AUDIO-VISUAL TRAINING EFFECT ON L2 PERCEPTION AND PRODUCTION OF ENGLISH /θ/-/s/ AND /ð/-/z/ BY MANDARIN SPEAKERS

Submitted by
Ying Li

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

School of English Literature, Language and Linguistics
Newcastle University

October 2015

Abstract

Research on L2 speech perception and production indicate that adult language learners are able to acquire L2 speech sounds that they initially have difficulty with (Best, 1994). Moreover, use of the audiovisual modality, which provides language learners with articulatory information for speech sounds, has been illustrated to be effective in L2 speech perception training (Hazan et al., 2005). Since auditory and visual skills are integrated with each other in speech perception, audiovisual perception training may enhance language learners' auditory perception of L2 speech sounds (Bernstein, Auer Jr, Ebehardt, and Jiang, 2013). However, little research has been conducted on L1 Mandarin learners of English.

Based on these hypotheses, this study investigated whether audiovisual perception training can improve learners' auditory perception and production of L2 speech sounds. A pilot study was performed on 42 L1-Mandarin learners of English (L1-dialect: Chongqing Mandarin (CQd)) in which their perception and production of English consonants was tested. According to the results, 29 of the subjects had difficulty in the perception and production of /θ/-/s/ and /ð/-/z/. Therefore, these 29 subjects were selected as the experimental group to attend a 9-session audiovisual perception training programme, in which identification tasks for the minimal pairs /θ/-/s/ and /ð/-/z/ were conducted. The subjects' perception and production performance was tested before, during and at the end of the training with an AXB task and "read aloud" task. In view of the threat to interval validity arising from a repeated testing effect, a control group was tested with the same AXB task and intervals as that of the experimental group. The results show that the experimental group's perception and production accuracy improved substantially during and by the end of the training programme. Indeed, whilst the control group also showed perception improvement across the pre-test and post-test, their degree of improvement was significantly lower than that of the experimental group. These results therefore confirm the value of the audiovisual modality in L2 speech perception training.

Declaration

I certify that all the material submitted in this work which is not my own work has been identified, and that no material is included which has been submitted for any other award or qualification.

Signed:

Ying Li

Date: 22. Oct, 2015

Acknowledgements

I would like to express my heartfelt gratitude to my academic supervisor S.J. Hannahs for his exceptional guidance and constant support throughout this project. Without his support and invaluable encouragement throughout the journey, this thesis could not have been finished.

I am grateful to my supervisor Jalal-eddin Al-Tamimi for guiding me in the study of acoustic phonetics, and particularly for his guidance on acoustic measurements and the preparation of the recording of the stimuli used in the study.

This study has also benefitted immensely from my former colleagues and students, who helped me with subject recruitment, as well as providing me with places and equipment for the audiovisual training and tests. I would like to thank them for all of the help they have given to me.

Furthermore, I would like to thank the IT service centre at Newcastle University for their guidance on statistical analysis of the collected data. Many thanks for their patience and support when I was struggling with statistical analysis.

Finally, a special thanks to my beloved family and friends for always being supportive!

List of Abbreviations

CET-4: College English Test-level 4

CAH: Contrastive Analysis Hypothesis

CPH: Critical Period Hypothesis

CQd: Chongqing dialect of Chinese

CP: Categorical Perception

CV: consonant-vowel

dB: decibel

DRT: direct realist theory

F1: the first formant

F2: the second formant

F3: the third formant

GA: general approach

HVTP: High Variability Phonetic Training

Hz: Hertz

IPA: The International Phonetic Alphabet

IELTS: International English Language Testing System

ISE: Interstimulus interval

LVT: Low Variability Training

L1: First Language

L2: Second Language

MT: motor theory

NA: Non-Assimilable

NLM: Native Language Magnet Theory

PAM: Perception Assimilation Model

PI: Perception Interference

RP: Received Pronunciation

S: subject

SC: Single Category

SLA: second language acquisition

SLM: Speech Learning Model

TC: Two Categories

UC: Uncategorised

UU: Both uncategorised

VC: vowel-consonant

VCV: vowel-consonant-vowel

VOT: voiced onset time

2AFC: 2 alternative forced choice

Table of Contents

List of Figures

List of Tables

Table 5.37 The calculated scores of the male and female subjects’ production of /θ/ in pre-test, mid-test 1, mid-test 2 and post-test 150

Chapter 1. Introduction

1.1 Introduction

This chapter presents an outline of the thesis. The theoretical background of the present study is introduced first, and is followed by the presentation of the main aims and content of the study. Thereafter, the organization of the thesis is briefly outlined.

1.2 Theoretical background of the study

Research on speech perception and production has involved a series of topics. Regarding speech perception, one of the most intensively studied domains would be how speech sounds are perceived, particularly regarding auditory modality. Some scholars posit that speech perception is humans' response to the acoustic signals of the sounds they hear, as discussed in Stevens and Blumstein (1981), Diehl and Kluender (1989), as well as Diehl, Lotto, and Holt (2004). Others believe it is the articulatory gestures, or intended articulatory gestures that play a significant role in listeners' perception of speech sounds (Fowler, 1981, 1984, 1986, 1989, 1994a, b, 1996; Liberman and Mattingly, 1989). In other words, speech perception is in fact the discovery of the articulatory gestures which generate the speech sounds (Best, 1994, 1995a, 1995b; Best et al., 1988; Liberman, 1985). The research on speech production, however, does not involve such controversial debates concerning how speech sounds are produced. Nonetheless, we may need to pay attention to the factors which are found to have a potential influence on the acoustic features of the produced speech signals (e.g., pitch or formant trajectory), such as phonetic environments, gender, age differences and so on (Perkell, 1990).

The relationship between speech perception and production is another topic of debate. It is widely supported that speech perception and production may be closely tied to each other (Williams and McReynolds, 1975; Jamieson and Rvachew, 1992; Watkins, Strafella, and Paus, 2003), or even innately linked to each other (Liberman, Cooper, Shankweiler, and Studdert-Kennedy, 1967; Liberman, 1985). For instance, Liberman and colleagues' Motor Theory (MT) indicates that speech perception and production share a common link and a common processing strategy (Liberman et al., 1967; Liberman and Mattingly, 1985; 1989; Hawkins, 1999; Liberman and Whalen, 2000). Speech perception is believed to involve access to the speech motor system (Liberman et al., 1967; Liberman, 1985). Based on this hypothesis, it is possible to speculate that training on speech perception can enhance language learners' production of the trained speech sounds, and vice versa. Nonetheless, no consensus has been achieved on this issue. For example, some studies found that speech perception training did not benefit speech production (Winitz and Bellerose, 1963; Guess, 1969). Others revealed that training on speech production could not improve language learners' perception of the sounds (Winitz and Bellerose, 1962). However, there are some studies which found that speech perception training enhanced language leaners' perception and production of speech sounds (Winitz and Priesler, 1965; Mann and Baer, 1971; Rvachew, 1994).

Concerning the research on second language (L2) speech perception and production, the age factor and language learners' first language (L1) experience has usually been given great attention with regard to the learners' ultimate L2 achievement. That is, the influence of learners' age on their L2 learning was such that younger L2 learners were shown to have more advantages than older ones. Their advantages show first in the brain system of youths as it relates to language acquisition (Bialystok and Hakuta, 1999). Moreover, they have comparatively less L1 experience than older learners, which may interfere with their L2 learning (Best, 1994, 1995a, 1995b; Best and Tyler, 2007). Another age-related factor would be language learners' onset age (AO) of L2 learning. The critical Period Hypothesis (CPH) may shed some light on this issue. According to the CPH, L2 learners are unable to achieve native-like proficiency level if they commence L2 study after the end of the "critical period" (Lenneberg, 1967), or the "sensitive period" (Oyama, 1976), which is often defined as the period of puberty. The CPH has regard to the maturational changes in the brain that relate to language acquisition. That is, as the brain matures, language learners' brains lose plasticity, affecting L2 learning after the "critical period". The CPH is supported by findings from some previous studies on speech perception (Mayo, Florentine and Buus, 1997; Shi, 2010) and production (Tahta, Wood, and Loewenthal, 1981; Flege, Munro, and MacKay, 1995). However, the CPH also suffered criticism from the perspective of theoretical models (i.e. Flege's Speech Learning Model; Best and colleagues' PAM/PAM-L2; Kuhl's Native Language Magnet theory, and so on) and experimental findings (i.e. Flege et al., 1995; Fullana and Mora, 2008; Yamada, 1995). Moreover, its validity may be compromised since there is no consensus on when the "critical period" ends.

Another influential factor in L2 learners' acquisition would be their L1 experience. As early as the 1950s, Lado (1957) noticed the influence of learners' L1 on their acquisition of an L2, and proposed the Contrastive Analysis Hypothesis (CAH). CAH predicts that the differences between language learners' L1 and L2 systems pose difficulties for their L2 study. For supporting evidence, one needs to look no further than Japanese speakers' failure in distinguishing English /ɹ/-/l/ (Best and Strange, 1992), which could be explained by the non-occurrence of these two sounds in the Japanese phonetic inventory. However, the CAH is open to criticism from various perspectives. For instance, it is argued that the phonetic systems are unique for each language, so they are not comparable one to another (Weinreich, 1953; Wardhaugh, 1970). Even if phonetic systems can be compared across different languages, it may not be the case that the dissimilarities between learners' L1 and L2 pose difficulty for their L2 acquisition (i e., Flege's Speech Learning Model).

Best and colleague's Perception Assimilation Model-L2 (PAM-L2) also examines the influence of learners' L1 on their perception of L2 sounds. An important hypothesis of PAM-L2 is that language learners are likely to assimilate unfamiliar non-native speech sounds to the most articulatorily-similar phones of their L1 phonetic inventory. The assimilation is predicted to occur in different ways depending on the degree of variance between the learners' L1 and L2 in terms of articulatory gestures. PAM-L2 agrees that younger L2 learners have more advantages in L2 speech learning than older ones, because of their comparatively little L1 experience. Nonetheless, it predicts that adult L2 learners can eventually learn L2 speech sounds which they initially have difficulty with (Best and Tyler, 2007).

Moreover, Kuhl's (1992, 1993, 1994) Native Language Magnet theory (NLM), the expanded version—NLM-e), and Iverson, Kuhl, Akahane-Yamada, Diesch, Tohkura, Kettermann, and Siebert's (2003) Perception Interference (PI) theory all investigate the influence of learners' early language experience (typically their L1) on their perception of L2 speech sounds in future life. According to NLM/NLM-e and PI, language related neural tissue changes with initial exposure to a language. Early life experience makes learners neutrally commit to the acoustic cues of the language (usually, their L1). As a result, adult learners are less sensitive to the acoustic cues of non-native speech sounds than infants. Nevertheless, like PAM-L2, both NLM/NLM-e and PI hold the view that adult learners can ultimately acquire non-native speech sounds that they initially have difficulty with.

Flege's (1981, 1987, 1988, 1991a, 1992a, b, 1995a, 2003) Speech Learning Model (SLM) discusses the influence of language learners' L1 on their perception and production of L2 speech sounds. In common with the hypothesis of PAM-L2, NLM/NLM-e and PI, SLM predicts that adult language learners can eventually learn L2 speech sounds with which they initially have difficulty. In the meantime, SLM points out that greater L2 experience serves to enhance language learners' capability for the perception and production of L2 speech sounds. Speech perception training is predicted to be able to enhance language learners' production of the speech sounds. Moreover, SLM holds a different view to CAH with regard to the difficulties arising from the difference between learners' L1 and L2. According to SLM, the more dissimilar the L1 and L2 sounds are, the more likely it is that the L2 learners can develop the new phonetic categories of the L2 sounds.

On the whole, PAM-L2, NLM/NLM-e, PI and SLM all predict that adult learners can ultimately learn L2 speech sounds if given sufficient input of the target L2 sounds. Supporting evidence can be found in previous studies such as Pisoni, Aslin, Perey and Hennessy (1982) and Jamieson and Morosan (1986, 1989). This hypothesis serves as a significant theoretical basis of the present study.

Moreover, articulatory information (i.e. visible articulatory gestures) was found to be able to facilitate language learners' perception of L2 speech sounds (Navarra and Soto-Faraco, 2007; Walden, Prosek, Montgomery, and Scherr, 1977). The McGurk effect (McGurk and MacDonald, 1976) can be seen as one of the embodiments of the influence of articulatory gestures on speech perception. In addition, successful lip reading training studies on speech perception have further illustrated the facilitating role of visual articulatory information in speech perception (Walden, Prosek, Montgomery, Scherr, and Jones, 1977; Walden, Erdman, Montgomery, Schwartz, and Prosek, 1981). An important premise of the prediction is that audiovisual, auditory and visual skills are integrated with each other in speech perception (Bernstein, Auer Jr, Ebehardt, and Jiang, 2013). This hypothesis inspired the design of the training programme of the present study, in which an audiovisual modality was employed.

Based on these hypotheses and findings, it might be logical to predict that audiovisual speech perception training, which provides L2 learners with visible articulatory information of specific L2 speech sounds, can benefit their auditory perception of L2 sounds. Given that it is widely accepted that speech perception and production can be closely connected (Williams and McReynolds, 1975), speech perception training may benefit learners' production of L2 speech sounds. Furthermore, other relevant factors, such as gender difference, motivation and the amount of time spent in learning the L2 language may lead to individual differences concerning the training results.

1.3 Main aims and content of the present study

The present study aimed to reveal: (1) whether audiovisual speech perception training can benefit L2 learners' auditory perception of L2 speech sounds; and (2) whether audiovisual perception training can benefit L2 learners' production of the same speech sounds.

A pilot study was carried out prior to the main study, with the aim of (1) revealing which English consonant(s) was (were) relatively more difficult for the subjects to perceive and produce; and (2) identifying suitable subjects who had difficulty in perceiving and producing the target English speech sounds. 42 university level students of the same English proficiency level were recruited in China. They were L1-Mandarin learners of English. They all speak Chongqing dialect of Chinese (CQd thereafter) as their L1 dialect. Their performance in the production and auditory perception of all English consonants was evaluated. According to the results, 29 (14 male, 15 female) subjects out of 42 were selected to join the main study as the experimental group. The target English speech sounds that were identified as the most challenging ones for these subjects were /θ/-/s/ and /ð/-/z/.

The experimental group received audiovisual perception training. The training programme consisted of 9 sessions, each session lasting about 35 minutes. The training programme, to a large extent, followed the High Variability Phonetic Training (HVPT) approach (Bradlow, Pisoni, AkahansYamada, and Tohkura, 1997; Lively, Logan, and Pisoni, 1993). Each training session included 120 "minimal pairs" (60 trials for each contrast), in which the target contrasts were embedded in initial, medial and final positions. The stimuli were audiovisually produced by 3 Received Pronunciation (RP thereafter) speakers (2 female, 1 male). Subjects of the experimental group were asked to identify which word they heard and/or watched in the recording occurred first in each "minimal pair". Immediate feedback of the correctness of their responses was given. In order to investigate the improvement in their perception and production during and at the end of the training programme, if any, the subjects' accuracy in the perception and production of the target contrasts was tested before the training programme (pre-test), at the end of the 3rd training session (mid-test 1), at the end of the 6th training session (mid-test 2), and at the end of the whole training programme (post-test). An AXB test with nonsense words as the stimuli was employed in the perception test. With regards to the production test, the subjects were asked to read 12 English sentences, in which the target contrasts were embedded in real words in initial, medial, and final positions. 4 RP speakers were asked to assess their performance in producing the target contrasts with a *10-point Likert Scale*. Qualitative data was collected with a questionnaire to investigate relevant factors (i e., age, AO etc.) that may influence the subjects' performance during perception and/or production.

Due to the same stimuli being employed in the perception test 4 times, it was necessary to examine whether there were repeated testing effects on the subjects' perception performance. Therefore, 20 subjects of similar profiles to that of the experimental group were recruited as the control group (see Appendix 15 for details). Their accuracy in the perception of the target English contrasts was tested 4 times with the same testing materials and intervals as that of the experimental group.

1.4 Outline of the thesis

The following is a brief overview of the structure of the thesis. In Chapter 2, the theoretical background on L2 speech perception and production is presented. Speech perception and production theories are discussed first, including how speech sounds are perceived and produced, followed by the different factors which may have a significant influence on learners' perception and production of L2 speech sounds. The relationship between speech perception and production is then reviewed. Relevant factors which may have an influence on language learners' perception and production of L2 speech sounds are discussed, such as age, L1 influence, gender, motivation, the amount of time spent on L2 learning. The hypotheses of the CPH, CAH, PAM-L2, SLM, NLM/NLM-e and PI are reviewed and compared, which form the primary theoretical basis of the present study. Then, the critical role of visual codes/ articulatory information in speech perception is discussed with the support of the McGurk effect (McGurk and MacDonald, 1976), and experimental evidence of lip-reading in speech perception. This is followed by the analysis of the integration of audiovisual, auditory and visual skills in speech perception. In addition to the discussion of previous perception training programmes, the chapter ends with a comparison of the two approaches that are frequently used in speech perception training programmes, namely High Variability

Phonetic Training (HVPT) (Logan, Lively, and Pisoni, 1991) and Low Variability training (LVT) (Strange and Dittmann, 1984).

Chapter 3 discusses the articulatory and acoustic features of English /θ/-/s/, /ð/-/z/, Mandarin /s/ and CQd /s, z/. First, the articulatory characteristics of /θ/-/s/, /ð/-/z/ in English are presented, followed by the acoustic properties of the two contrasts. The background information on Mandarin and CQd is briefly introduced. This is then followed by a comparison of the phonetic inventories of English, Mandarin and CQd consonants. It is found that of the two contrasts, /θ, ð/ neither exist in Mandarin nor CQd; /z/ occurs in English and CQd but not Mandarin; whereas only /s/ exists in English, Mandarin and CQd. The acoustic characteristics of Mandarin /s/ are presented with findings in previous studies. Due to the lack of references on the acoustic properties of CQd /s, z/, 6 subjects' production of the two sounds are analysed using the Praat Programme (Boersma and Weenink, 2013).

Chapter 4 presents the content of the pilot study. The methodology is outlined first, which includes the recruitment of subjects, the preparation of stimuli and the design of the tasks. Then the results from the perception and production tests are presented. The results from the pilot study are discussed with regard to the literature on the acquisition of L2 speech sounds.

The main study is presented in Chapter 5. As for the pilot study, the methodology is introduced first, followed by the main study. The results of perception and production tests (pre-test, mid-test 1, mid-test 2, and post-test) are then presented. The subjects' perception and production accuracy in the pre-test is compared with that in the mid-test 1, mid-test 2, and post-test. Qualitative data about the participants was collected with a questionnaire are then presented. The results are analyzed with a *repeated measures ANOVA*.

Chapter 6 summarizes and discusses the results of the main study with a focus on the different theories/models of L2 acquisition that may account for the present findings. The implications of the study for L2 learning are also briefly discussed, followed by the limitations of the current study and areas for future research.

1.5 Conclusion

This chapter has provided an overview of the thesis and introduced the theories/models that serve as the theoretical basis of the present study. The main aims and content of the present study were presented, along with an outline of the thesis.

Chapter 2 Literature Review

2.1 Introduction

This chapter primarily reviews the existent theories and models of speech perception and production, particularly regarding L2 speech perception and production. The hypotheses of these theories and models are discussed with the reference to findings from previous studies. Moreover, the relationship between speech perception and production is discussed. Factors which may have an impact on language learners' perception and production of L2 speech sounds are presented and discussed. Given that the present study involves audiovisual perception training, the literature on the significance of visual articulatory information in the perception of speech sounds is reviewed. Some previous studies on phonetic training are discussed. Audiovisual trainings approaches that were carried out by former scholars are presented and discussed, with an emphasis on the two frequently used approaches in speech perception training.

2.2. Speech perception

2.2.1 What is speech perception?

The definition of speech perception, as reviewed by Klatt (1989), varies from one author to another, particularly in terms of different finite and a small number of linguistically defined attributes (for example, features, phones, phonemes, syllables, words). Sebastián-Gallés (2005) defines speech perception as the process that happens between the perception of an acoustic wave and the discovery of the meaning of words, which involves the transition of physical sound waves to neural patterns that represent the meaning of words. Similarly, Hawkins (1999) describes speech perception as a listener's task of understanding the meaning of what a speaker said. It is revealed that during the process of speech perception, the auditory system records the sound vibrations generated by the speaker, which is followed by the translation of these vibrations into a sequence of sounds that are then perceived by the listener. Although being different from one author to another, a common hypothesis of these views would be that perception begins with a speech signal being well-composed and fit for analysis (Remez, 2008). The present study focuses on the perception of phonetic segments, namely the two English contrasts /θ/-/s/, /ð/-/z/.

2.2.2 How speech sounds are perceived

In the research on how speech sounds are perceived (auditory only), typically there are two different views. The first view is that speech perception is humans' responses to acoustic stimuli (Raphael, 2005). For instance, the general approach (GA) of Diehl et al. (2004) notes that speech sounds are perceived through the recovery of the acoustic signals of the sounds. This is congruent with the acoustic-auditory theory of speech perception, which posits that speech perception and production are linked through the perceived acoustic signals and auditory feedback mechanism of speech production (Stevens and Blumstein, 1981; Diehl and Kluender, 1989). Speech signals include a number of acoustic cues or acoustic properties such as duration, static and dynamic spectral features, periodicity, noise, intensity, and so on, which differentiate one speech sound from another belonging to a different phonetic category (Escudero and Boersma, 2004). The integration of these acoustic cues is predicted to play a significant role in the differentation of phonological contrasts. For instance, voice onset time (VOT) is widely adopted as the primary cue for the distinguishing of voiced stops from voiceless ones (Kuhl and Miller, 1975; Dooling, Okanoya, and Brown, 1989; Flege, 1991c). Tenseness and duration are usually employed in the categorization of different vowels (Peterson and Barney, 1952; Peterson and Lehiste, 1960; Bohn and Flege, 1990). Nevertheless, in Remez (2008), acoustic cues are predicted to be more appropriate for use in speech analysis than as a medium in speech perception.

The second view, however, indicates that the perception of speech sounds is the recovery of articulatory gestures, or intended articulatory gestures (Best, 1994; Best and Tylor, 2007; Liberman et al., 1967). It was found that synthetic speech cannot be produced across different contexts unless its acoustic patterns are modified. In other words, synthetic speech does not sound natural on its own, and can be better perceived if in an appropriate context. Based on this finding, it is hypothesized that instead of perceiving acoustic signals, an underlying process is progressing when perceiving speech sounds (Cooper, Delattre, Liberman, Borst, and Gerstman, 1952; Liberman, Delattre, and Cooper, 1952). Based on this finding, Liberman and colleagues' Motor Theory of speech perception (MT) indicates that the goal of speech perception is the recovery of the articulatory gestures of target speech sounds (Liberman et al., 1967; Liberman, 1985, 1989, 2000). One of the essential hypotheses of MT is that instead of identifying the acoustic patterns that are the product of articulatory gestures, language listeners perceive, or "reconstruct" (Hawkins, 1999) speech sounds via the identification of vocal tract movements, which are used in the production of the sounds. The revised version of MT suggests that what listeners perceive is the intended vocal tract gestures, or intended articulatory gestures. This refers to the abstract control units that can give rise to linguistically relevant vocal tract movements (Liberman and Mattingly, 1985, also see Hawkins, 1999).

Similarly, direct realist theory (DRT) (Fowler 1981, 1984, 1986, 1989, 1994a, b, 1996) agrees with the hypothesis that the object of speech perception is articulatory gestures rather than acoustic events. However, different from MT, DRT asserts that the articulatory objects of perception are phonetically structured, vocal tract movements, or gestures" rather than intended articulator gestures as hypothesized by MT. The acoustic signals of speech sounds play a medium role in listeners' identification of the articulatory gestures (Fowler, 1981, 1984, 1986, 1989, 1994a, b, 1996).

Moreover, the McGurk effect provides more supporting evidence for the view that the perception of speech sounds is in fact the perception of (intended) articulatory gestures. It has been shown that when the auditory component of one sound is paired with the visual component of another sound, it can lead to the perception of a third sound (McGurk and MacDonald, 1976; Nath and Beauchamp, 2011). Therefore, articulatory gestures are incorporated into speech perception. In addition, infants' initial ability to perceive the speech sounds of different languages is attributed to their discovery of simple articulatory gestures (Best, 1994, 1995a, 1995b; Best et al., 1988).

This view is also supported by findings on the experimental level. For instance, Liberman et al. (1967) found that listeners successfully perceived /d/ both in /di/ and /du/, despite the fact that /d/ displays different F2 trajectories in the two different vowel contexts. Similarly, in Fowler, Brown, Sabadini, and Weihing (2003), the subjects were asked to imitate and choose the "words" they heard. The stimuli were words of different syllabic structures (vowel-consonant-vowel (VCV) and consonant-vowel (CV)), which were produced by a model speaker. The duration of the vowel in the stimulus words varied unpredictably from one to another. It was discovered that the subjects' response time in the choice task exceeded that in the imitation task by 26ms. Fowler et al. (2003) attributed this finding to the listeners' rapid extraction of articulatory gestures from speakers in the choice task. Findings from coarticulation experiments, such as that of Viswanathan, Magnuson, and Fowler (2010) further confirmed the view that the finite goal of speech perception is in fact the discovery of the articulatory gestures.

Another debate related to speech perception is the variation of ways in which speech sounds of different categories are perceived. According to Fry's (1966) Categorical Perception (CP) theory (cited by Barkat-Defradas, Al-Tamimi, and Benkirane, 2003), speech sounds are perceived either categorically or continuously. If a speech sound perceived with a peak occurs "at the midrange of a continuum" (Macmillan, 1987), or between phonemic categories (Liberman, Harris, Hoffman, and Griffith, 1957), it is viewed as categorical perception. Otherwise a speech sound is predicted to be perceived continuously (Macmillan, 1987; Liberman, Harris, Hoffman, and Griffith, 1957). According to CP, "the continuous, variable, and confusable stimulation that reaches the sense organs is sorted out by the mind into discrete, distinct categories whose members resemble one another more than they resemble members of other categories" (Harnad, 1990; Liberman et al., 1957; Liberman, Harris, Hoffman, Eimas, Lisker, and Bastian, 1961a, b). It was discovered that consonants are likely to be perceived categorically. That is, "listeners perceive speech sounds which differ from each other in terms of equal steps along a continuum as belonging to either one or another category" (Bradlow, 2008). For instance, Liberman et al. (1957) reported that when listening to consonants in a continuum from /b/ to /d/ to /g/, the subjects' judgement of the heard stimuli were /b/, /d/ or /g/, rather than anything in between. Therefore, Liberman et al. (1957) noted that phoneme boundaries for consonants are sharp and stable. In contrast, vowels are found likely to be perceived continuously, or non-categorically. That is, vowels are likely to be heard arbitrarily along a continuum (Pickett, 1999). However, the study in Pisoni and Tash (1974) provided counterevidence to this view. In their study, the subjects' reaction time in the identification of speech sounds from a speech continuum increased as tokens moved closer to the phoneme boundary. Moreover, in Samuel (1977), listeners were successfully trained to discriminate within-category tokens in stop consonant continua.

In addition, the perception of a speech sound is found to largely depend on its nearby context(s). It was discovered that two speech sounds can be easily labelled if they are presented in singleton, but relatively harder to be distinguished if embedded in contexts with other speech sounds. For example, in Jamieson and Morosan (1989), the subjects successfully distinguished /ð/ from /θ/ when they were presented as isolated forms, but performed less successfully when they were contained in normal, full syllables and in the context of words. The vowel context has been shown to to influence listeners' perception of consonants. For instance, Mann and Repp (1980) found that when synthetic fricative noise of a /ʃ/-/s/ continuum is followed by /a/ or /u/, listeners perceive more instances of /s/ in the context of /u/ than in the context of /a/. Miller and Liberman (1979) examined the subjects' perception of stops and glides. They found that the subjects' identification of /b/ and /w/ was largely dependent upon the duration of the following vowels, with more stops being identified if they were followed by longer vowels. Miller and Liberman (1979) indicate that due to longer vowels involving a slower speech rate, this results in a greater range of transitional durations being compatible with the stop category. Similar findings were obtained by Diehl et al. (1980), Miller (1987) and Summerfield (1981). Furthermore, the perceived length of an acoustic segment is found to be inversely related to the length of its adjacent acoustic segments. That is, the longer the adjacent context is, the shorter the target speech sound is perceived to be (Diehl et al., 2004). Mann (1980) reported that in the identification of /da/-/ga/, which vary in the trajectory of F3, the subjects gave more /ga/ responses when following /al/ than following /ar/. This is consistent with the findings presented in Mann (1986) and Fowler et al. (1990).

2.3. Speech production

2.3.1 What is speech production?

Speech production is typically defined as the way in which speech sounds are produced by human articulatory organs (Taylor, 1974). It is "the only part of language which is directly 'physical', and demands neuromuscular programming" (Scovel, 1988). Speech sounds are "created by modifying the volume and direction of a flow of air using various parts of the human respiratory system" (Davenport and Hannahs, 2010). The process of speech production is fulfilled in terms of passing the air through the vocal tract (see Figure 2.1), which results in an acoustic output (Taylor, 1974; Shadle, 1985; Pickett, 1999). The acoustic output of speech sounds can be modified by the speaker by varying the volume of air that flows through the vocal tract and also by articulatory gestures (Shadle, 1985).

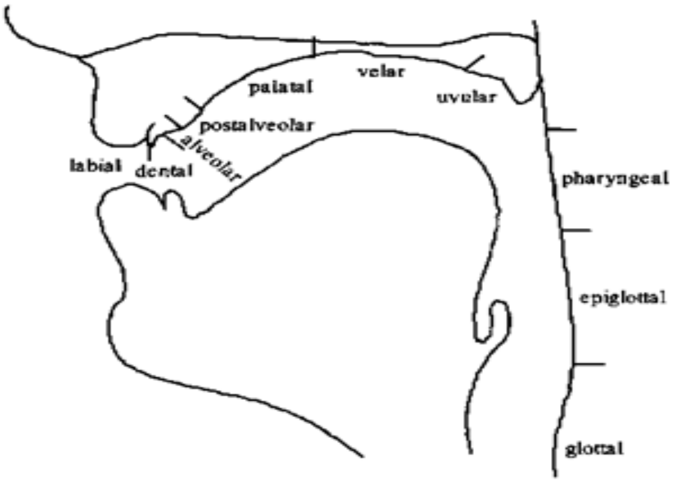


Figure 2.1 Target regions of the upper and posterior parts of the vocal tract of an ideal reference speaker (Ladefoged, 2007).

Figure 2.1 above depicts the articulators involved in speech production. The movement of the articulators determines the resonance properties of the acoustic output. The acoustic properties of produced speech sounds change according to the change in the vocal tract shape (Kent and Read, 2002). Speech sounds are composed of a number of articulatory properties/components, which are independent of one another. The produced output of speech sounds, therefore, can be viewed as the combination of these properties/components (Davenport and Hannahs, 2010). Accordingly, one speech sound can be differentiated from others in terms of distinctive acoustic properties. However, the mapping between articulation and acoustic properties has been found to be nonlinear (Stevens, 1972). For instance, speakers can produce voicing with different possible glottal widths. This is because instead of being singleton, speech sounds are usually produced as a continuum or a stream (Sweet, 1877, cited by Wood, 1993), which includes a considerable overlap of linguistic units (Löfqvist, 2010). Therefore, the influence of context on speech production can never be ignored.

2.3.2 Coarticulation of speech production

Coarticulation is a common phenomenon in speech production. It refers to the influence of one articulatory segment on its neighbouring segment(s) of the same utterance. It is shown in the form of overlapping movements in the production of adjacent speech sounds (Nittrouer and Studdert-Kennedy, 1987; Pickett, 1999). On the acoustical level, it is shown in the form of "temporal smoothing of a sequence of presumably inherent phonetic gestures in adjacent (or in some sense distant) entities" (Fujimura and

Erickson, 1997). In other words, it is the transition between neighbouring vowel(s) and consonant(s), which involves “the adjustments in vocal tract shape made in anticipation of a subsequent motion” (Chomsky and Halle, 1968). Coarticulation presents in the form of anticipatory (right-to-left) and retentive (or carryover, left-to-right). In anticipatory coarticulation, the articulation of a sound is influenced by its adjacent following sound. For instance, due to the influence of following rounded vowels, the alveolar fricative /s/ is likely to be produced with the articulatory gesture of lip rounding when producing *stew* (Kent and Read, 2002). Retentive coarticulation, however, operates in the opposite direction. That is, the pronunciation of a target speech sound is affected by its adjacent preceding sound. For example, when an English vowel is followed by a nasal stop, it is usually pronounced as nasalized, despite there being no nasal vowel in English. For instance, the mid back round vowel [ɔː] is pronounced as nasalized [ɔ̃ː] in *north*.

Coarticulation appears frequently in running speech in different phonetic environments – consonant contexts for vowel targets (influence of consonants on a target vowel) (Holt et al., 2000; Lindblom and Studdert-Kennedy, 1967; Nearey, 1989), vowel contexts for consonant targets (influence of vowel contexts on the target consonant) (Holt, 1999; Mann and Repp, 1981), as well as vowel contexts with vowel targets (influence of vowel contexts on the target vowel) (Fowler, 1981). Consequently, instead of being independent, phonetic properties of a speech sound vary according to the context(s) that they are embedded in. For instance, Wilde (1995) examined the effects of vowel context on labiodental consonants (the configurations for /fu/ and /u/), and found that the tongue body and blade “anticipate” the articulation of the following vowel. In Pickett (1999), the intensity of F1 is shown to be weaker when adjacent to /m/ than to /b/. Zue (1985) reported that /ð/ is likely to be realized in a stop-like manner when preceded by a consonant. Moreover, the acoustic features of /a/ in CV syllables (like /ba/-/ga/) are found to be different from those in VC (vowel-consonant) syllables (such as /al/-/ar/) (Diehl, Lotto, and Holt, 2004).

The degree of coarticulation depends on the extent to which articulatory gestures share the same articulator(s). For example, in the production of a speech sound involving two overlapping gestures that share the same articulator(s), there is the highest degree of spatial overlap (Farnetani, 1997). Moreover, the articulatory gestures of a following speech sound may lead to change in the articulatory/acoustic characteristics of a target speech sound. For instance, it has been shown that lip-rounding leads to a decrease of all formant frequencies, which particularly affects the trajectory of F2 (Ashby and Maidment, 2005; Pickett, 1999). Bell-Berti and Harris (1979) found that, when paired with rounded vowels /u, o/, the identification of /θ, ð/ depends more on the phonetic portion than on the portion of friction. Being adjacent to rounded vowels can also result in lowering the frequency of all sibilants by 300-500 Hz (Bell-Berti and Haarris, 1979). Klatt (1974) (as reviewed by Wilde, 1995) revealed that the high frequency intensity of /s/ increased by 6 dB when it preceded rounded vowels. Nevertheless, the coarticulatory lip-rounding effect is found to be speaker-dependent in Wilde (1995). Farnetani (1997) elegantly described coarticulation in different situations, which include the main articulators and muscles involved in coarticulation, the overlapped movements in contiguous segments, and acoustic consequences of the movement, as shown in Table 2.1.

Articulator	Level of discritption		
	Myomotoric	Articulatory	Acoustic
LIPS	Orbicularis Oris/Orisorius	Lip rounding/ spreading	Changes in F1, F2 and F3
TONGUE	Genioglossus and other extrinsic and intrinsic lingua muscles	Tongue front/ back, high/ low displacement	Changes in F2, F1 and F3
VELUM	(Relaxation of) Levator Palatini	Velum lowering	Nasal Formants and changes in Oral Formants
LARYNX	Posterior Cricoarytenoid/ Interarytenoid, Lateral Cricoarytenoid	Vocal fold abduction/ adduction	Aperiodic/ Periodic signal Acoustic duration

Table 2.1 Coarticulation (adapted from Farnetani, 1997).

The phenomenon of coarticulation tells us that the acoustic/articulatory properties of a produced speech sound could be affected by its neighbouring contexts. Therefore, in the present study, when analysing the subjects' accuracy in the perception of a target speech sound, it is necessary to detect whether their perception performance was significantly affected by the neighbouring (vowel) context(s) of the target speech sounds. Specifically, the variable *context* will be coded as a within-subject factor to investigate whether the subjects perform differently when the target contrasts are embedded in different vowel contexts.

2.3.3 Speaker-variation in speech production

In speech production, speaker differences may lead to variation in the speech signals, which can be observed in terms of formant frequency, amplitude, and so forth. This has been shown to result from many factors. For instance, speakers who come from different social classes, with different dialects, may show different degrees of variation in speech signals when producing the same speech sound. Gender and age differences would be another two significant factors that lead to variation in speech signals. Moreover, different sounds are possibly produced with the same acoustic patterns by different speakers (Peterson and Barney, 1952; Fant, 1973). Speech sounds produced by female adults and children are found to display higher frequencies than those produced by male adults. This is because female and children have a relatively smaller larynx and shorter vocal tracts than male adults. For instance, the spectral peaks of /s, z/ are found to vary from 5 to 10 kHz if produced by female adult speakers, whereas they range from 4 to 9 kHz if produced by male adult speakers (Shadle, Badin, and Moulinier, 1991)

2.4 The relation between speech perception and production

As a long-standing issue, the relation between speech perception and production has given rise to many different views. In a number of studies, speech perception and production were found to be closely tied to each other (Williams and McReynolds, 1975; Jamieson and Rvachew, 1992; Watkins, Strafella and Paus, 2003), or even innately linked to each other (Liberman et al., 1967; Liberman, 1985). For example, children who suffer from functional articulation disorders are likely to have difficulties in speech perception (Jamieson and Rvachew, 1992; Broen, Strange, Doyle, and Heller, 1983; Hoffman, Daniloff, Bengoa, and Schuckers, 1985; Morgan, 1984; Raaymakers and Crul, 1988; Rvachew and Jamieson, 1989; Winitz, 1969). In other studies, however, speech perception and production are viewed as two independent modalities (Locke, 1988; Schwartz and Leonard, 1982).

Liberman and colleagues' Motor theory (MT) shed some light on the relation between speech perception and production. One of MT's hypotheses is that speech perception and production share a common link and a common processing strategy (Liberman et al., 1967; Liberman and Mattingly, 1985; 1989; Hawkins, 1999; Liberman and Whalen, 2000; Galantucci, Fowler, and Turvey, 2006). Cooper (1979) illustrated the link between speech perception and production with the finding that the VOT of /pi/ and /ti/ is slightly reduced by speakers after adapting to acoustically presented /pi/.

Furthermore, Bell-Berti, Raphael, Pisoni, and Sawusch (1979) compared the acoustic cues the subjects employed in the perception and production of /i/-/e/ and /ɪ/- /ɛ/, and found that the strategies they adopted in the perception of these vowels were the same as those used in the production of them. Specifically, tongue height and tenseness were employed as the critical cues (Bell-Berti et al., 1979).

Moreover, MT postulates, speech perception involves access to the speech motor system (Liberman et al., 1967; Liberman and Mattingly, 1985). It was found that audiovisual speech can activate both the cerebellum and cortical motor areas which are involved in the planning and production of speech sounds (Skipper, Nusbaum, and Small, 2005). Meanwhile, both auditory and visual speech perception have been shown to be able to facilitate the excitability of the motor systems involved in speech production (Watkins et al., 2003). Further supporting evidence comes from the discovery of *Mirror neurons,* which enable speakers to learn and imitate the observed sounds which they have previously encountered (Di Pellegrino, Fadiga, Fogassi, Gallese, and Rizzolatti, 1992). Monkeys' premotor cortex was found to discharge both when they perform a specific action and when they hear a sound caused by the action (Kohler, Keysers, Umiltà, Fogassi, Gallese, and Rizzolatti, 2002). A similar phenomenon was also discovered in humans (Fadiga, Fogassi, Pavesi, and Rizzolatti, 1995; Strafella and Paus, 2000; Kohler et al., 2002; Rizzolatti and Craighero, 2004). In other studies, an overlap between cortical areas was found, which can be activated both during speech production and when listening to speech sounds (Pulvermüller, Huss, Kheri, Moscoso del Prado Martin, Hauk, and Shtyrov, 2006; Wilson, Saygin, Sereno, and Iacoboni, 2004). These findings led to the assumption that a motor plan can be elicited from the observed articulatory information during the production of corresponding speech sounds (Skipper et al., 2005, 2006). This has been confirmed by findings from experimental studies. For instance, in Fadiga, Craighero, Buccino, and Rizzolatti (2002), when listeners hear utterances that include lingual consonants, their muscle activity in the tongue is enhanced. Similarly, Watkins et al. (2003) found enhanced muscle activity in listeners' lips both when listening to speech and seeing speech related lip movements.

Another piece of evidence for the close relationship between speech perception and production can be found in the influence of coarticulation on speech perception and production. As discussed in section 2.3.2 above, coarticulation changes the articulatory gestures, and consequently modifies the acoustic properties of a target speech sound in different phonetic environments. As a result, the perceived speech sounds may vary accordingly, as shown by the findings in Mann and Repp (1980) as well as those of Miller and Liberman (1979).

