised or natural altered but focuses upon directing the learners' attention upon category boundary locations by beginning with presentation of stimuli pairs furthest from the category boundary and finishing with presentation of stimuli pairs near this boundary (Jamieson & Morosan, 1986). PF training originally also made use of the contrast in one context and it can therefore be argued that using multiple vowel positions in PF training as in the present work can also be seen borrowing from the HVPT technique. However, it was decided that in order to find sufficient minimal pairs to train and test, the findings of Lively et al (1993) (which highlighted the importance of multiple voices over having the contrast to be trained in multiple word positions) rendered this potential borrowing less important.

As previously noted, the present work found these techniques in their classic forms (or possibly nearer to classic form in the case of the single speaker PF technique as multiple phonetic positions were used in training) to be equally successful at improving perceptual ability (and equally as unsuccessful at improving pronunciation ability). It therefore appears that despite the difference in focus, they produce very similar results and work equally well. In other words, the variability in voices and the variability in spectral quality through using a synthesised vowel continuum are equally beneficial, but, again as previously noted, using both in one stimulus may create a task that is too difficult. Although the present work suggests that single speaker PF and multiple speaker HVPT appear to achieve comparable results, the best way to make use of both techniques together if so desired therefore appears to have separate training blocks for each technique within one training programme (see, e.g., Wang & Munro, 2004).

Examining the perceptual training results, overall those undergoing the single speaker PF and multiple speaker HVPT techniques retain the benefits of training over time whilst the other techniques do not retain the benefits present immediately after training. Practice with the contrasts using all perceptual training techniques appears to be enough to score more highly in the immediate post-test rather than the pre-test. However, as previously noted, the single speaker HVPT technique does not appear to offer sufficient variability to result in the establishment of phonetic categories, and the multiple speaker PF technique appears to be too difficult to establish phonetic categories in the long term.

The single speaker PF and multiple speaker HVPT training techniques are only separated in terms of efficacy when they are used alongside pronunciation training. Examining the perceptual results, removal of half of the training time on PF training appears to be slightly more detrimental to the benefits of this technique than removal of half of the training time on HVPT, as undergoing HVPT + pronunciation training produces perceptual results comparable to having undergone perceptual training only in terms of comparison with native speakers. It is therefore possible that hearing multiple voices, even for a short time, is enough to benefit from the multiple speaker HVPT technique, whereas for the PF technique more training time is required on each pair of points on a continuum to effectively establish category boundaries. The results therefore suggest that the HVPT technique could be more efficient and could result in greater gains over shorter periods of training, a possibility which could be investigated in future work.

8.2.2.2 Transfer Between Training Modalities

The results from Chapter 6, which revealed little evidence of transfer from perceptual training to pronunciation, are inconsistent with the results from, for example, Bradlow et al (1997 – HVPT training) and Rochet and Chen (1992 – PF training) and the lack of evidence for this or for transfer of pronunciation training to perception is in disagreement with a number of subsequent studies. For example, Leather (1997) had one group of native Dutch speakers perceptually trained on Putonghua tones and tested their pronunciation after training and had a second group undergo pronunciation training and tested their perception after training and found evidence that training in one modality resulted in performance improvements in the other. Furthermore Gómez Lacabex et al (2008) trained native speakers of Spanish on full vowel-schwa minimal pairs with one group undergoing perceptual training and the second undergoing pronunciation training and found that both types of training had an equally positive effect on perceptual performance. In a second study Gómez Lacabex and García Lecumberri (2010) similarly trained two groups of participants and found both types of training had an equally positive effect on pronunciation performance (see also Aliaga-Garcia, 2010, and for mixed results in terms of transfer, Iverson, Pinet & Evans, 2012).

It is unclear why the present results disagree with these findings, in particular given that the present work has used the same PF and HVPT perceptual techniques which did result in pronunciation improvements in previous work. However, modification of the pronunciation training technique may result in the perception gains noted in the studies above; as there was less strong (although present to some extent) evidence for the benefit of pronunciation training for pronunciation, particularly in terms of generalisation, it is less likely that there would be any benefit to transfer to perception. For example, increasing the number of participants or the modification of the pronunciation training technique in particular may lead to more conclusive results from the pronunciation training and analysis of participant productions. It is possible that the pronunciation training technique used in the present work was too simple and that a technique with, for example, more interaction with the experimenter, more immediate comparison of both minimal pair sounds together rather than treating each sound individually, or which uses different modalities such as AV training (see Section 8.4 below) and pronunciation feedback (McCandliss et al (2002) found the use of feedback in perceptual training to be important and it seems reasonable that this would be important for pronunciation training also) may be more successful. However, Hattori (2009) specifically administered a far more detailed and complex pronunciation only training technique (using many of the above suggestions) on the English /r/-/l/ contrast to native speakers of Japanese and found improved participant pronunciation of the contrast without having any effect on their perceptual abilities. The lack of transfer in this direction is in contrast to the findings of Leather (1997) and the combined results of Gómez Lacabex et al (2008) and Gómez Lacabex and García Lecumberri (2010) yet somewhat consistent with the present results, suggesting that the issue of transfer of training across domains requires further investigation.

8.2.2.3 Perception/Pronunciation Learning vs. Perception/Pronunciation Links

Whilst finding little evidence of transfer of training across domains, the results from Chapter 6 do agree with another aspect of the findings of Bradlow et al (1997). The authors found that whilst there was an improvement in pronunciation by their perceptually trained participants overall, there was no correlation between perceptual learning and pronunciation learning, the presence of which would have demonstrated that those participants who improved most perceptually also improved most in pronunciation. However, participant perceptual and pronunciation performance were closely linked prior to training, showing that it was the amount of learning in each domain which differed in their individual participants (Bradlow et al, 1997; see also, e.g., Schneiderman, Bourdages & Champagne, 1988 and Iverson et al, 2012 for similar findings). Similarly, in the current data, a Pearson's correlation between participant average improvement in perception and native speaker average identification accuracy improvement of participant productions reveals only a modest positive relationship ($r = .207$, $p = .030$, 1-tailed). In addition, a Pearson's correlation between participant average improvement in perception and native speaker average rating improvement of participant productions is not significant ($r = -.101$, $p = .183$, 1-tailed). However, prior to training participant perceptual abilities are significantly related to their production abilities as measured by both native speaker identification accuracy ($r = .486$, $p <.000$) and z-transformed native speaker ratings ($r = -.479$, $p <.000$). Thus, whilst the present work found no evidence of a close link in perception and pronunciation learning, the data are in agreement with Bradlow et al (1997) that "...the processes of learning in the two domains [= perception and production] appear to be distinct within individual subjects." (p.2307). In the present work, this distinction of learning within individual subjects is likely to be at least partly due to the differing training techniques used by the participants, as demonstrated by the previously described significant group effects in the perception results in particular.

The agreements and contrasts between the present findings and those of other studies suggest that the relationship between perception and production is unclear. It may be prudent to consider a distinction between perception and pronunciation ability (the relationship between which is demonstrated by the significant correlation between participant perception and pronunciation abilities prior to training found in the present work and, e.g., Bradlow et al, 1997) and perceptual and pronunciation learning (the lack of a link is suggested by the non significant relationships found in the present work and, e.g., Bradlow et al, 1997), and consider these distinct from a general link between the processes of perception and production (the presence of which is suggested by the studies which have found that training in one modality transfers to another, e.g. Bradlow et al 1997; Leather, 1997; Gómez Lacabex et al, 2008; Gómez Lacabex & García Lecumberri, 2010; Aliaga-Garcia, 2010. Studies such as the present work and

Hattori (2009) which found no transfer suggest that the nature of this link is unclear). Further confusion about the links between perception and production is provided by the findings of, for example, Goto (1971) and Sheldon and Strange (1982) who found that their native speaking Japanese participants could pronounce but not perceive the English /r/-/l/ contrast. This is in contrast with the findings of Rochet (1995) which suggested that pronunciation was dependent upon perception. Further work should therefore continue to examine perception-production links, with and without training in order to clarify the nature of these links.

8.2.2.4 Implications for L2 Learning Models

The results from Chapter 6 can also be examined with reference to two main theories of second language and cross language speech in adults. The Speech Learning Model (SLM, e.g. Flege, 1995b) and the Perceptual Assimilation Model (PAM, e.g. Best, 1995) extended to the PAM-L2 by Best and Tyler (2007) are in agreement that perception of L2 sounds is partly shaped by experience in perceiving the L1. In extending the PAM to the PAM-L2 to consider L2 learners Best and Tyler (2007) note that the SLM and PAM-L2 are also in agreement that perceptual learning mechanisms and processes used in learning the L1 sound system remain intact throughout life and therefore that phonetic categories can be modified with exposure to an L2. The present results are therefore consistent with these models in that perceptual learning did occur with the use of some of these training techniques (see also, e.g., Hattori, 2009, for similar conclusions).

In contrast, the present results appear at first glance to be in disagreement with other aspects of the predictions of these models. The SLM suggests that without accurate perceptual representations of L2 sounds, pronunciation of L2 sounds will be inaccurate but if a new category is established for an L2 vowel through experience, it should also be produced accurately (Flege, 1995b) and according to Bradlow et al (1997) the PAM (and, by extension, the more recent PAM-L2) would make similar predictions. Formation of new phonetic categories can be blocked by, in SLM terms, an L2 phonological category being perceived as equivalent to an L1 phoneme, a process known as equivalence classification (e.g. Flege, 1995b); or in PAM terms blocked by an L2 phonological category being perceptually assimilated to the L1 phoneme or a new phoneme in a number of patterns (e.g. Best 1995). For example, potential patterns of equivalence classification or assimilation for the /u/-/y/ contrast examined by the present work are suggested by, for example, Flege (1987) who found that /y/ was a 'new' sound pronounced accurately by all but the most inexperienced of L2 French speaking participants, whereas French /u/ was never pronounced accurately by the participants, having been blocked by the process of equivalence classification to English /u/ (or in PAM terms /u/ was assimilated to a native category and /y/was assimilated as an uncategorisable (and therefore non-native or new) speech sound (Best, 1995)). The results of Macdonald (2006) largely supported this except that for inexperienced Scottish learners of French, French /y/ was just as an acceptable exemplar of their English /u/ as French /u/ (perhaps due to a more fronted /u/ in their English). For these participants, both members of the contrast had been subject to equivalence classification (e.g. Flege, 1995b) or had been assimilated to their native /u/ as equally good or acceptable members of that category, in a single-category assimilation (e.g. Best, 1995).

As little previous work has been carried out on the nasal /ɑ̃/-/ɔ̃/ contrast it is unclear what patterns of equivalence classification or assimilation may have been expected to occur. However, the results from Chapter 4 (Section 4.1) suggest that untrained inexperienced and experienced participants have great difficulty discriminating one member of this contrast from the other. It is unclear whether these nasal vowels are 'new' (or assimilated as uncategorisable) to native speakers of English or are sufficiently akin to sounds which do exist to be classified as 'similar' (or assimilated to a native category) and therefore be subject to equivalence classification (Best, 1995; Flege, 1995b). However, the poor discrimination performance is predicted by the PAM in both cases with either a both uncategorisable or single category assimilation pattern.

Given that some perceptual learning occurred in the present study, it can be argued that category formation through training was not fully blocked by these processes. In addirion, the present results suggest that as well as the perceptual representations of the sounds had becoming more accurate, this also transferred to pronunciation to a lesser extent as predicted by these models. Furthermore, the SLM in particular predicts that adults can learn to produce foreign vowels in a nativelike manner if they are not like any native vowel (see Flege, 1987), and there was slight evidence that the difference between some trained participants and native speakers was not significant.

A closer examination suggests that the present perceptual results being more convincing than the pronunciation results would also be predicted by the present models. For example, as Bohn and Flege (1997) note, the accurate perception and pronunciation of and new category formation of foreign vowels which are not similar to any L1 vowel will only happen through extensive L2 input. It is likely that the participant experience of classroom teaching along with the training in the present study was not sufficient L2 input in terms of quality or quantity for full category formation to occur. Bohn and Flege (1997) examined the perception-production relationship between the L2 English /ɛ/-/æ/ contrast for L1 speakers of German and found that inexperienced L2 learners may differentiate a new vowel contrast perceptually whilst being unable to differentiate the contrast in production. More specifically, Bohn and Flege (1997) found that the 'experienced' group of native German speakers living in America could produce the 'new' L2 English vowel /æ/ to a native standard, whereas the 'inexperienced' group could not. Although both groups had learned English in school for approximately the same amount of time (experienced M = 7.6 years, inexperienced M = 6.6 years), the experienced group had been living in America for at least 5 years with a mean of 7.5 years, whereas the inexperienced group consisted of recent arrivals with a mean residence of 0.6 years. It can therefore be argued that when the SLM hypothesis regarding pronunciation states that 'the pronunciation of a sound eventually corresponds to the properties represented in its phonetic category representation.' (Flege, 1995b, p. 239), this 'eventually' is likely to represent a number of years of intensive L2 input not experienced by the present participants.

Furthermore, Bohn and Flege (1997) also found, similarly to the present results, a lack of a relationship between perception and production of L2 vowels and suggest that the perception-production relationship for L2 sounds may differ for different classes of sounds (e.g. vowels vs. consonants) or may differ according to perceived similarity between L1 and L2 sounds.

8.2.2.5 Implications for Speech Perception Models

More broadly, the lack of link between perception and production suggested by no transfer of training benefit across domains in the present work appears to be in disagreement with the three main general accounts of speech production and perception, which all state that perception and production are closely linked[2]. The Motor Theory (e.g. Liberman & Mattingly, 1985), proposes that humans perceive intended speech gestures which are neuromotor commands to the articulators in the acoustic signal of speech and that this depends upon a specialised module which is used for both perception and pronunciation. The Direct Realist Theory (e.g. Best, 1995, the PAM has its basis in this) suggests that it is the actual (rather than intended) speech gestures which are perceived in the acoustic signal of speech, that there is a general perceptual system for all noise events rather than a specific language module and that perception and production are closely linked within this general system.

Finally, psychoacoustic models or the General Approach (see, e.g. Diehl, Lotto & Holt, 2004) state that listeners perceive speech directly from the acoustic signal using general perceptual and auditory mechanisms. This approach also posits a strong link between perception and production with Diehl et al (2004) describing the relationship as "Production follows perception, and perception follows production". (p.167). Regarding the former, Diehl et al (2004) note that there is a tendency for the sounds of languages to be placed as far apart within the phonetic space as possible in order that speech is intelligible even under more difficult listening conditions, known as the dispersion principle. The requirement for this dispersion to be maintained shapes production through the auditory enhancement hypothesis (e.g. Diehl & Kluender, 1989) by the fact that the gestures necessary to produce phonemes combine to make each phoneme as distinct as possible (Diehl et al, 2004). Regarding the latter, consistencies in production such as context dependencies will be in the acoustic signal and listeners will learn (via general perceptual learning mechanisms) to make use of these consistencies to judge the phonemic content of speech (Diehl et al, 2004).

[2] An alternative viewpoint is that the initial correlation between participant perception and production

Given that the wealth of evidence from such contrasting theories generally demonstrates close perception-production links, a change in pronunciation training methodology in the present work in order to firstly achieve greater pronunciation benefits (see 8.2.2.2 for suggested changes) could subsequently result in more convincing support or contrast to the work described above.

8.2.3 Concern for Pronunciation Accuracy and Pronunciation Success

Moving on to examine the results described in Chapter 7, the mixed findings regarding the relationship between motivation/concern for pronunciation accuracy and actual pronunciation accuracy are largely in agreement with those cited in Chapters 1 and 3, in particular the finding of a strong relationship by Elliott (1995) versus the finding of a limited relationship by Smit (2002). It is likely that studies should be in agreement in their measure of concern for pronunciation accuracy and indeed of pronunciation accuracy itself in order to produce more consistent results. In the present work, again, as the correlations were based upon pronunciation results, it is possible that a modified pronunciation training technique (including such modifications as suggested in 8.2.2.2) would result in stronger relationships.

Furthermore, it was suggested that the slightly more mundane nature of the training with the two techniques which did result in a link between concern for pronunciation accuracy and pronunciation success (PF training and pronunciation training always used the same task and voice) required a high level of concern for pronunciation accuracy to effectively engage in the process. The other techniques experienced variability in task, voice or both, perhaps creating enough interest to override a low of concern for pronunciation accuracy. Administration of an end of training questionnaire investigating how much participants had enjoyed the process should therefore be considered for future work looking at motivation/concern for pronunciation accuracy and pronunciation success whilst comparing techniques. This would allow for testing of whether or not participants undergoing the potentially more mundane techniques enjoyed the process less than those who underwent other training conditions. Flege (1995a) administered such a questionnaire when comparing identification vs. discrimination training techniques and as the techniques produced essentially the same results he found it worth considering an advantage for identification training based upon the fact that participants enjoyed this more than discrimination training.

8.2.4 Counteracting Age Related Effects?

The introduction to this thesis concerned nativelike performance by late learners of second languages and suggested that participation in training could contribute to this ability. The comparison of individual results with native speaker performance detailed in Chapter 6 were therefore of interest in terms of ascertaining whether anyone at all within the present sample performed to a native level (within the native range) both at initial pre-test and then with training.

Across all perception and pronunciation tests, only one participant scored within the native range in at least one of the post-training tests in both contrasts for all of the measures which was an improvement on this participant's performance pre-training. In addition training resulted in more participants scoring within the native range at least once than in pre-testing and training. It can therefore be argued that, for some participants, training has begun to counteract age-related effects on perception and pronunciation of their L2 French to a native level in some instances, and therefore also that the general aim of improving intelligibility through training has been met. However, no participant scored at a native level in all possible cases either before or after training, again emphasising both how rare nativelike performance in late learners can be and that motivation and training may be necessary but are certainly not sufficient to attain nativelike pronunciation by such late learners.

A further consideration in terms of counteracting age related effects is that the starting point for the present work was the general agreement that accurate pronunciation is one of the most difficult aspects of a second language to acquire, particularly if beginning to learn a language after early childhood, and that evidence of nativelike pronunciation by such late learners therefore suggests that it is possible to counteract such age related effects. However, a recent study by Abrahamsson and Hyltenstam (2009) firstly identified L2 learners (with Spanish as their L1) who sounded like native speakers of Swedish according to a native speaker panel, and then gave them further tests of perception, pronunciation and language tests such as grammaticality judgements. None of the late learners performed at a native level across all tests. It therefore appears that even those who can pass for native speakers do not necessarily master all aspects of L2 learning to a native level. At the same time, it is interesting to note that the highest performing late learner in this study only deviated from native speaker norms in phonetic aspects of speech production and perception thereby again highlighting the difficulty of this aspect of L2 acquisition.

8.3 Future Directions

8.3.1 Methodological Changes/Enhancements

There are a number of potentially beneficial changes to the training techniques used in the present studies which are worthy of consideration when attempting to train non-native speech contrasts and these should be taken into consideration when designing future training studies. One potential methodological enhancement is the number of vowels trained at one time. It is possible that simply selecting the most difficult vowel contrasts of a language to train may be less effective than training a number of vowels or vowel contrasts encompassing the vowel space of the L2. For example, whilst using the HVPT technique, Nishi and Kewley-Port (2007) trained native speakers of Japanese on either three difficult English vowels or a set of nine English vowels which covered the entire vowel space (all monopthongs). Although the authors found that perceptual training using the three difficult vowels resulted in small improvements and generalisations, this did not transfer to untrained vowels (the authors noted that McClaskey Pisoni and Carroll (1983) found that training the voicing contrast in stop consonants generalised to other stop contrasts with different places of articulation) and the training was not retained in the long term. Training using nine vowels was equally successful for all vowels and was retained in the long term. The authors concluded that training with small subsets containing the more difficult vowels may interfere in learning a complete vowel set (see also Nishi & Kewley-Port, 2008 for similar results with native speakers of Korean).

A further methodological change is potentially how the 'fading' aspect of PF training from easy-difficult is achieved. It need not be, for example, through alteration of formant values to make the contrasts more or less similar, but an alteration of the phonological contexts in which the contrasts appear dependent upon how easy or difficult the contrasts are to perceive in these contexts. How the /u/-/y/ contrast is categorised and how well it is discriminated by native speakers of American English has been found to be dependent upon context in which the vowels occur (Simon, Chambless & Alves, 2010; Levy & Strange, 2008; Levy, 2009). Simon et al (2010) therefore suggest a natural fading technique whereby teaching this contrast to native speakers of English should begin by using minimal or near minimal pairs with the vowels in consonantal contexts where the contrast can be most easily perceived. Once participants can perceive the contrast in the easiest context(s), the authors note that teaching or training can then move on to using more difficult pairs (Simon et al, 2010). This natural fading can also be easily combined with the HVPT technique (if beneficial and not resulting in too much variability) by presenting these pairs using multiple voices.

A potential enhancement to pronunciation training is an alteration to the voice used for participants to imitate. A number of studies have found participant pronunciation improvement after training sessions where the voice used in training is that of the participant altered to sound like a native speaker (e.g. Martin, 2004; Felps, Bortfield & Gutierrez-Osuna, 2009). The authors of both studies agree that this would be of particular benefit if the participant does not have access to a native speaker. Martin (2004) suggested that the strength of this technique may be that the participant having their own voice to imitate rather than that of another speaker whose voice may be very different (e.g. due to gender and/or pitch). In addition, it may be of benefit to use such an altered voice to perceptually train difficult L2 contrasts.

A final potential beneficial change to the techniques used in the present work is modality of training. Whilst using the HVPT technique, a number of studies over the past ten years have investigated the importance of using visual cues in training and have found an advantage for audiovisual training as opposed to audio only training. (e.g. Hardison, 2003; Hazan, Sennema, Iba & Faulkner, 2005). In audiovisual (AV) training the speaker's face is seen onscreen as well as their voice being heard by participants. For example, Hardison (2003) found that whilst auditory (A) HVPT training did result in a generalisable improvement in perception and pronunciation of the English /r/-/l/ contrast by native speakers of Japanese, AV training resulted in significantly greater improvement in perceptual accuracy from pre-test to post-test, greater generalisation to new words and voices and more beneficial transfer to pronunciation. Further perceptual advantages for AV over A training were found by Hazan et al (2005) for training native speakers of Japanese on the English /v/-/p/-/b/ contrast (and generalisation to new words and voices) but, in contrast to Hardison (2003), not the /r/-/l/ contrast. The authors suggested that this could be because the /r/-/l/ contrast is less visually distinct. However, the authors also examined transfer to pronunciation from perceptually training the /r/-/l/ contrast and found a significant benefit of AV over A training. Further work attempting to ascertain an optimal training technique should therefore consider using AV training stimuli. In addition, AV training stimuli could also be used with the PF training technique.

8.3.2 Broader Considerations: Second Language vs. Foreign Language Learning

A number of researchers have raised concerns that the general finding of an advantage for early learners acquiring their L2 in the L2 country or in an educational immersion setting (second language learning) is being applied to students of foreign languages learning their L2 solely in foreign language classes whilst living in their L1 country (foreign language learning, such as being experienced by the present participants) which by definition is more limited in the quality (i.e. lack of access to native speakers) and quantity (i.e. hours of exposure) of L2 input (Muñoz, 2006, 2008). One main area of concern on the part of such researchers is that applying the 'earlier is better' finding from more naturalistic L2 acquisition to classroom settings may influence educational policies without any evidence that this tenet stands true in such a learning environment (Muñoz, 2006, 2008, see also Moyer, 2004, for similar considerations regarding the importance of learning environment). Of more relevance to the present work (and any other work attempting to train or teach L2 contrasts with limited input), is that researchers such as Muñoz (2006, 2008) and Fullana (2006) propose that any research on age-related effects on L2 acquisition which uses Foreign Language (FL) learners as opposed to L2 learners should focus on rate of learning rather than ultimate attainment; and should focus on optimal realistic levels of attainment rather than attaining nativelike levels, primarily due to the fact that foreign language learning generally occurs during a finite length of time (Muñoz, 2006).

With regard to general language abilities Muñoz (2006) and with specific reference to L2 perception and pronunciation, Fullana (2006) examined language English L2 attainment in Spanish/Catalan bilinguals grouped according to when they had commenced English instruction and found that the older the learner was at the beginning of instruction, the faster their initial rate of learning. In addition, the younger learners (aged 2-6) eventually caught up to a degree with the older learners but did not generally surpass the older learners in terms of ability across the timescale of the study. In other words, in a foreign language learning environment there was no advantage for the early learners. In addition, none of the participants performed at nativelike levels. Muñoz (2006) concluded that the confusion surrounding age related effects in second language acquisition may at least be in part due to the lack of distinction drawn between foreign language learning in a classroom setting versus L2 learning in a naturalistic setting (or in a school immersion setting where the second language input is believed to be comparable). Caution should therefore be taken when examining age related effects on L2 acquisition and applying experimental conclusions to instructed language learning settings, in other words, research regarding age related effects may be less relevant to the participants used in the present study, or what should be expected of them in terms of ability and improvement through training.

At the same time, as noted in Chapter 1, Bongaerts and his colleagues found evidence of nativelike pronunciation among both late second language and late foreign language learners. Bongaerts et al (2000) compared the results of finding nativelike pronunciation amongst advanced Dutch learners of English and French from previous studies carried out by Bongaerts and his colleagues (Bongaerts et al, 1995, 1997; Palmen et al, 1997; Bongaerts et al 1999) along with the Bongaerts et al (2000) findings of nativelike pronunciation of Dutch by those who had acquired Dutch in an immersion setting in the Netherlands with little formal instruction. As previously noted, the performance of the nativelike participants who had received formal instruction was in the upper end of the native range whereas the nativelike participants who acquired Dutch in the immersion setting performed within the lower end of the native speaker range. It is therefore likely that researchers in both the second and foreign language fields would agree with the conclusions of Bongaerts et al (2000) that formal instruction along with a high quality and quantity of target language input are both important in promoting accurate pronunciation of that language.