Moreover, although Fowler (1986) agrees that speech perception and production are not independent from each other, she suggests that instead of being connected by a mediating link, speech perception and production belong to an integrated system. Fowler’s (1986) assumption may be disproved by relevant neurological findings. For instance, *canonical neuron* and *visual-tactile neurons* have been proven to link humans’ perceptions of the surrounding world (not only speech) and their physical reactions to it (Murata, Fadiga, Fogassi, Gallese, Raos, and Rizzolatti, 1997; Rizzolatti, Fadiga, Fogassi, and Gallese, 1997). Therefore, it seems speech perception and production may be more likely to be linked to each other rather than share an integrated system.

However, even if speech perception has been proven not to be independent from speech production, no consensus has been achieved regarding whether speech perception precedes or follows its production. One of the widely supported views is that the accurate perception of L2 speech sounds is at least one necessary component for L2 learners’ accurate production of them (Flege, 1995a, b; Flege, Bohn, and Jang, 1997; Wode, 1996). Consequently, language learners’ perception ability usually surpasses and precedes their production ability (Flege, 1988). Supporting evidence for this view comes from Flege et al. (1999). In their study, although the L2 leaners’ perception and production abilities were found to be nearly asymptotic, they displayed better performance in the perception of some vowels than the production of them.

On the contrary, other scholars indicate that speech production precedes perception. Some L2 learners have been found to be able to accurately produce L2 speech sounds but failed to correctly perceive them (Zampini and Green, 2001). For example, some Japanese learners of English were found to be able to produce English /ɹ/-/l/ without being able to accurately perceive the contrast (Sheldon and Strange, 1982; Smith, 2001). Baker and Trofimovich (2006) reviewed these studies and gave two possible explanations for these findings, namely (1) speech perception and production develop independently; or (2) correct production is indispensable for the accurate perception of speech sounds. The hypothesis that speech production precedes speech perception is congruent with conceptualizations in L1 development, which suggests that input (perception) and output (production) of lexicons are two separate underlying representations in children's linguistic system development (Locke, 1988; Schwartz and Leonard, 1982).

Nevertheless, some scholars insist that speech perception and production are interdependent and develop simultaneously (Best, 1995a, b). Best and colleagues' PAM/PAM-L2 claims that accurate perception of speech sounds is achieved through the discovery of articulatory gestures that produce the sounds (i.e., tongue movements; lip movements). Perception and production are always aligned. Thus speech perception neither surpasses nor precedes speech production. Learners' ability in perceiving and producing L2 speech sounds develops in synchrony, and depends on how difficult it is to discover the articulatory differences between the L1 and L2 sounds (Best, 1995a, b; Best and Tyler, 2007). This view is supported by the McGurk effect (McGurk and MacDonald, 1976), which shows listeners adopted both visual codes and auditory signals in speech perception.

Experimental trainings on speech perception and production also provide us with inconsistent, or even conflicting results. For instance, early experimental studies revealed that speech production training could not change language learners' perception capability (Winitz and Bellerose, 1962; Harrelson, 1969). Others found speech perception training does not benefit language learners' production performance (Winitz and Bellerose, 1963; Guess, 1969). For instance, Guess and Baer (1973) revealed that little change occurred in either speech perception or production when the opposite modality was trained. In Williams and McReynolds (1975), although production training both improved the subjects' perception and production performance, perception training only enhanced the subjects' perception of the target speech sounds rather than their production. However, Williams and McReynolds' (1975) finding was contradicted by Jamieson and Rvachew (1992), in which four children with functional articulation disorders were trained to identify 2 synthesized fricatives. Twenty 6-minute sessions of sound identification training were carried out without production training. The post-test results indicated that three of the subjects' ability in the production of the target phonemes was significantly improved, which predicts a transferred beneficial effect of speech perception on speech production. Similarly, in Bradlow et al. (1997), Japanese speakers underwent perception training with English /ɹ/-/l/ as the target contrast. It turned out that both the subjects' accuracy in the perception and production of the contrast was improved. Similar findings were obtained by Lambacher, Martens, Kakehi, Marasinghe, and Molholt (2005), Winitz and Priesler (1965), Mann and Baer (1971), as well as Rvachew (1994).

On the whole, no consensus has been achieved on the relationship between speech perception and production. It is still an unresolved question concerning whether speech perception training can improve subjects' ability in producing the trained speech sounds, or vice versa.

2.5 L2 speech perception and production

The research on L2 speech perception and production belongs to the domain of second language acquisition (SLA). Therefore, findings from studies on SLA may shed some light on this discussion. In the following sections, details of some factors that were shown to have an influence on learners' perception and production of L2 speech sounds are discussed.

2.5.1 Age factor

One of the intensively studied factors that may have an influence on language learners' L2 proficiency would be the age of onset (AO) of L2 learning. With regard to the age of onset of L2 learning, the traditional view is "the earlier the better" (Lenneberg, 1967). The critical Period Hypothesis (CPH) is one of the representative theories on this issue. It claims that language learners are unable to achieve native-like proficiency level if they begin L2 learning after the end of a "critical period" (Lenneberg, 1967), or "sensitive period" (Oyama, 1976). The assumption of the CPH rests on the maturational changes in brain structures that relate to language acquisition. In other words, language learners lose brain plasticity as the brain matures, which impedes their L2 learning (Lenneberg, 1967; Scovel, 1969, 1988; Patkowski, 1980, 1990).

It is suggested that language learners' AO of L2 learning may affect their perception and production of L2 speech sounds. In perception, for instance, Mayo et al. (1997) conducted a perception test in noise with early bilinguals (who began learning L2-English before 6 years of age), late bilinguals (who commenced L2-English study after 14 years of age), and monolingual American-English speakers. All the bilinguals were native Mexican Spanish-speaking subjects. It turned out that monolinguals and early bilinguals benefited significantly from the context, and showed a higher level of comprehension in noise than late bilinguals. Similarly, Shi (2010) revealed that in the perception of sentences presented with noise, the effect of noise and the integrated effect of reverberation and context were meditated by the subjects' AO of L2 learning. Specifically, "very late" bilingual learners' perception performance was significantly poorer than late and early bilingual learners. However, inconsistent results were found by Fullana and Mora (2008). Fullana and Mora (2008) examined early starters (AO<8 years old) and late starters' (AO>8 years old) perception of English voicing contrasts in word-final position with an AXB task. All the subjects were bilingual speakers of Catalan and Spanish who studied English as a foreign language. Surprisingly, the "late starters" showed better perception performance than the "early" ones. Moreover, in Yamada (1995), although the Japanese speakers with a younger AO of English study performed better in the perception of /ɹ/-/l/, nevertheless a *t-test* failed to prove the statistical significance of AO as a function on the subjects' perception performance.

Regarding L2 speech production, Tahta et al. (1981) examined the degree of foreign accent among 109 immigrants in UK. According to the results, only those who were exposed to English before 6 years old were assessed to be "accent-free", or native-like. Similarly, in Flege et al. (1995), Italian speakers who commenced L2-English study from 3 to 13 years of age produced English final /t/-/d/ as accurately as native English speakers did, whereas those who started L2-English learning between 15 and 21 years old did not produce the contrast in a native-like manner. Flege, Yeni-Komshian, and Liu (1999) reported that L1-Korean of L2-English speakers' degree of foreign accents increased as their age of arrival in United States increased. However, in Bongaerts, Summeren, Planken, and Schils (1997), L1-Dutch speakers were found to have achieved a native-like English accent, despite not commencing L2-English study until after 12 years of age. Similarly, Birdsong (2007) examined the French vowel length and VOT of stops produced by 22 late L1-English learners (AO≥18 years old). Acoustic measurements showed that 2 of the subjects' performance fell in the range of native-like levels. In Fullana and Mora (2008), acoustic data obtained from the L2-English subjects' production results failed to yield significant differences as a function of AO on their production performance. Even in Flege, Takagi, and Mann (1995), there were several individual L2 learners, whose AO was after puberty and who produced the voicing contrasts /p/-/b/ and /k/-/g/ within the native English speaker range. Therefore, the debate continues concerning the role of AO on the acquisition of L2 speech sounds.

In addition, the validity of the CPH may be compromised due to the fact that no consensus has been achieved regarding when the "critical period" or "sensitive period" ends. Although the end of puberty is typically defined as the end of the “critical period”, when this is exactly varies from one version to another e.g. 9 years old (Penfield and Lamar, 1959), 12 years old (Scovel, 1988), 11-14 years old (Lenneberg, 1967), and 15 years old (Patkowski, 1990). Furthermore, it was found that the “critical period” may end at different ages in different linguistic domains. For instance, in speech perception, language-specific biases are found to begin from infancy, develop through childhood, and become drastic in adults (Best, 1994). Therefore, instead of during puberty, language learners may have lost sensitivity to L2/non-native speech sounds in the first year of life (Best and Tyler, 2007). In speech production, however, it was hypothesized that if language learners commence L2 learning later than 15 years of age, few of them can manage to speak the L2 without a detectable foreign accent (Oyama, 1976; Flege and Fletcher, 1992). Moreover, the “critical period” for phonology is found to end sooner than that of morphology or syntax (Long, 1990; Hurford, 1991). Due to this disagreement, it would be difficult to decide whether an L2 learner’s AO was before or after the “critical period”/ “sensitive period”.

Furthermore, the age factor may display a negative influence on L2 learners’ acquisition of L2 sounds. This is caused by cognitive aging, which includes gradual decline of working memory, executive control, speech sound processing, and the inhibition of task-irrelevant information (Hakuta, Bialystok, and Wiley, 2003). The biological aging process in the brain is predicted to start at 20 years old, which may result in the progressive loss of cognitive functions related to linguistic performance (Birdsong, 2005, 2006, 2007). On this point, younger learners may have more advantages than older ones in L2 learning. However, compared with children, adult learners are predicted to have the advantage of being mature in cognitive ability, which can benefit their L2 learning (Taylor, 1974; Ausubel, 1964).

Another advantage of younger learners over older ones comes from the influence of L1 experience on L2 learning (see the discussion in section 2.5.2 below). Since being younger means they would have comparatively less L1 experience than older ones, younger learners may suffer less L1 interfere on the perception and/or production of L2 speech sounds (Best, 1994, 1995a, 1995b; Best and Tyler, 2007; Flege, 1981, 1987, 1988, 1989, 1991a, 1992a, b, 1995a, b, 2003).

2.5.2 The influence of language learners' L1 on L2 perception and production

Language learners' L1 experience would be another factor that may significantly influence their perception and production of L2 speech sounds.

Regarding L2 speech perception, for instance, young infants (usually in the first half-year) have been found to be highly sensitive in the perception of non-native speech sounds. Adults, however, show less sensitivity in the perception of speech sounds that are not from their L1 (Best, 1994, 1995a, 1995b; Best and McRoberts, 2003). This is attributed to the interference of adult learners' L1 experience (Best, 1994, 1995a, 1995b; Best and Tyler, 2007), which is also called "deafness" by Sebastian-Galles (2005). For instance, Japanese speakers are found to often have difficulty in distinguishing English /ɹ/-/l/ (Best and Strange, 1992). This is typically explained by the non-occurrence of /ɹ/-/l/ in Japanese (Sebastian-Galles, 2005). Moreover, speakers of different L1s are reported to perhaps depend on different acoustic cues in the identification of non-native speech sounds. Iverson et al. (2003) examined Japanese, German and American adult subjects' underlying perception spaces of English /ɹ/-/l/. German and American adult subjects were found to be dependent upon the trajectory of F3 in the identification of the contrast, which is the crucial acoustic cue for its categorization. Japanese adult subjects, however, were revealed to be more sensitive to the trajectory of F2, which is an irrelevant acoustic cue for the categorization of the contrast. Another piece of evidence comes from Flege and Hillenbrand (1986), in which Swedes and Finns were found to be able to identify tokens in a continuum from /pis/ to /piz/, despite their L1 having no /z/. Nevertheless, they only used vowel duration in the perception of the tokens, whereas native English speakers employed the vowel context and the fricative duration of /s, z/ as cues. Similar findings are reported by Bohn (1995), Flege et al., (1997), Bradlow (1995), Fox, Flege, and Munro (1995), Gottfried and Beddor (1988). Moreover, it was revealed that the identification of boundaries and discrimination peaks differed among speakers with different L1 experiences (Lisker and Abramson, 1967; Abramson and Lisker, 1970; Elman, Diehl, and Buchwald, 1977; Williams, 1977). Concerning the explanation of these findings, Werker and Tees (1984) note that listeners' language experience is likely to "maintain and perhaps enhance natural boundaries that coincide with phonemic boundaries and to downgrade natural boundaries that are linguistically non-functional" (Diehl et al., 2004). This hypothesis is in accordance with McAllister, Flege, and Piske's (2002) "feature" hypothesis. That is, features of the L2 that are not used to signal L1 phonological contrasts may pose difficulties for L2 learners' perception of L2 speech sounds based on these features, which consequently would be shown in their production of these sounds (McAllister et al., 2002; McAllister, 2007).

With regard to L2 speech production, incorrect pronunciation of consonants and/or vowels in part cued foreign accent (Flege, 1995a). Individuals who commence L2 learning at about 7 years old are found to be able to speak an L2 without a detectable foreign accent. However, late learners, such as those whose AO of L2 learning is after 15 years of age, seem unable to speak an L2 with a native-like proficiency level (Oyama, 1976; Flege and Fletcher, 1992). This is attributed to the fact that phonetic inventories vary across different languages. As a result, listeners' sensitivity to speech sounds is attuned to the phonetic inventory of their L1, which reduces their sensitivity to L2 speech sounds (Logan et al., 1991; Lively et al., 1993; Pisoni et al., 1982, Best, 1994a, b). Meanwhile, due to L1 interference, language learners may be unaware, or even be ignorant of certain properties of L2 speech sounds which are phonetically important for the production of the sounds (Yamada and Tohkura, 1992). Additionally, L2 speakers of different L1s are found to realize the same L2 sound as different sounds from the phonetic inventory of their L1. In the production of fricative dental /θ/ and /ð/, for instance, Japanese speakers are observed to replace them with /s/-/z/ (Picard, 2002); Italian speakers are found to realize them as either /s/-/z/ or /t/-/d/, depending on their English proficiency level, and Russian speakers are reported to be likely to realize /θ/ as /t/ (Flege, 1995a, b).

In addition, given the fact that the rules of syllable organization vary from one language to another, language learners of different L1s may have different syllable-processing strategies (Cutler, Mehler, Norris, and Segui, 1983, 1986; Flege, 1989; Flege and Davidian, 1984; Flege and Wang, 1989). L2 learners of different L1s may have different difficulties in perceiving and/or producing L2 speech sounds in different syllable positions (Sheldon and Strange, 1982; James, 1988; Wieden, 1990). For instance, Morosan and Jamieson (1989) revealed that perception training on word-initial allophones did not benefit the subjects' ability to perceive the allophones in medial or final syllable positions, which may suggest that language learners learn L2 sounds "syllabically". Flege (1989) reported that Chinese subjects achieved near-perfect rates in the identification of unedited English words with /t/-/d/ in word-final position, whereas they had poor perception performance when the final release bursts of the stimuli were removed. This was explained by the non-occurrence of /t/-/d/ in word-final position in Chinese. Moreover, L1-Romance speakers are reported to have potential difficulty in the perception and production of voiced English consonants in word-final position. This can either be ascribed to the non-occurrence of word-final consonants in the learners' L1 (Spanish and Italian), or the lack of voiced consonants in word-final position (as in Catalan). Specifically, these speakers may either delete the word-final consonants, or devoice the final voiced consonant (Fullana and Mora, 2008). Similarly, in Bada (2001), Japanese speakers are reported to have difficulty in the production of devoiced /d/ in word-final position, which is attributed to the non-occurrence of devoiced /d/ in word-final position in Japanese. Flege and Davidian (1984) note that most native English speakers' productions of voiced /b, d, g/ and voiceless /p, t, k/ in word-final position were heard as the intended sounds. In contrast, a few of the Spanish and Taiwanese speakers omitted to produce these stops, and more than one-third of their /b, d, g/ tokens were devoiced. This is explained by the phonological differences between English and Taiwanese. Specifically, in Taiwanese, /p, t, k/ are permitted to be in word-final position, whereas /b, d, g/ are not allowed to be in word-final position (Cheng, 1968). In Spanish, however, voiced stops are typically devoiced in "utterance-final" position (Flege and Davidian, 1984). Moreover, the few word-final stops that do not exist in Spanish words were found to be omitted by the Spanish speakers (Harris, 1969).

Following the experimental findings discussed above, several models/theories which investigate how L1 phonetic systems interfere with language learners' perception and production of L2/non-native speech sounds are discussed below.

2.5.2.1 Contrastive Analysis Hypothesis (CAH)

Lado's Contrastive Analysis Hypothesis (CAH) systematically explores the influence of learners' L1 on L2 acquisition. It rests on a comparison between the learners' L1 and L2 systems (Lado, 1957). As Ellis (1985) notes, CAH was developed based on the hypotheses of Behaviourism applied to cross-language acquisition. According to behaviourists, the influence of learners' L1 on the acquisition of an L2 shows in the form of *positive transfer*, *negative transfer* and *zero transfer*. *Positive transfer* occurs when language learners' habitual responses in the L1 are similar to the new skills acquired in the L2, which are predicted to be able to facilitate their L2 acquisition. By contrast, *negative transfer* occurs when the habitual responses of language learners' L1 are contrary to that of the L2. This is predicted to hinder language learners' L2 acquisition. *Zero transfer* occurs when the habitual responses of language learners' L1 has no relationship with that of the L2, and this is also predicted to pose difficulty in L2 acquisition (Stockwell and Bowen, 1983; reviewed by Ellis, 1985). Based on these hypotheses, CAH predicts that the similarities between learners' L1 and L2 systems facilitate their L2 learning, whereas the differences pose difficulties for their L2 learning. CAH is divided into a *strong* and *weak form.* The *strong form* serves to predict the potential errors/difficulties that may occur during language learners' L2 learning, whereas the *weak form* helps diagnose the errors/difficulties that appear during their L2 acquisition (Wardhaugh, 1970).

CAH is supported by findings from early studies on SLA (Robinett and Schachter, 1983; Banathy, Trager, and Waddle, 1966; Berger, 1952; Lado, 1957), and studies on L2/ non-native speech perception and production in recent years (e.g., Shih and Kong, 2011; Zhang et al., 2012). For instance, due to the lack of retroflex fricatives in the phonetic inventory of Taiwanese, Guoyu-Taiwanese bilinguals have been shown to have difficulty in the acquisition of alveolar vs. retroflex fricatives (Shih and Kong, 2011). Native Cantonese speakers were found to have difficulty in distinguishing Mandarin alveolar from retroflex affricates, and the aspirated alveolar from retroflex affricate contrasts. This was explained by the fact that these sounds do not exist in Cantonese (Zhang et al., 2012). Similarly, Tutatchikova (1995) reported that native English speakers are likely to produce Mandarin alveolar-palatal fricatives as post-alveolar fricatives, and Mandarin retroflex fricatives as alveolar fricatives, because of the non-occurrence of these sounds in English.

However, CAH has also been criticized from different perspectives, especially with regard to the *strong form* hypothesis. Weinreich (1953) and Wardhaugh (1970) argue that phonetic systems are unique in individual languages, and so they cannot be compared with each other. Even though some of the phonological rules are similar or the same across the two languages, the combination of these rules in one language may vary from that in another. Regarding the hypothesis that the differences between language learners' L1 and L2 lead to difficulties in their L2 learning, Flege's (1995a, b) Speech Learning Model (SLM) holds the opposite view. SLM suggests that the more similar the phonetic features between learners' L1 and L2, the more difficult they may find it to acquire the L2 sounds. For instance, contrary to findings in Tutatchikova (1995), Chang, Yao, Haynes and Rhodes (2011), found that L1-English speakers could distinguish Mandarin retroflex and alveolar-palatal fricatives. Furthermore, L1-English speakers were shown to be likely to produce alveolar fricatives as post-alveolar fricatives, despite the fact that the English phonemic inventory contains alveolar fricatives.

Another limitation of CAH that ought to be pointed out is that even if the systems of two languages can be compared in a specific linguistic domain, CAH does not provide us with a specific standard based upon which the degree of similarity or difference between language learners' L1 and L2 can be evaluated. Furthermore, the extent to which the hypothesized "difficulty" or "facilitation" are predicted to result from the differences/similarities between the learners' L1 and L2 is not specifically quantified.

Nonetheless, the *weak form* of CAH is supported by some studies. Wardhaugh (1970) points out that the *weak form* contributes to the analysis of the errors of L2 learners' performance in L2 acquisition. Some studies conducted by Ritchie (1968) and Carter (unpublished) may provide supporting evidence for this prediction. Specifically, Russians are found to be likely to pronounce *think* as *tink*, whereas L1 French speakers tend to substitute *think* with *sink.* Through the comparison of the phonological systems of Russian, French and English, it was found that the lack of /θ/ in Russian and French phonological systems might be the reason. However, this explanation would be controversial considering that Russian and French substitute /θ/ with different sounds.

On the whole, CAH is controversial, especially concerning its *strong form*. Its *weak form*, however, as suggested by Wardhaugh (1970), is arguably helpful for the explanation of language learners' errors or incapability in the learning of an L2.

2.5.2.2 Perception Assimilation Model-L2 (PAM-L2)

The Perception Assimilation Model-L2 extended the hypotheses of the Perception Assimilation Model (PAM) (Best, 1994, 1995a, b). PAM explores the influence of language learners' L1 on their perception of non-native speech sounds. According to PAM, being influenced by their L1, language learners tend to assimilate unfamiliar non-native speech sounds to the most articulatorily-similar sounds of their L1 phonetic inventory. Listeners' success in the identification of L2 speech sounds is attributed to their discovery of the articulatory gestures of the sounds (Best, 1994, 1995a, 1995b). PAM predicts, whether language learners' L1 inhibits, aids, or does not affect their discrimination of L2/non-native speech sounds depends on how the target non-native sounds relate to the corresponding L1 sounds in terms of articulatory gestures (Best, 1994, 1995a, 1995b). Based on these predictions, PAM-L2 examines the influence of L1 experience on the perception of L2 speech sounds. An important prediction of PAM-L2 is that L2 learners, even adults, can learn to perceive L2 speech sounds, but the level of success may vary depending on the assimilation between the L1 and L2 sounds following one of the four possible outcomes (Best and Tyler, 2007):

1. Only one phonetic category will be permanently assimilated as an equivalent to listeners' L1 phonetic category. L2 listeners are predicted to have minimal difficulty in the perception of L2 sounds, because they tend to equate an L1 sound to a correlated L2 sound on phonological level, despite the two sounds being phonologically different. For instance, L1-English listeners are frequently found to equate French /r/ with English /ɹ/, though the two sounds are different from each other on a phonological level (Best and Tyler, 2007).

2. Two L2 sounds are perceived to be in the same phonetic category, while one is perceived as a better exemplar than the other. It is predicted that with further exposure to a target L2 sound, a new phonetic category can be developed for the initially deviant phone (Best and Tyler, 2007). This prediction is consistent with the hypotheses of SLM and NLM/NLM-e, all of which attach great importance to the amount of input or experience of L2, as discussed in sections 2.5.2.3 and 2.5.2.4.

3. Two L2 speech sounds are assimilated to a single L1 phonetic category, and are perceived to an equal degree as either good or poor. This hypothesis leads to the prediction that the most difficult situation in L2 speech perception would occur in minimal pairs, in which two words are different from each other by only one contrasting sound. If subjects can successfully distinguish one sound from another in a minimal pair, their competence in the perception of the target speech sound would be assumed to be good (Best and Tyler, 2007).

4. No assimilation occurs. This is termed as uncategorised in PAM. That is, the L2 sounds can be perceived without being assimilated to listeners' L1 phonetic categories. In this situation, whether new phonetic categories can be established depends on whether listeners can successfully perceive the two sounds. If uncategorised L2 phones are deviant from one to another within L1 phonetic space, they are predicted to be comparatively easier to be perceived. Otherwise the L2 phones will be difficult to distinguish (Best and Tyler, 2007).

PAM-L2 admits the advantage of child L2 learners over adult L2 learners in the perception of L2 speech sounds, as they have comparatively less L1 experience to interfere with their identification of L2 sounds. However, PAM-L2 suggests that adult L2 learners are neither uniformly poor at perceiving all the L2 sounds, nor incorrigible in the perception of the L2 sounds which they initially have difficulty with (Best, 1994). They are predicted to be able to learn the L2 sounds (Best, 1994, 1995a, 1995b).

2.5.2.3 Speech Learning Model (SLM)

Flege's Speech Learning Model (SLM) (Flege, 1981, 1987, 1988, 1991a, 1992a, b, 1995a, b, 2003) examines the constraints of L1 experience on learners' perception and production of L2 speech sounds. It ascribes speakers' foreign accent in L2 production to their inaccurate perception of the sounds. L2 learners' failure in the perception and/or production of L2 speech sounds is assumed to arise from prior experience (typically this would be their L1 experience), rather than from the loss of neural plasticity in language learning as claimed by CPH. SLM proposed an extensive set of assumptions and hypotheses. Several hypotheses that are closely related to the present study are summarized and discussed below.

1. Learners' capacity in speech learning remains intact throughout their life (Flege, 1995a, b). In this connection, SLM holds a totally opposite view to CPH, which claims that after the "critical period", language learners will be unable to achieve a native-like proficiency level in L2 acquisition. SLM agrees that language learners' age plays a critical role in the acquisition of L2 speech sounds. Specifically, with the increase of age, language learners' L1 experience increases accordingly. Consequently, L1 phonetic segments become more and more powerful "attractors" of L2 phonetic segments (Flege, 2003). However, L2 learners are predicted to be able to eventually create new L2 categories if given sufficient L2 input. Numerous studies on L2 speech perception and/or production training have illustrated this view (e g., Hazan, Sennema, and Faulkner, 2005; Iverson and Evans, 2009). On this point, it is congruent with the hypothesis of PAM-L2, which also predicts that even adult L2 learners can eventually learn L2 speech sounds in a native-like manner.

2. The perceived similarity or phonetic space between L1 and L2 phonetic categories determines whether a new L2 category can be formed (Flege, 1987, 1995a, b). SLM predicts that the accuracy of language learners' production of L2 sounds varies over time as a function of their perceived relation between sounds in their L1 and L2

phonetic inventories. In the production of sounds which are perceived to be similar but not identical to their counterparts in the L1, learners are predicted to realize the L2 sound as the L1 counterpart. In the production of more dissimilar L2 sounds, however, L2 learners may struggle in the early stages in that they may substitute the sound for an L1 sound. However, once they perceive the high degree of dissimilarity, they are expected to perceive and produce the L2 sounds with a high degree of accuracy. Therefore, the greater the perceived phonetic dissimilarity between an L2 speech sound and the closest L1 sound, the more likely it is that L2 learners can distinguish these sounds. This is opposite to the hypothesis proposed by CAH, which claims that the dissimilarities between learners' L1 and L2 pose difficulty for their acquisition of the L2. For instance, English /ɹ/ is phonetically more dissimilar from Japanese /ɾ/ than English /l/. Japanese speakers are expected to perform better in the acquisition of English /ɹ/ than /l/ according to SLM. Yet, they should have better performance in the learning of English /l/ than /ɹ/ according to CAH. Some studies found that Japanese speakers displayed better performance in the identification of English /ɹ/ than /l/ (Sheldon and Strange, 1982; Flege et al., 1995), which confirms the hypothesis of SLM.

Moreover, SLM predicts that if an L2 sound is similar to or identical to its counterpart in the L1, learners may be able to perceive the acoustic differences, but may not be able to use the perceived acoustic differences in the production of the sound. This process is termed "equivalence classification". The L2 sounds which are classified as similar will be assimilated to a diaphone – the sound category which accounts for both the L1 and L2 sounds. Nevertheless, new sound categories in the L2 are predicted to be establishable if the L2 learners are given sufficient input of the L2 sound (Flege, 1987, 1991a, 1992, 1995a, b), which leads to the next hypothesis of SLM.

3. The amount of language experience plays a significant role in language learners' perception and production of L2 speech sounds (Flege, 1981, 1987, 1988, 1991a, 1992a, b, 1995a, b, 2003). Greater L2 experience is predicted to be able to enhance language learners' capability in perceiving and producing L2 speech sounds. SLM predicts that L2 speech learning is a long journey, which requires a large amount of native-speaker input to be successful (Flege, 2003). For instance, MacKain, Best, and Strange (1981) found that inexperienced Japanese subjects performed on a near-chance level in the perception of English /ɹ/-/l/ contrasts, while those who lived in the USA for 28 months and with 55% daily use of English performed on a native-like level. This finding is consistent with that presented in Bohn and Flege (1992). Moreover, Mortreux (2008)

investigated L1-French speakers' production of English /t, d, n/. The three sounds are typically produced as labial-dentals in French, but apical-alveolars in English. Acoustic and articulatory data revealed that advanced learners showed a shift in the production of the sounds in French and English, whereas the beginners produced the sounds similarly in both languages. However, some studies provide counterevidence to this hypothesis. In Fullana and Mora (2008), for example, exposure to English failed to show a significant effect on L1-Catalan and L1-Spanish speakers' competence in perceiving and producing voiced English contrasts in word-final position. Moreover, Munro (1993) examined L1-Arabic adults' production of English vowels. The goodness rating and acoustic measurements revealed that most of the subjects produced the vowels differently from native English speakers, although they had lived in the USA for 1-27 years. In the meantime, the subjects' length of residence in the USA was found to be non-significant for their accuracy in the production of the English vowels.

4. L2 speech perception precedes its production. L2 speech sounds cannot be produced accurately unless they are perceived accurately. Thus, L2 speech sounds can be produced only as accurately as they are perceived (Flege, 1987, 1995, 2003). This hypothesis has been disproved by findings in some previous studies. For instance, Sheldon and Strange (1982) and Goto (1971) reported that some Japanese subjects are able to produce identifiable /ɹ/-/l/ tokens, even though they cannot identify them from native English speaker's production. Similarly, Yamada, Strange, Magnuson, Pruitt, and Clarke (1994) reported that some subjects' ability to produce English /ɹ/-/l/ exceeded their ability to perceive them.

On the whole, Flege's SLM provides the present study with further theoretical evidence regarding the influence of language experience on language learners' acquisition of L2 speech sounds.

2.5.2.4 Native Language Magnet Theory (NLM)/NLM-e

The Native Language Magnet theory (NLM), which is also known as Neural Commitment theory (Kuhl, 1992, 1993, 1994), examines the constraint of language learners' early L1 experience on their perception of L2 speech sounds. NLM is based on the hypothesis of Kuhl's Perception Magnet Effect (PME). PME suggests that "linguistic experience alters the perceived distances between speech stimuli", which "warps" the perception space underlying speech, and results in the formation of "mirror the phonological categories of the ambient language" (Kuhl, 1994). A 'prototype', which is a key term in PME, refers to the best and most representative instances of phonetic categories and serves as a perception magnet for other sounds in the category. A prototype usually attracts its surrounding sounds in terms of pulling other members of the category toward it. This leads to L2 listeners' difficulty in discriminating the prototype from its surrounding sounds. Non-prototypes (poor instances of categories) do not have this function. Moreover, perceived distances between phonetic categories differ from one language learner to another, which results in the formation of different perception "maps" in L2 speakers' minds (Kuhl, Williams, Lacerda, Stevens, and Lindblom, 1992; Miller, 1994).

Based on the hypotheses of PME, NLM posits that infants' ability to discriminate speech sounds becomes increasingly committed to their native language with the increase of age. In other words, their perception "map" is tuned to their L1 as they begin to develop prototypes of L1 speech sounds at an early stage (i.e., before 6 months old). Future learning is predicted to be greatly affected by the initial mapping of speech sounds. The tuned "map", therefore, poses difficulty for their perception of L2 speech sounds later in life (Kuhl, 1994; Kuhl et al., 1992). Moreover, the effect of L1 interference is predicted to be progressively stronger as language learners' L1 experience increases. In this connection, it is congruent with the hypotheses of PAM-L2 and NLM, all of which attribute language learners' failure in the perception/production of L2 sounds to the constraints of their L1 experience, rather than to the loss of neural plasticity as claimed by CPH.

Kuhl, Conboy, Coffey-Corina, Padden, Rivera-Gaxioda, and Nelson (2008) expanded NLM into NLM-e with five guiding principles. Figure 2.2 presents the overview of NLM-e (adapted from Kuhl et al., 2008):

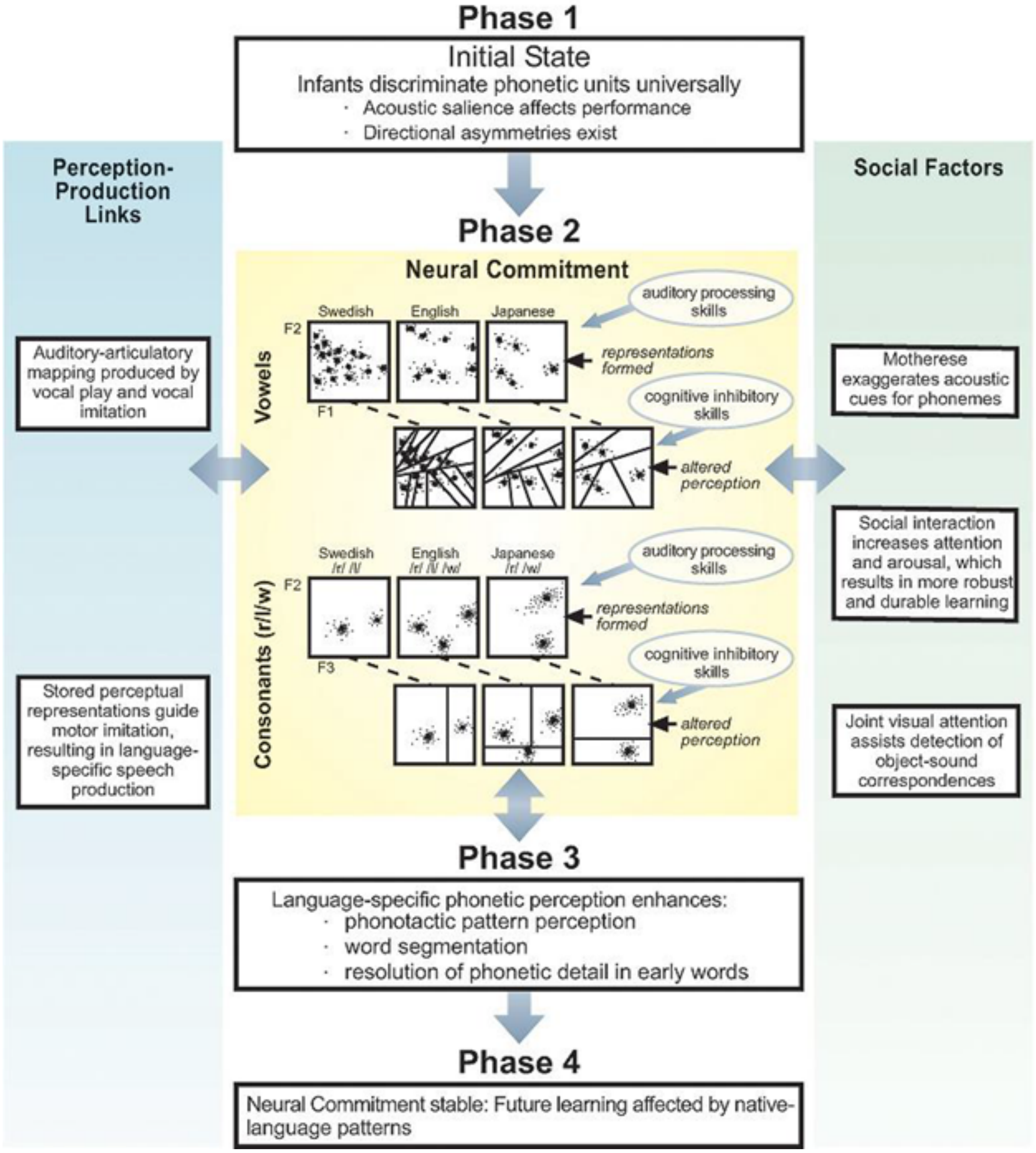

Phase 1
Initial State
Infants discriminate phonetic units universally
· Acoustic salience affects performance
· Directional asymmetries exist
Perception-Production Links
Social Factors
Phase 2
Neural Commitment
Swedish
English
Japanese
F2
F1
Vowels
auditory processing skills
representations formed
cognitive inhibitory skills
altered perception
Auditory-articulatory mapping produced by vocal play and vocal imitation
Motherese exaggerates acoustic cues for phonemes
Social interaction increases attention and arousal, which results in more robust and durable learning
Swedish /r/ /l/
English /r/ /l/ /w/
Japanese /r/ /w/
F3
Consonants (r/l/w)
Stored perceptual representations guide motor imitation, resulting in language-specific speech production
Joint visual attention assists detection of object-sound correspondences
Phase 3
Language-specific phonetic perception enhances:
· phonotactic pattern perception
· word segmentation
· resolution of phonetic detail in early words
Phase 4
Neural Commitment stable: Future learning affected by native-language patterns

Figure 2.2 Native Language Magnet theory expanded (NLM-e) (adapted from Kuhl et al., 2008). Phase two includes data from studies on Swedish (Fant, 1973), English (Dalston, 1975; Flege et al., 1995; Hillenbrand, Getty, Clark, and Wheeler, 1995) and Japanese (Iverson et al., 2003; Lotto et al., 2004).The following are the main hypotheses of NLM/NLM-e:

1. “Distributional patterns and infant directed speech are agents of change” (Kuhl et al., 2008). According to this principle, early phonetic perception can be induced by infants’ sensitivity to distributional properties, which is phase 2 in Figure 2.2. Evidence in support of this principle can be found in Kuhl et al. (1992), Maye, Werker, and Gerken

(2002), McMurray and Aslin (2005). In previous studies, adult-directed speech (without exaggerations in producing speech sounds) and infant-directed speech (with exaggeration in producing speech sounds) are compared in the guidance of infants' language learning. As a result, infant-directed speech is proven to be more effective than adult-directed speech (Bernstein-Ratner, 1984; Kuhl, Andruski, Chistovich, Chistovich, Kozhevnikova, Ryskina, and Lacerda, 1997; Burnham, Kitamura, and Vollmer-Conna, 2002; Liu, Kuhl, and Tsao, 2003; de Boer and Kuhl, 2003). Based on this finding, NLM indicates, the best way for adult learners to circumvent L1 constraints on the acquisition of L2 speech sounds is to recapitulate infants' experience of L1 learning. That is, to receive exaggerated L2 input with "multiple instances by multiple speakers, and massed listening experience" (Kuhl et al., 2008; also see review in Flege, 2003).

2. "Language exposure produces neural commitment that affects future learning" (Kuhl, Conboy, Coffey-Corina, Padden, Rivera-Gaxiola, and Nelson, 2008), which is phase 4 in Figure 2.2. It is predicted that neural tissues which relate to language coding change with initial exposure to a language. These changes affect language learners' subsequent ability to learn the phonetic scheme of a new language (Kuhl, 2000a, b; 2004). Compared with adults, infants' ability to learn more than one language, therefore, is due to their un-fully developed neural network. On this point, it is similar to the hypotheses of CPH in relation to the influence of the neural system on learners' acquisition of an L2. However, contrary to CPH, NLM-e hypothesizes that even adults can eventually learn L2 speech sounds.

3. "Social interaction influences early language learning at the phonetic level" (Kuhl et al., 2008). As shown in Figure 2.2, in the initial stage, infants are born with the ability to learn different languages. Social interaction reduces their sensitivity to phonetic cues, which are not available in their surrounding language environment (phase 2). To demonstrate this prediction, Kuhl, Tsao and Liu (2003) examined the influence of social interaction on infants' acquisition of Mandarin. One group of infants was exposed to passive Mandarin materials (television or specially designed audiotape). Another group of infants was not exposed to any Mandarin input. The results showed that the two groups of infants did not perform much differently in the perception of Mandarin speech sounds. The learning of L2 speech sounds is predicted to be similar to infants' acquisition of L1 sounds. Thus it might be possible to speculate that social interaction with L2 speakers could benefit language leaners' acquisition of L2 speech sounds.

4. “The perception-production link is forged developmentally” (Kuhl et al., 2008) (see the left-hand part of Figure 2.2). Infants imitate, and are guided to “match” the sounds they hear with the sounds they produce. The sounds are then stored in their memory. The perception patterns stored in memory serve to guide their production. During the process, language-specific patterns of speech perception are predicted to emerge before that of speech production (Boysson-Bardies, 1993). This prediction is consistent with that hypothesized by SLM, which also emphasizes the significant effect of speech perception on the production of speech sounds.

5. “Early speech perception predicts language growth” (Kuhl et al., 2008). It is hypothesized that infants’ language performance – both native and non-native languages were tested – at seven months predicts their future language abilities. For example, Tsao, Liu, and Kuhl (2004) investigated 28 6-month-old infants’ discrimination of /y/-/u/ with a head-turn task. The infant subjects’ language abilities were measured again at 13, 16 and 24 months of age, and the infants’ language perception ability at 6 months positively correlated to their later language outcomes over the next 18 months.

To sum up, NLM/NLM-e serves as another piece of theoretical evidence for the present study, which further examines the influence of language learners’ L1 on their acquisition of L2 speech sounds.