It is also worthy of note that as well as having effects on target language achievement, the distinction between foreign and second language learning is also an important one in terms of language learning motivation. As noted in Chapter 3, the Socio-Educational model of motivation was criticised for being too dependent upon identification with the target language community and the Canadian language learning context. For example, Dörnyei (2001) suggested that integrativeness may be less relevant in foreign language contexts due to lack of contact with the target language community (in contrast to second language learning of French or English in Canada). The studies described in Chapter 3 of the present work suggested a close link between integrativeness and pronunciation accuracy, and many used foreign as opposed to second language learners, however future research using foreign language learners in instructed settings should take note of this distinction when considering how language learning motivation may contribute to counteracting age related effects.

8.4 Contributions and Conclusions

The present work has demonstrated that perceptual training and, to a lesser extent, pronunciation training can be beneficial to those experiencing difficulties with difficult L2 French vowel contrasts, with the optimal technique for obtaining improvement in both modalities appearing to be a combination of high variability perceptual training and pronunciation training (across the same timescale as training in one modality alone). Motivation level, or more specifically concern for pronunciation accuracy as measured by the pronunciation attitude inventory, appears only to be relevant in specific circumstances, speculated to be when training is more monotonous in nature. The findings of this work appear to be largely in agreement with theories of second language speech learning, that age effects or the influence of the L1 in perceiving or pronouncing the L2 can be counteracted to some extent by extra L2 input or experience. However, the lack of correlation between perceptual and productive learning compared with finding a perception-production correlation before training along with a lack of transfer between training modalities have resulted in a less good fit with general theories of speech perception and their stand on perception-production links.

More specifically, the experiment described in Chapter 5 provides, as far as can be ascertained, only the second comparison of the existing perceptual fading and high variability training techniques, and is the first to use L2 French vowels and is the first to address the comparison in terms of transfer to pronunciation and in terms of longer term retention. Chapter 6 takes this comparison and contribution further in suggesting that halving the time spent on training each modality in order that a little time can be spent on each is equally as beneficial for improving performance in each domain. Furthermore, the present work adds to the current small body of work which attempts to address the relationship between motivation/concern for pronunciation accuracy and actual accuracy in pronunciation. In addition, the present work has contributed further evidence that the L2 French /u/-/y/ contrast is difficult to perceive and produce for native speakers of English, has been among the first to empirically note the difficulties that native speakers have with the nasal /ɑ̃/-/ɔ̃/contrast and has contributed to the debate regarding identification versus discrimination testing and training. Overall, this thesis has contributed to the refinement of lab based training techniques for improving perception and pronunciation of difficult L2 contrasts towards a native level. However, there is still much room for improvement of these techniques and the refinement should continue from here by considering the suggested methodological modifications and broader considerations outlined above.

9 References

Abrahamsson, N., & Hyltenstam, K. (2003). Age of Onset and Nativelikeness in a Second Language: Listener Perception Versus Linguistic Scrutiny. *Language Learning, 59(2),* 249-305.

Abu-Rabia, S. & Kehat, S. (2004). The Critical Period for Second Language Pronunciation: Is there such a thing? Ten case studies of late starters who attained a native-like Hebrew accent. *Educational Psychology, 24 (1),* 77-98.

Aliaga-Garcia, C. (2010). Measuring perceptual cue weighting after training: A comparison of auditory vs. articulatory methods. In Dziubalska-Kołaczyk, K., Wrembel, M. & Kul, M. (Eds), *New Sounds 2010: Proceedings of the 6th International Symposium on the Acquisition of Second Language Speech,* pp 12-18, Poznań, Poland.

Baayen, R.H. (2008). *Analyzing Linguistic Data: A Practical Introduction to Statistics Using* R. Cambridge: Cambridge University Press.

Baayen, R.H. (2011). languageR: Data sets and functions with "Analyzing Linguistic Data: A practical introduction to statistics" R package version 1.1. http://CRAN.R-project.org/package=languageR

Bates, D., Maechler, M. & Bolker, B. (2011). lme4: Linear mixed-effects models using S4 classes. R package version 0.999375-42. http://CRAN.R-project.org/package=lme4

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

Best, C.T. & Tyler, M.D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In Bohn, O-S., & Munro, M. (Eds), In Bohn, O-S., & Munro, M. (Eds), *Language Experience in Second Language Speech Learning: In honor of James Emil Flege,* pp. 13-34, Amsterdam, John Benjamins.

Birdsong, D. (2003). Authenticité de prononciation en Français L2 chez des apprenants tardifs Anglophones: Analyses segmentales et Globales. *Acquisition et Interaction en Langue Etrangère, 18,* 17-35.

Birdsong, D. (2007). Nativelike pronunciation among late learners of French as a second language. In Bohn, O-S., & Munro, M. (Eds), *Language Experience in Second Language Speech Learning: In honor of James Emil Flege,* pp. 99-116, Amsterdam, John Benjamins.

Birdsong, D., & Molis, M. (2001). On the Evidence for Maturational Constraints in Second-Language Acquisition. *Journal of Memory and Language, 44,* 235-249.

Bognar, E. & Fujisaki, H. (1986). Analysis, synthesis and perception of the French nasal vowels. *Proceedings of ICASSP 86, Tokyo,* 1601-1604.

Bohn, O-S & Flege, J.E. (1997). Perception and production of a new vowel category by second-language learners. In A. James & J. Leather (Eds) *Second-language speech: Structure and process.* Berlin: Walter de Gruyter, pp. 53-74.

Bongaerts, T. (1999). Ultimate Attainment in L2 Pronunciation: The Case of Very Advanced Late L2 Learners. In: Birdsong, D. (Ed.), *Second Language Acquisition and the Critical Period Hypothesis,* Mahwah NJ, Lawrence Erlbaum, pp133-160.

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. In Singleton, D. & Lengyel, Z. (Eds.), *The age factor in second language acquisition,* pp. 30-50, Clevedon: UK, Multilingual Matters.

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,* 447-465.

Bongaerts, T., Mennen, S. & van der Slik, F. (2000). Authenticity of pronunciation in naturalistic second language acquisition: the case of very advanced late learners of Dutch as a second language. *Studia Linguistica, 54,* 298-308.

Bradlow, A. R. (2008). Training non-native language sound patterns: Lessons from training Japanese adults on the English /r/-/l/ contrast. In Hansen, J. and Zampini, M. (Eds.), *Phonology and Second Language Acquisition,* John Benjamins, 287-308.

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

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(5),* 977-985.

Calbris, G. (1978). Nasales et Orales Dites Correspondantes. *IRAL, 16(4),* 297-312.

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

Catford, J.C. & Pisoni, D.B. (1970). Auditory vs. Articulatory Training in Exotic Sounds. *The Modern Language Journal, 54(7),* 477-481.

Cole, R., Yan, Y., Mak, B., Fanty, M., and Bailey, T. (1996). The contribution of consonants versus vowels to word recognition in fluent speech. *Proceedings of the ICASSP'96,* pp. 853–856.

Csizér, K. & Dörnyei, Z. (2005). The internal structure of language learning, motivation and its relationship with language choice and learning effort. *Modern Language Journal, 89,* 19-36.

De Bot, K. (1983). Visual feedback of intonation: Part 1. Effectiveness and induced practice behaviour. *Language & Speech, 26,* 331-350.

Derwing, T.M., & Munro, M.J. (2005). Second Language Accent and Pronunciation Teaching: A Research-Based Approach. *TESOL Quarterly, 39(3),* 379-397.

Derwing, T.M., & Munro, M.J. (2009). Putting accent in its place: Rethinking obstacles to communication. *Language Teaching, 42(4),* 476-490.

Derwing, T.M., Munro, M.J. & Wiebe, G. (1998). Evidence in Favor of a Broad Framework for Pronunciation Instruction. *Language Learning, 48(3),* 393-411

Diehl, R.L. & Kluender, K.R. (1989). On the objects of speech perception. *Ecological Psychology, 1,* 121-144.

Diehl, R.L., Lotto, A.J. & Holt, L.L. (2004). Speech Perception. *Annual Review of Psychology, 55,* 149-179.

Dörnyei, Z. (1990). Conceptualising motivation in foreign language learning. *Language Learning, 40,* 47-78

Dörnyei, Z. (1994). Motivation and motivating in the foreign language classroom. *Modern Language Journal, 78,* 273-284.

Dörnyei, Z. (2001). *Teaching and Researching Motivation.* Harlow, Pearson Education Ltd.

Dowd, A., Smith, J. & Wolfe, J. (1998). Learning to Pronounce Vowel Sounds in a Foreign Language Using Acoustic Measurements of the Vocal Tract in Real Time. *Language and Speech, 41(1),* 1-20.

Drager (2011). Speaker Age and Vowel Perception. *Language and Speech*, 54, 99–121.

Elliott, A.R. (1995). Field Independence/Dependence, Hemispheric Specialisation and Attitude in Relation to Pronunciation Accuracy in Spanish as a Foreign Language. *Modern Language Journal, 79(3),* 356-371.

Felps, D., Bortfeld, H. & Gutierrez-Osuna, R. (2009). Foreign accent conversion in computer assisted pronunciation training. *Speech Communication, 51,* 920-932.

Flege, J.E. (1987). The production of "new" and "similar" phones in a foreign language: evidence for the effect of equivalence classification. *Journal of Phonetics, 15,* 47-65.

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

Flege, J.E. (1995b). Second language speech learning: Theory, findings and problems. . In: Strange, W. (Ed.), *Speech Perception and Linguistic Experience, Issues in Crosslinguistic research.* York Press, Timonium, MD, pp 233-277.

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

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

Flege, J.E., MacKay, I.R.A., & Meador, D. (1999). Native Italian speakers' perception and production of English vowels. *Journal of the Acoustical Society of America, 106(5),* 2973-2987.

Flege, J.E., Yeni-Komshian, G.H., & Liu, S. (1999). Age Constraints on Second-Language Acquisition. *Journal of Memory and Language, 41,* 78-194.

Fullana, N. (2006). The Development of English (FL) Perception and Production Skills: Starting Age and Exposure Effects. In Muñoz, C. (2008) (Ed.). *Age and the Rate of Foreign Language Learning,* Clevedon, UK, Multilingual Matters, pp 41-64.

Gardner, R.C. (1985). *Social Psychology and Second Language Learning: The role of attitudes and motivation.* Edward Arnold, London.

Gardner, R.C. & Lambert, W.E. (1959). Motivational variables in second language acquisition. *Canadian Journal of Psychology, 13,* 266-272

Gardner, R.C. & MacIntyre, P.D. (1991). An instrumental motivation in language study: Who says it isn't effective? *Studies in Second Language Acquisition, 13,* 57-72

Gardner, R.C., Tremblay, P.F. & Masgoret, A-M. (1997). Towards a full model of second language learning: An empirical investigation. *Modern Language Journal, 81,* 344-362.

Garrott, C. L. (2006). *L1 (English) to L2 (French) transfer: The question of nasalization.* Unpublished MSc Thesis, University of Virginia State. http://www.eric.ed.gov/PDFS/ED493645.pdf Accessed 30th March 2012.

Gómez Lacabex, E., García Lecumberri, M.L. (2010). Investigating training effects in the production of English weak forms by Spanish learners. In Dziubalska-Kołaczyk, k., Wrembel, M. & Kul, M. (Eds), *New Sounds 2010: Proceedings of the 6th International Symposium on the Acquisition of Second Language Speech,* pp 137-143, Poznań, Poland.

Gómez Lacabex, E., García Lecumberri, M.L., & Cooke, M. (2008). Perception of English Vowel Reduction by Trained Spanish Learners. In Rauber, A.S., Watkins, M.A. & Baptista, B.O. (Eds) *New Sounds 2007: Proceedings of the Fifth International Symposium on the Acquisition of Second Language Speech,* pp 294-299, Florianópolis, Brazil.

Goto, H. (1971). "Auditory perception by normal Japanese adults of the sounds "L" and "R"," *Neuropsychologia*, *9*, 317-323.

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

Gregg, R.J. (1960). *A Students' Manual of French Pronunciation.* Toronto, MacMillan.

Handley, Z., Sharples, M. & More.D. (2009). Training Novel Phonemic Contrasts: A Comparison of Identification and Oddity Discrimination Training. *Online Proceedings of the 2009 International Speech Commumication Association (ISCA) Workshop on Speech and Language Technology in Education (SLaTE 2009).* http://www.eee.bham.ac.uk/SLaTE2009/papers/SLaTE2009-02-v2.pdf Accessed 28th March 2012

Hattori, K. (2019). Perception and Production of English /r/-/l/ by Adult Japanese Speakers. Unpublished PhD Thesis, University College London (UCL). http://discovery.ucl.ac.uk/19204/1/19204.pdf Accessed 29th June 2012

Hardison, D. (2003). Acquisition of second-language speech: Effects of visual cues, context, and talker variability. *Applied Psycholinguistics, 24,* 495-522.

Hardison, D. (2005). Second-language spoken word identification: Effects of perceptual training, visual cues, and phonetic environment. *Applied Psycholinguistics,26,* 579-596.

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*(3), 360-378.

Hirata, Y. (2004). Computer assisted pronunciation training for native English speakers learning japanese pitch and durational contrasts. *Computer Assisted Language Learning, 17*(3-4), 357-376.

Hyltenstam, K. & Abrahamsson, N. (2003). Maturational Constraints in SLA. In Doughty, C. & Long, M. (Eds), *The Handbook of Second Language Acquisition.* Oxford, Blackwell.

Ingram, J. & Pittam, J. (1987) Auditory and acoustic correlates of perceived accent change: Vietnamese schoolchildren acquiring Australian English, *Journal of Phonetics, 15,* 127-145.

Ioup, G., Boustagui, E., El Tigi, M., & Moselle, M. (1994). Reexamining the Critical Period Hypothesis: A case study of successful adult SLA in a naturalistic environment. *Studies in Second Language Acquisition, 16,* 73-98.

Iverson, P., Hazan, V. & Bannister, K. (2005). Phonetic training with acoustic cue manipulations: A comparison of methods for teaching English /r/-/l/ to Japanese adults. *Journal of the Acoustical Society of America, 118(5),* 3627-3278.

Iverson, P., Pinet, M. & Evans, B.G. (2012). Auditory training for experienced and inexperienced second-language learners: Native French speakers learning English vowels. *Applied Psycholinguistics, 33,* 145-160.

Jacewicz, E. (2002). The Perception-Production Relationship in the Acquisition of Second Language Vowel Contrasts. *Journal of Language and Linguistics, 1(3),* 314-337

Jamieson, D.G. (1995). Techniques for training difficult non-native speech contrasts. In Elenius, K.- Branderud, P. (Eds), *Proceedings of the XIIIth International Congress of Phonetic Sciences. Stockholm, Sweden, 13-19 August, 1995*. Stockholm: KTH / Stockholm University. Vol. 4. pp. 100-107.

Jamieson, D.G. & Moore, A.E. (1991). Generalisation of new speech contrasts trained using the fading technique. *Proceedings of the XIIth International Congress of the Phonetic Sciences, 5,* 286-289.

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

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

Kewley-Port, D., Burkle, T. Z., and Lee, J. H. (2007) Contribution of consonant versus vowel information to sentence intelligibility for young normal-hearing and elderly hearing-impaired listeners. *Journal of the Acoustical Society of America, 122,* 2365-2375.

Klatt, D.H. (1980). Software for a cascade/parallel formant synthesiser. *Journal of the Acoustical Society of America, 67(3),* 971-995.

Lambacher, S.G., Martens, W.L., Kakehi, K., Marasinghe, C.A. & Moholt, G. (2005). The effects of identification training on the identification and production of American English vowels by native speakers of Japanese. *Applied Psycholinguistics, 26(2),* 227-247.

Leather, J. (1997). Interrelation of perceptual and productive learning in the initial acquisition of second-language tone. In James, A. & Leather, J. (Eds.), *Second Language Speech: Structure and Process,* 75-101, Berlin, Walter de Gruyter.

Lengeris, A. & Hazan, V. (2010). The effect of native vowel processing ability and frequency discrimination acuity on the phonetic training of English vowels for native speakers of Greek. *Journal of the Acoustical Society of America, 128(6),* 3757-3768.

Lenneberg, E. (1967). *Biological Foundations of Language.* New York, John Wiley.

Levy, E.S. (2009). On the assimilation-discrimination relationship in American English adults' French vowel learning. *Journal of the Acoustical Society of America, 126(5),* 2670-2682.

Levy, E.S. & Law, F.F. II (2010). Production of French vowels by American-English learners of French: Language experience, consonantal context, and the perception-production relationship. *Journal of the Acoustical Society of America, 128(3),* 1290-1305.

Levy, E.S. & Strange, W. (2008). Perception of French vowels by American English adults with and without French language experience. *Journal of Phonetics 36(1),* 141-157.

Li, P., Sepanski, S. & Zhao, X. (2006). Language history questionnaire: A Web-based interface for bilingual research. *Behavior Research Methods, 38(2),* 202-210.

Liberman, A.M. & Mattingly, I.G. (1985). The motor theory of speech perception revised. *Cognition, 21,* 1-36.
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(3),* 1242-1255.

Lively, S.E., Pisoni, D.B., Yamada, R.A., Tohkura, Y. & Yamada, T. (1994). Training Japanese listeners to identify English/r/and fl/. III. Long-term retention of new phonetic categories. *Journal of the Acoustical Society of America, 96(4),* 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(2),* 874-886.

Logan, J.S., & Pruitt, J.S. (1995). Methodological Issues in Training Listeners to Perceive Non-Native Phonemes. . In: Strange, W. (Ed.), *Speech Perception and Linguistic Experience, Issues in Crosslinguistic research.* York Press, Timonium, MD, pp 351-377.

Long, M. 1990. Maturational Constraints on Language Development. *Studies in Second Language Acquisition, 12,* 251–285.

McCandliss, B.D., Fiez, J.A., Protopapas, A., Conway, M. & McClelland, J.L. (2002). Success and failure in teaching the [r]–[l] contrast to Japanese adults: Tests of a Hebbian model of plasticity and stabilization in spoken language perception. *Cognitive, Affective & Behavioral Neuroscience, 2(2),* 89-108.

McClaskey, C. L., Pisoni, D. B., & Carrell, T. D. (1983). Transfer of training of a new linguistic contrast in voicing, *Perception & Psychophysics, 34,* 323–330. Cited in Nishi, K. & Kewley-Port, D. (2007). Training Japanese Listeners to Perceive American English Vowels: Influence of Training Sets. *Journal of Speech, Language, and Hearing Research, 50,* 1496-1509.

McClelland, J.L. (2001). Failures to Learn and Their Remediation: A Hebbian Account. In McClelland, J.L. & Siegler, R.S. (Eds.) (2001). *Mechanisms of Cognitive Development – Behavioural and Neural Perspectives,* Mahwah, Erlbaum.

Macdonald, R.M.M. (2006). *The Production of New Vowels by Scottish Learners of French.* Unpublished MSc dissertation, Department of Linguistics and English Language, University of Edinburgh.

MacKay, I.R.A. (1997). *Phonetics: The science of speech production.* Boston, MA, Allyn&Bacon.

Major, R.C. (2001). *Foreign Accent: The Ontogeny and Phylogeny of Second Language Phonology.* Mahwah, NJ, Lawrence Erlbaum.

Martin, C.C. (2004). Autre méthode pour apprendre la prononciation d'une seconde langue. *Proceedings from JEP-TALN-RECITAL, Fez, Morocco.* http://aune.lpl.univ-aix.fr/jep-taln04/proceed/actes/jep2004/Martin.pdf
Accessed 21st March 2012.

Morosan, D.E. & Jamieson, D.G. (1989). Evaluation of a technique for training new speech contrasts: Generalisation across voices but not word-position or task. *Journal of Speech and Hearing Research, 32(3),* 501-511.

Moyer, A. (1999). Ultimate Attainment in L2 Phonology: The Critical Factors of Age, Motivation and Instruction. *Studies in Second Language Acquisition, 21,* 81-108.

Moyer, A. (2007). Do Language Attitudes Determine Accent? A Study of Bilinguals in the USA. *Journal of Multilingual and Multicultural Development, 28(6),* 502-518.

Muñoz, C. (2006). The Effects of Age on Foreign Language Learning: The BAF Project. In Muñoz, C. (2008) (Ed.). *Age and the Rate of Foreign Language Learning,* Clevedon, UK, Multilingual Matters, pp 1-40.

Muñoz, C. (2008). Symmetries and Asymmetries of Age Effects in Naturalistic and Instructed L2 Learning. *Applied Linguistics, 29(4),* 578-596.

Munro, M.J. & Derwing, T.M. (2001). Modeling Perceptions of the Accentedness and Comprehensibility of L2 Speech: The Role of Speaking Rate. *Studies in Second Language Acquisition, 23,* 451-468.

Nearey (1989). Static, dynamic and relational properties in vowel perception. *Journal of the Acoustical Society of America, 85(5),* 2088-2113. Cited in Rochet, B.L. (1995). Perception and Production of Second-Language Speech Sounds by Adults. In: Strange, W. (Ed.), *Speech Perception and Linguistic Experience, Issues in Crosslinguistic research*. York Press, Timonium, MD, pp. 379–410.

Newport E.L., (2002). Critical Periods in Language Development. In L Nadel (Ed.), *Encylopedia of Cognitive Science,* London, Macmillan Publishers Ltd/Nature Publishing Group.

Nishi, K. & Kewley-Port, D. (2007). Training Japanese Listeners to Perceive American English Vowels: Influence of Training Sets. *Journal of Speech, Language, and Hearing Research, 50,* 1496-1509.

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.

Oxford, R.L. & Shearin, J. (1994). Language learning motivation: Expanding the theoretical framework. *Modern Language Journal, 78,* 12-28.

Palmen, M-J., Bongaerts, T. & Schils, E. (1997). L'authenticité de la prononciation dans l'acquisition d'une langue étrangère au delà de la période critique: des apprenants néederlandais parvenus à un niveau très avancé en français. *Acquisition et Interaction en Langue Etrangere, 9,* 173-191.

Pimsleur, P. (1963). Discrimination Training in the Teaching of French Pronunciation. *Modern Language Journal, 47(5),* 199-203.

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

Polat, N. (2011). Gender differences in motivation and L2 accent attainment: an investigation of young Kurdish learners of Turkish. The *Language Learning Journal, 39(1),* 19-41.

Polka, L. (1992). Characterising the influence of native language experience on adult speech perception. *Perception & Psychophysics, 52(1),* 37-52.

Polka, L. (1995). Linguistic influences in adult perception of non-native vowel contrasts. *Journal of the Acoustical Society of America, 97(2),* 1286-1296.

Purcel, E.T. & Suter, R.W. (1980). Predictors of pronunciation accuracy: A Reexamination. *Language Learning, 30(2),* 271-287.

R Development Core Team (2010). R: A language and environment for statistical computing. R Foundation for Statistical Computing,Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

Rochet, B.L. (1995). Perception and Production of Second-Language Speech Sounds by Adults. In: Strange, W. (Ed.), *Speech Perception and Linguistic Experience, Issues in Crosslinguistic research*. York Press, Timonium, MD, pp. 379–410.

Rochet, B.L. & Chen, F. (1992). Acquisition of the French VOT contrasts by adult speakers of Mandarin Chinese. In Ohala, J.J. et al (Eds) *ICSLP 92 (International Conference on Spoken Language Processing) Proceedings, 1,* Edmonton, University of Alberta, 273-276. Cited in Rochet, B.L. (1995). Perception and Production of Second-Language Speech Sounds by Adults. In: Strange, W. (Ed.), *Speech Perception and Linguistic Experience, Issues in Crosslinguistic research*. York Press, Timonium, MD, pp. 379–410.

Schairer, K.E. (1992). Native Speaker Reaction to Non-Native Speech, *The Modern Language Journal*, *76(3),* 309-319.

Schneiderman, E., Bourdages, J. & Champagne, C. (1988). Second-Language Accent: The Relationship Between Discrimination and Perception in Acquisition. *Language Learning, 38(1),* 1-19.

Scovel, T. (1969). Foreign accents, language acquisition, and cerebral dominance. *Language Learning, 19*, 245–253.

Scovel, T. (1988). *A time to speak: A psycholinguistic inquiry into the critical period for human speech.* Rowley, MA: Newbury House.

Scovel, T. (2000). A Critical Review of the Critical Period Research. *Annual Review of Applied Linguistics, 20*, 213–223.

Sheldon, A., & Strange, W. (1982). "The acquisition of /r/ and /l/ by Japanese listeners of English: Evidence that speech production can precede speech perception," *Applied Psycholinguistics*, *3*, 243-261.

Shinohara, Y. (2012). Identification and discrimination training for English /r/-/l/ perception by Japanese adults. ESRC Centre for Research on Bilingualism, University of Bangor, 2012 Conference Abstracts: Bilingual & Multilingual Interaction. http://bilingualism.bangor.ac.uk/conference/documents/ConferenceBooklet20120306SG-10.pdf
Accessed 28th March 2012

Simon, E., Chambless, D & Alves, U.K. (2010). Understanding the role of orthography in the acquisition of a non-native vowel contrast. *Language Sciences, 32,* 380-394.

Smit, U. (2002). The interaction of motivation and achievement in advanced EFL pronunciation learners. *IRAL, 40,* 89-116.

Smit, U. & Dalton, C. (2000), Motivation in advanced EFL pronunciation learners. *IRAL, 38,* 229-246.

Smith, L.C. & Baker, W. (2010). L2 Dialect on Learning German Vowels: The Case of Northern German and Austrain Dialects. IN Dziubalska-Kołaczyk, K., Wrembel, M & Kul, M. (Eds). *Proceedings of the 6th International Symposium on the Acquisition of Second Language Speech New Sounds 2010.* Poznań, Poland, pp 451-456.

Strange, W. (1995). Cross-language studies of speech perception: A historical review. In Strange, W. (Ed.) (1995). *Speech Perception and Linguistic Experience, Issues in Crosslinguistic research*, pp 3-45, York Press, Timonium, MD.

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

Suter, R.W. (1976). Predictors of pronunciation accuracy in second language learning. *Language Learning, 26,* 233-253.