2.5.2.5 Perception Interference (PI)

Similarly to NLM-e, Iverson et al.’s (2003) Perception Interference (PI) theory also investigates how learners’ early language experience influences their future learning of L2 speech sounds. According to PI, learners’ low-level perception processing is altered by early language experience (typically their L1), which interferes with the formation and adaptability of higher-level linguistic representations, and results in the loss of sensitivity towards non-native speech sounds, specifically in terms of being unable or less likely to perceive critical acoustic cues of non-native speech sounds. This hypothesis is in accordance with the prediction of NLM/NLM-e, which indicates that early language experience makes adults neutrally committed to a particular network structure in language processing. The neural change is predicted to be irreversible in later life. However, it is still possible for learners to learn, or become tuned to L2 speech sounds. This view is the same as that hypothesized by PAM/PAM-L2, SLM and NLM/NLM-e.

Let us take the study of Iverson et al. (2003) as an example again, which was mentioned earlier in this chapter. The acoustic cues employed in the perception of English /ɹ/-/l/ by Japanese, German, and American adults were compared. It turned out that Japanese adults were most sensitive to F2, which is irrelevant to the categorization of the contrast. The German and American adults, however, were found to be sensitive to F3, which is a more critical acoustic cue for the categorization of English /ɹ/-/l/. Iverson et al. (2003) infer that early language experience interferes with language learners' perception of non-native speech sounds in adulthood. The extent to which adults can perceive non-native speech sounds in terms of critical acoustic cues depends on the degree of interference arising from the difference between their L1 and the target non-native speech sounds.

2.5.2.6 Comparative analysis of PAM-L2, SLM, NLM/NLM-e and PI

	PAM-L2 (Best, 1995a; Best and Tyler, 2007)	SLM (Flege, 1995a, b)	NLM/NLM-e (Kuhl, 1991, 1992, 1993b, Kuhl et al., 2008)	PI (Iverson et al., 2003)
Initial learning stage	L1 categories	L1 categories	L1 neutral mappings	L1 categories/cues
Learning mechanisms	Same as in L1	Same as in L1	Same as in L1	Interference with L1 cues
Perception development	Recognition of categories	Creation of new L2 and/or mapping on same L1 categories	New L2 categories	Sensitivity to critical/noncritical cues
Prediction	Speech category assimilation	Speech category formation/merging	Creation of New L2 categories	Acoustic cues of L1 to interfere with L1 cues
Final learning stage	Depending on L1 vs. L2 articulatory differences	Depending on Age of Learning and L2 experience	Depending on L1 vs. L2 experience	Depending on degree of L1 vs. L2 acoustic interference

Table 2.2 Comparison and comparative of PAM-L2, SLM, NLM/NLM-e, and PI (adapted from Escudero, 2005 and Giannakopoulou, 2012).

PAM-L2, SLM, NLM/NLM-e and PI all investigate the constraints of learners' L1 and age on how they perceive/produce L2 speech sounds. Table 2.2 compares and contrasts these models. In the initial stage, learners' L1 categories (PAM-L2, SLM), or neural commitment to L1 categories (NLM/NLM-e, PI) interferes with their learning of L2 speech sounds. PAM-L2, SLM, and NLM/NLM-e all predict that L2 learning mechanisms are the same as those for the L1. PI, however, highlights the interference of L1 cues on learners' perception of L2 speech sounds.

Regarding predictions concerning the stages of perception development, PAM-L2 views the process as the learners' reorganization and assimilation of categories. However, both SLM and NLM/NLM-e predict that there is a procedure in which learners create new categories for the non-native speech sounds through category formation or merging. PI can be viewed as an extension version of NLM-e. According to PI, acoustic cues from the learners' L1 will interfere with their perception of L2 sounds, specifically in terms of adopting cues which are different from those of native speakers in the perception of L2 speech sounds.

Applying this to the present study, the subjects might assimilate /θ/ and /ð/ to the most articulatorily-similar sounds in their L1/L1-dialect (/s/ and /z/ (as predicted by PAM-L2)), or form new categories in the realization of the non-native sounds /θ/ and /ð/ (as predicted by SLM and NLM/NLM-e). Nonetheless, the cues that the subjects adopt in the differentiation of /θ/ and /ð/ might be different from those of native English speakers (as predicted by PI).

With respect to the final learning stage, all the models hypothesize that L2 learners can eventually learn L2 speech sounds, though it is predicted that the learning results may depend on the influence of L1 in different aspects. PAM-L2 notes, the degree of articulatory differences between an L2 sound and its counterpart in learners' L1 plays a critical role in determining the extent to which the learners can successfully learn the L2 sound. SLM views learners' age of L2 learning and the amount of L2 experience as the decisive factors regarding their successful acquisition of L2 speech sounds. Similarly, NLM/NLM-e also predicts that language experience is significant for learners' success in learning L2 speech sounds. However, PI notes that it is the degree of acoustic interference between the phonetic categories of learners' L1 and L2 that plays an essential role.

One of the most significant common hypotheses of these models would be that language learners, irrespective of their age, can eventually learn L2 speech sounds that they initially have difficulty with. Supporting evidence can be found from experimental studies. Pisoni et al. (1982) examined the possibility of altering adults' perception mechanism in stop consonants categorization with a laboratory training approach. The subjects were asked to identify the presented stimuli by deciding which phonetic categories they belong to. It turned out that the subjects' perception mechanism of phonetic categorization was modified by training. Similarly, Jamieson and Morosan (1986) successfully trained Canadian francophone adults to distinguish English /ð/ from /θ/ with synthetic and naturally produced stimuli. Jamieson and Morosan (1986, 1989) and Morosan and Jamieson (1989) managed to increase Chinese speakers' sensitivity in the perception of word-final /t/ and /d/ through training. In Strange and Dittman (1984), L1-Japanese subjects' accuracy in the auditory perception of English /ɹ/-/l/ was also improved after being trained. In Flege (1989), although Chinese speakers of English performed poorly in the perception of /t/-/d/ in word-final position, a non-significant increase in sensitivity to the contrast was found after a small amount of training with feedback. After presenting the subjects with more training trials, however, a slightly larger and more significant effect was obtained.

Aliaga-García and Mora (2009) conducted six 2-hour perception and production training sessions with advanced adult L1-Catalan and L1-Spanish learners of L2-English. The target contrasts were /b/-/p/, /t-/d/, /i:/-/r/, and /æ/-/ʌ/, which are reported to be difficult for Catalan and Spanish learners to distinguish. High Variability Phonetic Training (HVPT) was adopted as the training approach. Various perception and production tasks were carried out in the training process. Identification, discrimination, phonetic transcription and exposure to native speakers' production of the sounds were employed as perception tasks. Imitation, reading aloud, dialogues, and tongue-twisters were adopted as production tasks. After training, the subjects' perception and production performance was significantly improved.

Iverson and Evans (2009) investigated how L1 categories interfere with language learners' acquisition of new vowels. L1-Spanish and L1-German adults were trained to distinguish English vowels. The phonetic inventory of Spanish includes 5 vowels, whereas the phonetic inventory of German contains 18 vowels. After being trained for 5 sessions, the Germans performed better than the L1 Spanish speakers in the discrimination of English vowels. However, after 10 additional training sessions, the

Spanish listeners' performance improved as much as that of the Germans. According to this result, Iverson and Evans (2009) predict that a larger phonetic inventory can facilitate learners' acquisition of new sounds, though it may not be a decisive factor regarding the learners' ultimate L2 learning achievement.

The studies mentioned above provide evidence in support of the common hypothesis of PAM-L2, SLM, NLM/NLM-e and PI. That is, language learners are able to learn L2 speech sounds eventually, even if they initially have difficulty with the perception and/or production of these sounds. Meanwhile, since all the studies include training programmes, which expose the subjects to the input of the target speech sounds, the significance of L2 input in the acquisition of L2 speech sounds is further confirmed.

2.6 The role of articulatory information in speech perception

'Articulatory information', or 'visual codes', refers to the visible articulatory gestures in the production of speech sounds. Articulatory information is found to be helpful in communication when an auditory signal is compromised (Jackson, 1988). Speech sounds have been shown to be perceived more accurately when visual articulatory information is used (Chen, 2001; Hirata and Kelly, 2010). For instance, Bernstein et al.'s (2013) study has shown that, in the perception of paired nonsense words and nonsense pictures, the subjects who underwent audiovisual training in which the articulatory information was provided, showed significantly higher accuracy levels than those who were auditorily trained. On the other hand, the subjects who were not trained had the same degree of accuracy as those who were auditorily trained. Moreover, it was found that when observing speakers' mouth movements, the listeners' auditory cortex was activated even in the absence of speech sounds (Calvert, Bullmore, Brammer, Campbell, Williams, McGuire, and David, 1997). Articulatory information has also been found to be primarily responsible for the activation of listeners' motor system, rather than auditory input (Skipper, Van Wassenhove, Nusbaum, and Small, 2007). It is reported to be able to facilitate hearing-impaired listeners and cochlear implant users' speech perception and comprehension (Grant and Seitz, 1998; Desai, Stickney, and Zeng, 2008). In speech perception, articulatory information has been shown to be helpful for language learners' perception of L2 speech sounds (Sumby and Pollack, 1954; Navarra and Soto-Faraco, 2007). For instance, Best and colleagues' PAM and PAM-L2 rest on the hypothesis that speech perception is realized through the discovery of articulatory gestures. Relevant theories and studies will be discussed in the following sections in support of the important role of articulatory information in speech perception.

2.6.1 The McGurk effect

The McGurk effect illustrates the effect that articulatory information has on speech perception. McGurk and MacDonald (1976) reported that providing adult listeners with a film of a speaker's lip movements of /da/ dubbed on /ga/ resulted in their identification of the syllable /da/. The reverse dubbing process resulted in the majority of listeners' reporting having heard /bagba/. However, the subjects made the correct response to the speech sounds in auditory modality. Thus, when the auditory component of one sound is paired with the visual component of another sound, it may lead to the perception of a third sound (Nath and Beauchamp, 2011).

The McGurk effect was found to exist for learners of different ages (Massaro, 1984; Massaro, Thompson, Barron and Laren, 1986; McGurk and MacDonald, 1976) though children of different ages seem to show the McGurk effect to different degrees (McGurk and MacDonald, 1976). McGurk and MacDonald (1976) conducted an experiment regarding the McGurk effect among children of 3-5 years old, 7-8 years old and adults. According to the results, the child subjects were less influenced by articulatory information than the adult subjects in speech perception. Similarly, in Massaro et al. (1986), child subjects were reported to be less influenced by visual codes than adult subjects in distinguishing /ba/-/da/. Massaro et al. (1986) attribute this finding to the developmental differences between adults and children regarding their sensitivity to visual information. As for infants, they seem to have an advantage compared to older language learners in speech perception, as they do not have the same level of L1 experience as the adults (e.g., Best, 1994). Nevertheless, they were found to show the same level of influence from the McGurk effect as older language learners (Rosenblum, Schmuckler, Johnson, 1997).

Since the McGurk effect is exhibited in language learners of different ages, it might be possible to speculate that visual codes can have an influence on language learners' identification of speech sounds regardless of their age. Moreover, what the McGurk effect reveals is that speech sounds can be best perceived using a bimodal modality (auditory and visual), and may be compromised if one of the modalities is absent (McGurk and MacDonald, 1976).

2.6.2 Lip-reading

The employment of lip-reading in speech perception would be another piece of evidence in support of the critical role of articulatory information in speech perception. Lip-reading is proved to be effective in helping hearing-impaired listeners' understanding of speech (Walden et al., 1977). It is, therefore, speculated to facilitate language learners' perception of L2 speech sounds.

The term "viseme", or "visual phoneme" is usually employed in the description of the features of phonemes concerning their particular facial/oral positions and mouth movements. Visemes represent speech units in the visual domain. A viseme of the same group may differ in manner and/or voicing features, but share the same place of articulation (Jackson, 1988). Therefore, as explained in the McGurk effect, any speech sounds which look the same in terms of visible articulatory information, belong to the same viseme (Fisher, 1968). For instance, phonemes /k/, /ɡ/, and /ŋ/ share the same viseme of velar stop. Nevertheless, other phonetic characteristics of each phoneme underlying one viseme, such as timing, duration, voicing, could be different from one to another. Yet, these characteristics cannot be captured only with visible articulatory information (Chen, 2001). Speech sounds of different visemes are easier to lip-read than those of the same viseme (Owens and Blazek, 1985; Massaro, Cohen, Gesi, Heredia, and Tsuzaki, 1993). Moreover, speech sounds produced at the back of the mouth are comparatively harder to lip-read than those produced at the front of the mouth (Gesi, Massaro, and Cohen, 1992).

Findings from some studies revealed that language learners' ability to lip-read can be improved with training (Gesi et al., 1992; Walden et al., 1977). For example, Walden et al. (1977) carried out 14 hours of concentrated lip-reading training with 31 hearing-impaired adults (with the help of hearing aids throughout the study). "Same-different" judgement and identification tasks were carried out to help the subjects' identification of some speech sounds. The target speech sounds included sounds of the same and different visemes. Training tasks ranged from easy ones (speech sounds of different visemes) to difficult ones (speech sounds of the same viseme, such as /b/, /p/, /m/). The training both resulted in the subjects' increase in the recognition of the number of visemes and improvement in within-viseme identification.

Moreover, lip-reading training was illustrated to be beneficial for language learners' perception of speech on the sentence level. Based on the results of Walden et al. (1977), Walden et al. (1981) further examined the transferred effect of consonant recognition at the syllable level on the sentence level. There were 3 groups of subjects in their study. The subjects in group 1 received auditory training. The subjects in group 2 underwent lip-reading training. The subjects in group 3, however, neither received auditory nor visual training. The training lasted 7 hours. After that, a two-week aural rehabilitation program was conducted among the subjects in group 1, group 2, and group 3. A perception test with audiovisual sentences as stimuli was conducted. The results, not surprisingly, show that the subjects in group 1 and group 2 performed better than those in group 3. Specifically, in the perception of the target speech sounds at the syllable and sentence level, the subjects' perception accuracy in group 2 improved by 10% and 28% respectively, whereas the subjects in group 1 improved by 7% and 23% respectively.

Massaro et al. (1993) reviewed the two studies and indicated that the training period in both Walden et al. (1977) and Walden et al. (1981) was quite short, only lasting for hours. Additional long-term retention tests may provide more valuable findings. Moreover, the vowel contexts adopted in these two studies only included a single vowel, /a/. In both Walden et al. (1977) and Walden et al. (1981), no evidence is provided in support of the subjects' improvement in the identification of the target consonants in more complex vowel contexts, which may lead to different results. In addition, the subjects' improved accuracy may be the result of repeated testing rather than the effect of training (Massaro et al., 1993).

Nevertheless, some studies in recent years replicated previous findings in support of the critical role of articulatory information in speech perception. For instance, Hirata and Kelly (2010) compared the training results of native English speakers' learning of Japanese vowels in 4 different modalities: "audio-only, audio-mouth, audio-hands and audio-mouth-hands", and found that lip movements significantly assisted the subjects' learning of the vowels, whereas hand gestures did not.

On the whole, studies on lip-reading in speech perception provide evidence in support of the view that articulatory information can assist listeners' perception of speech sounds.

2.6.3 Factors affecting language learners' employment of articulatory information in L2 perception

According to the evidence provided above, articulatory information can facilitate learners' perception of L2 speech sounds. The extent to which this facilitating effect manifests, however, is shown to be mainly dependent upon the language learners' age and the articulatory features of their L1 (Hazan et al., 2005).

The age factor discussed here is different from that mentioned in section 2.5.1 above, in which the younger the learners' AO is, the higher L2 proficiency they may achieve (Liberman, 1957; Oyama, 1976). By contrast, studies from Massaro et al. (1986) suggest that in consonant perception, compared with adult language learners, 6-10-year-olds are less likely to be influenced by visual information than adults. This is because "the sensitivity to certain acoustic cues increases within the first 10 years of life" (Mayo and Turk, 2004).

Regarding language learners' L1, the number of visemes in their phonetic inventory[1] and whether it is a tone language[2] are both factors found to be influence L2 learners' employment of articulatory information in L2 speech perception. It is predicted that L2 learners may lose sensitivity to even salient visual cues that are irrelevant to their L1, just like they lose sensitivity to acoustic cues which do not exist in the phonetic inventory of their L1. Specifically, L2 learners might be able to notice the articulatory difference of L2 sounds, but cannot correlate them with corresponding phonetic labels (Hazan et al., 2005). For instance, Sekiyama et al. (2003) reported that articulatory information displayed the same level of influence on 6-year-old English children as on Japanese speakers of the same age. Nonetheless, developmental visual influence was found among English speakers but not among Japanese speakers. Sekiyama et al. (2003) identify one of the important reasons as the fact that Japanese has a relatively lower degree of articulatory information than English, which leads to Japanese speakers' loss of sensitivity to certain visual cues that do not exist in Japanese. In de Gelder and Vroomen's (1992) study, Chinese speakers displayed a lower degree of usage of

[1] The number of visemes in their phonetic inventory refers to the number of identifiable 'visual categories' which are pronounced with visual movements.

[2] Tone information is not visually observable.

articulatory information in the perception of /ba/-/da/ than Dutch listeners. This is explained by the fact that Chinese is a tone language, and so Chinese speakers rely less on visual cues in speech perception than Dutch speakers (Sekiyama, 1997; Sekiyama and Tohkura, 1993). Likewise, Ortega-Llegaria, Faulkner, and Hazan (2001) found that Spanish speakers did not employ visual cues that disambiguated contrasts which are phonemes in English but allophones in Spanish. Therefore, they concluded that "visual features have different weights when cueing phonemic and allophonic distinctions" (Ortega-Llegaria et al., 2001). More recently, however, Want et al. (2009) found that although both Mandarin and Korean speakers displayed lower accuracy than native English speakers concerning the visual perception of labiodentals, they achieved a native-level of performance in auditory and audiovisual modalities. Nonetheless, in the identification of interdentals, the Mandarin subjects showed poorer performance in auditory and audiovisual modalities, but greater audiovisual-fusion in the perception of incongruent audiovisual materials than the Korean subjects. Thus, it is hypothesized that listeners are able to use non-native visual cues in the perception of non-native speech sounds (Want et al., 2009).

Hazan et al. (2006) identified three types of visual speech categories according to their occurrence in language learners' L1 and L2: (1) a visual category that exists in both the L1 and L2; (2) a visual category that occurs in the L2 but not the L1; (3) a visual category that occurs in both the L1 and the L2, but is used in different phonetic distinctions in the L1 and the L2 (also see Wang Behne, and Jiang, 2009). Hazan, Sennema, Faulkner, Ortega-Llebaria, Iba, and Chung (2006) predict that due to the influence of L1 experience, L2 learners may lose sensitivity to visual categories which do not exist in their L1. Consequently, they may find difficulty for their perception of these speech sounds, specifically in terms of being unable to associate these sounds with their corresponding visual categories. Accordingly, language learners may not have difficulty in the perception of the L2 sounds for which the visual categories occur in their L1. Nonetheless, audiovisual training was predicted and illustrated to be able to facilitate L2 learners' correlation of non-native speech sounds with their visual categories (Hardison, 2003, 2005a, b; Hazan et al., 2005).

2.7 The relation of audiovisual integration, auditory and visual skills in speech perception

The McGurk effect, as discussed above, provides evidence in support of the view that articulatory information is significant in speech perception. However, this may also raise the question of whether audiovisual integration in speech perception is an independent skill involving listeners' ability to process auditory or visual speech codes alone (Ranta, 2010). Findings in previous studies may shed some light on this issue.

Grant and Seitz (1998) hypothesize that audiovisual integration is independent from auditory and visual skills in speech perception. In their study, hearing impaired subjects were presented with auditory, visual and audiovisual stimuli. Both congruent (auditory codes that are synchronized with visual codes) and discrepant (auditory codes that are not synchronized with visual codes) stimuli were employed. It was revealed that subjects relied more on visual information when the amount of auditory input was not enough. Therefore, Grant and Seitz (1998) claimed that the amount of audiovisual integration could neither be predicted from auditory-only nor visual-only performance. This hypothesis is supported by findings from DiStefano (2010), in which the subjects were audiovisually trained to perceive bilabial, alveolar and velar contrasts with degraded stimuli. As a result, the subjects' perception performance was only improved in audiovisual conditions, but not in auditory-only or visual-only conditions. Similarly, both James (2009) and Gariety (2009) conducted auditory training with their subjects. Auditory and audiovisual tests were conducted. Auditory training was found to only improve the subjects' perception ability in the auditory modality but not in the audiovisual modality.

Based on these findings, we might be inclined to agree that audiovisual integration of speech perception is independent from auditory and visual skills in speech perception. However, findings on the neural system of human beings provide counterevidence to this view. For instance, cortical operations are found to be potentially multisensory (Ghazanfar and Schroeder, 2006). Sams, Aulanko, Hämäläinen, Hari, Lounasmaa, Lu, and Simola (1991) reported that visual codes of articulatory information have an entry in the auditory cortex. Similar evidence is available from Calvert et al. (2000), in which magnetoencephalography was used to detect the changes in cortical processing of audiovisual and visual speech stimuli. Congruent (acoustic /iti/, visual /iti/) and incongruent (acoustic /ipi/, visual /iti/) audiovisual stimuli were presented in the audiovisual experiment. Only visual components of these stimuli were presented in the visual experiment. The subjects' auditory cortex was found to be activated bilaterally both in audiovisual and visual experiments. Moreover, Schwartz, Basirat, Ménard, and

Sato's (2012) *Perception for Action Control Theory* views speech perception as a multisensory processing approach in the human brain. It argues that what language listeners' perceive are perceptually shaped gestures, which are called perceptuo-motor units. Perceptuo-motor units are characterised by both the articulatory coherence of gestural nature and the perception value of auditory and/or visual templates. The employment of multisensory modalities in speech perception is further illustrated by Sato, Troille, Ménard, Cathiard, and Gracco (2013). In their study, the synchronization of the silent articulation of a syllable, and concordant auditory and/or visually ambiguous speech stimuli were found to facilitate the listeners' identification of the stimuli. Therefore, we might be able to speculate that, instead of being independent from each other, audiovisual integration is linked with auditory and visual skills in speech perception.

2.8 Other factors affecting L2 speech perception and production

In addition to the influence of language learners' age and L1 experience, there are other factors that may have an impact on their perception and production of L2 speech sounds. Findings from the domain of second language acquisition (SLA) may shed some light on this issue. Several frequently examined factors in SLA are discussed below.

2.8.1 Gender

The gender difference has been shown to have great importance in the SLA research. Although few researchers specifically examined the influence of gender on learners' perception and production of L2 speech sounds, findings from SLA have provided us with valuable evidence. Specifically, female and male language learners are reported to be different from each other in learning styles (Reid, 1987; Powell and Baters, 1985; Kaylani, 1996) and learning strategies (Oxford, Nyikos, and Ehrman, 1988; Oxford, 1993), and as a result, in L2 learning achievement (Asher and Garcia, 1969). Some previous studies found that female learners are better than male learners in L2 learning in terms of maturing earlier, and consequently being more serious about their studies (Clark and Trafford, 1995; Wright, 1999). Female learners are also found to make greater use of strategies in vocabulary learning than males (Catalan, 2003).

However, findings from other studies suggest that gender does not display a significant effect on language learners' acquisition of an L2. For example, in Piske, MacKay, and Flege (2001), the variable gender did not show a significant independent effect on the subjects' production of L2 speech sounds. Similarly, Tercanlioglu (2005) investigated the effect of the gender difference in language learning, and failed to find a significant difference between the performance of males and females. In some other studies, gender was not identified as a significant predictor, particularly in the research on L2 accent (Flege and Fletcher, 1992; Elliott, 1995). One explanation for these findings is that the interaction with other factors, such as AO and amount of L2 experience, may have neutralized the effect of gender (Piske et al., 2001).

Nonetheless, there are some studies in which the male subjects outperformed the females. For example, in Fullana and Mora (2008), the male subjects achieved a higher correctness rate than the females in the perception of voiced English contrasts in word-final positions. Moreover, male subjects have been proven to have higher visual-spatial ability than females by a substantial body of evidence (Bouchard and McGee, 1977; Harris, 1978; Goldstein et al., 1990). In Sanders et al. (1982), for example, the male subjects outperformed the female subjects in the task of mentally rotating three-dimensional arrays of cubes. This advantage of males may benefit them in learning L2 speech sounds in an audiovisual modality.

On the whole, no consensus has been achieved regarding the influence of gender on L2 speech perception and production.

2.8.2 Motivation

Motivation, according to Gardner's (1985) socio-educational model, is an internal attribute of an individual that can be influenced by external forces. A truly motivated L2 learner, as hypothesized by Gardner (1985), should possess 3 characteristics: (1) integrativeness – desire to interact with the target language group; (2) positive attitudes toward learning – can be measured by L2 teachers and L2 courses; (3) positive motivation – the desire to learn the L2. The importance of motivation in SLA can be seen from previous studies. Taylor (1974) suggests that the adults' failure in L2 acquisition is largely due to their lack of strong motivation and a positive attitude towards L2 learning. Moreover, Lenneberg (1967) attributes children's higher phonological proficiency in some studies to their neural flexibility (that is, compared to adults), yet MacNamara (1973) gave a more convincing explanation for this finding. That is, compared with adults, children have stronger motivation to sound similar to their peers, because they hope to be accepted as the same cultural group by their peers. Similarly, Flege (1987) notes that the extent to which a second language speaker sounds like a native speaker is greatly decided by how strongly he or she desires to produce similar sounds to those of native speakers. Accordingly, if older L2 beginners possess a high level of motivation in L2 learning, even if their AO is beyond the critical period, they may achieve a high level of L2 proficiency. As suggested by Marinova-Todd et al. (2000), older leaners' (adults) success in L2 learning is attributed to their high level of motivation. Furthermore, in some studies, motivation was revealed to be a significant predictor concerning the accuracy of L2 pronunciation (Suter, 1976; Purcell and Suter, 1980; Elliott, 1995). Nevertheless, in other studies, motivation was found to be non-significant for language learners' perception and/or production of L2 speech sounds. For instance, in Oyama (1976) and Thompson (1991), no evidence was found concerning the influence of motivation on the subjects' foreign accent in the production of L2 speech sounds.

2.8.3 The amount of time spent on L2 learning

The amount of time that the learners spend on L2 learning is frequently viewed as a decisive factor for their L2 proficiency. Typically, the more time the learners spend in the learning of an L2, the higher the L2 proficiency level they are predicted to achieve (Cumming, 1994; Carroll, 1969). Although Carroll (1969) agrees with the view that beginning L2 learning at an early age is beneficial for learners, yet he attributes this to the fact that compared with older learners, the younger L2 learners have more time for L2 learning. Studies conducted by Flege and his colleague found that the degree of a learners' L2 accent is largely decided by the amount of time that the learner spent in the target language country (Riney and Flege, 1998). Moreover, Purcell and Suter (1980) asked their subjects (L2 English learners) to estimate the amount of time they spent speaking English with native English speakers, and compared it with the length of time of their residence in the USA. The interaction of these two variables turned out to be the third most important predictor regarding the degree of subjects' L2 foreign accent.

However, in Flege and Fletcher (1992), which studied L1-Spanish learners of English, the amount of daily English use was reported to be non-significant for the degree of foreign accent. Likewise, in Elliott (1995), traveling to Spanish-speaking countries and the number of Spanish-speaking relatives displayed little or even no effect on the L2-Spanish speakers' pronunciation of Spanish. In Thompson (1991), L1-Russian learners of English were asked to estimate the amount of time they used English at home, work and with friends. Although the amount of English-language use was found to be positively correlated with the subjects' English accent, it was not a significant predictor in a *multiple regression* analysis.

2.8.4 General factors

In addition to the factors discussed above, individual variances in language learning strategies (Ellis, 1985), cognitive abilities (Skehan, 1998), intelligibilities (Munro and Derwing, 1995) and so on, may also lead to language learners' differences in L2 speech perception and/or production performance. Moreover, language learners were reported to vary significantly in terms of lip-reading skills (Demores, Bernstein, and DeHaven 1996), the degree of sensitivity to visual cues (Sennema, Hazan and Faulkner, 2003), as well as the ability to integrate auditory and visual information in speech perception (Grant and Seitz, 1998). These differences may, in part, explain why in some previous L2/non-native speech perception/production training experiments, subjects of the same or very similar background (e.g. regarding age, gender, L1, L2 proficiency level) performed differently in the post-training test, despite the fact that they had undergone the same training programme (e g., Bradlow et al., 1997; Grant and Seitz, 1998; Hazan et al., 2005; Bernstein et al., 2013).

2.9 Main L2 speech perception training approaches

Both theoretical hypotheses (PAM/PAM-L2, SLM, NLM/NLM-e, PI) and experimental findings (e g., Logan et al., 1991; Lively et al., 1993; Hazan et al., 2005; Bradlow, 2008) predict that language listeners' capability in the perception and production of L2 speech sounds can be eventually improved with sufficient input of the target L2 speech sounds. Therefore, phonetic training, and particularly perception training, may help language learners' acquisition of unfamiliar L2 sounds.

Rvachew (1994) summarized the development of perception training approaches from the early stages to recent years. The early approaches to speech perception training were "ear training" (Van Riper, 1963), in which identification tasks were frequently employed, such as identifying the correct version of the target speech sound from the incorrect ones. "Ear training" aims to guide the subjects to recognize the distinctive elements that define the target speech sounds, and aims to internalize the recognition, with the expectation that the subjects will then produce the trained speech sounds more accurately. “Ear training” was criticized due to the fact that it separates the production from the perception phase. A recently developed proposal for perception training is to restructure the subjects’ underlying phonological contrasts and rules, which usually involves tasks with minimal pairs (Winitz, 1985). Winitz (1985) proposed a sound discrimination training approach. The approach involves presenting the subjects with recordings of naturally produced words, which contrast with each other in terms of distinctive features. This approach, however, was proved to be unsuccessful by Winitz and Bellerose (1967). More recently, two different approaches have been widely used in speech perception and production training. The first one is the High Variability Phonetic Training (HVPT) approach, which suggests the use of naturally produced stimuli (Lively et al., 1993; Logan et al., 1993; Bradlow et al., 1997). In contrast, another approach advocates the use of synthesized stimuli in the training tasks – the Low Variability training approach (Strange and Dittmann, 1984; Jamieson and Rvachew, 1992). Let us have a look at these two approaches.

2.9.1. High Variability Phonetic Training

One of the most frequently employed approaches in speech perception training is High Variability Phonetic Training (HVPT). It directs language learners’ attention towards relevant phonetic cues by providing them with stimuli of high-variability in different phonetic contexts (Lively et al., 1993; Logan et al., 1993; Bradlow et al., 1997). “Natural variability” is the key principle of HVPT. According to Logan et al. (1993), to achieve “variability”, the subjects should be provided with a wide range of stimuli produced by multiple speakers. Using a large number of stimuli is predicted to be more likely to facilitate their perception and/or production performance than a small number of stimuli. Moreover, exposing the subjects to input from multiple speakers of the target language is suggested to be more effective than using stimuli from a single speaker. The design of HVPT aims to teach the subjects which acoustic and/or articulatory cue(s) is (are) reliable for the discrimination of the target speech sounds. Receiving input from a wide range of examples, which are produced by different speakers, is predicted to enable language learners to form robust categories of the target speech sounds (Pickett, 1999). Typically, identification tasks with minimal pairs as the stimuli are adopted by the HVPT training approach (Pisoni and Lively, 1995). With regards to what may be considered “natural”, Logan et al. (1991) indicate that the use of synthetic speech can be misleading, or provide the subjects with incomplete information about the target speech sounds. Therefore, naturally produced stimuli would be a better choice. The HVPT approach was revealed to be successful in some previous studies (Bradlow et al., 1997; Handley, Sharples, and Moore, 2009)

For supporting evidence, let us take the study of Aliaga-García and Mora (2009) as an example again. In their study, the HVPT approach was adopted to train L1-Catalan and L1-Spanish learners of English in the perception and production of /b/-/p/, /t-/d/, /l/-/r/, and /æ/-/ʌ/. The stimuli used in the perception tasks were naturally produced by multiple speakers. The target speech sounds were embedded in multiple phonetic contexts. After six two-hour training sessions, the subjects' performance was significantly improved, both when perceiving and producing the target speech sounds. Similarly, Iverson and Evans (2009) also employed an HVPT approach in their study. The stimuli were English words read by five different British English speakers, two male and three female. Ten sets of minimal pairs were prepared in 4 clusters, which contained the target English vowels, thus yielding a total number of 140 stimuli. The stimuli were produced twice by each speaker. Therefore, the subjects were directed to multiple speakers and stimuli. Not surprisingly, the training results were quite successful. Both the Spanish and German listeners learned the target English vowels with high accuracy. Similarly, Lambacher et al. (2005) used HVPT as a baseline in the perception training of Japanese speakers' perception and production of several American English vowels. After 6 weeks of perception training with identification tasks, the subjects showed significant improvement both in the perception and production of the target speech sounds.

Nonetheless, Iverson et al. (2005) argued that the HVPT may be compromised by being not able to solve the problem of perception interference, such as L1-Japanese speakers' perception of English /ɹ/-/l/. In order to make up for this deficiency, Iverson et al. (2005) combined HVPT with other techniques as complimentary methods in their study, such as Perception Fading (Jamieson and Morosan, 1986), which turned out to be successful.

2.9.2 Low Variability Training

Unlike HVPT, the stimuli used in Low Variability Training (LVT) usually only include one or two minimal pairs of the target speech sounds, which are produced by only one speaker. The stimuli are usually synthesized and exaggerated with a synthesizer. In the training process, typically, the "more exaggerated" stimuli are presented first, followed by the "less exaggerated" ones. The last presented stimuli are usually the naturally produced ones, which are not synthesized. Progressively decreasing the acoustic distance between the stimuli aims to direct the subjects' attention towards the critical acoustic cues, which they are supposed to rely on in the perception of the target speech sounds (Bradlow, 2008).

The LVT approach was successfully adopted by some former studies in speech perception training. For instance, Strange and Dittmann (1984) trained Japanese speakers to perceive English /ɹ/-/l/ with an LVT approach. The subjects were presented with only one minimal pair of the target contrast (*rock-lock*), which was produced by one speaker and was synthesized with a computer. As a result, all the subjects' perception performance gradually improved over 14 to 18 training sessions. McCandliss, Fiez, Protopapas, Conway, and McClelland (2002) also employed an LVT approach to train Japanese adults' perception of English /ɹ/-/l/. In their study, the stimuli were two synthetic continua ranging from *rock* to *lock* and *road* to *load*, which were recorded by one native English speaker. Similarly, in Strange and Dittmann (1984), the subjects were trained with a limited number of synthetic stimuli in the perception of English /ɹ/-/l/. As a result, they achieved a significant improvement in the perception of the target contrast in synthetic speech, despite the fact that no significant improvement was found in the perception of naturally produced words. Nonetheless, as discussed by Bradlow (2008), the post-training test results were limited to the identification of the "trained" stimuli. Therefore, it is possible that the subjects may not be able to identify the contrast if they are embedded in novel words of different phonetic environments.

Both HVPT and LVT approaches are widely employed in speech perception training. Given that both HVPT and LVT have some disadvantages, it would be helpful to consider ways in which they can neutralize or circumvent these disadvantages. The choice of the training approach should be dependent upon the specific purpose of a study. HVPT can be both used in audiovisual and auditory training for speech perception and production (e.g., Hazan et al., 2005; Bradlow et al., 1997), whereas LVT is more likely to be employed in auditory speech perception training (e g., Strage and Dittmann, 1984). Moreover, previous studies on perception training contributed to the choice of training approach in the present study.

2.10 Former studies on perception training

Considering that the present study involves a perception training programme, it would be helpful to review similar studies carried out by previous scholars. Findings in these studies inspired the design of the present study.

Let us first take a look at some previous applications of audiovisual training, which may shed some light on the employment of visual cues in speech perception and production. For example, Hazan et al. (2005) investigated whether L2 learners could be trained to make better use of visual cues in the perception of novel speech sounds with two studies, which also explored whether audiovisual training could be more effective than simply an auditory modality regarding its impact on improving the subjects' perception of the trained speech sounds. In their first study, the subjects were 39 adult Japanese leaners of English. The target speech sounds were English labials /b/-/p/ and labiodental /v/, which were visually distinct from each other. The subjects' capability in the perception of the speech sounds was tested before and after a perception training programme. The tests were carried out in auditory, visual and audiovisual conditions with a 3AFC identification task. 21 of the subjects went through auditory training, while the remaining 18 subjects were audiovisually trained. The test results suggested that audiovisual training could be more effective than auditory training in improving the perception of the target contrasts.

In their second study, 62 adult Japanese speakers who learned English as a foreign language were recruited. Compared with the target speech sounds of the first study (/b/-/p/-/v/), the target contrast employed in this study was less visually distinct, namely /l/-/ɹ/. The subjects' accuracy in the perception of the contrast was auditorily, audiovisually and visually tested before and after a perception training programme. The training programme included 10 sessions of auditory, visual and audiovisual training. The subjects were divided into three groups and respectively experienced the three different training conditions. According to the results, the auditorily trained subjects did not perform better than the audiovisually trained ones, despite the fact that all the subjects of the three groups' accuracy in the perception of /l/-/v/ improved. However, the following test on the production of /l/-/ɹ/ showed that the audiovisually trained subjects achieved greater improvement than the other two groups.

Based on the findings, Hazan et al. (2005) suggested that audiovisual training is more effective than only an auditory modality when the visual cues of the target speech sounds are sufficiently salient. Training the subjects with the visual facial gestures of the speaker can benefit their production of the trained speech sounds, even if the target contrasts are low on visual distinctions. The findings inspired the selection of the target contrasts of the present study. That is, if the target contrasts are saliently different from each other in terms of visible articulatory gestures, the subjects may achieve greater perception and production improvement.

Similarly, Hardison (2003) conducted a perception training programme for L1-Japanese speakers' perception and production of English /l/-/ɹ/. Auditory, visual and audiovisual training effects on the subjects' perception and production of the contrast were compared. In the first experiment, 16 adult native Japanese speakers who learned English as a foreign language were selected (8 were assigned to the group for audiovisual training; another 8 were assigned to the group for auditory-only training). Another 8 subjects were recruited as a control group. The control group only participated in the pre-test, post-test and generalization tests without being trained. The stimuli employed in the test and training were minimal pairs, which contrasted /l/-/ɹ/ in different phonetic positions of various vowel contexts. The stimuli were produced by multiple General American English speakers, and were auditorily and visually recorded. 100 words contrasting /l/-/ɹ/ in 9 different phonetic environments were selected from the perception testing materials for the production test. Stimuli used in the generalization tests were novel words, which were produced by a familiar and an unfamiliar speaker. An identification task was adopted both in the perception training and the tests. The training phase included 15 sessions, with 30 minutes per session.

The overall results indicate that compared with the auditory modality, audiovisual training led to a significantly greater improvement in the subjects' perception and production of English /l/-/ɹ/. The control group showed non-significant improvement from pre-test to post-test. Moreover, speaker difference and phonetic environments were both shown to be statistically significant for the subjects' perception and production performance. Specifically, the subjects received a higher perception score for final singletons and clusters than in initial position. Their particular difficulty was shown in the perception of initial clusters with the vowel contexts /u, o/. In production, however, the subjects obtained higher scores for initial singletons and contexts with /ɑ, ɑɪ/ than in any other phonetic environments. This result was explained by the influence of the Japanese utterance-initial flap. Another significant finding was that the audiovisually trained subjects displayed similar accuracy in the perception of the target contrast produced by the familiar speaker to that by the unfamiliar speaker. The rest of the subjects, however, performed significantly better in the perception of the target contrast produced by the familiar speaker than by the unfamiliar one.

In the second experiment, 8 Korean speakers who learned English as a foreign language were divided into 4 groups: 2 auditory-only training groups who were trained either with multiple speakers or with a single speaker, and 2 audiovisual training groups who were trained either with multiple speakers or with a single speaker. The training and testing materials and procedures were the same as in experiment 1. A control group of 8 Korean speakers participated in the tests without being trained. The overall results were similar to those for experiment 1. That is, the audiovisually trained subjects showed a significantly greater improvement than the auditorily trained group and the control group. An interesting finding comes from the generalization tests. The subjects who received the training from a single speaker displayed comparable levels of accuracy with those who were trained with multiple speakers. Moreover, the Korean subjects' most challenging phonetic environments in the perception of /l/-/ɹ/ were found to be the final singleton with /i, ɪ/. This was attributed to the fact that Korean has a syllable-final non-velarized lateral.

Findings in Hardisonn (2003) further demonstrated the critical role of visual articulatory information in speech perception and production, which is consistent with Hazan et al. (2005). Moreover, as discussed in Hardisonn (2003), phonetic environments and speaker differences both have a significant effect on the subjects' perception/production performance. Thus, it is necessary to consider the influence of the two factors on the subjects' perception/production performance in the present study.