Tees, R.C. & Werker, J.F. (1984). Perceptual flexibility: Maintenance or recovery of the ability to discriminate non-native speech sounds. *Canadian Journal of Psychology, 38,* 579-590.

Terrace, H.S. (1963). Discrimination learning with and without "errors." *Journal of Experimental Analysis of Behavior, 6,* 1-27. Quoted in Jamieson, D.G. & Morosan, D.E. (1986). Training non-native speech contrasts in adults: Acquisition of the English /ð/-/Θ/ contrast by francophones. *Perception & Psychophysics, 40,* 205-215.

Thompson, M., Hazan, V. (2010). The impact of visual cues and lexical knowledge on the perception of a non-native consonant contrast for Colombian adults. In Dziubalska-Kołaczyk, K., Wrembel, M. & Kul, M. (Eds), *New Sounds 2010: Proceedings of the 6th International Symposium on the Acquisition of Second Language Speech,* pp 493-498, Poznań, Poland.

Tranel, B. (1987). *The Sounds of French: An Introduction.* Cambridge University Press, Cambridge.

Tremblay, P.F. & Gardner, R. C. (1995). Expanding the motivation construct in language learning. *Modern Language Journal, 79,* 505-520.

Vieru-Dimulescu, B. & Boula de Mareüil, P. (2006). Perceptual identification and phonetic analysis of 6 foreign accents in French. *Proceedings of INTERSPEECH 2006 – ICSLP, Sep 17-21, Pittsburgh, Pennsylvania,* 441-444.

Wang, X. & Munro, M.J.(2004). Computer-based training for learning English vowel contrasts. *System, 32,* 539-552.

Wayland, R.P. (2007). The relationship between identification and discrimination in cross-language perception. In Bohn, O-S., & Munro, M. (Eds), *Language Experience in Second Language Speech Learning: In honor of James Emil Flege,* pp. 201-218, Amsterdam, John Benjamins.

Wayland, R.P. & Li, B. (2008). Effects of two training procedures in cross-language perception of tones. *Journal of Phonetics, 36(2),* 250-267.

Zhang, Y., Kuhl, P.K., Imada, T., Iverson, P., Pruitt, J., Stevens, E.B., Kawakatsu, M., Tohkura, Y. & Nemoto, I. (2009). Neural signatures of phonetic learning in adulthood: A magnetoencephalography study. *Neuroimage, 46,* 226-240.

10 Appendices

Appendix A: List of minimal pairs tested to ascertain difficult L2 French contrasts for L1 speakers of English

/u/-/y/	/ɑ̃/-/ɔ̃/	/ɑ̃/-/ɛ̃/
bouche/bûche	ambre/ombre	attente/atteinte
boue/bu	angle/ongle	cendre/ceindre
boule/bulle	arrivant/arrivons*	cran/crains
bourreau/bureau	bande/bonde	dans/dain
dessous/dessus	blanc/blond	dépendre/dépeindre
doux/du	chantant/chantons*	dépens/dépeins*
joue/jus	croyant/croyons	détendre/déteindre
jour/jure	dent/dont	éprendre/épreindre
loup/lu	devant/devons	étant/étain
mou/mue*	disant/disons	étendre/éteindre*
moule/mule	donnant/donnons	gendre/geindre
nous/nu	grande/gronde	grand/grain
pour/ pur	langue/longue	menthe/mainte
pousse/puce	massant/maçon	parent/parrain
roue/ rue*	passant/passons	plan/plein
rougi/rugi	prenant/prenons	plante/plainte
rousse/russe	savant/savons	roman/romain
sourd/sur	vivant/vivons	Soudan/soudain
tous/tue	voulant/voulons	temps/tin
vous/vu	voyant/voyons	venant/venin

/ɔ̃/-/ɛ̃/	/ɛ/-/ɛ̃/
bon/bain	aide/Inde
fondre/feindre	bouquet/bouquin
font/fin	bourrait/bourrin
front/frein	brai/brin
long/lin*	claie/clin
longe/linge	craie/crin
marron/marin	entrait/entrain
mon/main	fait/faim
non/nain	malais/malin
plomb/plein	messe/mince
pondre/peindre	moulait/moulin
pont/pain	paie/pin
raison/raisin	sais/sain
rond/rein	salait/salin
son/saint	sec/cinq
songe/singe*	tais/tain
songer/singer	tannais/tanin*
ton/taint	trait/train
tondre/teindre	vais/vin
vont/vain	

/ɑ/-/ɑ̃/	/ɔ/-/ɔ̃/
as/anse	coq/conque
âtre/antre	cotte/compte
bas/ban	dock/donc
cas/camp	loge/longe
gars/gant	motte/monte
gâter/ganter	noce/nonce
las/lent	pop/pompe
lasse/lance	pote /ponte
lasser/lancer	robe/rhombe
mas/mens	rode/ronde*
mâte/mante*	sobre/sombre
pas/pan	sodée/sonder*
passe/pense	troque/tronque
passer/panser	
pâte/pente	
ras/rang	
tas/tant	
tâter/tenter	
tâtons/tentons	

* Minimal pairs used in practice trials

Appendix B: Instructions provided for AXB tasks

Initial screen:

Thank you very much for agreeing to take part in this study.

This study consists of a number of trials. In each trial you will be played three French words. Your task is to decide whether the second word you hear is the same as the FIRST word or the same as the THIRD word. The words in each trial will only be played once, so listen carefully!

If you think the second word is the same as the FIRST word, press 1. If you think the second word is the same as the THIRD word, press 3. The next trial will not begin until you have responded.

Press the SPACEBAR to begin.

Discrimination Screen:

Press '1' if the second word is the same as the first word.

Press '3' if the second word is the same as the third word.

<u>Appendix C:</u> Minimal Pairs used in comparing identification and discrimination testing

/u/-/y- Contrast			/ɑ̃/-/ɔ̃/ Contrast		
Word	**Word Freq. Per Million**	**Log Transformed Word Freq. Per Million**	**Word**	**Word Freq. Per Million**	**Log Transformed Word Freq. Per Million**
bourreau	6.71	1.46	allant	29.00	.83
bureau	97.77	1.95	allons	90.06	1.99
cou	64.39	1.51	branche	32.06	1.81
cul	36.87	.09	bronche	1.23	1.57
écrou	1.90	.72	dansant	5.19	.28
écru	.16	-.28	dansons	.52	-.80
four	18.52	-.46	massant	.35	1.27
fur	19.65	.47	maçon	2.94	1.29
joule	1.84	1.46	parlant	29.06	.26
Jules	6.23	1.08	parlons	12.00	.79
route	168.42	.84	repandre	6.87	2.23
rut	1.58	1.86	repondre	72.06	.20
toupie	1.87	1.48	voyant	30.10	.27
Tupi	.42	1.56	voyons	36.19	-.38

Appendix D: Onscreen instructions for 2AFC perceptual training and testing

Initial Screen

Thank you for participating in this experiment.

Here you will be played a number of words in French. You are given two options as to what the word you have been played could be. One option is on the left of the screen, and the other option is on the right. The words will only be played once for each trial, so listen carefully. There is a sheet in front of you with a spelling guide in case some of the words are unfamiliar.

Press '1' if you think the word you have been played is the word on the left. Press '2' if you think the word you have been played is the word on the right.

Press the SPACEBAR to continue.

Identification Screen

Press '1' if you think the word on the left has been played.

Press '2' if you think the word on the right has been played.

<u>Appendix E:</u> Spelling Guide: Sound to spelling mappings for stimuli provided to participants

SPELLING GUIDE

Despite other differences you may see on the screen, the words on the left and right of the screen will only differ by one vowel sound.

EITHER

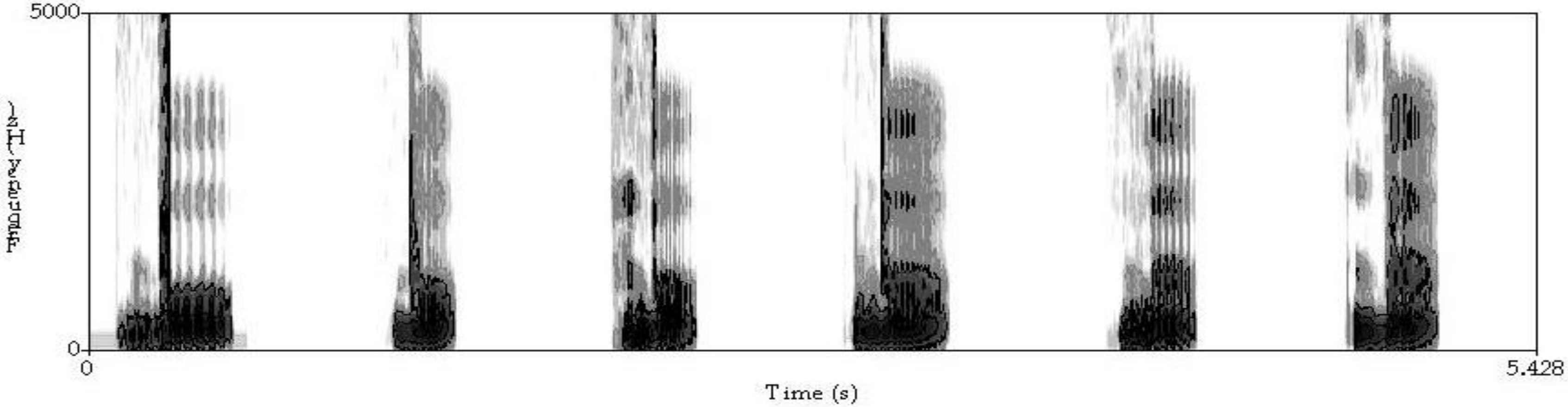
The difference will be between the vowel sound in the French word *on* ('we') and the vowel sound in the French word *an* ('year'). The vowel sound in the word *on* is spelled with the letters 'on' and the vowel sound in the word *an* is spelled with the letters 'an' or 'en'. If you are not familiar with any of the words you hear or see, try to decide whether or not you heard the 'on' sound or the 'an/en' sound where these letters are in the words onscreen.

OR

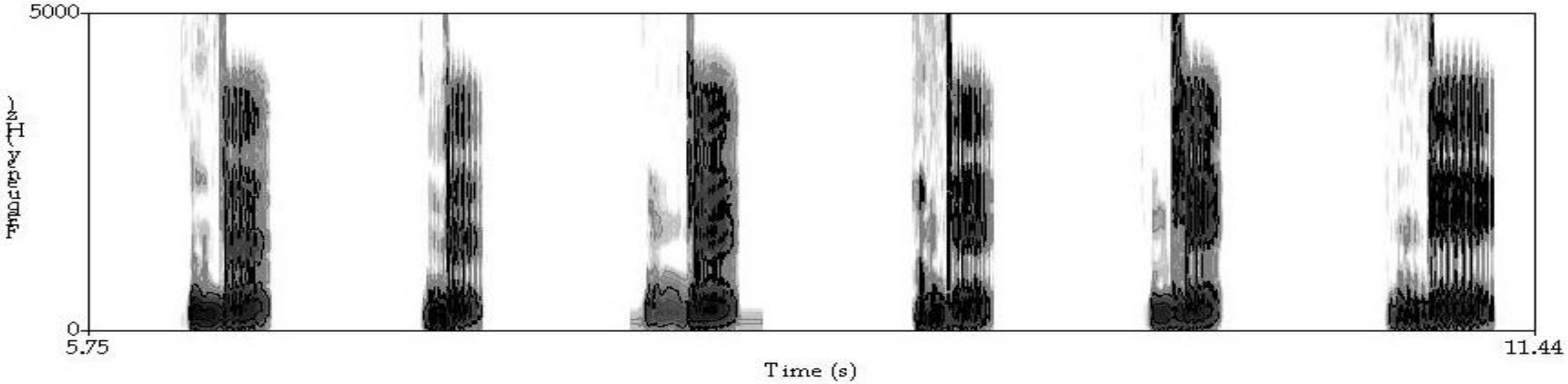
The difference will be between the vowel sound in the French word *vous* ('you' polite/plural) and the vowel sound in the French word *vu* (past participle of *voir* 'to see'.). The vowel sound in the word *vous* is spelled with the letters 'ou' and the vowel sound in the word *vu* is spelled with the letter 'u' only. If you are not familiar with any of the words you hear or see, try to decide whether or not you heard the 'ou' sound or the 'u' sound where these letters are in the words onscreen.