More recently, Lidestam, Moradi, Pettersson, & Ricklefs (2014) compared the effects of audiovisual and auditory-only training modalities on listeners' auditory perception of speech sounds in-noise. 60 adult Swedish speakers were randomly divided between an audiovisually trained, auditorily trained, and non-trained group. The training materials were Swedish consonants in /a/ contexts and monosyllabic words. The audiovisually trained subjects were provided with the speaker's face image, larynx image and auditory production of the stimuli. Only auditory recordings were presented to the auditory-only training group. The non-trained group were only shown a movie clip of the speaker's production. The subjects' ability in the auditory identification of speech sounds in-noise was tested before and after training. As a result, only the audiovisually trained group's accuracy significantly improved. Similar studies were performed by Bernstein, Auer Jr, Eberhardt, & Jiang (2013) and Moradi, Lidestam, and Ronnberg (2013). Although these studies focus on the effect of audiovisual training on the auditory perception of speech sounds in-noise, rather than on listeners' perception/production of L2 speech sounds, they provide us with further evidence regarding the critical role of articulatory cues in speech perception.

Apart from audiovisual training, some studies on auditory perception training also shed light on the design of the present study. For example, Logan, Lively and Pisoni (1991) conducted a phonetic training programme concerning six adult L1-Japanese speakers' perception of English /l/-/r/. In their study, the subjects' accuracy in the auditory perception of the target contrast was individually tested before and after the training programme with an identification task. Three of the subjects were tested twice before being trained, so as to assess whether repeated exposure to the words used in pre-test might lead to improvements in their performance in post-test (i.e. a repeated testing effect). A total of 136 minimal pairs which contrast /l/-/ɹ/ in different phonetic positions were employed as the perception training materials, which were naturally produced by 5 speakers without synthesization. The subjects were asked to identify the stimulus presented from a minimal pair with a 2AFC task. They were given immediate feedback on the correctness of their responses. The training included 15 sessions, with approximately 40 minutes per session. In addition to the post-test, two generalization tests were carried out for 3 of the subjects. This aimed to assess the degree to which the training generalized to novel words.

The overall results showed that all the subjects achieved different degrees of improvement in post-test compared to in pre-test. In addition, several sub-findings of this study inspired the design of the present study. First of all, it seems there was no repeated testing effect in this study. The subjects who were tested twice prior to the training did not show improvement in the perception of the target contrast, although the same stimuli were employed (their mean accuracy was 77.0% in the first pre-test, while it was 76.5% in the second pre-test.).

Secondly, the subjects performed differently in different phonetic environments. Specifically, they showed significantly better performance when the contrast was embedded in word final position and intervocalic position than in initial position (singleton and initial cluster). Moreover, from pre-test to post-test, greater improvement was achieved in the perception of the target contrast in initial cluster and intervocalic environments than in the other two phonetic environments. The subjects' perception accuracy as a function of phonetic environment was also displayed in the training phase. They showed significantly greater accuracy when /l/-/ɹ/ were embedded in final singleton and final cluster positions, whereas lower accuracy in initial singleton, initial cluster positions, as well as in intervocalic position. Given the fact that the subjects of Logan et al.'s (1991) study were L1 Japanese learners of English, it might be interesting to explore whether language learners of other L1 background perform differently when a target contrast is embedded in different phonetic environments. The present study, therefore, takes phonetic environments into consideration in the analysis of the subjects' perception performance.

In addition to the testing results, Logan et al. (1991) also analysed the results from the training phase. It was found that the subjects' perception accuracy improved significantly from week 1 to week 2. Their improvement from week 2 to week 3, however, was not significantly reliable. Based on this result, it would be interesting to have a look at the subjects' degree of improvement, if any, during the training programme. Inspired by this finding, the present study is designed to examine the subjects' perception and production improvement during and at the end of the training programme.

Furthermore, in the generalization tests, Logan et al. (1991) found that the subjects displayed higher accuracy in the perception of stimuli which were produced by a familiar speaker (whose voice the subjects heard in earlier tests before the generalization test) they had heard during the training phase than with a "new" speaker (whose voice the subjects didn't hear in earlier tests before the generalization test), despite the fact that all the stimuli in the tests were novel. This finding is consistent with that in Mullennix et al. (1989), in which the listeners were found to recall lists of spoken words produced by a single speaker more accurately than lists produced by multiple speakers. Logan et al. (1991) indicate that listeners have encoded detailed speaker-specific information in long-term memory, which facilitated their identification of the target speech sounds from the familiar speaker. Therefore, it is best that speakers employed in the production of training materials are different from those in the production of testing stimuli, so as to minimize bias.

Lively, Logan & Pisoni (1993) extended Logan et al.'s (1991) study on training Japanese listeners' auditory perception of English /l/-/ɹ/, which aimed to reveal the significance of variability in perception learning and robust category formation. 6 Japanese speakers who learned English as a foreign language were selected to join their first experiment. The training and testing procedures were identical to those in Logan et al. (1991). The training materials were minimal pairs of /l/-/ɹ/, embedded in initial singleton, initial consonant clusters, and intervocalic positions. They were produced by multiple native English speakers. The subjects' accuracy in the perception of the target contrast was tested before and after 15 days of perception training. Additional generalization tests were carried out at the end of the training programme, which aimed to further detect whether the subjects could generalize the training effect to perceive new words that contained the target sounds. The findings of experiment 1 replicated findings in Logan et al. (1991). From pre-test to post-test, the subjects displayed a significant increase in perception accuracy and a decrease in response time. The same trend was found during the training sessions. More importantly, the subjects generalized to the perception of new words, which were produced by new speakers, without being significantly effected by speaker difference.

The second experiment in Logan et al. (1991) included another 6 Japanese speakers who studied English as their foreign language. The materials used in the pre-test, post-test and generalization tests were identical to those in experiment 1. The training materials, however, embedded /l/-/ɹ/ in five different phonetic environments, and were produced by a single speaker. According to the results, although the subjects' accuracy improved from pre-test to post-test as well as during the training sessions, their performance in the generalization tests with an unfamiliar speaker was not as good as with a familiar speaker. Lively et al. (1993) indicate that this was because the subjects developed "talker-specific, context-dependent representations for new phonetic categories by selectively shifting attention toward the contrastive dimensions of the non-native phonetic categories." Therefore, variability in phonetic environments and speakers plays an important role in phonetic training (Lively et al., 1993). However, as discussed above, the finding in Hardisonn (2003) concerning the effect of speaker variability is at odds with this finding – the subjects who received training from a single speaker generalized to the perception of new words as successfully as those who were trained with multiple speakers. This inconsistency may be caused by the fact that the subjects in Hardisson (2003) were audiovisually trained, while those in Lively et al. (1993) received auditory training only. Visual articulatory cues may have facilitated the subjects' successful performance in generalization tests.

Nonetheless, both the studies by Lively et al. (1991) and Lively et al. (1993) bear the limitation of having a small sample size. Only 6 subjects participated in the training and tests. Some of the effects observed in their studies might be due to the small sample size. In Lively et al. (1993), only 3 subjects participated in the generalization tests. It would be more convincing if a larger sample size had been included in their studies. Furthermore, due to the fact that the subjects lived in the United States for several months, they may have received some exposure to English outside of the laboratory (Lively et al., 1994). Their successful perception performance, therefore, may not be totally attributed to the training programme.

On the whole, the phonetic training programmes discussed above were successful regarding their overall effect on the subjects' perception/production of the target contrasts. Nonetheless, the above studies lack long-term retention tests. It is not clear whether or how long the training effect lasted. However, some studies on phonetic training have addressed this limitation.

For instance, Lively, Pisoni, Yamada, Tohkura, & Yamada (1994) extended the study in Logan et al. (1991) by further testing the long-term retention effect of the training programme. The subjects in this study included an experimental group (19 adult Japanese speakers) and a comparable control group (23 adult Japanese speakers). Subjects of the experimental group participated in the tests and training sessions, whereas subjects of the control group only attended the tests in the study. The target contrast was English /l/-/ ɹ /, which is widely known to be difficult for Japanese speakers to acquire. The stimuli, tasks and procedures adopted in the pre-test, post-test, generalization test as well as training sessions were identical to that in Logan et al. (1991). The study also followed the "natural variability" principle of the HVPT approach (Logan et al., 1991). In addition, two follow-up tests were carried out 3 and 6 months after the training to assess the long-term retention effect. The stimuli used in the tests were new words, which did not occur in the training sessions. Both the experimental group and the control group participated in the long-term retention tests. The overall results indicate that the experimental group showed significant improvement in the post-test compared with in the pre-test, whereas the control group did not. Moreover, the experimental group's performance was maintained at a similar level to that in the post-test. Apart from the overall findings, it is revealing to take a look at some of the sub-findings in this study.

First of all, neither time of test nor the interaction between time of test and phonetic environment was found to display a significant effect on the control group's perception of English /l/-/ɹ/. This finding matched that in Logan et al. (1991). It may demonstrate that the experimental group's improvement from pre-test to post-test and long-term retention tests was not caused by repeated testing. Given that no repeated-testing effect was found in Lively et al. (1994) and Logan et al. (1991), repeated-testing was included in the present study.

Secondly, some of the findings in the study replicated those of Logan et al. (1991). For instance, the experimental group's perception accuracy varied across different phonetic environments. Specifically, the subjects performed better when /l/-/ɹ/ were embedded in final position than in initial and intervocalic positions. Moreover, although all the subjects of the experimental group achieved different degrees of improvement during and at the end of the training sessions, a significant improvement was found between training week 1 and week 2, but not between week 2 and week 3. In addition, the subjects' perception accuracy was found to be significantly higher when new words were produced by a familiar speaker than by an unfamiliar speaker, which suggests the significance of speaker variability in phonetic training.

Regarding the long-term retention tests, 3 months after the training, the subjects' perception accuracy decreased by only 2%. Also, a non-significant decrease was observed in the generalization test. After 6 months without training, the subjects' accuracy still remained 4.5% above the pre-test levels. One of the long-term goals of phonetic training is to develop new phonemic categories for non-native speech sounds which are robust and permanent (Lively et al., 1994). Thus, Lively et al. (1994) suggest that HVPT is an effective means of phonetic training.

The study carried out by Bradlow, Akahane-Yamada, Pisoni, and Tohkura (1999) provides us with additional information concerning long-term retention effects and phonetic training. In their study, 11 adult native Japanese speakers were selected to join the perception training programme. 9 of them returned for the 3-month follow-up test. Another 7 subjects who did not participate in the perception training were recruited to be the control group. The target contrast was the intensively studied English /l/-/ɹ/. The stimuli for training were minimal pairs of /l/-/ɹ/ in different phonetic environments, which were produced by multiple speakers. The training phase lasted for 15 days, with 3 training sessions per day. A 2AFC identification task was carried out in the training sessions. An identification task was employed to detect the subjects' accuracy in the perception of the target contrast in the pre-test, post-test and the 3-month follow-up test. In the production test, the subjects were asked to produce a set of 55 English /l/-/ɹ/ minimal pairs. The target contrast was embedded in a variety of phonetic environments. The subjects' production accuracy was then evaluated by native speakers of General American English. The results showed that subjects in the experimental group achieved a significant improvement both in the perception and production of the target contrast, whereas the control group did not. Moreover, in the 3-month follow-up test, the experimental group maintained a level comparable to in the post-test.

One of the commonalities of the studies reviewed above is that they emphasize the importance of "variability" in the choice of training and testing materials, as well as speakers, which follows the HVPT approach. The findings of the studies indicate that perception training can facilitate learners' acquisition of difficult non-native phonetic categories. The acquired phonetic categories were revealed to be robust, as proven by the results of long-term retention tests.

The studies discussed above are the successful examples. However, there are some studies that were not so successful. For instance, Strange and Dittmann (1984) also trained Japanese speakers to perceive the English liquid consonants /l/-/ɹ/. In their study, 8 female adult native speakers of Japanese were recruited to join the training programme. The stimuli used for training were 10 *rock-lake* synthetic series. The stimulus materials adopted in the test included (1) 16 real-speech minimal pairs that embedded /l/-/ɹ/ in different phonetic environments, which were produced by a male American English speaker; (2) 10 *rock-lock* synthetic series produced by a male American English speaker; (3) 10 rake-lake synthetic series produced by a female American English speaker. Identification and discrimination tasks were carried out in the training sessions. The subjects were given immediate feedback regarding the correctness of their responses. The subjects' capability in the perception of English /l/-/ɹ/ was tested before and after the training phase with identification and discrimination tasks.

According to the results, in the pre-test, the subjects performed best when /l/-/ɹ/ was embedded in word-final position, whereas worst when they were in consonant clusters. This result replicated the findings in Sheldon and Strange (1982), despite the fact that their mean accuracy was fairly low in the perception of the target contrast. During the training phase, the subjects' performance was characterized by gradual improvement across sessions. The greatest improvement was found in the first several sessions. Moreover, 7 out of the 8 subjects showed improvement in the "more demanding" identification and oddity discrimination tasks in the post-test. Regarding perception, 5 out of the 7 subjects also achieved improvement in the identification and oddity discrimination of an acoustically dissimilar *rake-lake* synthetic series. However, none of the subjects showed significant improvement in the perception of /l/-/ɹ/ in natural speech words, which embedded the contrast in word-initial position. On the whole, although there was significant improvement regarding the subjects' perception of the target contrast in synthetic speech, this training programme failed to extend the training effect to natural speech. This may be because the stimuli adopted in Strange and Dittmann (1984) were synthetic singletons of /ɹ/-/l/, which ignored the spectral and durational differences between /l/ and /r/ in different phonetic environments (Dissoway-Huff et al., 1982; Lehiste, 1964).

According to the training and testing materials, the training approach adopted in Strange and Dittmann (1984) seems to have followed LVP. The stimuli were synthetic rather than naturally produced, which may have led to the less successful results compared to other perception trainings (e.g., Hazan et al., 2005; Lively et al., 1993; 1994; Longan et al., 1991). Therefore, the present study, to a large extent, followed the principles of HVPT, which was regarded as more likely to guarantee the successfulness of the study. Considering that the HVPT approach bears the limitation of being unable to solve the problem of perception interference (Iverson et al., 2005), the present study made some changes in the design of the training stimuli (see chapter 5, section 5.3.4 for details).

2.11 Conclusion

This chapter reviewed the research on speech perception and production, as well as L2/non-native speech perception and production. Relevant models and theories, such as CPH, CAH, PAM/PAM-L2, NLM, PME and PI, were discussed with regard to findings in previous studies. The importance of visual codes/articulatory information in speech perception was analysed. Potential factors that may have influenced learners' perception and production of L2 sounds were discussed. Additionally, two of the frequently employed approaches in speech perception training were introduced. Some perception studies carried out by previous scholars were reviewed, which inspired the design of the present study.

Chapter 3 Articulatory and acoustic features of English /s, θ, z, ð/, Mandarin /s/, and CQd /s, z/

3.1 Introduction

So far the literature related to the present study has been reviewed. The models/theories of speech perception and production, which are relevant to this study, were discussed with regard to the findings from previous studies. This chapter primarily discusses the articulatory and acoustic characteristics of English /θ, s, ð, z/, Mandarin /s/, and CQd /s, z/. It aims to examine the degree of similarity and/or difference between English /θ/, Mandarin /s/, and CQd /s/, as well as that between English /ð/ and CQd /z/. First of all, the articulatory and acoustic features of English /θ/-/s/ and /ð/-/z/ are compared and discussed. After that, the phonological systems and phonetic inventories of the subjects' L1 (Mandarin) and L1-dialect (CQd), as well as the articulatory and acoustic features of the target speech sounds in Mandarin and CQd are explored. It is widely supported that language learners' failure in perceiving and/or producing L2 speech sounds is mainly due to the influence of their L1 experience (Lado, 1957; Best, 1994; Flege, 1995a, b; Kuhl, 1991; Iverson et al., 2003). Therefore, it is necessary to examine the phonetic inventories and phonological systems of the subjects' L1 and L1 dialect. Meanwhile, it will also be helpful to examine the articulatory and acoustic properties of the target speech sounds in the subjects' L1 and L1 dialect, so as to compare them with that of English.

3.2 What counts as "similar" and "dissimilar" sounds across the L1 and L2?

On the research on L1 influence on the acquisition of L2 speech sounds, the similarity/ difference between sounds in the L1 and L2 was ascribed great importance regarding the language learners' achievement in the acquisition of L2 speech sounds. For example, as discussed in chapter 2, PAM-L2 suggests that the articulatory similarity/difference between language learners' L1 and L2 plays a critical role in their acquisition of L2 speech sounds. SLM mentioned the "phonetic space" concerning the comparison of whether the sounds in the L1 and L2 are similar or not. Similarly, SLM, NLM/NLM-e and PI investigate the similarity/difference between L1 and L2 sounds from the perspective of acoustic properties. However, CAH as a model that can be applied to different fields of L2 acquisition, only mentioned the words “similar” and “different” without providing us with more specific information on the definition of what counts as “similar” and “different”. Therefore, it seems there is not a unified criteria concerning the definition of “similar” and “different” with regard to L1 and L2 speech sounds, which makes it difficult to decide whether two sounds are similar or different to each other. Nonetheless, since the articulatory gestures and acoustic properties are frequently adopted in the description of the characteristics of speech sounds, it is necessary to examine the characteristics of the target contrasts of the present study, so that we can evaluate to what degree they are similar to or different from each other.

3.3 Articulatory characteristics of English /θ/-/s/ and /ð/-/z/

Consonants are typically labelled in terms of voicing, place and manner of articulation (Pickett, 1999; Ashby and Maidment, 2005). Voicing refers to whether a speech sound is produced with vibration of the vocal cords. Place indicates “the physical place in the mouth where the sound is produced, or the location where the airstream is obstructed in the vocal tract” (Ranta, 2010). Manner depicts the articulators’ physical orientation when producing speech sounds (Ladefoged, 2006; Ranta, 2010). According to the International Phonetic Alphabet chart (hereafter, the IPA, see Appendix 12), /s, z/ are alveolar fricatives, whereas /θ, ð/ are dental fricatives. Meanwhile, /s, θ/ are voiceless, while /z, ð/ are voiced.

Although being defined by the IPA chart, the place of articulation of the four speech sounds may vary across different speakers (Li, Edwards, and Beckman, 2007). For instance, /θ, ð/ are typically described as interdentals, which means the tongue blade rises to in between the upper and the lower teeth (Prator and Robinett, 1985; Wang et al., 2009). Sometimes /θ, ð/ are pronounced as dentals (Taylor, 1976; Ladefoged, 1996). That is, the front of the tongue is placed against the back part of the upper teeth. The articulatory gestures of English /s, z/ are typically described as alveolar (Toda and Honda, 2003). Specifically, the tongue tip is placed against the alveolar ridge with a narrow groove in the tongue directing a jet of air towards the teeth (Ladefoged, 1996). Hence, /θ, ð/ are pronounced in a more frontal area in the mouth than /s, z/. The shape of the tongue in the articulation of /θ, ð/ is flat, whereas it is curved in the articulation of

/s, z/ (Ladefoged, 1996). In order to demonstrate the visible differences between /θ, ð/ and /s, z/ in terms of articulatory gestures, RP speakers (see the Methodology part in Chapter 5 for detailed information about the RP speakers) were asked to produce /θ, ð/ as interdental, and /s, z/ as alveolar. Thus /θ/ and /s/ and /ð/ and /z/ belong to different visemes. The salient visible articulatory differences served as the basis of the audiovisual perception training in the present study.

3.4 Acoustic properties of English /θ/-/s/ and /ð/-/z/

As discussed above, both /θ/-/s/ and /ð/-/z/ are fricatives. Fricatives are produced when the vocal tract is narrowly constricted somewhere along its length. When the air is forced through the constriction, it becomes a turbulent flow. This results in the production of random and noise-like sounds, also known as frication (Shadle, 1990; Wilde, 1995). In spectra, fricatives are characterized by high frequency aperiodic excitation (Wilde, 1995), random fluctuations in amplitude, as well as a broad range of frequencies, which is similar to white noise (Pickett, 1999). Compared with other speech sounds, which can also generate noise (such as stops and affricates), the duration of the noise generated by fricatives is longer. The lengthy interval of aperiodic energy characteristic of fricatives is assumed to distinguish them as a sound class (Kent and Read, 2002).

Moreover, the place of articulation of speech sounds display corresponding acoustic properties. It was revealed that the spectra characteristics of fricatives, specifically the frequency locations of poles and zeroes, depend on the vocal tract configuration and the location of the sound source of the fricatives (Heinz, and Stevens, 1961; Pickett, 1999; Kent and Read, 2002). The main front-cavity resonance has been found to be around 7 to 8 kHz for dental obstruents, whereas around 4 to 5 kHz for alveolar obstruents (Jongman, Wayland, and Wong, 2000; Stevens, 1998). In a number of vowel environments, F2 was shown to be lower for dentals than for alveolars in English (Cao, 2002; Fowler, 1994b; Olive, Greenwood, and Coleman 1993) as well as other languages, such as Malayalam (Stevens, Keyser, and Kawasaki 1986) and O’odham (Dart, 1991).

The voicing feature can also be understood in terms of in acoustic properties. The production of voiced sounds is a procedure of repeated opening and closing of the glottal slit between the vocal cords in the larynx, through which periodic pulses of airflow are produced and which results in periodic sounds. Voiced sounds are differentiated from voiceless sounds in the posture of the vocal cords as well. The vocal cords are held close during the constriction interval for the production of voiced fricatives, whilst they remain wide apart when producing unvoiced fricatives. As a result, less airflow goes through the vocal cords when producing voiced fricatives, while comparatively more airflow goes through the vocal cords when articulating voiceless fricatives. Consequently, voiced fricatives display weaker intensity than voiceless fricatives (Pickett, 1999). Yet, voiced fricatives show relatively greater amplitude than their voiceless counterparts (Kent and Read, 2002).

Ladefoged and Maddieson (1986) classified fricatives into stridents (high intensity fricatives (or obstacle fricatives), such as /s, z/) and non-stridents (low-intensity fricatives (or no-obstacle fricatives), such as /θ, ð/). "Stridency" includes both acoustic and articulatory properties of speech sounds. The quality of "stridency" is achieved "by directing a concentrated jet of air against an obstacle" (Wilde, 1995). According to Chomsky and Halle (1968), stridents are generated by forcing the air stream to go through a complex impediment. Therefore, stridents possess more intense noise energy than non-stridents (Taylor, 1974).

It was found that the relative friction region of non-strident /θ/ spreads out the whole spectra above 1kHz, whereas /s/ displays a friction region of above 3.5 kHz (Harris, 1958). Moreover, stridents present identifiable peaks below the frequency range of 10 kHz in spectra, while non-stridents display more or less flat spectra (Fant, 1960; Flanagan, 1972). Therefore, stridents /s/ and /z/ could be differentiated from non-stridents /θ/ and /ð/ with their frication portion (Harris, 1958; Manrique and Massone, 1981; Heinz and Stevens, 1961). Stridents also show longer duration (Hughes and Halle, 1956) and greater amplitude than non-stridents (Shadle, 1985). In Wilde (1995), the high-frequency amplitude difference between /θ/ and /s/ is found to be approximately 12-18 dB, with a mean amplitude difference of 13.7-19.9 dB when measured at the fricative's midpoint, and 12-21.5 dB when measured at the right edge of the fricative.

However, Shadle (1990) doubts this kind of "simplified" classification, and divides fricatives into those for which (1) the noise is generated from the upstream face of the "obstacle", which includes placing the teeth at approximately right angles to the jet axis, such as /s, z/; and (2), those for which the noise is generated by the jet all along the "wall", such as the hard palate and lips, like /θ, ð/ (Shadle, 1990). The location of the generation of the sound affects the amount of noise that is generated (Shadle, 2012). Typically, the frequency of fricatives is more intense in high frequency areas (above 2.5 kHz) than in low frequency ranges (Pickett, 1999).

Fricatives can also be classified as sibilants (including alveolar and palato-alveolar) and non-sibilants (including labiodentals and dentals) (Ladefoged, 1993). /s, z/ are classified as sibilants, whereas /θ, ð/ are non-sibilants. Sibilants possess more energy at higher frequency ranges than non-sibilants. Meanwhile, non-sibilants display relatively less intensity than sibilants (Wilde, 1995). Moreover, the amount of hissing noise in a speech sound is hypothesized to be a significant cue in differentiating alveolars from dental fricatives (Ladefoged, 1993). Although sibilants are also different from non-sibilants in terms of amplitude, this difference is revealed to be unreliable in distinguishing stridents from non-stridents. For instance, Behrens and Blumstein (1988b) reported that the decrease of the amplitude of the stridents /s, ʃ/ resulted in subjects' increased response of /f, θ/. Nevertheless, the increase of the amplitude of /f, θ/ did not result in the subjects' increased response of /s, ʃ/.

Table 3.1 shows some acoustic properties of English /θ/-/s/ and /ð/-/z/ reported by previous studies. English /θ/-/s/, /ð/-/z/ are different from each other in terms of frequency range, strongest frequency range, amplitude range, inherent duration, vowel transition, relative intensity, relative spectra length, and spectra shape. The specific acoustic features of a speech sound are both speaker-dependant and context-dependent (Ashby and Maidment, 2005; Soli, 1981; Jongman, Wayland and Wong, 2000).

	Frequency range	Strongest frequency range	Amplitude range	Inherent duration	Vowel transition	Relative intensity	Relative Effective Spectra Length	Spectra shape
/θ/	1.5—8.5 kHz (a; b)	around 5 kHz (g)	54 dB (d); 54.7 dB (i); 42—52dB (e)	110 ms (h)	downward F2 (g)	low(b)	Relatively long(b)	Relatively flat spectrum with no clear dominating peak (b, e)
/s/	above 4 kHz (a); 4-6 k Hz (f); 4—8 kHz (c)	at and above 4 kHz (g); 5—8 kHz (c)	65 dB (d); 64.9 dB (i); 57—68dB (e)	125 ms (h)	no vowel transition (g)	high(b)	Relatively short(b)	Well-defined, distinct shape with a primary spectra peak in high frequencies (b, e)
/ð/	Similar as that of /θ/ (a; b)	around and above 5 kHz (g)	66 dB (d); 62.7 dB (i)	50 ms (h)	downward F2 (g)	low(b)	Relatively long(b)	Similar as /θ/ (b, e)
/z/	4-6 kHz (f)	at and above 4 kHz (g)	70 dB (d); 67.7 dB (i)	75 ms (h)	no vowel transition (g)	high(b)	Relatively short(b)	Similar as /s/ (b, e)

Table 3.1 Acoustic properties of English /θ/-/s/ and /ð/-/z/ (a. Hughes and Halle (1956); b. Stevens (1960), reviewed by Kent and Read (2002); c. Manrique and Massone (1981); d. Jongman (1989); e. Behrens and Blumstein (1988a); f. Bitar (1993); g. Pickett (1999: 140); h. From Kent and Read (2002: 182); i. Jongman et al. (2000).

On the whole, English /θ/ and /s/, /ð/ and /z/ both differ from each other in terms of articulatory gestures and acoustic properties. These differences may contribute to the subjects' perception and production of the two contrasts. Their articulatory differences serve as the basis of the audiovisual training in the present study.

3.5 Background information on Mandarin and CQd

Mandarin (originally from Portuguese *Mandarim*) is the official language of the People's Republic of China, and is often called *Putonghua, Chinese*, or *Standard Chinese* (Coblin, 2000; Norman, 1988). It is the national language used in education, the media, formal situations as well as all governmental and official transactions in Mainland China (Bennan and Yang, 2004; Norman, 1988). The majority of educated people in Mainland China speak Mandarin as their L1.

For geographical and historical reasons, people from different areas of China speak different dialects, which differ from one another in terms of pronunciation, vocabulary and/or grammar (Escure, 1997). The dialects in China are mainly divided into minority dialects and Mandarin dialects. Minority dialects are spoken by the minority population, such as Tibetan. Mandarin dialects are spoken by about 70% of China's Han population to the north of the Yangtze River. Mandarin is usually thought to be Pekingese-based, particularly in terms of pronunciation, though historically Mandarin was found to have developed from the dialects of various provinces (Coblin, 2000). The classification of Chinese dialects varies from one schema to another. Here are two of the frequently cited versions. The first one is from Ramsey (1987), which divides Chinese dialects into 7 groups based on provincial boundaries (see Table 3.2). Another widely used version comes from the *Language Atlas of China* (Kurpaska, 2010) (see Table 3.3).

Dialect group	Where spoken
Mandarin	All of North and Southwest
Wu	Coastal area around Shanghai, Zhejiang
Gan	Jiangxi
Xiang	Hunan
Hakka	Widely scattered from Sichuan to Taiwan
Yue	Guangdong, Guangxi (and overseas communities)
Min	Fujian, coastal areas of South

Table 3.2 Seven groups of Chinese dialects (Adapted from Ramsey and Robert, 1987).

Groups		Provinces/Cities
1. Mandarin Dialect Group	a) Northern	Hebei (including Beijing,Tianjin, Henan, Shandong, Jilin, Heilongjiang, Northwestern part of Anhuui, Northeastern part of Jiangsu, Eastern part of Inner Mongolia
	b) Northeastern	Shanxi (including Taiyuan), Shanxi (including Xi'an), Gansu (including Lanzhou), Xinjiang, Ningxia, Qianghai, Western part of Inner Mongolia
	c) Southeastern	Central Jiansu (including Yangzhou), Central Anhui (including Hefei), Southeastern Hubei, Northern Jiangxi
	d) Southwestern	Sichuan (including Chengdu and **Chongqing**), Guizhou, Yunnan, Hubei (except southeastern corner), Northwestern part of Hunan, Northwestern part of Guangxi (including Guilin)
2. Wu Dialect Group		Zhejiang, Southern Jiangsu (including Shanghai, Suzhou, Wuxi), Southeastern Anhui
3. Xiang Dialect Group (=Hunanese)		Jiangxi (including Changsha)
4. Gan Dialect Group (=Jiangxi)		Jianxi (including Nachang), Southern Anhui, Southeastern Hubei
5. Kejia Dialect Group (=Hakka)		Communitites scattered in Saichuan, Jiangxi, Hunan, Guangdong (including Meixian), Guangxi, Fujian, and Taiwan
6. Yue Dialect Group (= Cantonese)		Guangdong (including Canton, Tishan, Zhongshan, Kaiping, Maca, Hong Kong, Southeastern Guangxi
7.Min Dialect Group	a) Northern Min	Southern Zhejiang, Northeastern part of Fujian (including Fuzhou)
	b) Southern Min (+Fukienese)	Southern part of Fujiang (including Xiamen (=Amoy), Northeastern Guangdong (including Chaozhou, Dongshan and Hainan Islands), Taiwan

Table 3.3 Dialect groups and the Provinces/cities in which they are spoken (adapted from Chan, 1987).

The subjects of the present study came from Chongqing China. As shown in Figure 3.1, Chongqing municipality is located in the southwest part of China. According to the classification of Chinese dialects from Ramsey (1987) and Chan (1987), CQd is one of the Mandarin dialects (see Table 3.2) and, more specifically, belongs to the group of Southwestern Mandarin dialects (see Table 3.3). Chinese linguists have classified Mandarin dialects into northern, northwestern, southwestern and eastern Mandarin.

Among them, southwestern Mandarin is spoken in Sichuan and other southwestern provinces (Norman, 1988). CQd is classified as one of the Sichuan dialects.

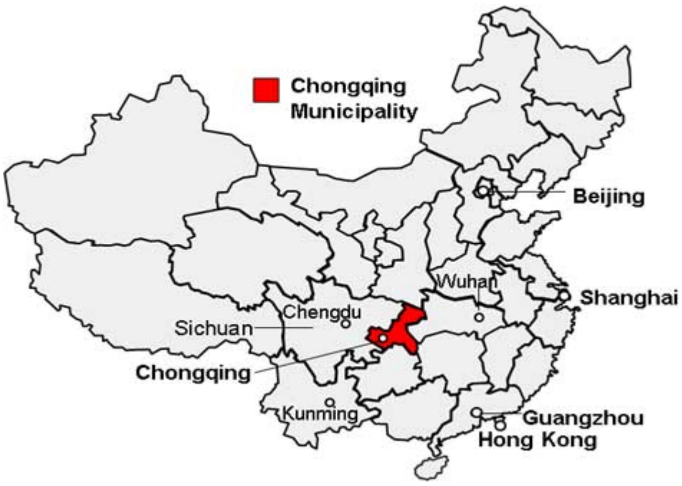


Figure 3.1 Geographical location of Chongqing Municipality (assessed from http://www.google.co.uk/search?q=chongqing+map+chinaandtbm=ischandtbo=uandsource=univandsa=Xandei=Uf7sUZyxBaqf0QWum4D4DQandsqi=2andved=0CC0QsAQandbiw=1466andbih=833. 22/07/2013).

Chinese is a phonologically and morphemically monosyllabic language, in which a syllable is viewed as a self-contained entity, and almost every syllable corresponds to a morpheme (Norman, 1988). A Mandarin syllable consists of an initial (or onset) and a final (see Appendix 9 for Mandarin initials and finals). The initial is the beginning consonant. Some syllables do not begin with an initial, and these are also described as beginning with a zero initial, such as /an/. The final refers to the rest of a syllable, which consists of a medial, a main vowel (the nucleus) and an ending (coda). Among them the medial and ending are optional, but the main vowel is obligatory (Norman, 1988). As a tone language, Mandarin has five tones at the word-level, as described in table 3.4 below (Suen, 1982). Instead of using IPA symbols, *pinyin* (Latin script) is used to transcribe Chinese characters (Ramsey, 1987). For instance, the Chinese character *山,* meaning *mountain* in English, can be transcribed as *shān*. *sh-* is the initial, *-an* is the final of the syllable, and its "high level" tone is represented by the horizontal bar.

Tone	Description	Pitch
1	high level	55
2	high rising	35
3	low rising	214
4	high falling to low	51
5	neutral	5

Table 3.4 Mandarin Tone description.

Dialects of Chinese possess most of the phonological rules of Mandarin pertaining to syllable structure, whereas show different degrees of variation regarding vocabulary and tones (Norman, 1988). Due to the tones and phonological rules of Mandarin dialects not being the focus of the present study, the following sections will mainly focus on the segmental phonetic inventories of Mandarin and CQd, and particularly on the target speech sounds of the present study.

3.6 Comparison and contrast of the consonant phonetic inventories of English, Mandarin and CQd

Given that the difference(s) between language learners' L1/L1-dialect and L2 phonetic inventories may influence their perception and/or production of L2 speech sounds (Lado, 1957; Best, 1994; Flege, 1995a, b; Kuhl, 1991; Iverson et al., 2003), it is needful to compare and contrast the phonetic inventories of the subjects' L1 (Mandarin), L1-dialect (CQd) and L2 (English). Due to the target speech sounds of the present study being consonants, only the consonant inventories of English, Mandarin and CQd are presented below.

	English	**Mandarin**	**CQd**
Plosive	/p/ /b/ /t//d//k//g/	/p/ /pʰ/ /t//tʰ//k/ /kʰ/	/p/ /pʰ/ /t//tʰ//k/ /kʰ/
Nasal	/m/ /n/ /ŋ/	/m//n/[1]//n/[2]/ŋ/[2]	/m//n/[2]/ŋ/[1]/ŋ/[2]
Trill			
Tap or Flap			
Fricative	/f/ /v/ /θ/ /ð/ /s/ /z/ /ʃ/ /ʒ/ /h/	/f//s//ɕ//ʂ//x/	/f//s//x//ɕ//z//v/
Affricate	/tʃ/ /dʒ/	/ts/ /tsʰ/ /tʂ/ /tʂʰ/ /tɕ//tɕʰ/	/ts/ /tsʰ//tɕ//tɕʰ/
Approximant	/w//ɹ//j/	/ɻ/	
Lateral Approximant	/l/	/l/	/l/

Table 3.5 Consonants in English, Mandarin and the CQd ([1] Used as a syllable onset (Initial). [2] Used as a syllable coda. From Qian, Liang and Soong (2009); Cheng (1966); Zhong (2005); Wang and Lee (1994); Edwards (1992); Fang and Ping-an (1992).

As shown in Table 3.5, some consonants are shared by English, Mandarin and CQd. For instance, the plosives /p, t, k/ occur across all three phonetic inventories, despite Mandarin and CQd also including /pʰ, tʰ, kʰ/ as plosives. As for nasals, although they all have /m/ /n/ /ŋ/ in their phonetic inventories, /n/ and /ŋ/ occur in different syllable positions in the three languages. /ŋ/ typically occurs word-finally in English and Mandarin, whereas it can be used as a syllable initial in CQd. Moreover, all three inventories do not possess a trill, tap or flap, and Mandarin and CQd have totally different affricates and approximants, but the same lateral approximant (/l/). The greatest variation can be found in the fricatives of the three languages. That is, only /f, s/ are the common fricatives. Regarding the target contrasts of the present study, English /θ/ and /ð/ neither exist in Mandarin nor occur in CQd. /z/ occurs in the phonetic inventories of English and CQd but not in Mandarin. /s/, however, exists in English, Mandarin and CQd. The following sections of this chapter will discuss the articulatory gestures and acoustic properties of Mandrain /s/ and CQd /s, z/, enabling comparison with these sounds in English.

3.7 Articulatory gestures and acoustic properties of Mandarin /s/ and CQd /s, z/

Due to language-related variation, the articulation of the same speech sound may vary across languages, particularly in terms of articulatory gestures. This may result in variation in acoustic properties (Toda and Honda, 2003). Speaker difference is another factor that may result in the variation of articulatory gestures for the production of the same speech sound (Pickett, 1999).

English /s/, as mentioned earlier in this chapter, is typically pronounced as alveolar (Toda and Honda, 2003) or dental (Taylor, 1976; Ladefoged, 1996), depending on the speaker. The articulatory gestures of /s/ in Mandarin display even more variation. It can be pronounced as dental (Chao, 1948; 1968; Lee, 2011; Tsai and Lee, 2003; Suen, 1982; Norman, 1988), alveolar (Chang, Haynes, Yao and Rhodes, 2009), apical or dental-alveolar (Lee, 1999). For instance, Norman (1988) described Mandarin /s/ as dental, which involves placing the tongue tip against the back of the upper teeth, with a point of articulation forward of the alveolar. Nevertheless, Hu (2008) examined the articulatory gestures of Mandarin /s/ in the vowel contexts /i, a, u/ with EGP and EMA. According to the results, the male speakers articulated /s/ as totally dental, whereas the female speakers articulated it as alveolar. As a native Mandarin speaker, the author personally noticed that Mandarin /s/ can either be pronounced with the tongue tip touching the back part of the upper teeth (dental) or against the alveolar ridge (alveolar). The production of Mandarin /s/ displays a constriction location close to the teeth, yet the exact constriction site is highly variable across speakers (Lee, 2011). Therefore, inter-individual differences may play an important role over language-related differences. In Toda and Honda (2003), 3/7 French speakers', 3/5 English speakers', 3/4 Chinese speakers', and 1/1 Swedish speakers' articulation of /s/ was found to be apical. A French and an English speaker pronounced /s/ as lateral and labio-dental respectively. Nevertheless, the averaged data of the acoustic results are not likely to vary significantly from the basic pattern of a specific speech sound (Toda and Honda, 2003).

Due to the variation of articulatory gestures, the acoustic properties of /s/ in English may vary from that in Mandarin (Davenport and Hannahs, 2010). Chang et al. (2009) compared some acoustic properties of Mandarin alveolar /s/ with that in English. The data was produced by heritage speakers[3] of Mandarin. It was found that the centroid frequency is 6006 Hz for Mandarin /s/, but 6133 Hz for English /s/. For the same stimuli, the majority of the subjects (both male and female) produced English /s/ with a lower mean value for peak amplitude frequency than Mandarin /s/, which ranges from about 100Hz to about 1400Hz. Nevertheless, in a few cases, English /s/ displays the same or even higher peak amplitude frequency than that of Mandarin /s/, which ranges from less than 100Hz to about 1000Hz (Chang et al., 2009). Furthermore, Mandarin /s/ is also found to display the highest energy peak at around 6.5kHz and a lower energy peak at 1.8kHz (Hu, 2008).

Compared with Mandarin, CQd is not widely studied, particularly in the domains of articulatory and acoustic phonetics. The available references indicate that /s/ and /z/ are typically pronounced as apical dental rather than alveolar as in English (Zhong, 2005; Zhou, 2012). As a native speaker, the author also noticed that /s, z/ are typically pronounced as apical dental in CQd, although there might be speaker differences.

On the acoustic level, due to the lack of references, basic acoustic analysis of CQd /s, z/ was performed using the Praat program (Boersma and Weenink, 2013). Six subjects were randomly selected from the experimental group of the present study to produce CQd /s, z/. The carrier sentence was: [pa] + *target word*+ [tu] [ts^{h}eu][laɪ] (11 carrier sentences * 3 repetitions per subject), which means *Read target word out* in English.

[3] The term 'heritage speaker' refers to "speakers who have had exposure to a particular language as a child, but who have shifted to another language for the majority of their communication needs" (Change et al., 2009).

Although there are 8 vowels in CQd, which include /ɿ, ɚ, a, o, e, i, u, y/, the legal combinations between /s, z/ and the vowels in CQd are /su/, /zu/, /so/, /zo/, /sɚ/, /zɚ/, /se/, /ze/,/sɿ/, /zɿ/, /sa/ (Zhong, 2005). Therefore, each subject was asked to read 11 carrier sentences (Appendix 13), 3 times for each sentence. A high quality recorder (Roland-05, with the settings: 16-bit mono channel and 44.1 KHz) was used for recording. The recording was carried out in the same room where the production test of the main study was conducted, with the same recording procedure.