<u>Appendix F:</u> Spectrogram examples of words with synthesised vowels and naturally produced words used in training

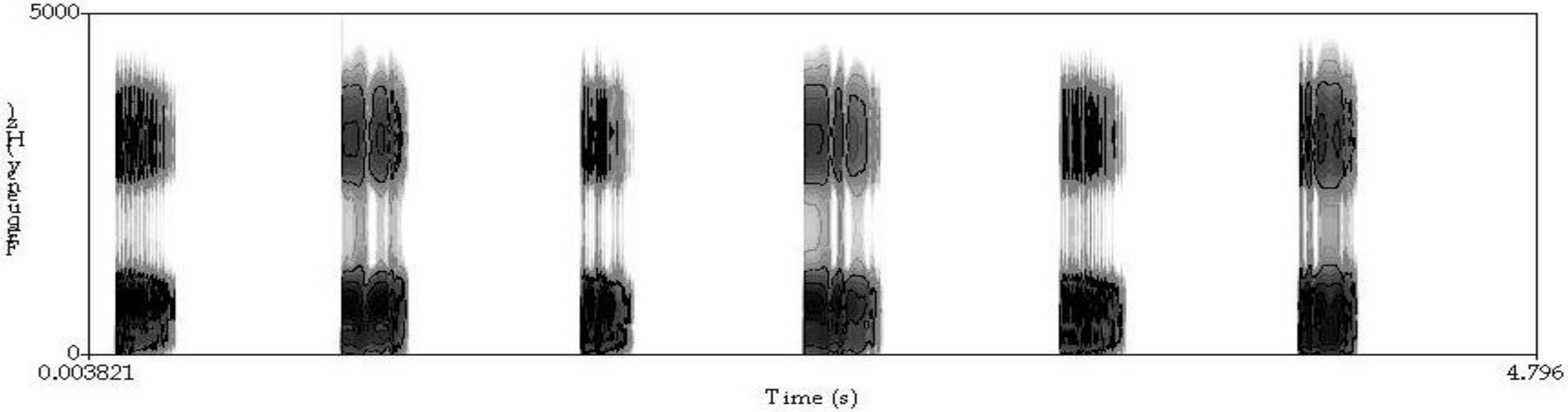
1. **PF Synthesised Vowel *bout* with Six Voices: Continuum points 1 (Male) – 6 (Female)**



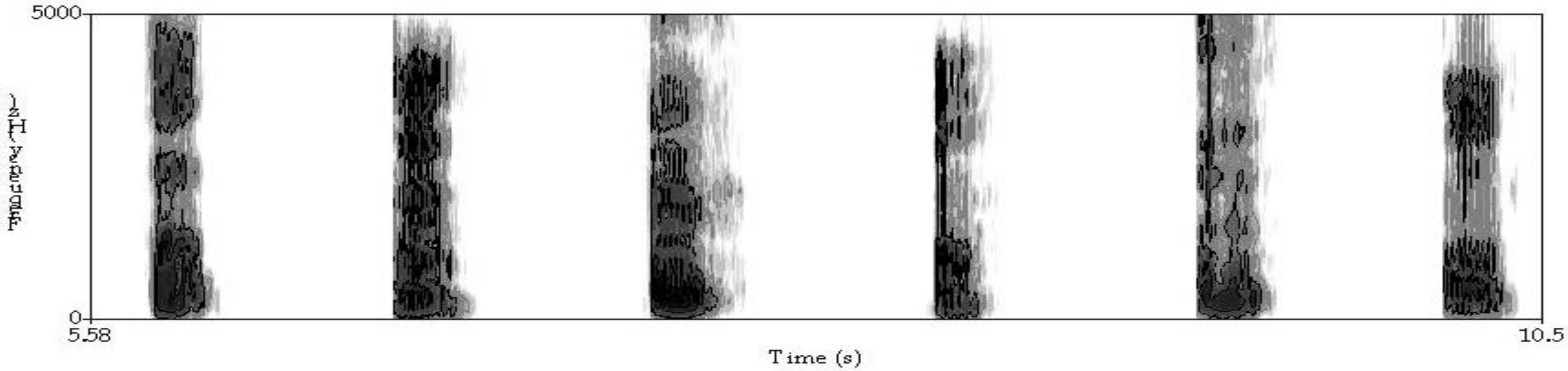
2. **PF Synthesised Vowel *bu* with Six Voices: Continuum points 7 (Female) – 12 (Male)**



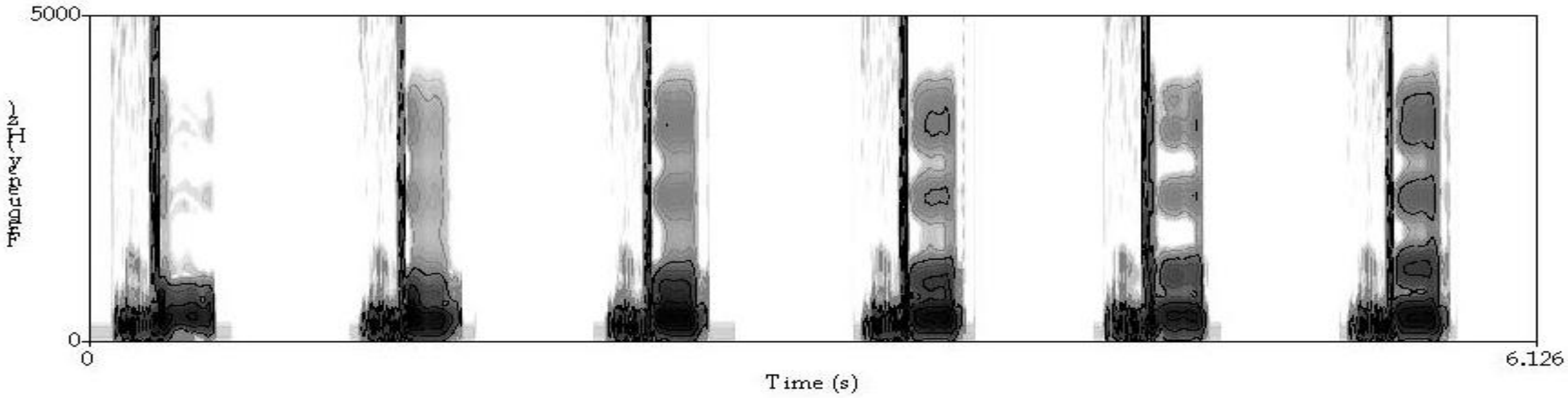
3. PF Synthesised Vowel *en* (thus fully synthetic) with Six 'Voices': Continuum points 1 ('Male') – 6 ('Female')



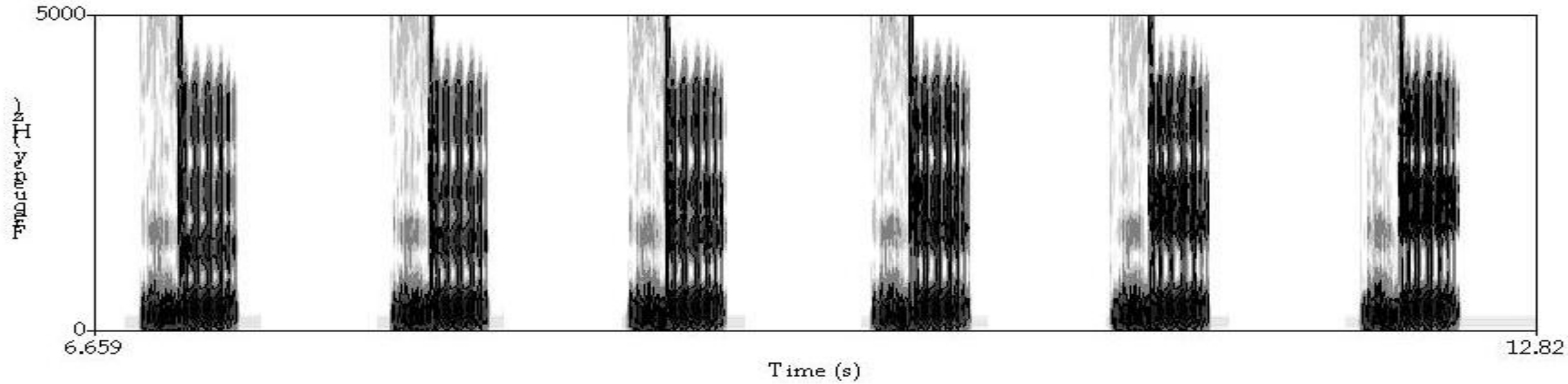
4. PF Synthesised Vowel *on* (thus fully synthetic) with Six 'Voices': Continuum points 7 ('Female' – 12 'Male')



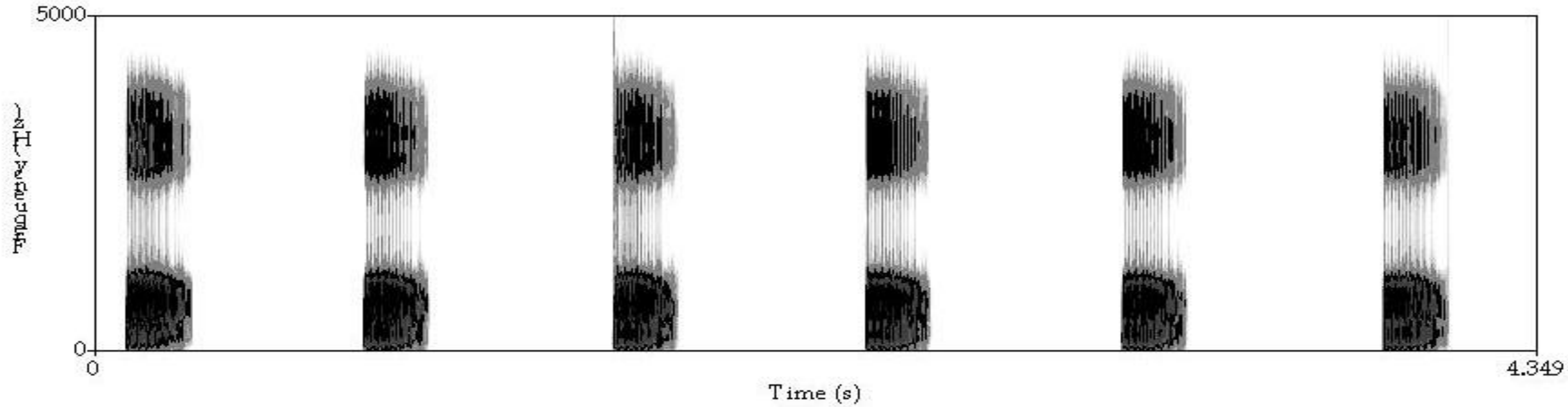
5. **PF Synthesised Vowel *bout* with One Voice (Male): Continuum points 1-6**



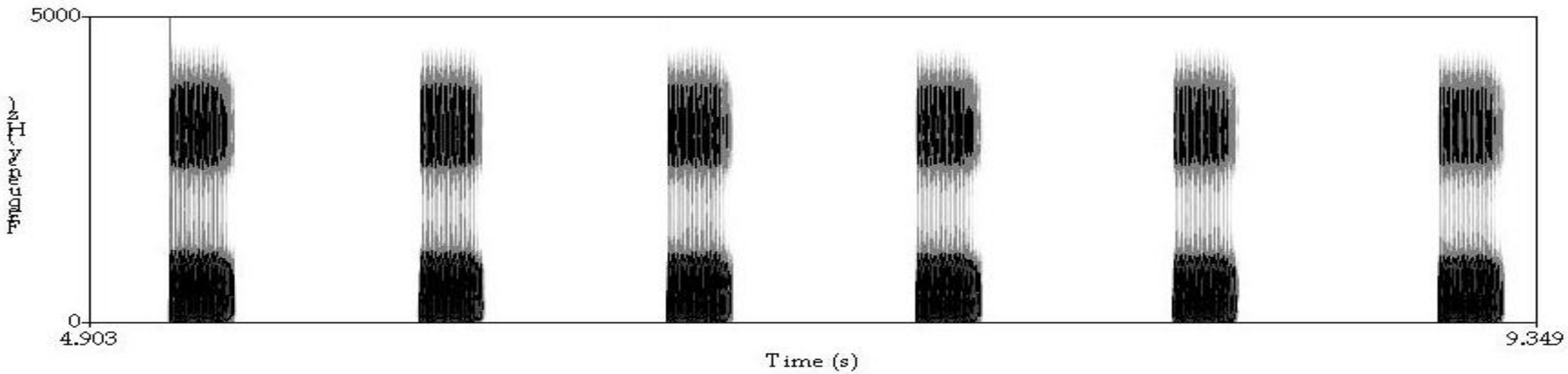
6. **PF Synthesised Vowel *bu* with One Voice (Male): Continuum points 7-12**



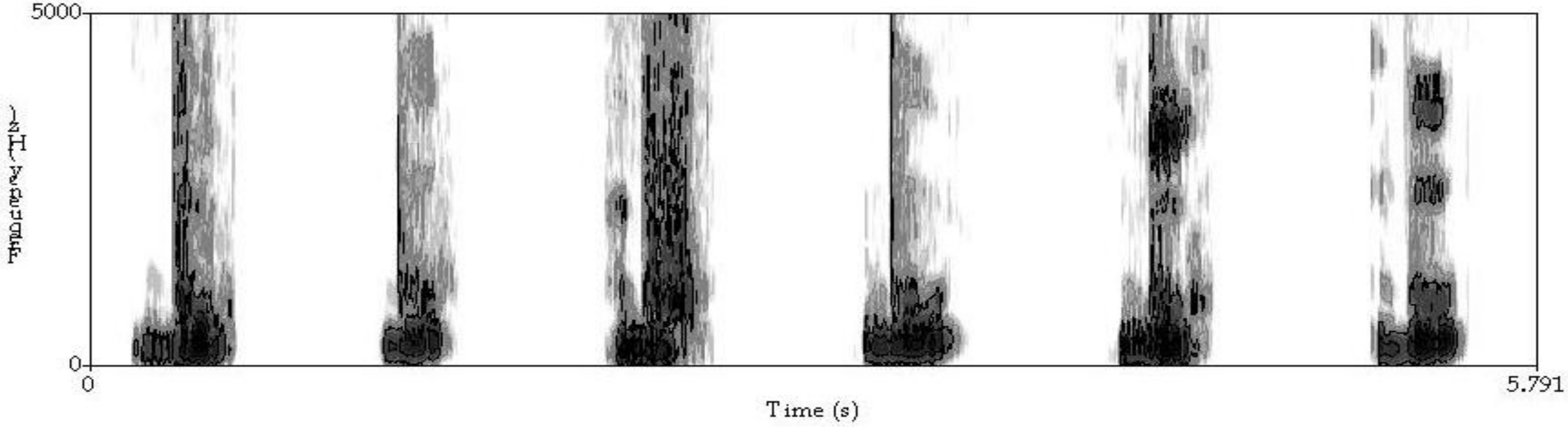
7. PF Synthesised Vowel *en* (thus fully synthetic) with One 'Voice' ('Male'): Continuum points 1-6



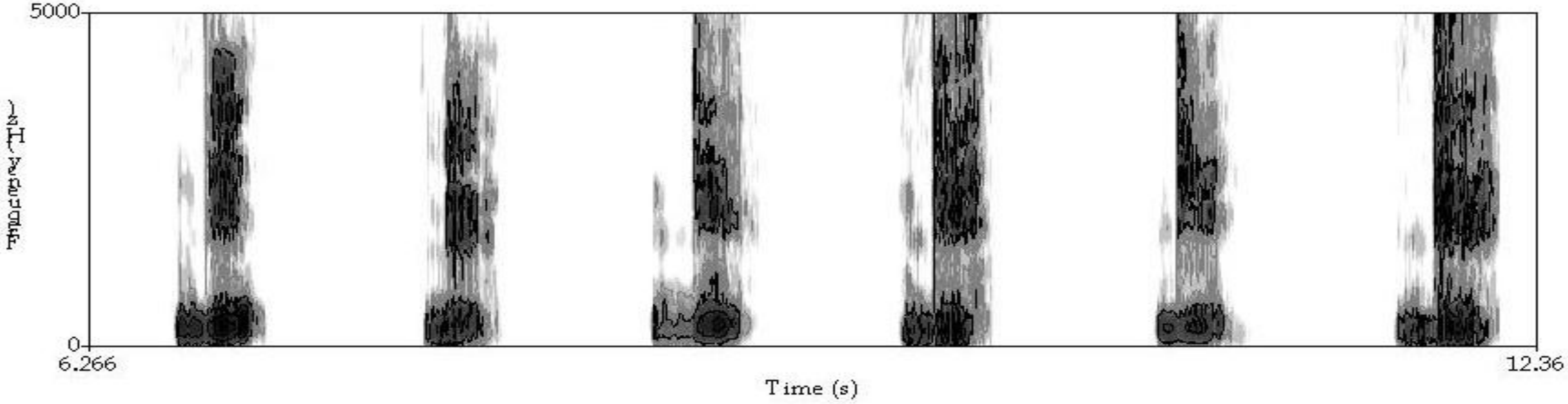
8. PF Synthesised Vowel *on* (thus fully synthetic) with One 'Voice' ('Male'): Continuum points 7-12



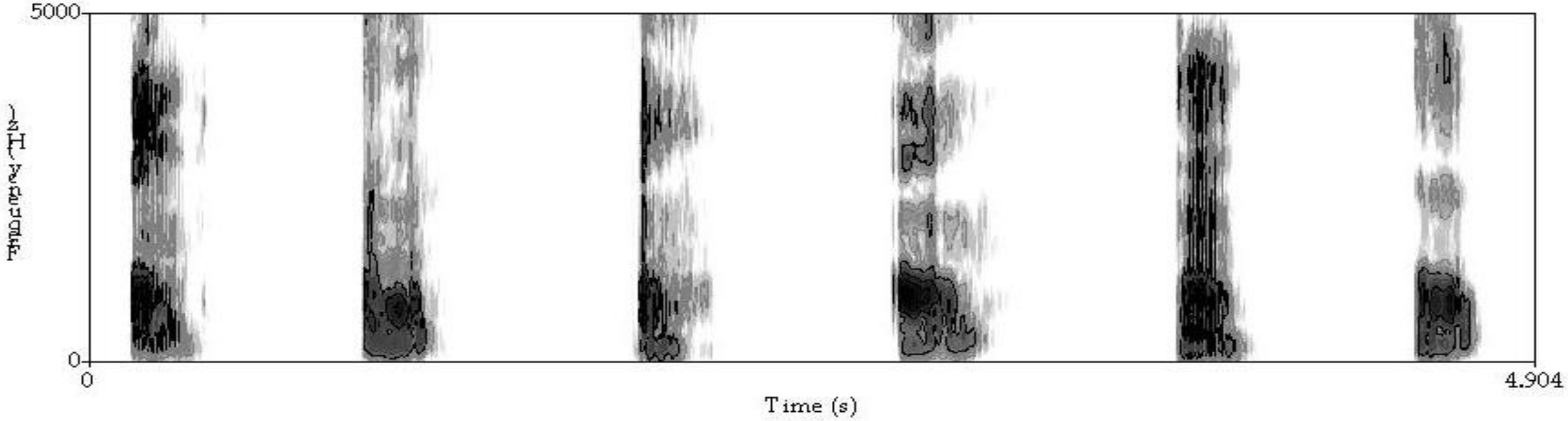
9. HVPT Natural *bout* with Voices 1 (Male) - 6 (Female)



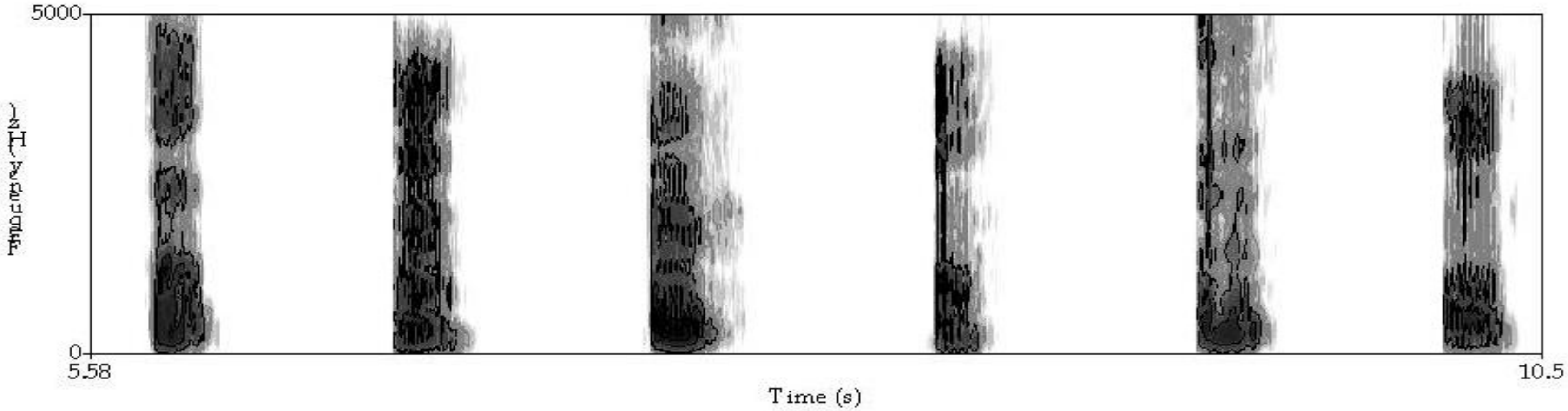
10. HVPT Natural *bu* with Voices 6 (Female) - 1 (Male)



11. HVPT Natural *en* with Voices 1 (Male) - 6 (Female)



12. HVPT Natural *on* with Voices 6 (Female) -1 (Male)



Appendix G: Language history questionnaire

L2 Language History Questionnaire (Version 2.0 - Short)[3]

Contact Information:
Name: ____________________ Email: ____________________
Telephone:____________________ Today's Date: ____________________

Please answer the following questions to the best of your knowledge.

PART A
1 Age (in years):

2. Sex (circle one): Male / Female

3. Education (degree obtained or school level attended):

4(a). Country of origin:

4(b). Country of Residence:

5. If 4(a) and 4(b) are the same, how long have you lived in a foreign country where your second language is spoken? If 4(a) and 4(b) are different, how long have you been in the country of your current residence? (in years)

6. What is your native language? (If you grew up with more than one language, please specify) If English, please state the variety of English.

7. Do you speak a second language (including those being learned at university)?

__YES my second language is/are ____________________.
__NO (If you answered NO, you need not to continue this form)

[3] [a] Li, P, Sepanski, S & Zhao, X (2006). Language history questionnaire: A Web-based interface for bilingual research. *Behavior Research Methods, 38(2),* 202-210.

8. If you answered YES to question 7, please specify the age at which you started to learn your second language(s) in the following situations (write age next to any situation that applies).

At home: __________

In school: __________

After arriving in the second language speaking country __________

9. How did you learn your second language(s) up to this point? (check all that apply)

(Mainly Mostly Occasionally) through formal classroom instruction.

(Mainly Mostly Occasionally) through interacting with people.

A mixture of both, but (More classroom More interaction Equally both).

Other (specify: ____).

10. List all foreign languages you know in order of most proficient to least proficient. Rate your ability on the following aspects in each language. Please rate according to the following scale (write down the number in the table):

Very poor Poor Fair Functional Good Very good Native-like

1 __________

2__________3__________4__________5__________6__________7__________

Language	Reading proficiency	Writing proficiency	Speaking fluency	Listening ability

11. Provide the age at which you were first exposed to each foreign language in terms of speaking, reading, and writing, and the number of years you have spent on learning each language.

Language	Age first exposed to the language			Number of years learning
	Speaking	Reading	Writing	

12. Do you have a foreign accent in the languages you speak (i.e. if you are a native speaker of English, how strong is your native English accent in your other languages?)? Please rate the strength of your accent according to the following scale (write down the number in the table – e.g. 1 = no trace of native English accent when speaking French):

No Accent	Very Weak	Weak	Intermediate	Strong	Very Strong
1 ________	2____________	3________	4__________	5________	6________

Language	Accent (circle one)	Strength
	Y N	
	Y N	
	Y N	
	Y N	
	Y N	

13. If there is anything else that you feel is interesting or important about your language background or language use, please comment below.

<u>Appendix H:</u> PAI questionnaire

*Experimental Questions
**Filler questions used for analysis of motivation in general
*** Not used for any analysis

<u>Language Learning Questionnaire</u>

Please answer all items using the following response categories (write your answer after the question):

5 = Always or almost always true of me

4 = Usually true of me

3 = Somewhat true of me

2 = Usually not true of me

1 = Never or almost never true of me

----

1. I try to learn some new French words every day.**
2. I'd like to sound as native as possible when speaking French.*
3. Studying French literature is not important to me.**
4. Acquiring proper pronunciation in French is important to me.*
5. I will never be able to speak French with a good accent.*
6. I prefer to work on my French assignments during the evening.***
7. I believe I can improve my pronunciation skills in French.*
8. I find French classes easier when I like the lecturer more.**
9. I believe more emphasis should be given to proper pronunciation in class.*
10. One of my personal goals is to acquire proper pronunciation skills and preferably be able to pass as a near-native speaker of the French language.*
11. I have a specific technique for learning French vocabulary.**
12. I try to imitate French speakers as much as possible.*
13. I wish that I had not chosen to study French at university.**
14. I am learning French because I'm interested in French culture.**
15. Communicating is much more important than sounding like a native speaker of French.*
16. There are some classes in my French course that I like more than others.***

17. Good pronunciation skills in French are not as important as learning vocabulary and grammar.*
18. I would not like to live in France for an extended period, other than when I have to for university.**
19. I like to listen to French music and watch French films.**
20. I want to improve my accent when speaking French.*
21. I try to read a French newspaper or watch or listen to the French news every day.**
22. French is the easiest of the subjects I am studying at university.***
23. I'm concerned with my progress in my pronunciation of French.*
24. Studying French will not be advantageous when finding a job after university.**
25. Sounding like a native French speaker is very important to me.*