In the recording of the female subjects, the Formant settings were: Maximum formant =5500.0Hz; Number of formants: 5.0; Window length =0.025s; Dynamic range =30.0 dB; Dot size =1.0 mm. The same settings were applied for the recording of the male subjects, except that the Maximum formant was set at 5000.0 Hz. This is because of the difference in the vocal tracts of females and males. Due to the energy of fricatives being mainly located at higher frequency ranges (Wilde, 1995), the view range of the spectrogram was set at 4000 to 10,000 Hz.

Before doing acoustic analysis, TextGrid was conducted in Praat for all the recordings, so that the target speech sounds (CQd /s, z/) would be visible. The frequency range, duration and mean intensity of each subjects' production of CQd /s, z/ in all the carrier sentences were then extracted from Praat. Given that there were only 6 subjects, the acoustic analysis of their production of the two speech sounds was performed mandatorily. Individual subjects' frequency range was obtained with the function "view spectra slice" in the Praat program. The mean intensity of the target sounds was obtained with the function "get intensity" in the Praat program. The duration of the two sounds was viewed from the bottom of the textgridded area of the sounds. All the collected acoustic data for individual subjects were then averaged. Table 3.6 shows the mean results of each subjects' production of CQd /s, z/ in terms of frequency range, duration and mean intensity.

Subject	Frequency range (kHz)		Duration (ms)		Mean intensity (dB)	
	/s/	**/z/**	**/s/**	**/z/**	**/s/**	**/z/**
Female 1	4-7	4-8	162.0	109.8	68.5	64.8
Female 2	4-7	4-8	156.5	115.1	63.2	59.9
Female 3	4-6	4-7	167.8	128.9	60.1	59.2
Male 1	4-6	4-8	131.6	87.5	66.9	64.0
Male 2	4-6	4-8	133.8	96.7	64.4	58.3
Male 3	4-6	4-6	127.3	107.4	63.8	62.9

Table 3.6 Frequency range, duration and mean intensity of /s, z/ in CQd.

As shown in table 3.6, CQd /s, z/ display very similar acoustic properties to English /s, z/ (see table 3.1) in terms of frequency range (from 4 to 8 kHz) and intensity. However, CQd /s, z/ show comparatively longer duration than English /s, z/, though there were speaker differences. Moreover, according to the spectra shown in the Praat program, both /s/ and /z/ in CQd show a well-defined distinct shape with a primary spectra peak at high frequencies, which is similar to that of English /s, z/.

3.8 Summary of the similarities/differences between the target contrasts

The key models (PAM-L2, SLM, NLM/NLM-e, PI), which serve as the theoretical basis of the present study, all attach great importance to the critical role of the similarity/difference between L1 and L2 speech sounds in the acquisition of the L2 sounds. In order to correlate the hypotheses of these models to the present study, it is necessary to evaluate the extent to which the target contrasts are similar/dissimilar to each other. Nevertheless, none of the models provide us with criteria regarding what counts as “similar” or “dissimilar” for the sounds in L1 and L2. As discussed in chapter 2, both articulatory and acoustic properties of speech sounds play significant roles in speech perception and production. The target contrasts' articulatory and acoustic differences/similarities discussed above are summarized and evaluated below, in order to enable us to build up a picture of the similarities and differences between the sounds in the subjects' L1, L1-dialect and L2.

Regarding articulatory gestures, English /θ/, Mandarin /s/ and CQd /s/ are quite similar to each other. Specifically, the three sounds share the same place of articulation if they are all produced as dental. The same situation applies to English /ð/ and CQd /z/, if they are both pronounced as dental. Nonetheless, this is largely dependent on speaker differences. If English /θ/ is produced as interdental as typically produced by native English speakers, then the articulatory gestures of English /θ/ would be distinct from Mandarin /s/ and CQd /s/. The same situation exists for English /ð/ and CQd /z/.

Comparatively, the articulatory gestures of English /s/, Mandarin /s/ and CQd /s/ seem to be less distinctive. They can share the same place of articulation either when produced as dental or alveolar. Nonetheless, English /s/ is typically produced as alveolar by native English speakers, while Mandarin /s/ and CQd /s/ are more likely to be produced as dental. In these circumstances, the places of articulation for /s/ in English and /s/ in Mandarin and CQd would be different. However, given the fact that the alveolar ridge (the place of articulation for alveolars) and the back of the upper teeth (the place of articulation for dentals) are quite close to each other, we might be able to assume that their places of articulation are similar to each other, particularly in comparison with the difference between dental/alveolar (Mandarin /s/, CQd /s, z/) and interdental (English /θ, ð/).

On an acoustic level, as reviewed in section 3.1 and 3.6 above, due to speaker differences and variation in stimulus materials, the acoustic data for the target contrasts varies slightly from one datum to another. Approximately, English /s, z/ shows quite similar acoustic characteristics to /s, z/ in Mandarin/ CQd in terms of frequency range, intensity and spectra shape, though with relatively shorter duration. However, /θ/ and /ð/ display quite different acoustic characteristics to /s/ and /z/ in English and Mandarin/ CQd (i.e. regarding frequency range, duration, intensity, and spectra shape). To sum up, the acoustic properties of English/θ, ð/ and Mandarin/ CQd /s, z/ are distinctive from each other, whereas English /s, z/ display similar acoustic characteristics to Mandarin/CQd /s, z/.

3.9 Conclusion

This chapter discussed the articulatory and acoustic features of English /θ/-/s/ and /ð/-/z/, followed by a comparison of the phonetic inventories for English, Mandarin and CQd consonants. Through the comparison, it was discovered that English /θ/ and /ð/ neither exist in Mandarin nor CQd. /z/ exists in CQd but not in Mandarin. /s/ is the only target speech sound that occurs in English, Mandarin and CQd. For the purpose of investigating the degree of similarity and the differences among the target contrasts in the subjects' L1, L1-dialect and L2, the articulatory and acoustic characteristics of Mandarin /s/ and CQd /s, z/ were compared and discussed.

Chapter 4 Pilot Study

4.1 Introduction

Due to the influence of L1 experience, adult L2 learners may have difficulty in the perception and/or production of some L2 speech sounds (Lado, 1957; Best, 1994; Flege, 1995a, b; Kuhl, 1991; Iverson et al., 2003). Nonetheless, not all adult language learners have been found to have difficulty in the perception and/or production of unfamiliar L2 speech sounds (Best, 1994). Therefore, a pilot study was carried out for the purpose of revealing which English sound(s) would be comparatively more difficult than others for the subjects to perceive and produce. In the meantime, the subjects who had difficulty both in the perception and production of the sounds were selected to participate in the main study. This chapter presents the methodology, results and discussion of the pilot study.

4.2 Methodology

4.2.1 Subjects

An Oxford quick placement test (Appendix 1) was conducted to select subjects of the same English proficiency level. 60 students (30 female, 30 male) from Yangtze Normal University were randomly recruited. Each subject was paid 20 Yuan as reasonable compensation for participating in the test. According to the test results, forty-two 19 to 23 years old subjects (mean age = 20.5 were selected to participate in the pilot study (20 male, 22 female; English proficiency level = B1 > B2, or intermediate). All of them were L1-Mandarin L2-English speakers. They all spoke CQd as their L1-dialect. None of them had lived or travelled in English-speaking countries. All the participants had been learning English as an L2 for about 6-8 years (mean = 6.59) (see Appendix 2). All of them were reported to have normal hearing, intellective ability, and were right-handed. They were reported to have neither speech nor language related dysfunctions or hearing problems. Each subject was further paid 50 Yuan to participate in the pilot study.

In order to measure the subjects' accuracy in the production of all the English consonants, a short text (Appendix 3) was used as the stimuli in the production test. It was an adaption of the passage *Comma Gets a Cure* (McCullough, Somerville, and Honorof, 2000). The revised version of the text contained 38 words (Appendix 3), which contained all the English consonants in initial and final positions. In some cases, a consonant was only embedded in the position(s) where it is legal in English. For instance, no English words with /ʒ/ in word initial position are available, thus it was embedded in medial and final positions. Most of the stimulus words were disyllables and frequently used (See Appendix 4 for their frequency of occurrence). Although some of the words (*goose, bathe, zoo, beige, dwell*) occurred comparatively less frequently than others, they occurred in the English textbooks used in middle schools in China. Therefore, the words in the passage were not assumed to be new or difficult for the subjects to read. This was also illustrated by the test results.

4.2.2 Preparation of stimuli and design of tasks

Stimuli for perception test: The stimuli in the perception test were prepared after the results of the production test were obtained. Given the fact that speech perception and production have been shown to be closely connected (Williams and McReynolds, 1975; Jamieson and Rvachew, 1992; Watkins, Strafella, and Paus, 2003), or even innately linked to each other (Liberman, Cooper, Shankweiler, and Studdert-Kennedy, 1967; Liberman, 1985), it was hypothesized that the subjects who had serious difficulty in the production of specific English sounds may be also struggle with the perception of those sounds. According to the results of the production test, most of the subjects realized English /θ/ as /s/, and /ð/ as /z/. Thus, it was predicted that these subjects may have difficulty with the perception of /θ/-/s/ and /ð/-/z/.

For the purpose of examining whether the subjects had difficulty in the perception of the two contrasts for which they displayed low accuracy in the production test, an AXB task was carried out. The stimuli were nonsense words that contained /θ/-/s/ and /ð/-/z/. The reason for the employment of nonsense words rather than real words was to avoid the influence of lexical knowledge on the subjects' perception of the target contrasts (e.g., Hazan et al., 2005). Specifically, /θ/-/s/ and /ð/-/z/ were embedded in initial, medial, and final positions of the nonsense words. The vowel contexts were /i, a, u/. The syllable structures were VC, VCV and CV, which were counter-balanced (Appendix 5). The design of the phonetic environments was based on the coarticulation effect on language listeners' perception of speech sounds, since the duration of a consonant could be affected by syllable stress, phonetic position in a word, as well as grammatical conditions (Pickett, 1999). Embedding the target contrasts in different phonetic environments served to vary their durations in different stimulus words. Furthermore, /i, a, u/ are phonetically distinct from each other in terms of vocal tract configurations, tongue height, backness, and lip-roundness (Hazan et al., 2005, 2006; Jongman, Wang and Kim, 2003; Wang et al., 2009). Therefore, /θ/-/s/ and /ð/-/z/ could show different acoustic characteristics in the three vowel contexts (Shadle Mair, and Carter, 1996), thus ensuring the subjects' responses were based on phonetic distinctions rather than auditory discrimination (Pisoni, 1973: W). Each contrast was embedded in 18 different nonsense words. In order to minimize the influence of the gender difference on the subjects' perception performance, each stimulus word was produced first by a female RP speaker, and then by a male RP speaker. This resulted in 36 stimuli for each contrast. The recorded nonsense words were then coded into 2 AXB tasks with a script, and were carried out with Praat program (Boersma and Weenink, 2013). Each stimulus was repeated 3 times, thus yielding a total of 108 stimuli for each contrast. Moreover, the order of the items was automatically randomized with the Praat program (Boersma and Weenink, 2013), so as to avoid bias. In addition, in order to ensure the subjects' responses to the stimuli were on a phonetic level rather than on an auditory level, the interstimulus interval (ISI) was 1,000ms (Pisoni, 1973; Werker and Logan, 1985).

Recording of the stimuli: As mentioned above, all the stimuli for the speech perception test were produced by a female and a male RP speaker. The female RP speaker was a 31-year-old Master's student. She was doing Linguistics at Newcastle University. She was born and grew up in the South of England. The male RP speaker was a lecturer at Newcastle University. He was teaching phonetics and was phonetically trained. He was born in Scotland and lived there until 7 years old. After that, he lived in various cities in England..

The recording was carried out in a soundproof booth with a high quality recorder (Roland-05) in the Speech and Language Science Department of Newcastle University. The settings of the recorder were 16-bit mono channel and 44.1 KHz for sampling frequency. The stimuli were sent to the RP speakers one week before recording, so as to ensure they had enough time to get familiar with them. During the recording process, the display of the stimulus materials was controlled using a laptop outside of the booth by the investigator (the author). It was connected to a screen inside the booth. Through the screen the speaker could see the stimuli. The stimuli were presented with Power Point one by one in a randomized order, so as to avoid list intonation (e g., Hazan et al., 2005). Moreover, the investigator personally observed the speakers' production of the stimuli. It was found that they produced /θ/ and /ð/ as interdental. As for /s/ and /z/, the place of articulation could not be easily observed. Nonetheless, the RP speakers reported that they articulated the two sounds (/s, z/) as lingual-alveolar.

Task for perception test: An AXB discrimination task was carried out in the speech perception test. There were three reasons for the choice of an AXB task over other frequently used tasks: (1) compared with other discrimination tasks, such as AX tasks (the same or different task), AXB tasks demand longer ISIs, which ensured the subjects' perception of the target speech sounds was through phonetic access (Pisoni, 1973; Pisoni and Lazarus, 1974; Carney, Widin, and Viemeister, 1977; Crowder, 1982; Werker and Tees, 1984; Werker and Logan, 1985; Best, McRoberts, and Goodell, 2001); (2) AXB tasks are comparatively more difficult than other tasks, so it was expected to be able to detect the subjects' capacity in the perception of the target contrasts (Best, 1994); (3) AXB tasks have a smaller risk of response bias compared with other kinds of tasks (e.g., AX tasks) (McGuire, 2010).

Task for production test: The subjects were asked to do a "read aloud" task. They were asked to read a short text (Appendix 3) aloud, and were recorded with a high quality recorder (Roland05).

4.3 Procedure

4.3.1 Pilot-for-pilot study

Due to the pilot study being carried out in China by a former colleague[4] of the author, a pilot-for-pilot study was conducted first in Newcastle by the author herself. The purpose was to detect potential problems that may occur in the pilot study. Three Master's students from the Business School of Newcastle University volunteered to join this study (2 female and 1 male, with a mean age of 22.3 years old and IELTS scores of 6.0, 6.5, and 6.0 respectively). A consent form was signed before the start of the study. All 3 subjects were L1 Mandarin speakers of L2 English. Their L1-dialect was CQd. At the time of the research, they had been in the UK for 5 months. Both the perception and production tests were carried out in the same sound-proof booth at Newcastle University, where the stimuli were recorded.

[4] The investigator who carried out the pilot study was a lecturer in the Physics School at Yangtze Normal University. He had little knowledge of English, which made him a good investigator for the pilot study, particularly for the production test, in which he would not be likely to guide the students' pronunciation.

Production test: The stimulus material (a short text) for the production test was sent to the subjects one week before the study. They were asked to get familiar with the text before the study. A laptop was used to present the stimuli to the subjects. The recorder (Roland-05) was put close to and in front of the subjects with the same settings as were used in the stimuli recording. To ensure the subjects were familiar with the text, they were asked to read it aloud twice before recording (the third reading was recorded). A hard copy of *The Oxford English Dictionary* was available in the room. The subjects were asked to look up any word they did not know during the preparation time.[5] The recording began when the subjects said they were ready. All the recordings were transferred to a laptop in a wav format.

A 28 year old male RP speaker (hereafter, RP4) was asked to evaluate the accuracy of their production. RP4 was a Master's student, who was studying Linguistics at Newcastle University. He was born in Scotland, but lived in London for more than 10 years since the age of 6. He was asked to pick out the words in which the consonant(s) was/were incorrectly produced, and transcribe the subjects' production with phonetic symbols. A *10-score Likert scale* (0: totally wrong; 10: totally correct) was employed to evaluate the degree to which the consonants were correctly/incorrectly produced. The phonetic transcriptions of the incorrectly produced words were also compared with the American English version. If a phonetic transcription of the incorrectly produced consonant was the same as that in American English, it was evaluated as correct. This was because the subjects' pronunciation could be influenced by American English input, such as American movies or songs, despite the fact that English teaching materials used in public schools in China follow the British English system.

After the evaluation of the subjects' production test was completed, it was found that the subjects' most serious problem was the production of English /θ/ and /ð/. Specifically, according to RP4, the subjects realized /θ/ as /s/ and /ð/ as /z/ (see Table 4.1).

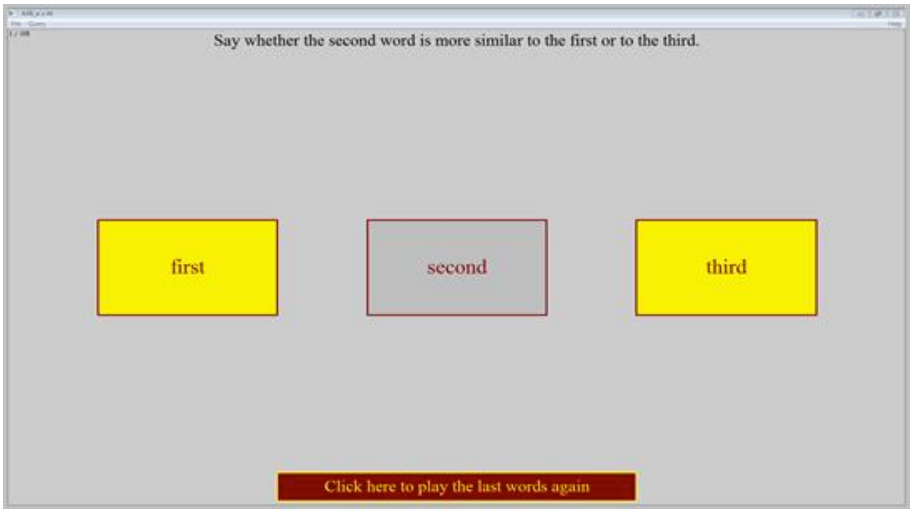
[5] It turned out no one looked up any word in the dictionary as they knew all the words in the text.

Incorrectly pronounced words	How many subjects incorrectly pronounced the word	British English transcription	American English transcription	Subjects' transcription of realisation	scores
Asia	1	/ˈeɪ.ʒə/	/ˈeɪ.ʒə/	/ˈeɪ. ʃə/	0
north	3	/nɔːθ/	/nɔrθ/	/nɔːs/	5
					0
					0
then(1st in the text)	3	/ðen/	/ðen/	/zen/	4
					0
					0
then(2nd in the text)	3	/ðen/	/ðen/	/zen/	0
					0
					0
mouth	3	/maʊθ/	/mɑʊθ/	/maʊs/	0
					6
					0
that	2	/ðæt/	/ðæt/	/zat/	0
					0
lunatic	1	/luː.nə.tɪk/	/ˈluː.nə.tɪk/	/ˈnuː.nətɪk/	1
itchy	1	/ˈɪtʃi/	/ˈɪtʃ.i/	/ˈɪŋtʃ.i/	0
thought	3	/θɔːt/	/θɔːt/	/sɔːt/	0
					0
					1
goose's	2	/guːsɪz/	/gusɪz /	/guːs/	0
					0
singing	1	/sɪŋ/	/sɪŋ/	/θɪŋ/	2
bathe	3	/beɪð/	/beɪð/	/beɪz/	0
					0
					0
gave	1	/geɪv/	/geɪv/	/gɪv/	0
the (1st in the text)	3	/ðə/	/ðə/	/zə/	0
					0
					3
the (2nd in the text)	3	/ðə/	/ðə/	/zə/	0
					0
					2
tune	1	/tjuːn/	/tun/	/tuŋ/	0

Table 4.1 Production test results in pilot-for-pilot study (The "British English transcription" and "American English transcription" were based on Cambridge Dictionaries Online (2015. See: http://dictionary.cambridge.org/) (0 = totally wrong; 10 = totally accurate).

Perception test: Given that speech perception and production are closely related to each other (Williams and McReynolds, 1975; Jamieson and Rvachew, 1992; Watkins, Strafella and Paus, 2003: Liberman et al., 1967; Liberman, 1985), based on the results of the production test, the subjects were predicted to have difficulty in the perception of English /θ/-/s/ and /ð/-/z/. Therefore, an AXB task with /θ/-/s/ and /ð/-/z/ as the target speech sounds was carried out.

An AXB task was presented with the Praat program (Boersma and Weenink, 2013). The subjects were asked to listen to three nonsense words in each trial, and decide whether the second word was the same as, or more similar to the first or the third by using the mouse to click on the appropriate symbol on the screen. When clicking on the "first" or the "third" button, a following trial was triggered. If the subjects wanted to listen to the current trial again, they could click on the red button on the bottom of the screen: "click here to play the last words again" (see figure 4.1 below). After clicking on this button, the trial is played again. The AXB task with /θ/-/s/ as the target contrast was presented first, then followed by that of the contrast /ð/-/z/. The subjects' responses were automatically recorded by the Praat program (Boersma and Weenink, 2013).



Say whether the second word is more similar to the first or to the third.

Figure 4.1 Screenshot of AXB test.

Subject	Percentage of correctness in perceiving /ð/-/z/	Percentage of correctness in perceiving /θ/-/s/
Female 1	44%	33%
Female 2	37%	30%
Male	33%	41%
Mean	38%	35%

Table 4.2 Perception test results of the pilot-study.

According to the results of the pilot-for-pilot study, it seems the subjects who had serious difficulty in the production of a specific English consonant also struggled with the perception of these sounds. Therefore, it was planned that the same procedure would be employed in the pilot study. That is, the subjects' production performance would be tested before the perception test. The target speech sounds of the perception test would be the ones that the subjects produced with comparatively low accuracy.

4.3.2 Pilot study

The pilot study was carried out in a quiet classroom at Yangtze Normal University of China by the investigator mentioned above. Prior to the study, a consent form was signed by the subjects. The production test was carried out first. After the evaluation of the subjects' production was completed, which was two days later, the perception test was then conducted.

1. Production test

A 'read aloud' task with the same stimulus material as that used in the pilot-for-pilot study was employed in the pilot study. The procedure was the same as that in the pilot-for-pilot study. All the recordings were transferred to a laptop in wav format with the subjects' names as the file names.

The recordings were then sent to RP4 for evaluation. RP4 was asked to evaluate the accuracy of the subjects' production with the same method as was employed in the pilot-for-pilot study. That is, the subjects' incorrectly produced words (consonants only) were phonetically transcribed. Some of the mistakes were slips of the tongue. For instance, one subject pronounced the first "she" in the reading text as "he". Incorrectly pronounced words like these were ignored. Each subject's mean score for the pronunciation of the incorrectly pronounced consonants was calculated and transformed into a percentage as follows: $(n_1 + n_2 + \dots + n_{x-1} + n_x)/g*10/100\%$ (n_x = the evaluated score of the x^{th} time that a consonant occurred in the reading text; g = the total number of times that the consonant occurred in the reading text). The lower the value, the lower the production accuracy was. For example, /θ/ occurred 3 times in the reading text. If a subject's score for the pronunciation of /θ/ in *thought, north* and *mouth* was evaluated to be 0, 10 and 10 respectively, his/her accuracy in the production of /θ/ would be: (0+10+10)/3*10/100%= 66.67%.

subject	/θ/ (%)	/ð/ (%)	/l/ (%)	/n/ (%)	/ʒ/ (%)	/f/ (%)
S1	13.3	50.6	100	100	50	100
S2	**83.35**	**90.6**	**100**	**100**	**100**	**100**
S3	40	66.7	100	100	50	100
S4	**90**	**93.8**	**100**	**100**	**100**	**100**
S5	13.3	49.4	100	100	100	100
S6	**86.7**	**88.1**	**100**	**100**	**100**	**100**
S7	43.3	32.5	91.7	100	50	100
S8	**93.3**	**87.5**	**100**	**100**	**100**	**100**
S9	66.7	36.3	91.7	100	50	90.9
S10	33	36.9	100	100	100	100
S11	**100**	**100**	**100**	**100**	**100**	**100**
S12	50	67.5	100	100	100	100
S13	66.7	62.5	100	100	50	100
S14	50	58.1	100	100	50	100
S15	40	51.2	9.2	100	100	100
S16	**86.7**	**97.5**	**10**	**100**	**100**	**100**
S17	**100**	**90.6**	**100**	**100**	**100**	**100**
S18	66.7	43.8	90	100	100	100
S19	**100**	**93.8**	**100**	**100**	**100**	**90.9**
S20	33.3	76.9	91.7	96.6	100	100
S21	33.3	53.1	100	100	100	100
S22	**86.7**	**81.3**	**100**	**96.6**	**100**	**100**
S23	30	39.4	100	100	100	100
S24	26.7	20.6	91.7	100	50	100
S25	23.3	33.8	100	100	50	100
S26	33.3	35	100	100	100	100
S27	50	31.9	100	100	100	100
S28	56.7	36.9	100	100	100	100
S29	**100**	**85.6**	**100**	**100**	**100**	**100**
S30	**86.7**	**93.8**	**91.7**	**100**	**100**	**100**
S31	53. 3	47.5	100	100	100	100
S32	**100**	**88.8**	**100**	**100**	**100**	**100**
S33	38.1	33.3	100	96.6	100	100
S34	34.4	30	100	100	100	100
S35	**91.3**	**93.3**	**100**	**100**	**100**	**100**
S36	42.5	50	100	100	100	90.9
S37	36.9	26.7	100	100	100	100
S38	30.6	40	100	100	100	100
S39	31.3	33.3	100	100	100	100
S40	27.5	23.3	100	100	100	90.9
S41	42.5	26.7	100	100	100	100
S42	35.6	43.3	100	96.6	100	100
Average	55.9	57.7	94.5	99.7	90.5	99.1

Table 4.3 Individual subjects' mean percentage accuracy in the production of /θ, ð, n, l, ʒ, f/ in the pilot study (subjects in the bold rows achieved high accuracy).

The evaluated results were then sent to the author to decide which speech sound(s) would be the target sounds for the following perception test. The sounds which were incorrectly produced by the subjects were /θ, ð, n, l, ʒ, f/, as shown in Table 4.3.

2. Perception test

According to the evaluated results of the production test, as shown in table 4.3 above, the subjects displayed the lowest accuracy in the production of English /θ/ and /ð/. Specifically, as in the pilot-for-pilot study, most of the subjects were found to realize /θ/ as /s/, and /ð/ as /z/ to different degrees. Therefore, /θ/-/s/ and /ð/-/z/ were selected to be the target contrasts in the perception test of the pilot study. An AXB test with the same stimuli as that employed in the pilot-for-pilot study was carried out in the study, which was presented with the Praat program (Boersma and Weenink, 2013). The procedure of the AXB test was the same as that in the pilot-for-pilot study.

subject	/θ/-/s/ (%)	/ð/-/z/ (%)	subject	/θ/-/s/ (%)	/ð/-/z/ (%)
S 1	35.18	37.04	**S 22**	**93.5**	**89.8**
S 2	**89.8**	**95.4**	S 23	34.26	37.04
S 3	39.81	38.89	S 24	32.41	31.48
S 4	**91.7**	**88.9**	S 25	36.11	32.41
S 5	64.81	67.59	S 26	33.33	33.33
S 6	**94.4**	**82.4**	S 27	43.52	45.37
S 7	39.81	33.33	S 28	40.74	27.78
S 8	**97.2**	**89.8**	**S 29**	**91.7**	**95.4**
S 9	42.59	25.93	**S 30**	**92.6**	**93.5**
S 10	32.41	45.37	S 31	36.11	39.82
S 11	**98.1**	**97.2**	**S 32**	**94.4**	**92.6**
S 12	32.41	53.71	S 33	54.63	48.15
S 13	46.3	48.15	S 34	53.7	35.19
S 14	34.26	38.89	**S 35**	**99.1**	**100**
S 15	55.56	68.52	S 36	33.33	38.89
S 16	**89.8**	**95.4**	S 37	51.85	51.85
S 17	**95.4**	**91.7**	S 38	37.04	37.04
S 18	40.74	45.37	S 39	40.74	48.15
S 19	**99.1**	**92.6**	S 40	37.96	50.93
S 20	39.81	48.15	S 41	46.3	28.7
S 21	48.15	39.82	S 42	44.44	35.19

Table 4.4 Individual subjects' accuracy in the perception of /θ/-/s/, /ð/-/s/ in the pilot study (mean accuracy for the correct perception of /θ/-/s/: 58.30%; average accuracy for the correct perception of /ð/-/s/: 58.10%; subjects in the bold rows achieved high mean accuracy).

As shown in table 4.4, the majority of the subjects displayed low accuracy in the perception of /θ/-/s/ and /ð/-/z/. Only 13 out of 42 subjects displayed accuracy of above 80% (shown in bold). The remaining 29 subjects' accuracy, however, was not satisfactory, ranging from about 25% (S9's accuracy in the perception of /ð/-/z/) to below 70% (S5's accuracy in the perception of the two contrasts). The mean accuracy of the subjects' perception of /θ/-/s/ and /ð/-/z/ was 58.30% and 58.10% respectively.

4.3.3 The selection of subjects and target contrasts for the main study

According to the test results in the pilot study, most subjects who had difficulty in perceiving the contrasts /ð/-/z/, /θ/-/s/ also incorrectly produced them (see Table 4.6). According to the phonetic transcriptions from RP4, the subjects who incorrectly produced /θ, ð/ realized /θ/ as /s/, /ð/ as /z/. This may be due to the lack of /θ/-/ð/ in the phonetic inventories of Mandarin and CQd, which is in accordance with the hypotheses of CAH; PAM/PAM-L2; SLM; NLM/NLM-e; PI. That is, L1 experience interferes with language learners' acquisition of L2 sounds. Additionally, the consonants in the two contrasts belong to different visemes with salient differences in articulatory information: /θ/ and /ð/ are produced as interdental, whereas /s/ and /z/ are produced as alveolar. Audiovisual training, which provides the subjects with visible articulatory information for the contrasts, was predicted to be able to facilitate their perception, and subsequently improve their production of the two contrasts (Owens and Blazek, 1985; Massaro et al., 1993). Therefore, /θ/-/s/ and /ð/-/z/ were selected as the target contrasts of the main study.

Regarding the choice of the subjects for the main study, those with an accuracy rate of below 80% in the perception of /θ/-/s/ and /ð/-/z/ in AXB tests, and with an average accuracy rate of below 80% in the production of /θ/ and /ð/ were selected. Thus, 29 subjects were selected to join the main study. Their number was recoded from S1 to S29 (see the first column of Table 4.6).

recoded number	gender	subject	perception of /θ/-/s/ (%)	production of /θ/ (%)	perception of /ð/-/z/ (%)	production of /ð/ (%)
S1	Male	S 1	35.18	13.3	37.04	50.6
	Female	**S 2**	**89.8**	**83.4**	**95.4**	**90.6**
S2	Male	S 3	39.81	40	38.89	66.7
	Female	**S 4**	**91.7**	**90**	**88.9**	**93.8**
S3	Male	S 5	64.81	13.3	67.59	49.4
	Female	**S 6**	**94.4**	**86.7**	**82.4**	**88.1**
S4	Male	S 7	39.81	43.3	33.33	32.5
	Female	**S 8**	**97.2**	**93.3**	**89.8**	**87.5**
S5	Male	S 9	42.59	66.7	25.93	36.3
S6	Male	S 10	32.41	33	45.37	36.9
	Female	**S 11**	**98.1**	**100**	**97.2**	**100**
S7	Male	S 12	32.41	50	53.71	67.5
S8	Female	S 13	46.3	66.7	48.15	62.5
S9	Male	S 14	34.26	50	38.89	58.1
S10	Female	S 15	55.56	40	68.52	51.2
	Male	**S 16**	**89.8**	**86.7**	**95.4**	**97.5**
	Male	**S 17**	**95.4**	**100**	**91.7**	**90.6**
S11	Female	S 18	40.74	66.7	45.37	43.8
	Male	**S 19**	**99.1**	**100**	**92.6**	**93.8**
S12	Female	S 20	39.81	33.3	48.15	76.9
S13	Female	S 21	48.15	33.3	39.82	53.1
	Male	**S 22**	**93.5**	**86.7**	**89.8**	**81.3**
S14	Female	S 23	34.26	30	37.04	39.4
S15	Female	S 24	32.41	26.7	31.48	20.6
S16	Female	S 25	36.11	23.3	32.41	33.8
S17	Female	S 26	33.33	33.3	33.33	35
S18	Female	S 27	43.52	50	45.37	31.9
S19	Female	S 28	40.74	56.7	27.78	36.9
	Male	**S 29**	**91.7**	**100**	**95.4**	**85.6**
	Male	**S 30**	**92.6**	**86.7**	**93.5**	**93.8**
S20	Female	S 31	36.11	53.3	39.82	47.5
	Female	**S 32**	**94.4**	**100**	**92.6**	**88.8**
S21	Female	S 33	54.63	38.1	48.15	33.3
S22	Male	S 34	53.7	34.4	35.19	30
	Female	**S 35**	**99.1**	**93.3**	**100**	**100**
S23	Male	S 36	33.33	42.5	38.89	50
S24	Male	S 37	51.85	36.9	51.85	26.7
S25	Male	S 38	37.04	30.6	37.04	40
S26	Male	S 39	40.74	31.3	48.15	33.3
S27	Female	S 40	37.96	27.5	50.93	23.3
S28	Male	S 41	46.3	42.5	28.7	26.7
S29	Female	S 42	44.44	35.6	35.19	43.3

Table 4.5 The subjects’ performance in the perception and production of /ð/ and /θ/ in the pilot study.

Even though adult L2 learners may have difficulty both in perceiving and producing L2 speech sounds, with sufficient input they are predicted to be able to learn the sounds eventually (PAM/PAM-L2; SLM; NLM/NLM-e; PI). Moreover, according to PAM/PAM-L2, SLM, NLM/NLM-e, and PI, language learners' L2 achievement is largely influenced by their L1 experience (Best, 1995; Flege, 1995a, b). Since all the subjects share the same L1 (Mandarin), L1-dialect (CQd) and L2 (English), they would be influenced by the same L1/L1-dialect in their perception and production of L2 speech sounds.

4.4 Discussion

The most significant finding in the pilot-for-pilot study and pilot study is that the majority of the subjects had serious difficulty in the perception and production of English /θ/-/s/ and /ð/-/z/. Some subjects also incorrectly produced /n/, /l/, /f/,/ʒ/, but with a much higher degree of accuracy compared with /θ/-/s/ and /ð/-/z/.

First of all, this finding is in accordance with the hypothesis of Lado's (1957) CAH, which predicts that the dissimilarity between language learners' L1 and L2 poses difficulty for their L2 acquisition. Among the contrasts /θ/-/s/ and /ð/-/z/, /s/ exists in the subjects' L1 and L1-dialect, /z/ occurs in their L1-dialect, whereas /θ/ and /ð/ do not occur in their L1 and L1-dialect phonetic inventories. Based on the hypothesis of CAH, the subjects were predicted to have difficulty in the perception and production of /θ/ and /ð/. Their low accuracy in the perception of /θ/-/s/, /ð/-/z/, and substitution of /θ/ with /s/, /ð/ with /z/ confirms this prediction. Moreover, this finding is congruent with that in Shih and Kong (2011) and Tutatchikova (1995), in which the subjects showed difficulty in distinguishing between retroflex fricatives due to the non-occurrence of these sounds in their L1 phonetic inventory.

However, some subjects also incorrectly pronounced /n/, /l/, /f/ and /ʒ/. According to RP4, the subjects realized /n/ as /l/, or /l/ as /n/ in word medial position. This may be because in CQd, /n/ is only used in coda rather than in initial position, and it is typically used in initial position in Mandarin. This finding suggests *negative transfer* from Mandarin to CQd as proposed by behaviourist approaches to L2 acquisition (see Ellis, 1985 for review). It seems the subjects' production of /n/ and /l/ was negatively influenced by their L1 and/or L1-dialect (CQd). Moreover, some of the subjects incorrectly produced /f/ and /ʒ/ by substituting the two sounds with /v/ and /ʃ/ respectively. Meanwhile, they did not show any difficulty in the production of /v/ and

/ʃ/. Given that /ʒ/ neither occurs in Mandarin nor CQd, the subjects' failure in the production of /ʒ/ might be explained by the hypotheses of CAH, which predicts that the differences between language leaners' L1 and L2 pose difficulties for their L2 learning. Nevertheless, it seems the CAH could not help explain the subjects' substitution of /ʒ/ with /ʃ/, because /ʃ/ neither exists in their L1 nor L1-dialect. Further examination of the stimuli for the production test may shed some light on this issue. Due to the lack of /ʒ/ in word-initial position, /ʒ/ was embedded in medial (*Asia*) and final (*beige*) positions of the stimuli. The subjects who incorrectly produced *Asia* correctly pronounced *beige*. The subjects' incorrect production of /ʒ/ in *Asia* might be caused by lexical knowledge, incorrect input, or the potential influence of orthography. Their incorrect production of /f/ may either be attributed to incorrect input or a slip of the tongue. For instance, they were found to only realize /f/ as /v/ in the production of *off*, but correctly produced it in *from*.

Moreover, the subjects' difficulty in the perception and production of /θ/-/s/ and /ð/-/z/ is also congruent with the common hypothesis of PAM-L2, SLM, NLM/NLM-e and PI. That is, due to the influence of L1 experience, language learners, particularly adults, may lose sensitivity in the discrimination of non-native speech sounds. However, it seems to be at odds with one of the hypotheses of SLM, which claims that the more dissimilar the L1 and L2 sounds are, the more likely language learners are to be able to develop new phonetic categories for the L2 sound (Flege, 1987). Given that SLM did not provide a specific standard for the definition of "dissimilarity", the target contrasts of the present study are compared in terms of articulatory gestures and acoustic properties. All the stimuli of /θ/ and /ð/ were produced as interdental, and /s/ and /z/ were produced as alveolar by the RP speakers. Thus, /θ/ and /s/, /ð/ and /z/ are quite different from each other in terms of articulatory gestures. On the acoustic level, /s/ and /z/ display a relatively higher frequency range and intensity, and a shorter spectrum length than /θ/ and /ð/. They are different form each other in terms of variances in amplitude ranges as well. Moreover, /θ/ and /ð/ display a relatively flat spectrum with no clearly dominating peak, whereas the spectra of /s/ and /z/ are well-defined with distinct shape and a primary spectra peak at high frequencies (see Table 3.1 in Chapter 3). Nevertheless, it seems these distinctive articulatory and acoustic differences between /θ/ and /s/, /ð/ and /z/ did not benefit the subjects' when it came to distinguishing them.

In addition, as discussed in chapter 3, the voiceless /s/ and voiced /z/ in English and voiceless /s/ Mandarin can either be produced as alveolar or dental depending on the speaker. Yet they are typically produced as apical dental instead of alveolar in CQd, which involves the tongue tip placed against the back of the upper teeth. Voiceless /θ/ and voiced /ð/ in English are typically produced as interdental by native English speakers, but they can also be pronounced as dental (Prator and Robinett, 1985; Wang et al., 2009). That is, the tongue tip is either placed against the back of the upper teeth (in the case of dental articulation), or in between the upper and the lower teeth (in the case of interdental articulation). On this point, the articulatory gestures of English /θ/, Mandarin /s/ and CQd /s/ are quite similar to each other, or even share the same place of articulation when they are all produced as dental. It is the same situation for English /ð/ and CQd /z/. In the production test, it was found that the majority of the subjects realized /θ/ as /s/, and / ð / as /z/. This finding may provide supporting evidence for the hypothesis of PAM-L2, which suggests that language learners tend to assimilate non-native speech sounds to the most-articulatory similar speech sounds of their L1 (Best and Tyler, 2007).

Regarding the subjects' production of English /s/ and /z/, RP4 gave them full scores with a *10-point Likert scale*. As discussed in Chapter 3, the articulatory gestures of English /s/ and /z/ are quite similar to or even identical to those in Mandarin and/or CQd, depending on the speaker. This finding may also be explained by the hypothesis of PAM-L2. That is, the subjects' native-like performance in the production of English /s, z/ may be because they assimilated English /s/ to Mandarin/ CQd /s/, and English /z/ to CQd /z/ in terms of articulatory gestures. CAH may also work for the explanation of this finding, which predicts that the similarities between language learners' L1 and L2 facilitate their acquisition of the L2. It might be tempting to speculate that this finding provides counterevidence to the hypothesis of SLM, which predicts that if an L2 sound is similar to or identical to its counterpart in language learners' L1, the learners may be able to perceive the acoustic differences, but are unable to use the perceived differences in the production of the L2 sound. According to RP4's assessment, the subjects did achieve native-like performance in the production of English /s, z/. Nevertheless, due to a lack of acoustic analysis, it was not clear whether the subjects' production of English /s, z/ displayed some acoustic differences to that of native English speakers.

Furthermore, most of the subjects who struggled with the perception of the target contrasts were found to have difficulty in the production of these sounds. In the meantime, for those whose perception accuracy was above 80%, their production performance was also satisfactory (and above 80%). This result may neither demonstrate the hypothesis that speech perception patterns emerge before speech production patterns (Kuhl et al., 2008; Flege, 1981, 1987, 1988, 1991a, 1992a, b, 1995a), nor answer the question of whether speech perception and production share a common link and a common processing strategy as hypothesized by Liberman and colleagues' MT. However, it does provide supporting evidence for the hypothesis that speech perception and production are closely connected (Williams and McReynolds, 1975; Jamieson and Rvachew, 1992; Watkins et al., 2003) or innately linked to each other (Liberman et al., 1967; Liberman and Mattingly, 1985).

Another finding in the pilot study was that not all the subjects had difficulty in the perception and production of /θ/-/s/ and /ð/-/z/. 13 out of 42 subjects' accuracy both in the perception and production of the two contrasts was above 80%. As predicted by Best (1994), not all adult L2 learners are uniformly poor at the perception/production of all L2 speech sounds. Nonetheless, the 13 subjects' good performance may be influenced by other factors, such as AO of L2-English learning, the amount of time spent on English learning, learning strategies, and so on. Given that the purpose of the pilot study was to select suitable subjects and target contrasts for the main study, no further investigation of this finding was carried out. The influence of these factors on the subjects' perception and/or production performance was investigated in the main study.