Appendix I: Full list of training and testing stimuli and their frequency

/u/-/y/ Contrast			/ɑ̃/-/ɔ̃/ Contrast		
Test/Word			**Test/Word**		
Pre-Post Perception	**Word Freq. Per Million**	**Log Transformed Word Freq. Per Million**	**Pre-Post Perception**	**Word Freq. Per Million**	**Log Transformed Word Freq. Per Million**
bourreau	6.71	1.46	allant	29.00	.83
bureau	97.77	1.95	allons	90.06	1.99
cou	64.39	1.51	branche	32.06	1.81
cul	36.87	.09	bronche	1.23	1.57
écrou	1.90	.72	dansant	5.19	.28
écru	.16	-.28	dansons	.52	-.80
four	18.52	-.46	massant	.35	1.27
fur	19.65	.47	macon	2.94	1.29
joule	1.84	1.46	parlant	29.06	.26
Jules	6.23	1.08	parlons	12.00	.79
route	168.42	.84	repandre	6.87	2.23
rut	1.58	1.86	repondre	72.06	.20
toupie	1.87	1.48	voyant	30.10	.27
Tupi	.42	1.56	voyons	36.19	-.38
Pre-Post Pronunciation			**Pre-Post Pronunciation**		
boule	23.58	1.37	amusant	9.94	1.00
bulle	4.42	.65	amusons	.23	-.64
bouter*	.16	-.80	angle*	45.16	1.65
buter*	3.81	.58	ongle*	6.06	.78
doux*	42.55	1.63	bande*	43.52	1.64
du*	7141.45	3.85	bonde*	.39	-.41
jour	568.13	2.75	chantant	8.48	.93
jure	18.03	1.26	chantons	.58	-.24
pour	5332.48	3.73	devenant	6.35	.80
pur	48.48	1.69	devenons	.61	-.21
roue	22.58	1.35	passant*	54.87	1.74
rue	260.97	2.42	passons*	7.52	.88
soude*	4.35	.64	semblant	27.45	1.44
sud*	79.16	1.90	semblons	.10	-1.00
Training			**Training**		
bouche	150.68	2.18	aimant	8.97	.95
bûche	2.84	.45	aimons	5.19	.72
bourre	3.61	.56	ambre	3.84	.58
bure	2.32	.37	ombre	121.87	2.09
bout	232.48	2.37	blanc	143.71	2.16
bu	21.77	1.34	blond	15.77	1.20
broute	.48	-.32	camp	43.48	1.64
brute	10.10	1.00	con	37.16	1.57
dessous	67.06	1.83	coupant	4.84	.68
dessus	258.42	2.41	coupons	.90	-.05
fou	66.42	1.82	devant	520.81	2.72
fût	440.46	2.64	devons	21.77	1.34
jouter	N/A	N/A	donnant	39.45	1.60
juter	.29	.0	donnons	5.06	.70

loup	17.26	1.24	en	10644.13	4.03
lu	38.29	1.58	on	4364.48	3.64
nous	3077.39	3.49	finissant	4.13	.62
nu	42.26	1.63	finissons	1.48	.17
poule	10.86	1.04	hante	2.65	.42
pull	6.68	.82	honte	47.97	1.68
rougi	2.35	.37	mettant	20.52	1.31
rugi	.32	-.49	mettons	7.39	.87
sou	9.19	.96	pouvant	32.45	1.51
su	55.65	1.75	pouvons	40.48	1.61
sourd	15.42	1.19	restant	17.03	1.23
sur	4209.61	3.62	restons	4.71	.67
vous	2476.78	3.39	vivant	60.84	1.78
vu	282.71	2.45	vivons	7.94	.90
Generalisation Test 1 (Perception)			**Generalisation Test 1 (Perception)**		
about	.48	-.32	adhérent	.94	-.03
abus	10.65	1.03	adhérons	.03	-1.52
broum	N/A	N/A	connaissant	7.77	.89
brume	21.45	1.33	connaissons	11.13	1.05
courée	N/A	N/A	dément	3.06	.49
curé	29.68	1.47	démon	9.26	.97
écoulé	2.81	.45	fumant	7.06	.85
éculé	.10	-1.00	fumons	.42	-.38
moule	7.26	.86	marchant	16.00	1.20
mule	2.71	.43	marchons	3.77	.58
échoue	1.94	.29	rassurant	5.71	.76
échu	.48	-.32	rassurons	.16	-.80
souk	.97	-.01	trouvant	13.10	1.12
suc	4.10	.61	trouvons	9.29	.97
Generalisation Test 2 (Perception)			**Generalisation Test 2 (Perception)**		
cour	99.52	2.00	achetant	1.52	.18
cure	10.23	1.01	achetons	.29	-.54
doucher	.26	-.59	croyant	13.87	1.14
duché	.68	-.17	croyons	7.97	.90
joue	79.16	1.90	disant	61.48	1.79
jus	14.65	1.17	disons	22.65	1.36
joute	.65	-.19	jetant	12.35	1.09
jute	2.03	.31	jetons	4.03	.61
moue	8.87	.95	pensant	21.97	1.34
mue	3.87	.59	pensons	6.55	.82
pousse	37.13	1.57	prenant	37.52	1.57
puce	2.55	.41	prenons	11.55	1.06
touffe	4.23	.63	savant	19.06	1.28
tuf	.71	-.15	savons	26.13	1.42
Pronunciation Generalisation Test			**Pronunciation Generalisation Test**		
boulot*	17.23	1.24	arrivant*	14.65	1.17
bulot*	N/A	N/A	arrivons*	4.45	.65
boute	1.16	.06	demandant	14.26	1.15
butte	4.39	.64	demandons	3.03	.48

couver	1.13	.05	entendant	8.77	.94
cuver	.84	-.08	entendons	6.71	.83
moufle	.48	-.32	grande*	458.48	2.66
mufle	3.90	.59	gronde*	2.52	.40
pou*	1.32	.12	langue*	105.42	2.02
pu*	267.74	2.43	longue*	124.45	2.09
rousse*	10.48	1.02	regardant	46.13	1.66
russe*	43.55	1.64	regardons	4.42	.65
tout	2718.48	3.43	voulant	12.39	1.09
tue	20.19	1.31	voulons	12.77	1.11

*Words chosen for pronunciation analysis.

Appendix J: Instructions for pronunciation testing

Initial Screen

Thank you for agreeing to take part in this experiment.

In this part of the study I need you to produce some French words.

The words I want you to produce will be displayed onscreen one at a time. Read them in a natural tone and at a steady pace into the microphone.

If you make a mistake in any of the readings, just read the word again.

Once you have read the word, press the SPACEBAR to continue to the next word.

If you need a break, you can delay pressing the SPACEBAR (the word list will not continue until you press the SPACEBAR).

Appendix K: Instructions for pronunciation training

Initial Screen

Thank you for participating in this experiment.

Here you are going to learn how to pronounce a number of French sounds.

You have some notes on the sheets in front of you which will accompany this session. Please read the front page now.

Then press the SPACEBAR to begin.

Front page of notes

These notes are for guidance only. Pay close attention to the native speaker.

The terms 'hard palate' and 'soft palate' are used in these notes. The hard palate is the hard area of the roof of your mouth, and the soft palate the soft area of the roof of your mouth towards the back of your mouth.

Pronunciation Instructions for /u/

/u/ – (written 'ou' or 'oû')

- This vowel is formed by raising the back of the tongue towards the soft palate as high as possible without producing audible friction, and by protruding the lips so as to leave only a small round opening.
- Ensure that the pronunciation is coming from well back in the mouth, and ensure that the lip rounding is in place even before any preceding consonants.
- Summary: tongue drawn back, lips pushed forward and rounded.
- Comparison with English: It is pronounced a bit like 'oo' in English words but the lips protrude more and have a much smaller opening (almost as if trying to whistle)

[Onscreen only] Keep your sheet in front of you as reference. Press the spacebar to hear the native speaker produce the sound three times again. Try mouthing it while the speaker is saying it.

Pronunciation Instructions for /y/

/y /– (written ‘u’ or ‘û’)

- This vowel is formed by raising the front of the tongue towards the hard palate as high as possible without producing audible friction, and by protruding the lips so as to leave only a small round opening.
- Summary: tongue pushed forward, lips pushed forward and rounded
- Comparison with English: There is no such vowel in English however, think of producing ‘ee’ in English but with your lips in the position for saying English ‘oo’, but make the lips protrude more and leave a much smaller opening (almost as if trying to whistle).
- Even try by saying 'ee' and then moving your lips (but not your tongue) to the 'oo'/whistling position.

[Onscreen only] Keep your sheet in front of you as reference. Press the spacebar to hear the native speaker produce the sound three times again. Try mouthing it while the speaker is saying it.

Pronuncation Instructions for /ɑ̃/

/ɑ̃/ – (written ‘an’, ‘en’, ‘am’, or ‘em’)

- This vowel is formed by opening the mouth fairly wide, keeping the lips in a neutral position (i.e. do not round them at all) and keeping the tongue as low as possible in the mouth. In addition, this vowel is a nasal vowel which means that the soft palate is also lowered which allows some breath/air to escape through the nose. This must all happen at the same time.
- Summary: mouth open, tongue in low position, lips neutral (i.e. do not round them at all), soft palate lowered so half of the air you produce speaking escapes through nose.
- Comparison with English: A non-nasalised ‘ɑ’ sounds a bit like the sound in English ‘father’. To nasalise this, try thinking of saying it through your nose.

[Onscreen only] Keep your sheet in front of you as reference. Press the spacebar to hear the native speaker produce the sound three times again. Try mouthing it while the speaker is saying it.

Pronunciation Instructions for /ɔ̃/

/ɔ̃/ – (written 'on' or 'om')

- This vowel is formed by opening the mouth no more than halfway, rounding the lips and keeping the tongue as low as possible in the mouth. In addition, this vowel is a nasal vowel which means that the soft palate is also lowered which allows some breath/air to escape through the nose. This must all happen at the same time.
- Summary: Mouth slightly open, tongue drawn back, lips pushed forward and rounded, soft palate lowered so half air you produce speaking escapes through nose.
- Comparison with English: A non-nasalised 'ɔ' sounds a bit like the sound in English 'or'. The nasalised 'ɔ' is should be pronounced with the tongue, jaw and lips in a position intermediate between this and English 'o'/'oh'(with lips rounded more closely and tongue higher than for 'ɔ') . To nasalise, try thinking of saying the non-nasalised vowel through your nose.

[Onscreen only] Keep your sheet in front of you as reference. Press the spacebar to hear the native speaker produce the sound three times again. Try mouthing it while the speaker is saying it.

Appendix L: Instructions for native speakers to identify and rate participant productions

ENGLISH:

Initial Screen

Thank you for participating in this experiment. You will be played a number of words in French. Some have been produced by people learning French, and some have been produced by native speakers of French. You have TWO TASKS.

FIRSTLY you will be given two options as to what the word you have been played could be. One option is on the left of the screen, and the other option is on the right. The words will only be played once for each trial, so listen carefully.

Press '1' if you think the word you have been played sounds most like the word on the left. Press '2' if you think the word you have been played sounds most like the word on the right.

SECONDLY you will then be told what word the speaker was trying to produce. Please rate the accuracy of the pronunciation on a scale of 1 (very accurate/nativelike) to 7 (very inaccurate/clearly not native) by pressing '1', '2', '3', '4', '5', '6', or '7'.

IMPORTANT: I am interested in the sounds that differentiate between, for example, 'tous' and 'tu'; and the sounds that differentiate between, for example 'devant' and 'devons'. Please pay particular attention to how accurately these sounds are produced when giving your rating.

Press the SPACEBAR to continue.

Identification Screen

Press '1' if you think the word played sounds most like the word on the left.

Press '2' if you think the word played sounds most like the word on the right.

Rating Screen

The word you just heard was supposed to be 'XXX'.

Please rate the accuracy of the pronunciation of this word on a scale of 1 to 7, where 1 means very accurate/nativelike and 7 means very inaccurate/definitely not a native speaker.

FRENCH:

Initial Screen

Merci de votre participation à cette expérience. Vous allez entendre quelques mots en français. Certains d'entre eux ont été prononcés par des gens qui sont en train d'apprendre le français, et d'autres par des personnes de langue maternelle française. Vous avez DEUX CHOSES à faire!

PREMIÈREMENT vous choisirez parmi deux mots celui que vous croyez avoir entendu. Un mot sera à la gauche de l'écran et un mot sera à la droite de l'écran. Vous n'entendrez chaque mot qu'une fois, il faut donc écouter attentivement.

Appuyez sur '1' si vous croyez que le mot que vous avez entendu ressemble le plus au mot à gauche. Appuyez sur '2' si vous croyez que le mot que vous avez entendu ressemble le plus au mot à droite.

DEUXIÈMEMENT on vous dira quel mot le locuteur a essayé de prononcer. Je voudrais que vous classiez la prononciation du mot sur une série de '1' (très exact/comme un locuteur natif) à '7' (très inexact/évidemment pas un locuteur natif) en appuyant sur '1', '2', '3', '4', '5', '6', ou '7'.

IMPORTANT : Je m'intéresse aux sons qui différencient, par exemple, 'tout' de 'tu' et aux sons qui différencient, par exemple, 'devant' de 'devons'. Donc, s'il vous plaît, faites surtout attention à ces sons quand vous faites vos classements.

Identification Screen

Appuyez sur '1' si le mot que vous avez entendu ressemble le plus au mot à gauche.

Appuyez sur '2' si le mot que vous avez entendu ressemble le plus au mot à droite.

Rating Screen

Le mot que vous venez d'entendre devait être 'XXX'

S'il vous plait, classiez la prononciation du mot sur une série de '1' (très exact/comme un locuteur natif) à '7' (très inexact/évidemment pas un locuteur natif).