4.5 Conclusion

This chapter presented the pilot study, which aimed to select suitable subjects and target contrasts for inclusion in the perception training in the main study. The methodology, procedure, and results were outlined. The results were discussed in light to relevant literature reviewed in Chapter 2. Based on the results of the pilot-study, 29 subjects who showed difficulty both in the perception and production of /θ/-/s/, /ð/-/z/ were recruited for the perception training programme in the main study. There were two reasons for selecting /θ/-/s/ and /ð/-/z/ as the target contrasts: (1) the majority of the subjects in the pilot study displayed difficulty in the perception and production of the two contrasts (n = 29 out of 42); (2) /θ/ and /ð/ are typically produced as interdental in English (Prator and Robinett, 1985; Wang et al., 2009), whereas /s/ and /z/ are frequently produced as alveolar by native English speakers. Thus the two contrasts belong to different visemes. The visible articulatory difference between the two contrasts served as the basis of the audiovisual training in the main study.

Chapter 5 Main study

5.1 Introduction

This chapter presents the main study. The research questions and hypotheses are presented first. After that, the methodology, procedure and results of the main study are outlined. For the analysis of the collected data, a *repeated-measures ANOVA* was carried out to detect the factors that may have an influence on the subjects' perception and/or production performance. Tables and graphs were adopted to help describe the results of the subjects' perception and production. The research questions are answered with the findings of the main study.

5.2 Research questions and hypotheses

For speech perception training, audiovisual training may be more effective than other modalities, such as a purely auditory modality, because it provides the subjects with the visual information for the target speech sounds' articulatory gestures (Hazan et al., 2005; Hardison, 2003; Lidestam et al., 2014). Although articulatory gestures have been shown to be able to facilitate listeners' perception of speech sounds (Summerfield, 1979, 1981, 1983; Breeuwer and Plomp, 1984; Massaro, 1987), debate still exists concerning (1) whether audiovisual perception training can facilitate the listeners' auditory perception; and (2) whether the training effect can be transferred to speech production. Therefore, based on the findings from the pilot study (see Chapter 4), and the theoretical background discussed in the literature review (see Chapter 2), the main study explores the extent to which, if at all, audiovisual perception training can facilitate the subjects' auditory perception and production of L2-English contrasts /ð/-/z/ and /θ/-/s/. Two research questions were formulated:

(1) *To what extent, if at all, can the subjects' capability in the auditory perception of English contrasts /θ/-/s/, /ð/-/z/ be improved by audiovisual perception training?*

(2) *To what extent, if at all, can the subjects' capability in the production of English contrasts /θ/-/s/, /ð/-/z/ be improved by audiovisual perception training?*

What makes this study different from previous ones is that (1) the training materials contain a larger number of stimulus words, which include a wide range of phonetic environments concerning vowel contexts and phonetic positions; (2) instead of only testing the subjects' perception and production performance in the pre-test and the post-test, two middle-tests were conducted in addition to the pre-test and post-test. The purpose was to examine the subjects' improvement, if any, for the auditory perception and production of the target contrasts during the training procedure.

Both theories/models (PAM/PAM-L2; SLM; NLM/NLM-E; PI) and findings from previous experimental studies (e.g., Bradlow et al., 1997; Hazan et al., 2005; Bernstein, et al., 2013) indicate that adult L2 leaners can eventually learn L2 speech sounds that they initially have difficulty with. The amount of L2 input is found to play an important role concerning their achievement in the acquisition of L2 speech sounds (Flege, 1981, 1987, 1988, 1991a, 1992a, b, 1995a, b, 2003). Articulatory gestures are found to be significant in speech perception (Best, 1994, 1995a, 1995b; Liberman et al., 1967; Cooper, et al., 1952). Therefore, providing the subjects with articulatory gestures of the target speech sounds is predicted to be able to facilitate their perception of these sounds. The target contrasts of the present study belong to different visemes (typically, /θ/ and /ð/ are produced as interdental, whereas /s/ and /z/ are produced as alveolar). Thus, the articulatory differences between /θ, ð/ and /s, z/ are saliently visible. Given that audiovisual integration, auditory and visual skills are found be integrated with each other in speech perception (Ghazanfar and Schroeder, 2006; Schwartz et al., 2012; Sams et al., 1991; Sato et al., 2013), audiovisual training is expected to benefit the subjects' perception of the target contrasts. However, due to individual differences, such as gender, individual intelligence, as well as other relevant factors as summarised in Chapter 2, the ultimate achievement in the perception and production of L2 speech sounds may vary across individual subjects. Therefore, for the first research question, it was hypothesized that with audiovisual perception training, the subjects' capability in the auditory perception of the target contrasts could be improved. The degree of improvement may vary across the subjects.

The second research question examines whether speech perception training can lead to the improvement of language learners' capability in the production of the speech sounds. So far, no consensus has been achieved on this issue. Nonetheless, it was found that speech perception and production are either closely linked (Williams and McReynolds, 1975; Jamieson and Rvachew, 1992; Watkins et al., 2003), or even innately connected (Liberman et al., 1967; Liberman and Mattingly, 1985, 1989; Hawkins, 1999; Liberman and Whalen, 2000). Moreover, Flege's SLM posits that language learners' failure in the production of L2 speech sounds is due to their inaccurate perception of these sounds (Flege, 1981, 1987, 1988, 1991a, 1992a, b, 1995a, b, 2003). Similarly, the hypotheses of NLM/NLM-e also hold that accurate speech perception can help correct production of the perceived speech sounds (Kuhl et al., 2008). In view of this theoretical background, concerning the second research question, it was hypothesized that audiovisual training may have transferred beneficial effects on the subjects' production of the target contrasts. Due to the influence of relevant factors (i e., AO, age, gender, etc.), their degree of improvement may also vary from one subject to another. Moreover, given that all the subjects did not show any difficulty in the production of /s, z/ in the pilot study, they were hypothesized to be able to produce the two sounds correctly in the main study.

5.3 Methodology

The main study included an audiovisual perception training programme. The HVPT approach is frequently adopted in audiovisual training. It has been shown to be effective, because it provides the subjects with stimuli of high-variability in different phonetic contexts (Logan et al., 1993; Lively et al., 1993; Bradlow and Pisoni, 1996; Handley et al., 2009). The training programme of this study, therefore, to a large extent, followed the principles of HVPT. The design of the training task and the preparation of the stimuli mainly followed the "natural variability" principle of HVPT (Logan et al., 1993), and as such sought to direct the subjects' attention towards the critical articulatory code by providing them with stimuli of high-variability (Bradlow et al., 1997; Lively et al., 1993).

However, HVPT may have the disadvantage of being unable to solve the problem of perception interference (Iverson et al., 2005). In order to make up for this limitation, the RP speakers were asked to produce /ð, θ/ as interdental, whereas /s, z/ as alveolar as they are typically produced by native English speakers. This could help the subjects to differentiate the two contrasts with distinctive articulatory gestures, despite not following the "variability" principle of HVPT.

The training effect on the subjects' perception and production of the target contrasts was tested before, during and at the end of the training programme. Qualitative data was collected with a questionnaire (Appendix 7) to detect the effect of relevant factors on the subjects' perception and production performance, if any.

The same stimuli were used 4 times in the AXB task in the perception tests. Although the order of the stimuli was different across these tests, there might still be bias concerning the repeated testing effect, or habituation to the evaluation test. In order to detect whether the subjects' perception and/or production improvement, if any, was because of the training programme rather than a repeated training effect, the performance of a control group was tested and compared with that of the experimental group.

5.3.1 Subjects

Experimental group: The 29 subjects (14 male, 15 female; mean age=20.03 years old; SD of age=0.89) who were found to have difficulty both in the auditory perception and production of English contrast /ð/-/z/ and /θ/-/s/ in the pilot study were selected. None of them was an English major.

Control group: For the purpose of selecting subjects of similar language background (i.e.AO of L2-English learning, years of L2-English learning, age, etc.), 57 students (30 female, 27 male) from Yangtze Normal University were randomly selected to complete a questionnaire (the same questionnaire as was completed by the experimental group). Their English proficiency level was also tested with an Oxford quick placement test (see Appendix 1). Subjects of similar profile to that of the experimental group were recruited to join the main study. They were 20 undergraduate students who were doing their Bachelor's Degree in different subjects at Yangtze Normal University (10 male and 10 female; mean age= 20.50 years old; SD of age=0.76). None of them was an English major. They were L1-Mandarin speakers of L2-English from Chongqing China, and so they spoke CQd as their L1-dialect. Their English proficiency level was intermediate, which was the same as that of the experimental group. Moreover, the subjects in the control group had a very similar profile to the experimental group regarding their age range, years of English learning, AO of English learning, primary purpose for learning English as an L2, the institute(s) in which they had been learning English, ways of learning English in their spare time, as well as opportunities for using English on a daily basis. Subjects of the control group were coded from S30 to S49.

5.3.2 Apparatus

(1) The Praat software program (Boersma and Weenink, 2013) was used to present an AXB task; (2) high quality digital recorders (Roland-05 and Roland-09) were employed

to record the stimuli and the subjects' production of the stimulus text/sentences; (3) a digital camera (Canon Legria, FS 37, with 41x zoom) was adopted for the audiovisual recording of training materials; (4) thirty desktops equipped with high quality headphones (JVC HA-RX700) in a quiet classroom were used for audiovisual training and perception tests. The production tests were conducted in the room where the training was carried out.

5.3.3 Stimuli

Stimuli for training (see Appendix 8**):** Based on the principles of HVPT, a large number of minimal pairs were prepared as the stimuli for training. Three versions of English "minimal pairs" were created with different stimulus words in each version. Due to a lack of English vocabulary items with the target contrasts in all possible word positions, most of the "minimal pairs" included one real word and one nonsense word (in some cases, both were nonsense words). That is, in each minimal pair, the pronunciation of the two words only differed in one sound, which was the sound of the target contrasts. For instance, in the "minimal pair" *sirty* and *thirty*, *thirty* was a real word, while *sirty* was a nonsense word. That is, /θ/ in *thirty* was substituted by /s/ in *sirty*. The stimulus words ranged from monosyllables to multisyllables, in which the target contrasts were embedded in various vowel and consonant environments. Each minimal pair was audiovisually repeated twice with the order of AB and BA (i e., *A sirty B thirty; A thirty B sirty*). The target contrasts were embedded in initial, medial and final positions of the stimuli, yielding 60 trials for each contrast and 120 trials in total in each training session. The order of the stimuli was randomized. The stimuli in sessions 1, 4 and 7 were based on the same set of words, but with different randomized orders. It was the same for sessions 2, 5, 8 and sessions 3, 6, 9. The purpose was to expose the subjects to multiple stimuli and direct their attention to discovery of the differences in the target contrasts.

"Minimal pairs" are predicted to be the most difficult situation in L2 speech perception, because the words are different from each other with only one contrasting sound (Best and Tyler, 2007). The identification task with minimal pairs as the stimuli, therefore, was expected to be effective at facilitating the subjects' perception of the target contrasts, and consequently, their production of these sounds.

Stimuli for perception tests (see Appendix 5): Stimuli for perception tests were the same as used in the pilot study.

Stimuli for production tests (see Appendix 6): Given that in the pilot study, the number of words that contained the target contrasts of the main study were not evenly distributed in the production test materials (see Appendix 3), 12 new sentences were created with /ð/-/z/, /θ/-/s/ embedded in different positions in different words (5 in initial, 5 in medial and 5 in final position). Among the stimulus words, *theatres* includes both /ð/ and /z/; *exist* contains both /z/ and /s/; *with* occurred 4 times. Thus there were a total number of 61 stimulus words. None of the stimulus words occurred in the training materials, so as to ensure the subjects' improved accuracy, if any, was not because of the repeated training experience. Due to there being a large number of words that contained the target contrasts in initial, medial, and final positions in the training materials, a limited number of words remained with /ð/-/z/ and /θ/-/s/ in different phonetic positions. Therefore, the vowel contexts of the stimulus words in the production test were not specified.

5.3.4 Recording

Considering that synthetic speech is predicted to be misleading, or to provide the subjects with incomplete information about the target phonetic category in speech perception learning (Logan et al., 1991), all the stimuli used in the training sessions were naturally produced. Thus the "natural" principle of HVPT was followed.

The stimuli employed in the perception test in the main study were the same as in the pilot study. The order of the stimuli, however, was re-randomized with the Praat program (Boersma and Weenink, 2013). Thus, only the stimuli used for the training sessions were recorded for the main study. In order to expose the subjects to different native English speakers' production of the target contrasts, following the HVPT principle of "variability", three RP speakers (RP1, RP2, RP3) were recruited to produce the training materials. RP1 (female, 31 years old) was the same female speaker who produced the stimuli used in the speech perception test. RP2 (female, 21 years old) and RP3 (male, 22 years old) were both from London, and were doing Bachelors' degrees at Newcastle University. All of them could read phonetic symbols. Each of them was paid 6 pounds and was given a gift for participation. Due to many of the stimulus words being nonsense words, the pronunciation of each word was recorded with phonetic symbols. The RP speakers were asked to read the stimulus words according to the phonetic transcriptions. The stimulus words in sessions 1, 4, and 7 were produced by

RP1. Stimuli in sessions 2, 5, and 8 were produced by RP2. RP3 produced the stimuli in sessions 3, 6, and 9. They were individually recorded in a quiet room at Newcastle University, where there were big windows with transparent and soundproof glass, so as to ensure it was bright enough to clearly show the RP speakers' face image.

The stimuli (Appendix 8) were sent to the RP speakers to get familiar with one week before recording. The recorder Roland-09 was used for auditory recording with the same settings as that in the pilot study. It was fixed close to and in front of the RP speaker's mouth to guarantee optimal recording quality. At the same time, a digital DVD camera (Canon Legria, FS 37) was used for visual recording. The camera was fixed in front of the speaker on the horizontal level of their faces. All the stimuli were printed on a piece of paper, which was fixed in front of the speaker and next to the DVD camera, so that the speakers could see them clearly. The speakers were asked to read each "minimal pair" twice. Following other researchers, such as Hazan et al. (2005) and Gesi et al. (1992), the RP speakers were told that their recordings would be used to teach L2-English speakers' production of the target contrasts.

As mentioned in section 5.3 above, considering that the HVPT approach bears the limitation of being unable to solve the problem of perception interference, the present study incorporated some changes to make up for this limitation. Specifically, the RP speakers were asked to exaggerate their pronunciation by producing /θ, ð/ as interdental, and /s, z/ as alveolar, so that the subjects could observe the tongue movements of /θ, ð/ more clearly in contrast with that of /s, z/. According to NLM/NLM-e (Kuhl et al., 2008), speech sounds produced with exaggeration can facilitate L2 learners' acquisition of these sounds. The RP speakers' exaggerated production was expected to be able to help the subjects' discrimination of the target contrasts.

During the recording process, a white background was set against the RP speakers with a fill light illuminated, so that the image of the speakers' front face could be seen clearly. The camera was zoomed in to ensure only the speakers' front face was captured, and so the subjects could observe the speakers' mouth movements. The recordings were then transferred to a computer. For the purpose of obtaining a high quality recording of the RP speakers' production, the sound recorded by the DVD camera was erased. The video channel and the auditory recording (obtained from the recorder Roland-09) were synchronised. After that, the synchronised audiovisual recordings of each "minimal pair" were cut and merged according to the randomized orders in each session (ISI=1000ms; Inter trial interval=3000ms). Each trial was displayed twice: the first time was the original production from an RP speaker, through which the subjects were expected to identify which word in a "minimal pair" occurred first. The second time, the correct answer for the trial was shown on the left-hand side of the image (see Figure 5.1 and 5.2 below. For anonymity, the RP speakers' face image is half covered). Providing the subjects with immediate feedback helps hold and increase their attention during the training process (McGuire, 2010).

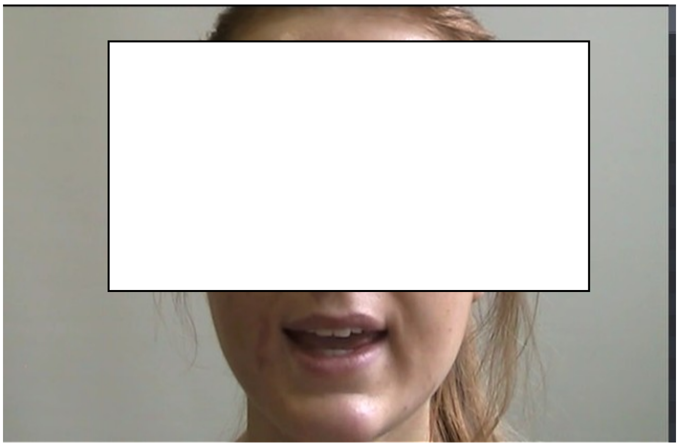


Figure 5.1 Screenshot of the first-time production of the "minimal pair" *A. sink B. think* from RP2 in training session 2, 5, 8.

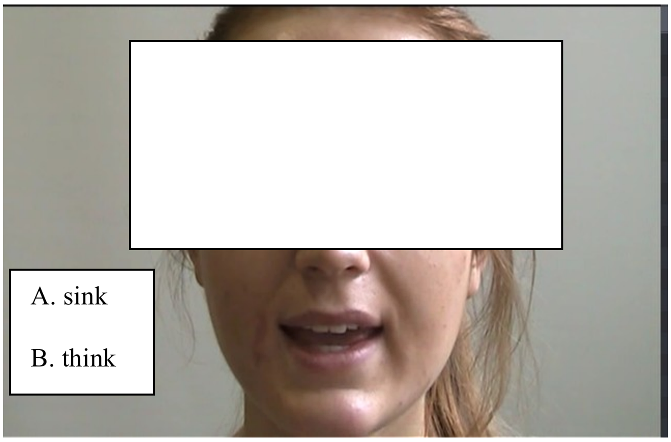


Figure 5.2 Screenshot of the second-time production of the "minimal pair" *A. sink B. think* from RP2 in training session 2, 5, 8.

5.4 Procedure

Prior to the training, all the subjects were asked to sign a consent form in addition to the one signed in the pilot study. Each of them was paid 50 Yuan to participate in the main study, as reasonable compensation. Table 5.1 below shows the timetable of the main study.

Time	Experimental group	Control group
Day 1	pre-test for speech production	Speech perception test (pre-test); Finish the questionnaire
Day 2	Finish the questionnaire; training session 1	
Day 3	training session 2	
Day4	training session 3	
Day5	mid-test 1 (for speech perception and production)	speech perception test (mid-test 1)
Day 6	rest	
Day 7	Training session 4	
Day 8	training session 5	
Day9	training session 6	
Day10	mid-test 12(for speech perception and production)	Speech perception test (mid-test 2)
Day11	rest	
Day12	training session 7	
Day13	training session 8	
Day14	training session 9	
Day 15	Post-test (for speech perception and production)	Speech perception test (post-test)

Table 5.1 The timetable of the main study.

Qualitative data of the study: Given that factors beyond L1 experience may influence language learners' achievement in learning L2 speech sounds, such as gender, age, etc. (Ausubel, 1964; Taylor, 1974; Bialystok, 1997; Bialystok and Hakuta, 1999; García Mayo and García Lecumberri, 2003), qualitative data was collected with a questionnaire (Appendix 7). This may give further information in addition to the quantitative data obtained from the perception and production tests. Specifically, the information concerning the subjects' age; AO of L2-English learning; primary motivation for learning English as an L2; in which institute(s) they had been learning English; as well as in which ways, if any, they had been learning English in their spare time was obtained from the questionnaire. All the questions were translated into Chinese to ensure the subjects could fully understand them. The questionnaires were handed out to the subjects by the investigator (the author) at the beginning of the first training session, and were collected after they were finished.

Training: The training programme included nine training sessions. A 2AFC (2 alternative forced choice) identification task was carried out in each session using the Praat program (Boersma and Weenink, 2013). Identification tasks are revealed to be able to heighten the subjects' sensitivity to the differences (articulatory and/or acoustic) between the target contrasts (Pisoni and Lively, 1995; McGuire, 2010). 2AFC tasks are predicted to be able to minimize the bias, as each choice can potentially be the right answer (McGuire, 2010). The training programme was carried out in a quiet classroom at Yangtze Normal University. There were 30 desktops in the classroom, each of which was equipped with high quality headphones (JVC HA-RX700). Each training session lasted about 35 minutes. Each subject was asked to sit in front of a desktop, and wear the headphone that was connected to the desktop. The stimuli were binaurally presented to the subjects via the headphone at a comfortable listening level (65—70dB), and visually presented via the monitor (33*20 cm) in front of them. The subjects were shown how to adjust the volume using the button on the headphones, so that they could adjust it by themselves if needed. All the instructions were given in Mandarin by the investigator to ensure the subjects had a clear understanding of what they were going to do. All the desktops were connected to and controlled by the "central computer" on the stage of the classroom. Therefore the investigator could control the display of the recordings.

Before the start of the first training session, the investigator demonstrated how to do the 2AFC task with an example – *A. sink B. think*. An answer sheet was handed out to the subjects before each training session. The subjects were told that some of the words would be new for them instead of being told that they would encounter nonsense words, so as minimize unnecessary concern. The subjects were then given 5 minutes to become familiar with the stimuli on the answer sheets. The investigator played the recordings via the "central computer". For each trial, the subjects were asked to circle on the answer sheet the right order of a "minimal pair" that they heard and/or watched: whether it was AB or BA. For example, whether it was (1) *A. sink B. think*, or (2) *A. think B. sink.* After 3000ms, the trial was automatically played again with the right answer on the left-hand side of a speaker's image (see Figure 5.2). When listening to/watching each trial for the second time, the subjects were asked to check their answer, and watch/listen to the recording carefully to see why their answer was correct or incorrect. After a 3000ms interval, the following "minimal pair" was played with the same procedure. The recordings of stimuli containing /ð/-/z/ were played first. After a 5 minute break, the recordings of stimuli containing /θ/-/s/ were played.

Tests for experimental group: In order to detect the subjects' improvement in the perception and production of the target contrasts, if any, during and at the end of the training programme, their performance was tested four times. That is, a pre-test (before the training programme) was administered, as well as mid-test 1 (at the end of the 3rd training session), mid-test 2 (at the end of the 6th training session), and a post-test (at the end of the training programme). Any subjects who achieved an accuracy of 90% or above in both in the perception and production of the target contrasts in mid-test 1 were dropped from the following training sessions and tests, because they were assumed not to need further training. The same principle was applied to mid-test-2. Moreover, the AXB test results in the pilot-study were adopted as the perception results of the pre-test, because the same stimuli were used both in the pilot study and the main study for the test, despite the stimuli being randomized in different orders. In the production test, however, the subjects were tested with the new stimuli used in the main study before being trained (Appendix 6), the results of which were employed as the pre-test results.

Tests for control group: The subjects' perception and production of the target contrasts was tested with the same tasks as were carried out by the experimental group. As shown in table 5.1 above, their performance was tested 4 times with the same intervals between tests as that of the experimental group. Before the test, they were asked to complete the same questionnaire as the experimental group.

Speech perception test: The same AXB task conducted in the pilot study was carried out in the main study. Each subject's responses were automatically recorded with the Praat Program (Boersma and Weenink, 2013), and were then extracted and saved on the computer. The tests were carried out in the same room where the training sessions were carried out. The subjects were tested individually to avoid being influenced by other subjects.

Speech production test: The procedure was the same as that in the pilot study, but with different stimulus materials (Appendix 6). Individual subjects were asked to read the stimulus sentences 3 times each, in the same room where the perception test was carried out.

5.5 Evaluation

In the perception test, all the subjects' responses in each trial of the AXB test were transferred to SPSS for further analysis.

In the production test, in order to minimize bias, the subjects' production results were evaluated by 4 raters. They were RP4 (the same rater who evaluated the subjects' production performance in the pilot study), RP5 (male, 27 years old), RP6 (female, 25 years old), and RP7 (female, 33 years old). Both RP5 and RP6 were from York, and were doing Linguistics for their Master's degree at the University of York. RP7 was a therapist for speech apraxia. She was born in Newcastle, but lived in London for 15 years. Each of them was paid 30 pounds to do the evaluation job.

After each test was finished, the recordings of the subjects' production in each session were sent to the 4 raters through drop-off. They did the evaluation job separately. For each test, individual subjects' recordings were put in one folder. The folders sent to the raters were named with randomized numbers in each test (the experimental group from No. 1 to No. 29; the control group from No. 30 to No. 49). For example, S1's folder was randomly named as 04 in the pre-test, 12 in mid-test 1, 11 in mid-test 2, and 20 in the post-test. This meant that the raters were not likely to know which subject's production they were evaluating, thus minimizing potential bias. After the evaluation of each session was finished, the evaluated results were sent back to the investigator (the author). Therefore, the raters knew which session they were scoring. However, the raters were not told that the experimental group went through audio-visual training, while the control group did not.

The method used for the evaluation of the subjects' production was the same as that in the pilot study. Instead of evaluating the correctness of a whole stimulus word, the 4 raters were asked to only focus on the subjects' production of /θ, s, ð, z/ in each stimulus word. Each rater was asked to evaluate the accuracy of all the produced stimulus words (3 repetitions of each stimulus word in one test). In each test, the highest accuracy score for each stimulus word amongst the three repetitions was adopted. After that, the accuracy scores given by the 4 raters were compared and checked. For the scoring of the same word, if the difference between the highest and the lowest scores was larger than 3 scale units, the highest and the lowest scores were ignored. The final result for each word's accuracy was the average score of the remaining scores. For instance, in the evaluation of *think* produced by S7 in the pre-test, the scores given by the 4 raters were 4, 3, 1, 5 respectively. The difference between the highest score 5 and the lowest score 1 was 4 (5-1=4), which was larger than 3 scale units, thus the scores 1 and 5 were ignored. The final accuracy score of S7's production of *think* was: (4+3)/2=3.5. If the difference between the highest and the lowest scores was 3 or less than 3 scale units, the 4 given scores were used. For example, S1's production of *thousand* in pre-test was given the scores: 2, 1, 2, 4. All four scores were adopted. Therefore, the final score of S1's production of *thousand* was (2+1+2+4)/4=2.25. In each test, the subject's final accuracy in the production of each target speech sound was the mean result for all the stimulus words that contained the sound, which was then transferred to percentage form[6].

In order to test whether the 4 raters' evaluation was reliable, a reliability test was carried out with SPSS. All the scores for each of the stimulus words that were given by the 4 raters were inputted in SPSS. It turned out that their evaluation was highly reliable both for the subjects' production of /θ/ (*Cronbach's Alpha*=0.931, p<0.001) and /ð/ (*Cronbach's Alpha*=0.923, p<0.001).

5.6 Results

The following sections (from section 5.6.1 to 5.6.5) present and interpret the subjects' perception and production results in mid-test 1, mid-test 2 and the post-test. The results in the post-test are compared with those of the pre-test, mid-test 1, and mid-test 2. The aim is to investigate whether subjects' accuracy in the perception and/or production of the target contrasts was improved during and at the end of the training programme. Qualitative data collected with the questionnaire (Appendix 7) are presented. A *Repeated-measures ANOVA* was performed to detect factors that had a significant effect on the subjects' perception and/or production performance. These results serve to answer the research questions of the present study.

[6] For instance, there were 15 stimulus words that contained /θ/. The average scores (from the scores given by the 4 raters) of S1's production of /θ/ in the pre-test were: 2.75 for *think*, 2.75 for *thousand*, 1.50 for *throne*, 3.25 for *three*, 2.75 for *theatres*, 3.00 for *Cathy*, 4.00 for *anything*, 2.25 for *athlete*, 3.00 for *method*, 2.50 for *wealthy*, 5.50 for *breath*, 3.50 for *teeth,* 4.50 for *cloth*, 3.75 for *fourth*, 7.00 for *bath.* S1's accuracy in the production of /θ/ in pre-test was: (2.75+2.75+1.50+3.25+2.75+3.00+4.00+2.25+3.00+2.50+5.50+3.50+4.50+3.75+7.00)/15/10*100%=34.33%

5.6.1 Perception test results of the experimental group

5.6.1.1 Overall results

The perception test results were obtained from the AXB tests conducted before the training programme (pre-test), at the end of the 3rd training session (mid-test 1), at the end of the 6th training session (mid-test 2), and at the end of the training programme (post-test). As shown in Figure 5.3 below, among the subjects in the experimental group, all the subjects' accuracy in the perception of the two contrasts improved to different degrees during and at the end of the training programme (see Appendix 10 for individual subjects' perception results in the four tests). Specifically, in the post-test, more than half of the subjects achieved an accuracy of above 90% in the perception of /θ/-/s/ (n=17) and /ð/-/z/ (n=16). With the exception of S29, whose accuracy was 79.63% in the perception of /ð/-/z/, the remaining subjects' accuracy was all between 80% and 90% in the perception of both of the target contrasts. Moreover, in mid-test 1, S10 achieved an accuracy of above 90% both in the perception and production of the target contrasts, and so she was dropped from the following training sessions and tests. Similarly, S3 was dropped from the last 3 training sessions and post-test. His accuracy was above 90% both in the perception and production tests in mid-test 2, thus there would have been little room for further improvement.

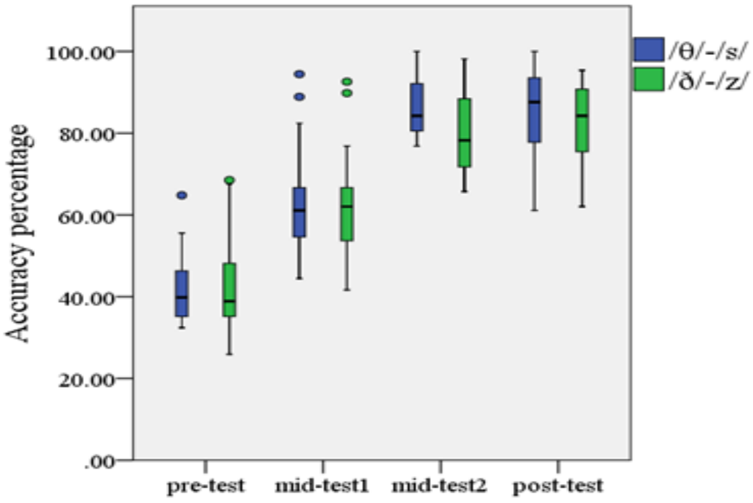


Figure 5.3 Boxplots of the experimental group's perception of /θ/-/s/ and /ð/-/z/ in the pre-test, mid-test 1, mid-test 2 and the post-test.

Table 5.2 presents the experimental group's improvement from pre-test to tests after-training in the perception of the target contrasts. It can be seen that in the perception of both of the contrasts, these subjects' largest degree of improvement occurred after the first three training sessions, with the mean percentage of 21.91% in the perception of /θ/-/s/, and 20.31% in the perception of /ð/-/z/. In the perception of both of the contrasts, their mean accuracy increased about 15% both after the 6th and the 9th training session. Moreover, it seems the subjects increased slightly more in the perception of /θ/-/s/ than for /ð/-/z/.

	Perception of /θ/-/s/			**Perception of /ð/-/z/**		
	(mid-test 1—pre-test)%	(mid-test 2—pre-test)%	(post-test—pre-test)%	(mid-test 1—pre-test)%	(mid-test 2—pre-test)%	(post-test—pre-test)%
N Valid	29	28	27	29	28	27
Mean	21.91	37.6	51.834	20.31	35.58	49.38
Median	21.3	37.5	53.7	21.3	36.11	49.07
Std. Deviation	6.61	5.81	6.51	4.73	5.63	6.12
Minimum	9.26	28.71	40.74	12.04	26.85	38.89
Maximum	38.89	49.08	65.74	29.63	47.22	64.81

Table 5.2. Experimental group's perception improvement – from pre-test to mid-test 1, from pre-test to mid-test 2, and from pre-test to post-test.

test	20%—<30%		30%—<40%		40%—<50%		50%—<60%		60%—<70%		70%—<80%		80%—<90%		90%—100%	
	/θ/-/s/	/ð/-/z/	/θ/-/s/	/ð/-/z/	/θ/-/s/	/ð/-/z/	/θ/-/s/	/ð/-/z/	/θ/-/s/	/ð/-/z/	/θ/-/s/	/ð/-/z/	/θ/-/s/	/ð/-/z/	/θ/-/s/	/ð/-/z/
pre-test	0	3	15	14	9	7	4	3	1	2	0	0	0	0	0	0
mid-test1	0	0	0	0	3	2	9	11	10	10	3	4	3	1	1	1
mid-test2	0	0	0	0	0	0	0	0	4	5	14	16	6	6	4	1
post-test	0	0	0	0	0	0	0	0	0	0	0	1	12	12	15	14

Table 5.3 The distribution of the number of subjects in different ranges of accuracy in the perception of /θ/-/s/ and /ð/-/z/ (experimental group).

More specifically, as show in Table 5.3 above, in the pre-test for the perception of /θ/-/s/, the majority of the subjects' accuracy was below 60% (n=28). Only S10's accuracy was above 60% (64.81%). In the perception of /ð/-/z/, the situation was even worse. There were 3 subjects whose accuracy was below 30%. Most of the subjects' accuracy was between 30% and 60%. Comparatively, S3 and S10 performed better than the rest of the subjects, with an accuracy of 37.59% and 68.52% respectively.

In mid-test 1, however, none of the subjects' accuracy was below 30% in the perception of /θ/-/s/ and /ð/-/z/. There were 3 subjects whose accuracy was between 40% and 50% in the perception of /θ/-/s/, whereas only 2 subjects' accuracy fell in this range of accuracy in the perception of /ð/-/z/. The majority of the subjects' accuracy was between 50% and 80% both in the perception of /θ/-/s/ (n=22) and /ð/-/z/ (n=25). Only S10's accuracy was above 90% in the perception of both of the contrasts. The remaining subjects achieved an accuracy of between 80% and 90%. The remaining subjects' accuracy was between 90% and 100%.

Similarly, further improvement was found in mid-test 2. Most of the subjects achieved an accuracy of between 60% and 90% both in the perception of /θ/-/s/ (n=24) and /ð/-/z/ (n=27).

In the post-test, except for S5, whose accuracy in the perception of /ð/-/z/ was 79.63%, the remaining subjects' accuracy was above 80% both in the perception of /θ/-/s/ and /ð/-/z/. Moreover, about half of them achieved an accuracy of above 90% in the perception of /θ/-/s/ (n=15) and /ð/-/z/ (n=14).

Given that there were two possible response choices in the AXB test, it was important to correct any potential bias in the subjects' responses (Hazan et al., 2005). The accuracy of the subjects' responses in the four tests was therefore converted to *d-prime*[7] scores (the signal detectability measure *d-prime*) to further examine their perception improvement. Individual subjects' responses were input to SPSS first. Their *hit-rate* and *false-alarm-rate* were computed by *Crosstabs*, which was then converted to *d'* scores by Excel. Due to most of the subjects' accuracy in the pre-test being below 50% (*hit-rate*

[7] The calculation of d' is by the formula *d'*= NORMINV(hit-rate,0,1) - NORMINV(false-alarm-rate,0,1) with Excel. The highest possible d' (greatest sensitivity) is 6.93, and the effective limit (using .99 and .01) is 4.65.

was less than *false-alarm-rate*), their *d'* scores were negative (see Table 5.3). Nonetheless, it was because the subjects had serious difficulty in the perception of the target contrasts that they were selected for inclusion in the main study. Thus the negative *d'* scores were reasonable.

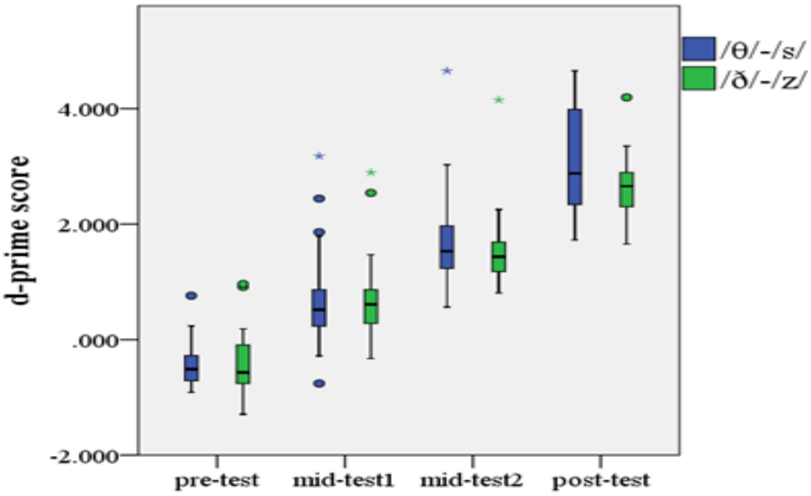


Figure 5.2 Boxplots of the experimental group's *d'* scores in the pre-test, mid-test 1, mid-test 2 and the post-test.

		Perception of /θ/-/s/				Perception of /ð/-/z/			
		pre-test	mid-test 1	mid-test 2	post-test	pre-test	mid-test 1	mid-test 2	post-test
N	Valid	29	29	28	27	29	29	28	27
	Missing	0	0	1	2	0	0	1	2
Mean		-0.44	0.71	1.72	3.09	-0.43	0.7	1.55	2.63
Median		-0.51	0.52	1.53	2.87	-0.569	0.61	1.43	2.69
Mode		-0.91	0.71	1.12	2.34	-0.1	0.33	0.91	2.54
Std. Deviation		0.41	0.84	0.83	0.93	0.55	0.7	0.65	0.59
Variance		0.17	0.71	0.7	0.86	0.3	0.49	0.42	0.34
Range		1.67	3.94	4.09	2.93	2.26	3.22	3.34	2.54
Minimum		-0.91	-0.76	0.56	1.73	-1.29	-0.33	0.81	1.66
Maximum		0.76	3.18	4.65	4.65	0.96	2.89	4.15	4.19
Sum		-12.64	20.68	48.22	83.48	-12.46	20.35	43.46	71.1

Table 5.4 The experimental group's calculated *d'* scores in the pre-test, mid-test 1, mid-test 2, and the post-test (S10 was dropped from mid-test 2 and the post-test; subject 3 was dropped from the post-test).

Figure 5.2 and Table 5.4 above provide us with a more detailed depiction of the subjects' improvement in accuracy in the perception of the two contrasts, and show that the experimental group's *d'* scores rise linearly from the pre-test to the post-test in terms of mean maximum, minimum value, and the range of most of the subjects' *d'* scores.

5.6.1.2 Pre-test vs. mid-test 1

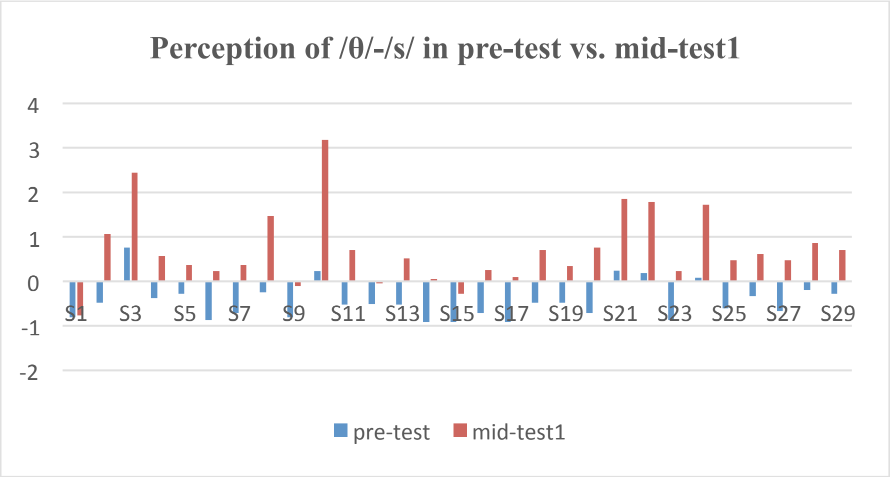


Figure 5.3 Individual subjects' *d'* scores in the perception of /θ/-/s/ in pre-test *vs.* mid-test 1 (experimental group).

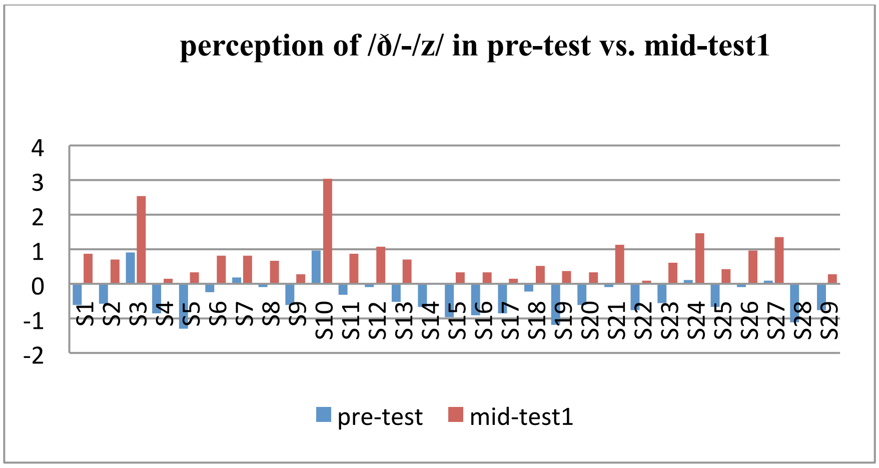


Figure 5.4 Individual subjects' *d'* scores in the perception of /ð/-/z/ in pre-test *vs.* mid-test 1 (experimental group).

As shown in Figure 5.3 and Figure 5.4 above, all the subjects showed improvement in the perception of both /θ/-/s/ and /ð/-/z/. The mean values of their *d'* scores were -0.43 and 0.71 in the perception of the two contrasts in the pre-test and mid-test 1 respectively. The majority of the subjects' *d'* scores were negative in the pre-test. This may illustrate that the subjects had serious difficulty in the perception of the target contrasts. In mid-test 1, however, except for S1's accuracy in the perception of /θ/-/s/ and S15's accuracy in the perception of the two contrasts (which was negative in terms of *d'*), the remaining subjects' *d'* scores were all above 0. Among all the subjects in the experimental group, S10 achieved the highest accuracy in the perception of /θ/-/s/ (3.18) and /ð/-/z/ (2.19).

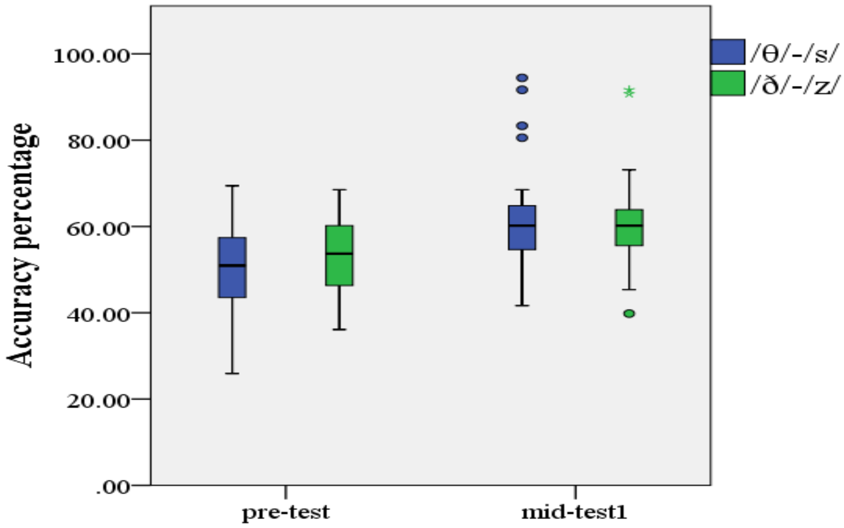


Figure 5.5 Boxplots of the experimental group's accuracy in the perception of the target contrasts in the pre-test vs. mid-test 1.

Figure 5.5 provides us with a visual depiction of the experimental group's perception improvement in mid-test 1 compared against the pre-test. Their mean accuracy was below the mean percentage, while it was more evenly distributed across the subjects in the perception of /θ/-/s/. However, the subjects' maximum and minimum percentage in the perception of /θ/-/s/ was slightly lower than that for /ð/-/z/ in mid-test 1. On the whole, after the first three sessions of audiovisual training, all the subjects' accuracy in the perception of the target contrasts was improved.

Among all the subjects, S10 achieved an accuracy of 94.44% (*d'*=3.18) and 92.59% (*d'*=3.03) in the perception of /θ/-/s/ and /ð/-/z/ respectively. Her accuracy in the production of the target contrasts was above 90% (see the production results in section 5.6.4). Therefore, S10 was assumed to not need further training, and was dropped from the following training sessions and tests.

5.6.1.3 Mid-test 1 vs. mid-test 2

After the 6th training session, the perception performance of the remaining 28 subjects in the experimental group was tested with the AXB task again (the order of the stimuli was randomized with the Praat program). Their *d'* scores further increased in mid-test 2 compared against mid-test 1. As show in Figure 5.6 and Figure 5.7, in the perception of the two contrasts, none of the subjects' *d'* score was below 0 in mid-test 2. In mid-test 2, their mean *d'* scores were 1.72 and 1.55 in the perception of /θ/-/s/ and /ð/-/z/ respectively, which were higher than in mid-test 1.

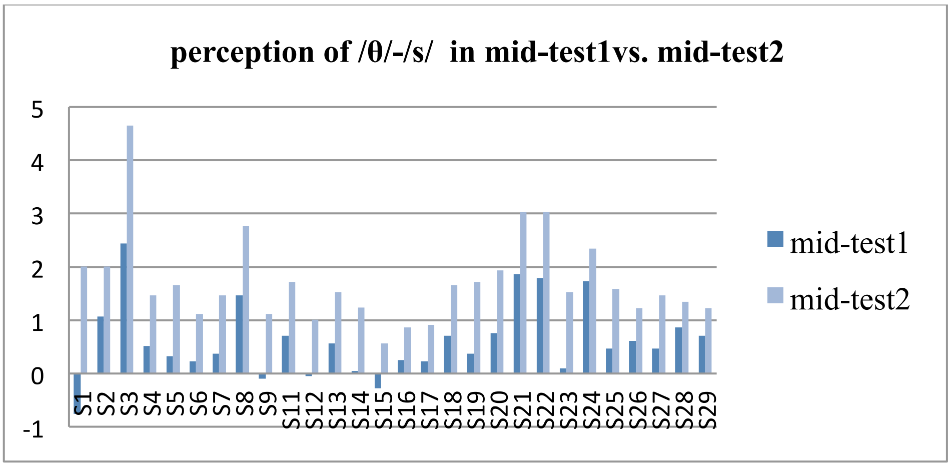
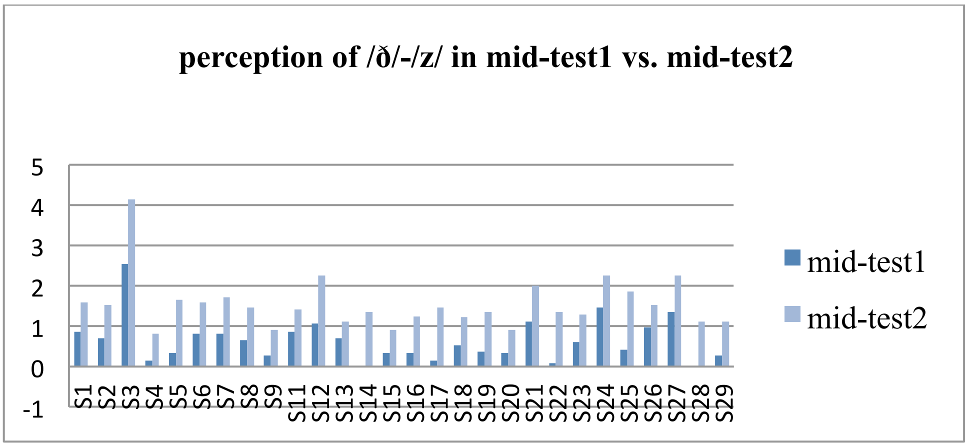


Figure 5.6 Individual subjects' *d'* scores in the perception of /θ/-/s/ in mid-test 1 vs. mid-test 2.



.

Figure 5.7 Individual subjects' *d'* scores in the perception of /ð/-/z/ in mid-test 1 vs. mid-test 2.

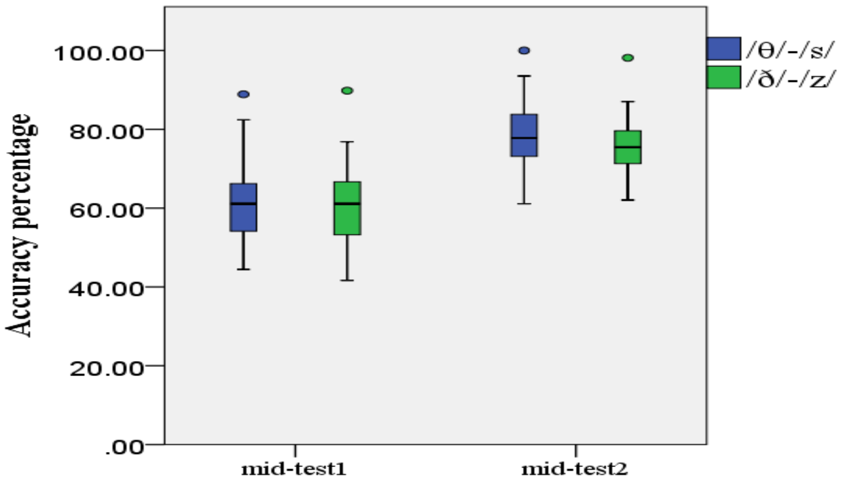


Figure 5.8 Boxplots of the experimental group's accuracy in the perception of the target contrasts in the pre-test vs. mid-test 1.

Figure 5.8 depicts the experimental group's perception improvement in mid-test 2 compared against mid-test 1 in terms of accuracy percentage. As show in the figure, their mean, maximum and minimum accuracy in the perception of the two contrasts all increased from mid-test 1 to mid-test 2. Nonetheless, it seems the subjects performed better in the perception of /θ/-/s/ than /ð/-/z/ in terms of maximum, medium as well as the range of the majority of the subjects' accuracy.

Moreover, S3 achieved the highest accuracy among the remaining subjects. That is, 98.15% (d'=4.65) in the perception of /θ/-/s/ and 100% (d'=4.15) in the perception of /ð/-/z/. In mid-test 2, his accuracy in the production of the two contrasts was above 90% (see the production results in section 5.6.4). Therefore, S3 was assumed not to need further training, and was dropped from the following training sessions and the post-test.

5.6.1.4 Mid-test 2 vs. post-test

At the end of the training programme, the remaining 27 subjects of the experimental group had been trained over 9 sessions. Then the AXB test was carried out again (the order of the stimuli was randomized with the Praat program). As show in Figure 5.9 and Figure 5.10, compared with in mid-test 2, the subjects further improved their accuracy in the post-test. Their mean *d'* score was 3.09 in the perception of /θ/-/s/, and 2.63 in the perception of /ð/-/z/, which were higher than in mid-test 2.

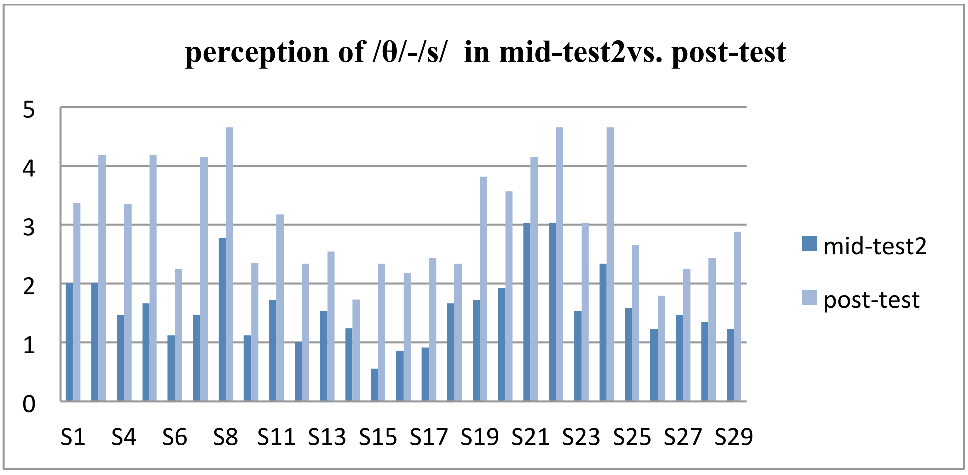


Figure 5.9 Individual subjects' *d'* scores in the perception of /θ/-/s/ in mid-test 2 vs. post-test (experimental group).

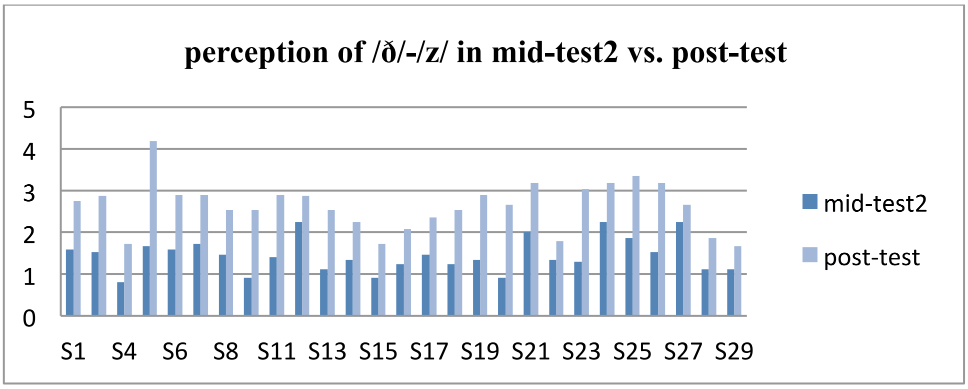


Figure 5.10 Individual subjects' *d'* scores in the perception of /ð/-/z/ in mid-test 2 vs. post-test (experimental group).

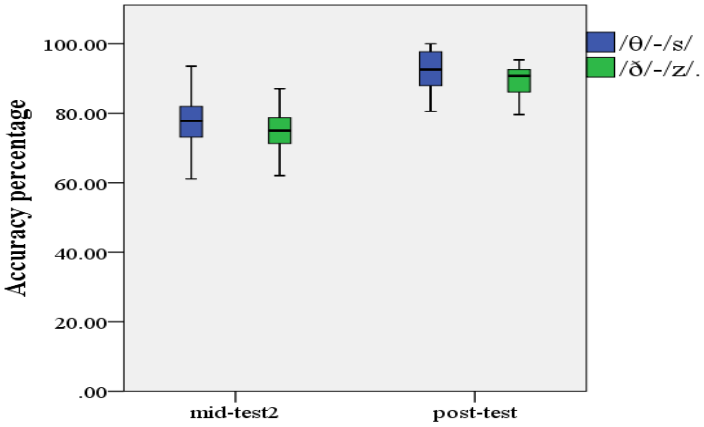


Figure 5.11 Boxplots of the experimental group's accuracy in the perception of the target contrasts in mid-test 2 vs. post-test.

The boxplots in Figure 5.11 show the experimental group's perception improvement in the post-test compared against mid-test 2 in terms of accuracy percentage. As in the comparison between the pre-test and mid-test 1, and between mid-test 1 and the post-test, the remaining subjects displayed much higher accuracy in the perception of the two contrasts in the post-test compared with in mid-test 2. Specifically, they displayed comparatively higher minimum, maximum, and mean scores as well as the range of the majority of the subjects' accuracy in post-test.

5.6.1.5 Statistical analysis of the perception test results of the experimental group

The results above indicate that the experimental group's accuracy in the perception of the target contrasts improved stably and linearly from pre-test to post-test. However, it was unclear whether their improvement could be attributed to the audiovisual training programme, because there were other factors that may have a significant impact on their perception performance. Therefore, it is necessary to have a look at the subjects' answers in the questionnaire.

Factors	**Answer**	**Num ber**	**Percent age**
Gender	male female	14 15	48.28% 51.72%
Years of English study	6 years 7 years 8 years	13 15 1	46.00% 46.90% 7.10%
AO	13 years old 14 years old	17 12	58.40% 41.60%
Age	19 20 21 22	9 12 6 2	31.80% 40.80% 20.30% 7.10%
Majority motivation	hobby The need to get high scores in English exams	1 28	2.70% 97.30%
Learn English in spare time	No Yes	4 25	14.20% 85.80%
Institute of English learning	Public school/university	29	100.00 %
Use English on a daily basis (except for study)?	No	29	100.00 %
Travelled/lived aboard?	No Yes	29 0	100.00 % 0.00%

Table 5.5 Data collected from the questionnaire (experimental group).

As shown in Table 5.5 above (see Appendix 2 for individual subjects' answers to each question), the age of the subjects in the experimental group ranged from 19 to 22 years old. They had been learning English from 6 to 8 years by the time of the present study. With respect to their primary motivation for learning English as an L2, except S3 who reported that he had been learning English as a hobby, the remaining subjects all reported that they studied L2-English primarily due to the need to get high scores in English exams. Moreover, 85.80% of the subjects learned English in their spare time. The amount of time and the ways in which they had been learning English in their spare time were quite similar to each other (see Appendix 2). Specifically, the majority of them reported that they read articles, did exercises, and recited vocabulary in English textbooks. A few of them (S10, S13, S14, S18, S20) also watched English movies or listened to English songs in addition to doing exercises. Only S3 spent 1-2 hours per day reading English newspapers, watching English movies and listening to English songs. Moreover, they all learned English at public schools and at the same university. The English educational system in public schools and universities all follow the British English system, which is embodied in their English textbooks. None of them had any chance to use English on a daily basis, or had ever travelled/lived in English speaking countries.

On the whole, there was no difference among the subjects concerning the factors of the institute in which they had been learning L2-English, whether they had any chance to use English on a daily basis, and whether travelled/lived abroad. Moreover, they were similar to each other in terms of years of L2-English learning, age, primary motivation for L2-English learning, the amount of time spent using English on a daily basis, as well as the ways in which they had been learning English. These factors, therefore, were not adopted as a *between-subjects factor* for statistical analysis regarding their influence on the subjects' perception performance.

Considering that each subject was tested with the same AXB test 4 times, a *repeated-measures ANOVA* was conducted to detect which factor(s) may have had a significant impact on their perception performance. Moreover, *sphericity* test (p<0.001) and *normal distribution* (p<0.05) results indicated that a *repeated-measures ANOVA* was appropriate for the analysis of the data. Given that there might be bias among the subjects' responses in the AXB test, the subjects *d'* scores were employed when performing the statistical analysis, which were coded as the dependent variable. S3 was dropped from the post-test, and so his *d'* scores in this test was coded as missing value.

Likewise, the responses of S10 were coded as a missing value in mid-test 2 and the post-test, because she was dropped from the two tests.

Based on the descriptive statistics, we would expect training to be a significant factor, as the aim was to explore whether the audiovisual training programme displayed a significant effect on the experimental group's perception performance. Given that each subject was tested 4 times with the AXB task, the factor *training* was coded into 4 levels, which were *non-trained*, *trained for 3 sessions*, *trained for 6 sessions,* and *trained for 9 sessions*. Moreover, the factor *phonetic environment* was further divided into two factors – *vowel context* and *phonetic positions*. The factor *vowel context* was also divided into 3 levels, in which the stimuli being embedded in */i/*, */a/* and */u/*. The factor *phonetic position* was divided into 3 levels, which were *initial*, *medial* and *final*.

Considering that the same subjects were tested 4 times, a *repeated-measures ANOVA* was carried out. That is, *training* (4 levels) * *vowel context* (3 levels) was defined as a within-subjects factor, with *gender* defined as a between-subjects factor; *training* (4 levels) * *phonetic position* (3 levels) was defined as a within-subjects factors with *gender* as the between-subjects factor.

5.6.2 Statistical analysis of the experimental group's perception test results

5.6.2.1 Factors of significant effect on the experimental group's perception of /θ/-/s/ and /ð/-/z/

factor	df and F-value	Sig.	Partial Eta Squared
training	F(1,25)=127.262	p<0.001	η2=0.853
gender	F(1,25)=154.389	p<0.001	η2=0.861
training * gender	F(1,25)=5.266	p=0.030	η2=0.714
phonetic position	F(2,50)=6.911	p=0.002	η2=0.855
phonetic position* training	F(6, 156)=2.339	p=0.034	η^2=0.083

Table 5.6 Factors which were significant for the experimental group's accuracy in the perception of /θ/-/s/ in the pre-test, mid-test 1, mid-test 2 and the post-test.

factor	df and F-value	Sig.	Partial Eta Squared
training	F(3, 75)=90.317	p<0.001	η2=0.749
gender	F(1, 25)=233.281	p<0.001	η2=0.903
training * gender	F(3, 75)=3.458	p=0.029	η2=0.657
phonetic position	F(2, 50)=34.346	p<0.001	η2=0.579
phonetic position* training	F(6, 150)=3.477	p=0.003	η^2=0.122

Table 5.7 Factors which were significant for the experimental group's accuracy in the perception of /ð/-/z/ in the pre-test, mid-test 1, mid-test 2 and the post-test.

1. *Training*

The audiovisual training effect was the key factor of the present study. It examined whether audiovisual training displayed a significant effect on the subjects' perception performance. As a within-subjects factor, *training* was found to display a significant effect on the subjects' perception performance for both of the contrasts (see Table 5.6 and Table 5.7).

In order to detect whether the more training sessions the subjects went through, the more likely they were to be able to perceive the target contrasts correctly, a *Post Hoc Test* was carried out with the subjects' *d'* scores in the perception tests. As shown in Table 5.8, the subjects' mean improvement in the perception of both /θ/-/s/ and /ð/-/z/ from the pre-test to mid-test 1, mid-test 2, and the post-test was statistically significant (*p<0.001*). As show in the column *Mean Difference*, the more training sessions the subjects received, the higher *d'* scores they received in the perception of the target contrasts.

(I) tests	(J) tests	Mean Difference (I-J)		Std. Error		Sig.		95% Confidence Interval			
								Lower Bound		Upper Bound	
		/θ/-/s/	/ð/-/z/	/θ/-/s/	/ð/-/z/	/θ/-/s/	/ð/-/z/	/θ/-/s/	/ð/-/z/	/θ/-/s/	/ð/-/z/
Pre-test	Mid-test 1	-1.148	-1.133	0.2	0.168	0	0	-1.545	-1.47	-0.751	-0.8
	Mid-test 2	-2.221	-2.04	0.202	0.17	0	0	-2.622	-2.38	-1.82	-1.7
	Post-test	-3.754	-3.257	0.204	0.171	0	0	-4.158	-3.6	-3.349	-2.92
Mid-test 1	Pre-test	1.148	1.133	0.2	0.168	0	0	0.751	0.8	1.545	1.47
	Mid-test 2	-1.073	-0.906	0.202	0.17	0	0	-1.473	-1.24	-0.672	-0.57
	Post-test	-2.605	-2.123	0.204	0.171	0	0	-3.01	-2.46	-2.201	-1.78
Mid-	Pre-test	2.221	2.04	0.202	0.17	0	0	1.82	1.7	2.622	2.38

test 2	Mid-test 1	1.073	0.906	0.202	0.17	0	0	0.672	0.57	1.473	1.24
	Post-test	-1.533	-1.217	0.206	0.173	0	0	-1.941	-1.56	-1.125	-0.87
Post-test	Pre-test	3.754	3.257	0.204	0.171	0	0	3.349	2.92	4.158	3.6
	Mid-test 1	2.605	2.123	0.204	0.171	0	0	2.201	1.78	3.01	2.46
	Mid-test 2	1.533	1.217	0.206	0.173	0	0	1.125	0.87	1.941	1.56

Table 5.8 *Post Hoc Tests* for the experimental group's perception performance in the four AXB tests.

Further supporting evidence is available from Table 5.9, which shows the estimated marginal means of the subjects' perception performance before being trained, after being trained for 3 sessions, 6 sessions, and 9 sessions. It can be seen that their mean scores, lower bound, as well as upper bound of scores in the perception of both /θ/-/s/ and /ð/-/z / all increased stably and linearly from pre-test to post-test. Figure 5.12 and Figure 5.13 provide us with a visual depiction of these findings.

training	Mean		Std. Error		95% Confidence Interval			
					Lower Bound		Upper Bound	
	/θ/-/s/	/ð/-/z/	/θ/-/s/	/ð/-/z/	/θ/-/s/	/ð/-/z/	/θ/-/s/	/ð/-/z/
Non-trained	-1.281	-0.451	0.438	0.367	-2.189	-1.207	-0.372	0.305
Trained for 3 sessions	2.825	2.537	0.374	0.413	2.05	1.686	3.601	3.389
Trained for 6 sessions	4.558	4.324	0.116	0.076	4.318	4.168	4.798	4.481
Trained for 9 sessions	5.58	5.156	0.105	0.112	5.362	4.926	5.797	5.386

Table 5.9 Estimated marginal means of the experimental group's perception of /θ/-/s/ and /ð/-/z/ across the four AXB tests.

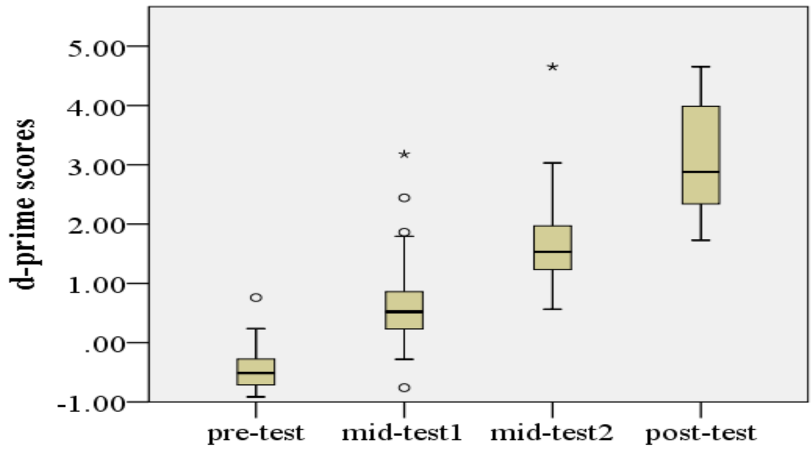


Figure 5.12 Boxplots of the experimental group's perception of /θ/-/s/ in the pre-test, mid-test 1, mid-test 2 and the post-test.

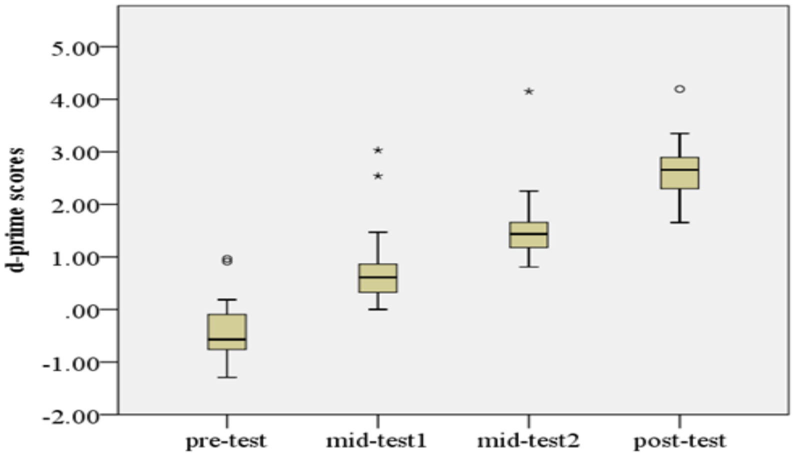


Figure 5.13 Boxplots of the experimental group's perception of /ð/-/z/ in the pre-test, mid-test 1, mid-test 2 and the post-test.

2. Gender and its interaction with *training*

As shown in Table 5.6 and Table 5.7, *gender* as a between-subjects factor displayed a significant effect on the experimental group's perception of the target contrasts. Moreover, its interaction with *training* was also shown to have a significant effect on their perception performance.

Table 5.10 below presents the comparison of the female and male subjects' mean *d'* scores in the perception of /θ/-/s/ and /ð/-/z/. In the perception of /θ/-/s/, at first, the female subjects' mean *d'* scores were similar to (in pre-test) and even slightly higher (in mid-test 1) than that of the male subjects. After being trained for 6 more sessions, the male subjects showed higher mean scores than the female. Moreover, as shown in Figure 5.14 below, the majority of the male subjects *d'* scores fell in a relatively higher range than those of the female subjects in mid-test 2 and the post-test. In the perception of /ð/-/z/, however, the male subjects performed better than the females from the pre-test to mid-test 1, mid-test 2 and the post-test. The boxplot in Figure 5.15 indicates that both in the pre-test and mid-test 1, the majority of the male subjects' *d'* scores fell in narrower but similar ranges to the majority of the female subjects. In mid-test 2 and the post-test, however, the majority of the male subjects' *d'* scores fell in higher ranges than the majority of the female subjects. On the whole, with the training sessions carried out, the male subjects performed better than the female subjects.

test	pre-test		mid-test 1		mid-test 2		post-test	
gender	female	male	female	male	female	male	female	male
mean *d'* score in perceiving /θ/-/s/	-0.43	-0.44	0.74	0.68	1.53	1.94	2.86	3.39
mean *d'* score in perceiving /ð/-/z/	-0.46	-0.39	0.64	0.77	1.42	1.71	2.45	2.87

Table 5.10 The female and male subjects of the experimental group's mean *d'* scores in the perception of /θ/-/s/ and /ð/-/z/.

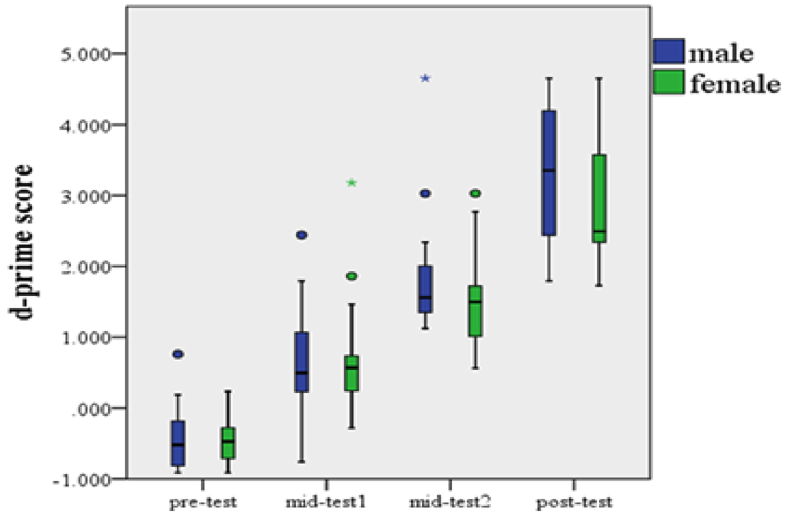


Figure 5.14 Boxplots of the *d'* scores in the perception of /θ/-/s/ for female and male subjects in the experimental group.

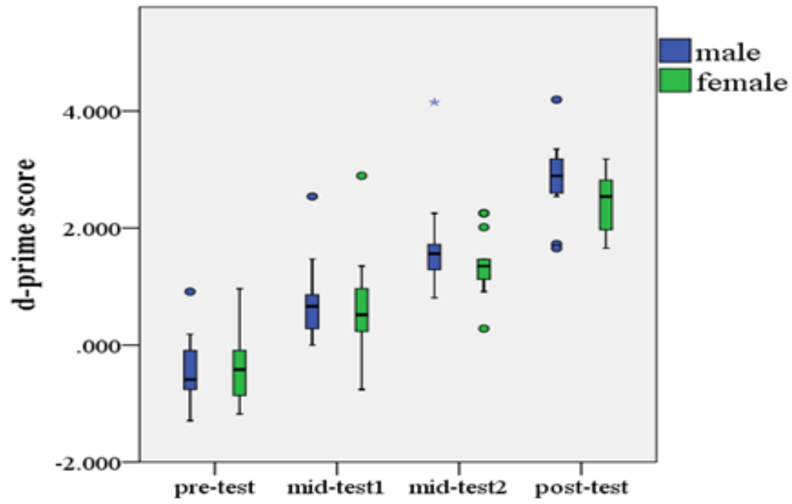


Figure 5.15 Boxplots of the *d'* scores in the perception of /ð/-/z/ for female and male subjects in the experimental group.

3. *Phonetic position* and its interaction with *training*

The factor *phonetic position* and its interaction with *training* were also found to be statistically significant for the experimental group's perception of the target contrasts.

According to the *Post Hoc Test* results (see Table 5.11), when /θ/-/s/ was embedded in initial position, the subjects achieved a higher mean *d'* score than in medial and final positions. Specifically, the mean difference was 3.72 between initial and medial positions, 5.32 between initial and final positions, and 1.60 between medial and final positions. The mean differences were all found to be statistically significant (p<0.005). On the whole, the subjects performed best when /θ/-/s/ was embedded in initial position, and worst when these phonemes were embedded in final position.

In the perception of /ð/-/z/ (see Table 5.12), the mean difference was 0.83 between initial and medial positions. However, this finding was revealed to be non-significant (p=0.101). The mean difference was 2.14 between the perception of /ð/-/z/ in initial and final positions, and 1.31 between medial and final positions, both of which were statistically significant (p<0.005). On the whole, the subjects performed better when /ð/-/z/ was embedded in initial and medial positions than when in final position.

(I) position	(J) positions	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
initial	medial	3.72	0.74	0	-5.06	-2.39
	final	5.32	0.68	0	-6.66	-3.97
medial	initial	-3.72	0.67	0	2.39	5.06
	final	1.6	0.68	0.021	-2.94	-0.25
final	initial	-5.32	0.71	0	3.97	6.66
	medial	-1.6	0.7	0.021	0.25	2.94

Table 5.11 *Post Hoc Tests* of the experimental group's perception of /θ/-/s/ in different phonetic positions.

(I) positions	(J) positions	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
initial	medial	0.83	0.51	0.101	-0.16	1.82
	final	2.14	0.49	0	1.15	3.13
medial	initial	-0.83	0.53	0.101	-1.82	0.16
	final	1.31	0.57	0.01	0.32	2.3
final	initial	-2.14	0.66	0	-3.13	-1.15
	medial	-1.31	0.75	0.01	-2.3	-0.32

Table 5.12 Post Hoc Tests of the experimental group's perception of /ð/-/z/ in different phonetic positions.

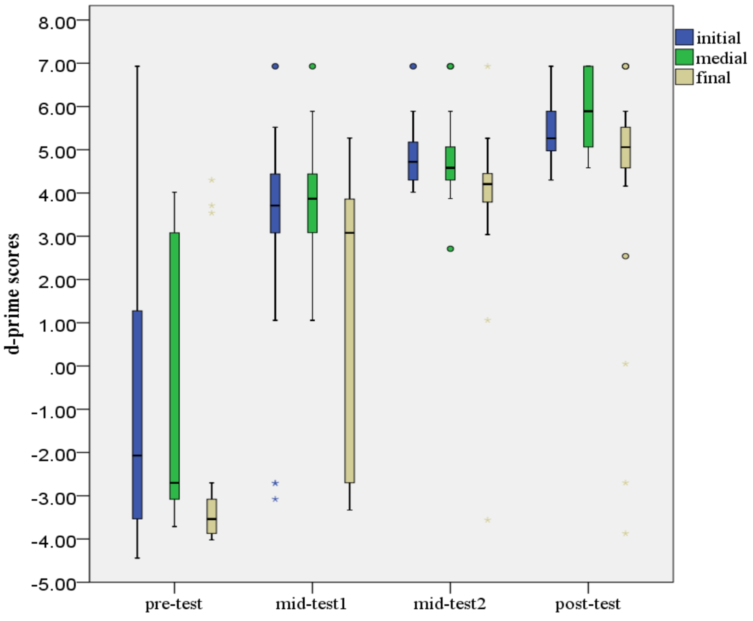


Figure 5.16 Boxplot of the experimental group's perception of /θ/-/s/ in different phonetic positions.

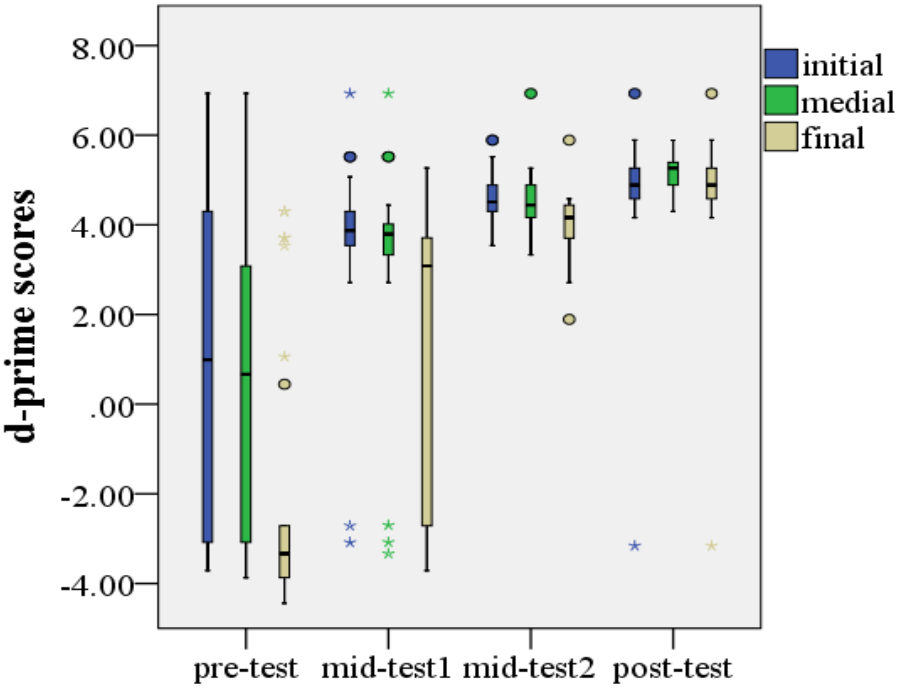


Figure 5.17 Boxplot of the experimental group’s perception of /ð/-/z/ in different phonetic positions.

Figure 5.16 and Figure 5.17 provide us with visual depictions of the findings. In the perception of /θ/-/s/ and /ð/-/z/, the subjects showed the lowest median when they were in final position. The majority of the subjects’ *d’* scores also fell in the lowest range when the target contrasts were embedded in final position.

5.6.2.2 Factors of non-significant effect on the experimental group’s perception of /θ/-/s/ and /ð/-/z/

According to the results from a *repeated-measure ANOVA*, the factors listed in Table 5.13 were found to be non-significant for the experimental group’s perception of /θ/-/s/ and /ð/-/z/ across the 4 AXB tests (*p>0.05*).

factor	df and F-value		Sig.		Partial Eta Squared	
Vowel context	/θ/-/s/	/ð/-/z/	/θ/-/s/	/ð/-/z/	/θ/-/s/	/ð/-/z/
	F(2, 50)=0.003	F(2, 50)=0.528	p=0.997	p=0.593	η^2=0.003	η^2=0.021
Vowel context *training	F(6, 150)=0.228	F(6, 150)=0.387	p=0.967	p=0.886	η^2=0.009	η^2=0.015
Vowel context * gender	F(2,50)=1.611	F(2, 50)=0.660	p=0.210	p=0.521	η^2=0.061	η^2=0.026
Vowel context * gender* training	F(6, 150)=0.997	F(6, 150)=0.960	p=0.430	p=0.455	η^2=0.038	η^2=0.037
phonetic position*gender	F(2, 50)=0.434	F(2, 50)=1.138	p=0.650	p=0.329	η^2=0.017	η^2=0.044
phonetic position*gender *training	F(6, 150)=0.211	F(6, 150)=1.297	p=0.973	p=0.262	η^2=0.008	η^2=0.049

Table 5.13 Factors which were statistically **non-significant** for the experimental group's perception of /θ/-/s/ and /ð/-/z/.

As a within-subjects factor, *vowel context* was found to not display a significant effect on the experimental group's perception of the target contrasts. Table 5.14 depicts the subjects' mean *d'* scores in the perception of /θ/-/s/ in different vowel contexts. The subjects' mean difference was 0.0006 between /i/ and /a/ contexts. It was 0.11 between /i/ and /u/ contexts, as well as between /a/ and /u/ contexts. However, all of these differences were statistically non-significant (*p>0.005*).

Similarly, in the perception of /ð/-/z/ (see Table 5.15), the mean difference between their *d'* scores was 0.18 between /i/ and /a/ contexts, 0.53 between /i/ and /u/ contexts, but 0.71 between /a/ and /u/ contexts. However, as in the perception of /θ/-/s/, these differences were statistically non-significant (*p>0.005*).

(I) Vowel context	(J) Vowel context	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
/i/	/a/	0.0006	0.46	1	-0.91	0.91
	/u/	-0.11	0.46	0.82	-1.02	0.81
/a/	/i/	0.0006	0.46	1	-0.91	0.91
	/u/	-0.11	0.46	0.82	-1.02	0.8
/u/	/i/	0.11	0.46	0.82	-0.81	1.02
	/a/	0.11	0.46	0.82	-0.8	1.02

Table 5.14 *Post Hoc Test* of the experimental group’s perception of /θ/-/s/ in different vowel contexts.

(I) Vowel context	(J) Vowel context	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
/i/	/a/	0.18	0.96	0.85	-1.72	2.08
	/u/	-0.53	0.96	0.58	-2.43	1.37
/a/	/i/	-0.18	0.96	0.85	-2.08	1.72
	/u/	-0.71	0.96	0.46	-2.61	1.19
/u/	/i/	0.53	0.96	0.58	-1.37	2.43
	/a/	0.71	0.96	0.46	-1.19	2.61

Table 5.15 *Post Hoc Tests* of the experimental group’s perception of /ð/-/z/ in different vowel contexts.

Moreover, *vowel context* and its interaction with *training* were also found to be non-significant for the subjects’ perception of the target contrasts. The interaction between *vowel context* and *gender*, as well as among *vowel context*, *gender* and *training* were also found to not display a significant effect on the subjects’ perception of the target contrasts (*p>0.05*).

Similarly, the interaction among *phonetic position*, *gender* and/or *training* was revealed to be non-significant for the subjects’ perception performance (*p*>0.05, see Table 5.13). As shown in Table 5.16, it seems there were some differences concerning male and female subjects’ perception of the two contrasts in different phonetic positions. However, their perception performance did not change in regular fashion with the change of phonetic positions.

Test	Phonetic position	Gender	Mean		Std. Deviation	
			/θ/-/s/	/ð/-/z/	/θ/-/s/	/ð/-/z/
Pre-test	Initial	male	1.672	-0.296	1.596	1.192
		female	0.513	-0.972	2.4	0.807
	Medial	male	-1.016	-0.446	2.82	1.077
		female	1.012	-1.07	2.224	2.801
	Final	male	-3.351	-3.518	1.372	0.315
		female	-1.535	-3.113	1.223	1.962
Mid-test 1	Initial	male	3.48	3.598	1.048	1.363
		female	3.614	3.515	1.099	0.676
	Medial	male	2.608	3.004	2.541	2.229
		female	3.497	2.484	2.103	1.06
	Final	male	0.985	2.173	1.201	1.116
		female	1.041	1.817	1.359	1.076
Mid-test 2	Initial	male	4.609	4.219	0.569	0.736
		female	4.522	4.191	0.46	1.063
	Medial	male	4.51	4.642	0.53	0.523
		female	4.375	4.582	0.411	0.532
	Final	male	3.958	4.081	0.713	1.394
		female	3.973	4.518	0.579	0.622
Post-test	Initial	male	5.862	5.088	0.833	1.134
		female	5.359	5.192	0.727	0.727
	Medial	male	5.212	5.01	0.448	0.89
		female	5.095	4.782	0.363	1.661
	Final	male	4.27	6.032	0.254	0.808
		female	5.139	5.448	0.733	0.876

Table 5.16 Perception of /θ/-/s/ and /ð/-/z/ in different phonetic positions across the 4 tests (in *d'* scores) for male and female subjects in the experimental group.

5.6.2.3 Statistical analysis results for the experimental group's perception of voiceless /θ/-/s/ and voiced /ð/-/z/

It seems the experimental group displayed a relatively higher degree of accuracy in the perception of voiceless /θ/-/s/ than voiced /ð/-/z/ in the AXB tests. The factor *contrast difference* was coded as a between-subjects factor in the *repeated-measures ANOVA*. According to the *Post Hoc Tests*, the experimental group's difference in the perception of the two contrasts was statistically non-significant (*F*(1, 53)=0.194*, p*=0.158, $\eta^{2=}$0.037).

5.6.3 Perception test results for the control group

5.6.3.1 Overall results

In order to detect whether there was a repeated testing effect jn the experimental group's progress in the perception of the target contrasts, the control group's perception of the target contrasts were tested after the study on the experimental group was done. As shown in Table 5.1, as for the experimental group, subjects in the control group were also tested 4 times without being trained – pre-test (day1), mid-test 1 (day 5), mid-test 2 (day 10), post-test (day 15).

As shown in Figure 5.18 below, similar to that of the experimental group, the control group's perception of the two contrasts improved during and at the end of the study, yet to much lower degree. Specifically, in the perception of both of the contrasts, their mean accuracy increased from about 57% in the pre-test to about 64% in the post-test (see Appendix 14 for perception test results for individual subjects in the control group).

Table 5.17 below provides us with more information on the subjects' improvement in the perception of the target contrasts. In the pre-test, the majority of the subjects' accuracy was between 50% and 60% in the perception of /θ/-/s/ (n=11) and /ð/-/z/ (n=13). However, the majority of the subjects' accuracy in the perception of the two contrasts gradually improved to the range between 60% and 70% from mid-test 1 to the post-test (n=13 in the perception of /θ/-/s/; n=18 in the perception of /ð/-/z/).

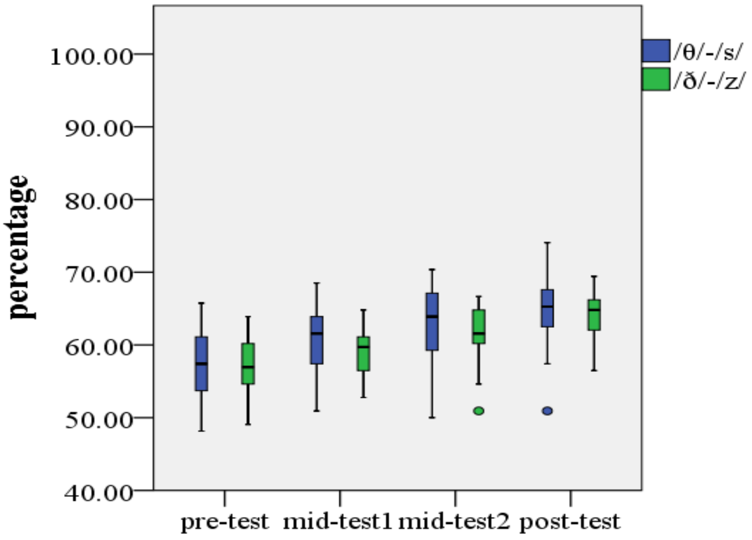


Figure 5.18 Boxplots of the control group's perception of /θ/-/s/ and /ð/-/z/ in the pre-test, mid-test 1, mid-test 2 and the post-test.

test	40%—<50%		50%—<60%		60%—<70%		70%—<80%		80%—100%	
	/θ/-/s/	/ð/-/z/	/θ/-/s/	/ð/-/z/	/θ/-/s/	/ð/-/z/	/θ/-/s/	/ð/-/z/	/θ/-/s/	/ð/-/z/
pre-test	2	1	11	13	7	6	0	0	0	0
mid-test 1	0	0	7	10	13	10	0	0	0	0
mid-test 2	0	0	6	4	13	16	1	0	0	0
post-test	0	0	3	2	13	18	4	0	0	0

Table 5.17 The distribution of the number of subjects in different ranges of accuracy in the perception of /θ/-/s/ and /ð/-/z/ (control group).

Considering that there might be potential bias in the AXB test, the subjects' accuracy was converted to *d-prime* scores with the same method employed for the experimental group. As can be seen from Figure 5.19 and Table 5.18, the control group's *d'* scores increased slightly from pre-test to post-test in terms of mean, maximum and minimum scores.

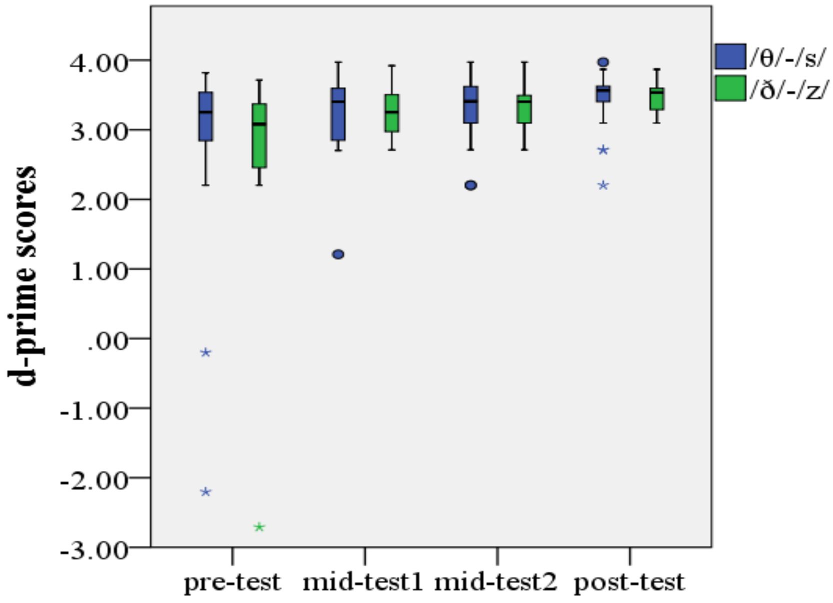


Figure 5.19 Boxplots of the control group's *d'* scores in the pre-test, mid-test 1, mid-test 2, and the post-test.

		Perception of /θ/-/s/				**Perception of /ð/-/z/**			
		pre-test	mid-test 1	mid-test 2	post-test	pre-test	mid-test 1	mid-test 2	post-test
N	Valid	20	20	20	20	20	20	20	20
	Missing	0	0	0	0	0	0	0	0
Mean		2.82	3.19	3.32	3.43	2.73	3.25	3.29	3.46
Median		3.25	3.4	3.41	3.57	3.08	3.28	3.31	3.49
Mode		2.71	2.71	3.6	3.6	2.2	3.17	3.1	3.6
Std. Deviation		1.47	0.6	0.48	0.44	1.37	0.34	0.33	0.23
Variance		2.16	0.36	0.23	0.19	1.89	0.11	0.11	0.05
Range		6.02	2.76	1.77	1.77	6.42	1.21	1.26	0.77
Minimum		-2.2	1.21	2.2	2.2	-2.71	2.71	2.71	3.1
Maximum		3.82	3.97	3.97	3.97	3.71	3.92	3.97	3.87
Sum		56.47	63.85	66.34	68.66	54.62	64.93	65.89	69.26

Table 5.18 The control group's *d'* scores in the pre-test, mid-test 1, mid-test 2, and the post-test.

As presented in section 5.6.2 above, all the subjects in the experimental group showed perception improvement linearly from pre-test to post-test. In the control group, however, some subjects' accuracy in the perception of the target contrasts did not improve linearly from pre-test to post-test. Specific examples, as presented below, can be obtained from the comparison between the pre-test and mid-test 1, between mid-test 1 and mid-test 2, as well as between mid-test 2 and the post-test.

From the pre-test to mid-test 1, as shown in Figure 5.20, in the perception of /θ/-/s/, the majority of the subjects displayed some improvement. However, S33, S37, S40 and S41's *d'* scores decreased in mid-test 1. Similarly, S37 and S45's *d'* scores decreased in mid-test 1 in the perception of /ð/-/z/ (see Figure 5.21), whereas S43 and S48's *d'* score did not change in the two tests.

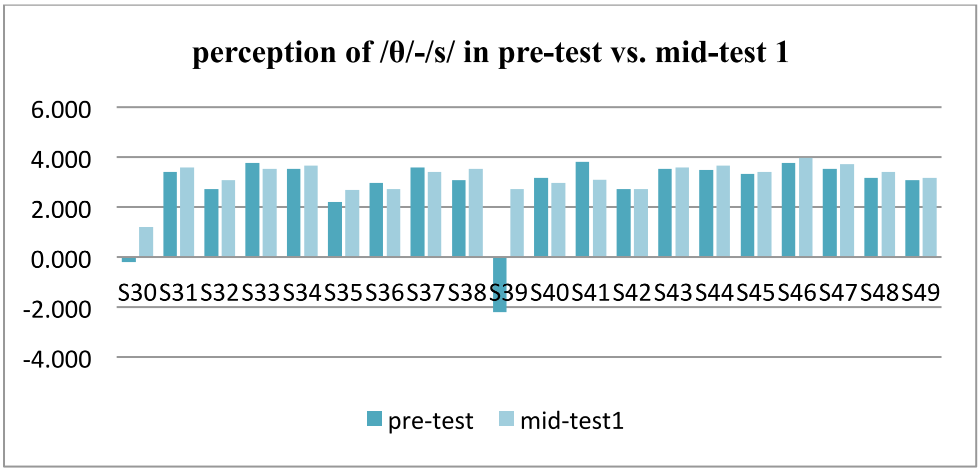


Figure 5.20 Individual subjects’ *d’* scores in the perception of /θ/-/s/ in pre-test vs. mid-test 1 (control group).

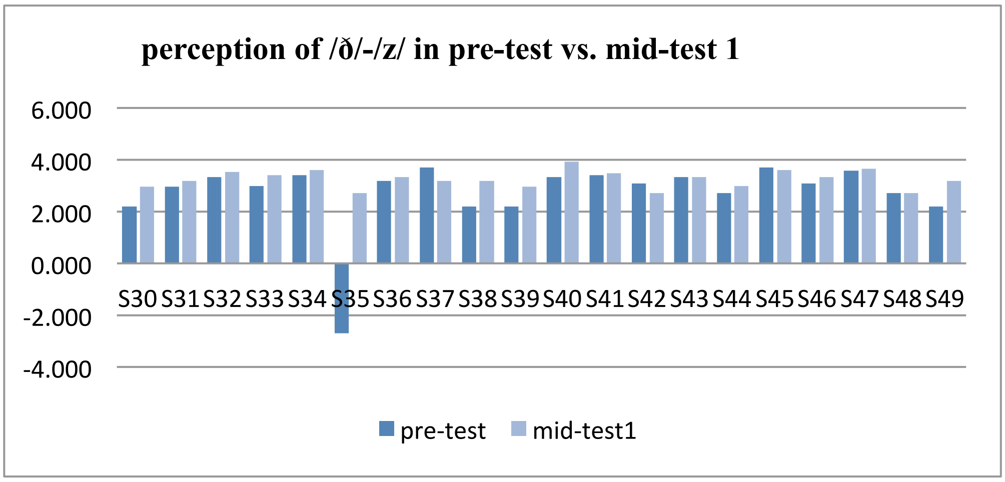


Figure 5.21 Individual subjects’ *d’* scores in the perception of /ð/-/z/ in pre-test vs. mid-test 1 (control group).

From mid-test 1 to mid-test 2, the majority of the subjects’ *d’* scores improved in the perception of both of the contrasts. However, in the perception of /θ/-/s/, S38 and S42’s *d’* scores decreased, while S40’s *d’* score remained the same in the two tests. In the perception of /ð/-/z/, S30, S40, S41, S47 and S49’s *d’* scores decreased with different degrees in mid-test 2, while S44 and S45’s *d’* scores remained the same in the two tests (see Figure 5.22 and Figure 5.23).

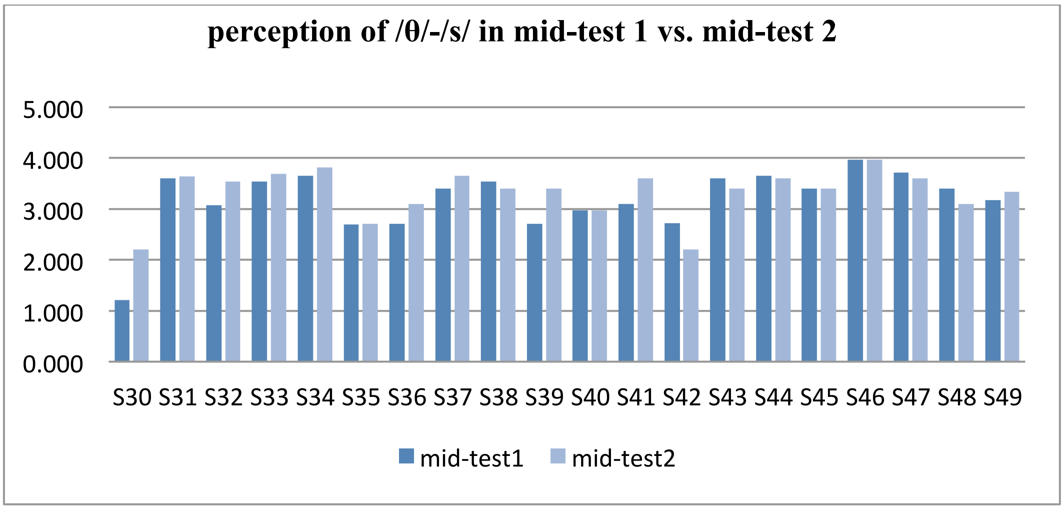


Figure 5.22 Individual subjects' *d'* scores in the perception of /θ/-/s/ in mid-test 1 vs. mid-test 2 (control group).

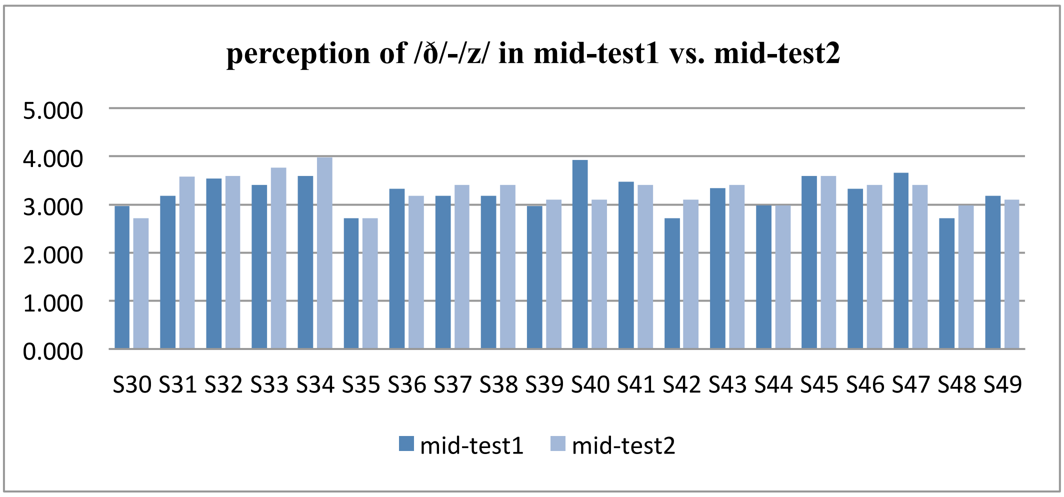


Figure 5.23 Individual subjects' *d'* scores in the perception of /ð/-/z/ in mid-test 1 vs. mid-test 2 (control group).

From mid-test 2 to the post-test, most of the subjects showed improvement in the perception of /θ/-/s/. S30, S31, S32, S35's *d'* scores, however, did not change in the two tests. Moreover, S37, S44, S46's *d'* scores decreased in the post-test compared with in mid-test 2. Similarly, in the perception of /ð/-/z/, S34's *d'* scores decreased, while S33, S38, S39, S42, S45's *d'* scores remained unchanged in mid-test 2 and the post-test (see Figure 5.24 and Figure 5.25).

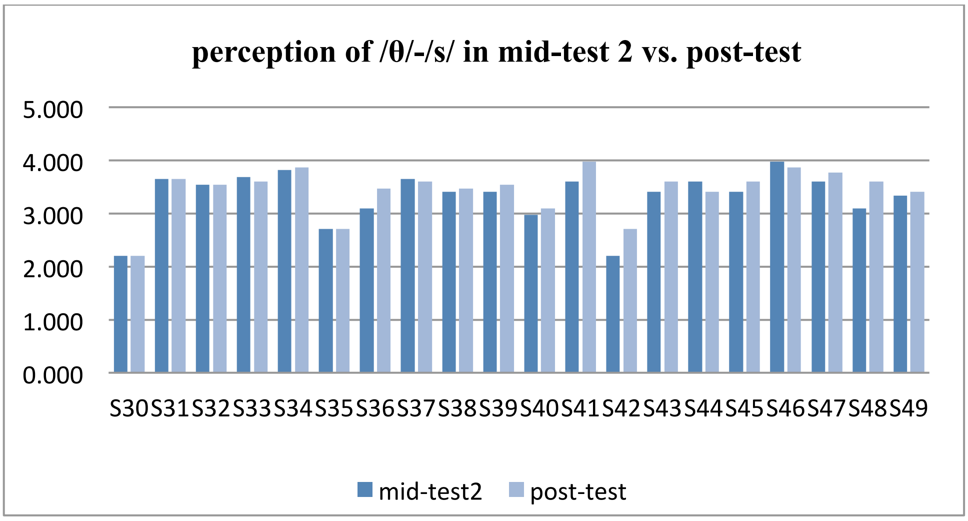


Figure 5.24 Individual subjects' *d'* scores in the perception of /θ/-/s/ in mid-test 2 vs. post-test (control group).

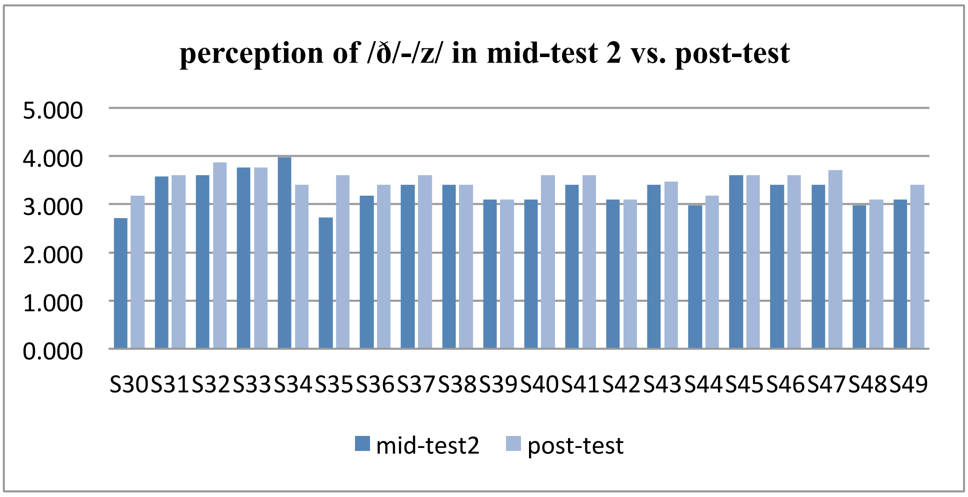


Figure 5.25 Individual subjects' *d'* scores in the perception of /ð/-/z/ in mid-test 2 vs. post-test (control group).

On the whole, the accuracy of all subjects in the control group in the perception of the target contrasts improved in the post-test compared with the pre-test, yet with a much lower degree than that for the experimental group. Furthermore, some of the subjects' improvement was not linear, which was different from the experimental group.

5.6.3.2 Statistical analysis of the control group's perception test results

As discussed above, the subjects of the control group all made some improvement in the perception of the target contrasts from pre-test to post-test, yet it was still unclear whether their improvement was statistically significant. It was also not clear which factors, if any, had a significant effect on their perception improvement. More importantly, given that the control group did not experience phonetic training, if their improvement was statistically significant, it would be due to a repeated testing effect. Accordingly, there would be a repeated testing effect on the experimental group's perception performance. Considering that it was the same group of subjects who were repeatedly tested with the AXB test, a *repeated-measures ANOVA* was employed for the statistical analysis.

Let us first have a look at the subjects' answers to the questionnaire. The intention was to select subjects of a similar profile to that of the experimental group. The subjects of the control group were similar to/the same as each other in terms of age, AO of L2-English learning, etc. (as listed in Table 5.20). Therefore, as for the experimental group, only *gender difference* was adopted as a between-subjects factor. In the meantime, *repeated testing experience* (*experience*, hereafter) was coded as a within-subjects factor to detect whether the control group's performance significantly benefited from the repeated testing experience. Given that each subject was tested 4 times with equal intervals, as were the experimental group, the factor *experience* was coded into 4 levels. They were *no experience* (pre-test), *experience 1* (mid-test 1), *experience 2* (mid-test 2) and *experience 3* (post-test). The *phonetic environment* as another within-subject factor was further divided into *vowel context* (*/i/*, */a/* and */u/*) and *phonetic positions (initial*, *medial* and *final)*.

Factors	Answer	Number	Percentage
Gender	male	10	50%
	female	10	50%
Years of English study	6 years	13	50.00%
	7 years	15	50.00%
AO	13 years old	17	30%
	14 years old	12	70%
Age	19	5	25.00%
	20	9	45.00%
	21	6	30.00%
Major motivation	hobby	0	0.00%
	The need to get high scores in English exams	20	100.00%
Learn English in spare time	No	0	0.00%
	Yes	20	100.00%
Institute of English learning	Public school/university	20	100.00%
Use English on a daily basis (except for study)?	No	20	100.00%
Travelled/lived aboard?	No	20	100.00%
	Yes	0	0.00%

Table 5.19 Data collected from the questionnaire (control group).

5.6.3.3 Statistical analysis of the control group's perception of /θ/-/s/.

As for the experimental group, statistical analysis was performed on the subjects' *d'* scores obtained from the 4 AXB tests to avoid bias. As shown in Table 5.20 and Table 5.21 below, the factor *experience* and *phonetic position* were found to be statistically significant for the control group's perception of /θ/-/s/ and /ð/-/z/ (*p<0.05*). The rest of the factors and their interaction with each other, however, were revealed to be non-significant for the control group's perception performance (*p>0.05*).

factor	df and F-value	Sig.	Partial Eta Squared
gender	F(1, 18)=1.548	p=0.229	η^2=0.079
experience	**F(3, 54)=3.884**	**p=0.014**	**η^2=0.177**
experience * gender	F(3, 54)=2.050	p=0.118	η^2=0.102
vowel context	F(3, 54)=3.886	p=0.509	η^2=0.037
vowel context * gender	F(3, 54)=0.756	p=0.477	η^2=0.037
experience * vowel context	F(6, 108)=0.541	p=0.111	η^2=0.135
experience * vowel context * gender	F(6, 108)=0.961	p=0.455	η^2=0.051
phonetic position	**F(2, 36)=59.984**	**p<0.001**	**η^2=0.760**
phonetic position * gender	F(2, 36)=1.155	p=0.326	η^2=0.060
experience * phonetic position	F(6, 108)=0.384	p=0.888	η^2=0.021
experience * phonetic position * gender	F(6, 108)=1.147	p=0.340	η^2=0.060

Table 5.20 Statistical analysis of the control group's perception of /θ/-/s/.

factor	df and F-value	Sig.	Partial Eta Squared
gender	F(1, 18)=0.944	p=0.171	η^2=0.101
experience	**F(3, 54)=2.872**	**p=0.045**	**η^2=0.138**
experience * gender	F(3,54)=2.353	p=0.082	η^2=0.116
vowel context	F(3, 54)=1.907	p=0.163	η^2=0.096
vowel context * gender	F(3, 54)=0.726	p=0.491	η^2=0.039
experience * vowel context	F(6, 108)=0.135	p=0.161	η^2=0.106
experience * vowel context * gender	F(6, 108)=0.618	p=0.715	η^2=0.033
phonetic position	**F(2, 36)=5.294**	**p<0.001**	**η^2=0.435**
phonetic position * gender	F(2, 36)=0.825	p=0.446	η^2=0.044
experience * phonetic position	F(6, 108)=928	p=0.389	η^2=0.082
experience * phonetic position * gender	F(6, 108)=0.310	p=0.930	η^2=0.017

Table 5.21 Statistical analysis of the control group's perception of /ð/-/z/.

1. Repeated testing effect (*experience*)

In the perception of /θ/-/s/, the mean difference between *no experience* (*d'* scores obtained from the pre-test) and *experience 1* (*d'* scores obtained from mid-test 1) was 0.280. Given *p>0.05* on a 0.05 significant level, the difference was not statistically significant. That means the subjects' performance in mid-test 1 did not significantly benefit from repeated testing experience due to the pre-test. Similarly, the mean difference between *experience 1* and *experience 2* was 0.091, yet it was not statistically significant either (*p*>0.05). In other words, the subjects' perception improvement in mid-test 2 was not significantly influenced by the testing experience in mid-test 1. In comparison, the mean difference between *no experience* and *experience 2* was revealed to be statistically significant (*p<0.05*). Thus, the repeated testing experience due to the pre-test was statistically significant for the control group's improvement in the perception of /θ/-/s/ in mid-test 2. The same situation was found between *no experience* and *experience 3,* between *experience 1* and *experience 3,* as well as between *experience 2* and *experience 3*. It was revealed that the more testing experiences the subjects had, the smaller the value of *p* was, and the greater the effect the *repeated testing experience* had on the subjects perception of /θ/-/s/. It was the same situation in the control group's perception of /ð/-/z/. On the whole, the more testing experiences the subjects had, the more they were likely to accurately perceive the target contrasts in the AXB test.

(I) experience	(J)experience	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval for Difference	
					Lower Bound	Upper Bound
no experience	experience 1	-0.28	0.224	0.228	-0.75	0.191
	experience 2	-0.37	0.259	0.17	-0.914	0.174
	experience 3	-0.583	0.257	**0.036**	-1.122	-0.043
experience 1	no experience	0.28	0.224	0.228	-0.191	0.75
	experience 2	-0.091	0.086	0.306	-0.271	0.09
	experience 3	-0.303	0.111	**0.014**	-0.537	-0.07
experience 2	no experience	0.37	0.259	0.17	-0.174	0.914
	experience 1	0.091	0.086	0.306	-0.09	0.271
	experience 3	-0.212	0.091	**0.031**	-0.403	-0.022
experience 3	no experience	0.583	0.257	**0.036**	0.043	1.122
	experience 1	0.303	0.111	0.014	0.07	0.537
	experience 2	0.212	0.091	0.031	0.022	0.403

Table 5.22 *Post Hoc Tests* of the control group's perception of /θ/-/s/ in the fours AXB tests.

(I) experience	(J)experience	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval for Difference	
					Lower Bound	Upper Bound
no experience	experience 1	-0.516	0.269	0.071	-1.081	0.049
	experience 2	-0.564	0.261	**0.045**	-1.112	-0.015
	experience 3	-0.732	0.3	**0.025**	-1.363	-0.101
experience 1	no experience	0.516	0.269	0.071	-0.049	1.081
	experience 2	-0.048	0.063	0.456	-0.18	0.084
	experience 3	-0.216	0.055	**0.001**	-0.332	-0.1
experience 2	no experience	0.564	0.261	0.045	0.015	1.112
	experience 1	0.048	0.063	0.456	-0.084	0.18
	experience 3	-0.168	0.064	**0.017**	-0.302	-0.034
experience 3	no experience	0.732	0.3	0.025	0.101	1.363
	experience 1	0.216	0.055	0.001	0.1	0.332
	experience 2	0.168	0.064	0.017	0.034	0.302

Table 5.23 *Post Hoc Tests* of the control group’s perception of /ð/-/z/ in the fours AXB tests.

2. *Phonetic position*

Phonetic position, as another within-subjects factor, was found to be highly significant for the control group’s perception performance. According to the *Post Hoc Tests* results (as shown in Table 5.24 and Table 5.25), the mean differences between the subjects’ accuracy for initial and medial positions were 0.077 in the perception of /θ/-/s/, and 0.064 in the perception of /ð/-/z/ respectively, both of which were statistically non-significant (*p>0.05*). In other words, the subjects did not display a significant difference in the perception of the target contrasts in initial and medial positions. However, in the perception of both of the target contrasts, the mean differences between their perception accuracy in initial and final positions, as well as between medial and final positions were detected to be statistically significant (*p<0.05*). That is, the subjects were more likely to have accurate perception of the target contrasts when they were embedded in initial and medial positions than in final position. Moreover, the interactions of *phonetic position* with *gender* and/or *experience* were found to be statistically non-significant for the control group’s perception performance.

(I) phonetic position	(J) phonetic position	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval for Difference	
					Lower Bound	Upper Bound
Initial	Medial	0.029	0.077	0.708	-0.133	0.192
	Final	0.839	0.101	**0**	0.628	1.05
Medial	Initial	-0.029	0.077	0.708	-0.192	0.133
	Final	0.81	0.088	**0**	0.624	0.995
Final	Initial	-0.839	0.101	**0**	-1.05	-0.628
	Medial	-0.81	0.088	**0**	-0.995	-0.624

Table 5.24 *Post Hoc Tests* of the control group's perception of /θ/-/s/ in different phonetic positions.

(I) phonetic position	(J) phonetic position	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval for Difference	
					Lower Bound	Upper Bound
Initial	Medial	0.064	0.086	0.465	-0.117	0.245
	Final	0.474	0.115	**0.001**	0.232	0.717
Medial	Initial	-0.064	0.086	0.465	-0.245	0.117
	Final	0.41	0.089	0	0.223	0.597
Final	Initial	-0.474	0.115	**0.001**	-0.717	-0.232
	Medial	-0.41	0.089	**0**	-0.597	-0.223

Table 5.25 *Post Hoc Tests* of the control group's perception of /ð/-/z/ in different phonetic positions.

3. Vowel context

Vowel context, as for the experimental group, was revealed to be non-significant for the control group's perception performance. According to the *Post Hoc Tests* results (see Table 5.26 and Table 5.27), the subjects' mean difference between the contexts /i/ and /a/ was 0.039 in the perception of /θ/-/s/, and 0.151 in the perception of /ð/-/z/. Their mean difference between the contexts /i/ and /u/ was 0.133 in the perception of /θ/-/s/, and 0.201 in the perception of /ð/-/z/. Between /a/ and /u/ contexts, however, their mean difference was 0.094 in the perception of /θ/-/s/, and 0.050 in the perception of /ð/-/z/. Nevertheless, all the differences were revealed to be statistically non-significant (*$p>0.05$*). Moreover, the interactions of *vowel context* with *experience* and/or *gender* were also detected to be statistically non-significant for the control group's perception performance.

(I) vowel	(J) vowel	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval for Difference	
					Lower Bound	Upper Bound
/i/	/a/	0.039	0.099	0.7	-0.17	0.247
	/u/	0.133	0.124	0.3	-0.128	0.394
/a/	/i/	-0.039	0.099	0.7	-0.247	0.17
	/u/	0.094	0.124	0.458	-0.166	0.354
/u/	/i/	-0.133	0.124	0.3	-0.394	0.128
	/a/	-0.094	0.124	0.458	-0.354	0.166

Table 5.26 *Post Hoc Tests* of the control group's perception of /θ/-/s/ in different vowel contexts.

(I) vowel	(J) vowel	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval for Difference	
					Lower Bound	Upper Bound
/i/	/a/	-0.151	0.11	0.189	-0.383	0.081
	/u/	-0.201	0.122	0.118	-0.458	0.056
/a/	/i/	0.151	0.11	0.189	-0.081	0.383
	/u/	-0.05	0.085	0.565	-0.228	0.129
/u/	/i/	0.201	0.122	0.118	-0.056	0.458
	/a/	0.05	0.085	0.565	-0.129	0.228

Table 5.27 *Post Hoc Tests* of the control group's perception of /ð/-/z/ in different vowel contexts.

5.6.3.4 Statistical analysis of the experimental and control group's perception test results

The experimental group and the control group achieved some perception improvement from pre-test to post-test, despite the fact that the control group did not experience audiovisual training. It is therefore necessary to determine whether the difference between the two groups concerning perception improvement was statistically significant, in order to thus reveal whether the experimental group's perception improvement can be largely attributed to the audiovisual training rather than the repeated testing experience. Therefore, *group difference* was coded as a between-subjects factor for further statistical analysis. All the subjects' *d'* scores in the four AXB tests were inputted into SPSS. A *repeated-measures ANOVA* was adopted for the analysis. As a result, the differences between the two group's *d'* scores in the perception of /θ/-/s/ ($F(1, 44)=289.539, p<0.001, \eta^2=0.868$) and /ð/-/z/ ($F(1, 44)=200.840, p<0.001, \eta^2=0.859$) were both statistically significant. Specifically, as shown in Table 5.28, the mean difference between the experimental group and the control group was 2.818 in the perception of /θ/-/s/, and 2.091 in the perception of /ð/-/z/. In other words, the experimental group significantly outperformed the control group. Moreover, the interaction between the factor *group* and *training* also displayed a significant effect on the subjects' perceptual performance (*$F(3,108)=204; p<0.001; \eta^2=0.878$*).

(I) group	(J) group	Mean Difference (I- J)		Std. Error		Sig.		95% Confidence Interval for Difference			
								Lower Bound		Upper Bound	
		/θ/-/s/	/ð/-/z/	/θ/-/s/	/ð/-/z/	/θ/-/s/	/ð/-/z/	/θ/-/s/	/ð/-/z/	/θ/-/s/	/ð/-/z/
experimental group	control group	-2.818	-2.091	0.166	0.124	0.00	0.00	-3.152	-2.34	-2.484	-1.842
control group	experimental group	2.818	2.091	0.166	0.124	0.00	0.00	2.484	1.842	3.152	2.34

Table 5.28 *Post Hoc Tests* of the experimental and the control group's perception test results.

5.6.3.5 Statistical analysis of the control group's perception of voiceless /θ/-/s/ vs. voiced /ð/-/z/

As for the experimental group, the factor *contrast difference* was coded as a between-subjects factor in a *repeated-measures ANOVA*, so as to detect whether the control group's difference in the perception of the two contrasts was statistically significant. The results showed that the control group's difference between the perception of the two contrasts was statistically non-significant (*$F(1, 38)=0.049, p=0.826, \eta^2=0.001$*).

5.6.4 Production test results

5.6.4.1 Overall results

The production test results were converted into an accuracy figure with the same method described in Chapter 4. Individual subjects' accuracy in the production of /θ/ and /ð/ is displayed in Table 5.29 below. All the results for production tests were obtained from the experimental group.

Subject	Pre-test (%)		Mid-test 1 (%)		Mid-test 2 (%)		Post-test (%)	
	/θ/	/ð/	/θ/	/ð/	/θ/	/ð/	/θ/	/ð/
S1	29.17	37.36	50	62.92	64.5	74.03	92.67	89.86
S2	31.17	38.61	57	63.89	76.33	70.14	86.5	81.11
S3	45.33	41.25	77.66	81.11	93.33	90.27	drop	drop
S4	33.17	39.58	51	47.5	71.5	71.25	81.17	78.19
S5	41.67	38.47	64.5	75.69	79.5	82.22	90.83	93.19
S6	64.5	36.11	79	52.92	81.83	65.28	89.67	82.36
S7	38.67	48.89	54.17	66.39	72	70.42	89.83	80.42
S8	40.67	40.14	56.5	49.44	67.33	63.61	80	75.83
S9	23.83	32.92	48.17	47.36	62.83	70.69	83.5	88.06
S10	49.83	42.22	92.17	90.28	drop	drop	drop	drop
S11	53.5	41.67	74.67	66.25	72.17	71.94	81	75.14
S12	34.33	41.39	61.83	47.78	74.83	63.19	79.67	65.69
S13	37	32.64	55.33	45.56	63.55	69.58	77.83	75.69
S14	26.5	42.08	33.17	51.25	58.5	62.85	74.67	73.33
S15	67.67	54.31	83.67	74.86	87	87.22	95.83	89.31
S16	28.67	42.78	42.5	49.86	48.33	53.33	58	66.38
S17	31	34.58	48.33	49.44	57.67	58.89	79.5	74.44
S18	31.67	33.06	49.67	43.19	60.33	50.56	68.17	68.06
S19	37	28.61	44.5	34.58	70.33	61.25	77.17	75
S20	61.67	47.78	72.83	53.06	82.33	67.78	85.17	74.17
S21	33	43.47	54.56	59.13	67	69.31	77.17	75.14
S22	30.67	35.69	43.83	51.25	77.17	60.31	86.5	76.53
S23	41.67	38.33	78.67	72.5	88	82.08	93.33	87.78
S24	31.83	36.81	84.33	76.25	82	82.36	85.5	87.22
S25	36.83	45.28	41.67	56.84	57.67	62.64	88	75
S26	32.17	47.78	50.83	59.83	80	68.06	85	73.06
S27	41.83	45.14	53.83	61.25	65	71.32	69.17	80.97
S28	34	54.31	47.33	64.86	55.67	80.28	79	86.39
S29	34.96	44.38	64.5	66.32	75	82.92	87	92.22

Table 5.29 Individual subjects' accuracy in the production of /θ/ and /ð/ (*drop: dropped from the test).

Due to S10's accuracy in the perception and production of the target contrasts being above 90% in mid-test 1, he was dropped from training sessions 4-9, mid-test 2 and the post-test. Similarly, S3 was dropped from the last 3 training sessions and the post-test, because her accuracy was above 90% in the perception and production of the target contrasts in mid-test 2.

test	20%—<30%		30%—<40%		40%—<50%		50%—<60%		60%—<70%		70%—<80%		80%—<90%		90%—100%	
	/θ/	/ð/	/θ/	/ð/	/θ/	/ð/	/θ/	/ð/	/θ/	/ð/	/θ/	/ð/	/θ/	/ð/	/θ/	/ð/
pre-test	4	1	15	12	6	14	1	2	3	0	0	0	0	0	0	0
mid-test 1	0	0	1	1	8	8	9	7	3	7	5	4	2	1	1	1
mid-test 2	0	0	0	0	1	0	4	3	7	11	9	7	6	6	1	1
post-test	0	0	0	0	0	0	1	0	2	2	7	12	13	10	4	2

Table 5.30 The distribution of the number of subjects in different ranges of accuracy in the production of /θ/ and /ð/.

In the pre-test, in the production of /θ/, most of the subjects' accuracy was below 60% (n=26). Only 3 subjects (S6, S15, S20) achieved an accuracy of between 60% and 70%. The subjects' performance was even poorer in the production of /ð/ – none of them managed to achieve an accuracy of above 60%.

In mid-test 1, more than half of the subjects' accuracy was between 30% and 60% (n=18 in the production of /θ/: n=16 in the production of /ð/). Moreover, there were some subjects whose accuracy was between 60% and 80% (n=8 in the production of /θ/; n=11 in the production of /ð/). S15 and S24 achieved an accuracy of between 80% and 90% in the production of /θ/, while only S3 achieved this level of accuracy in the production of /ð/. Furthermore, S10 was the only subject whose accuracy was above 90% both in the production of /θ/ (92.17%) and /ð/ (90.28%).

In mid-test 2, in the production of /θ/, only S16's accuracy was below 50% (48.33%). 4 subjects' accuracy was between 50% and 60%; 16 subjects' accuracy was between 60% and 80%; another 6 subjects achieved an accuracy of between 80% and 90%; S3 achieved an accuracy of 93.33%. In the production of /ð/, 3 subjects' accuracy was between 50% and 60%. More than half of the subjects' performance was between 60% and 80% (n=18). Moreover, 6 subjects achieved an accuracy of between 80% and 90%. Nevertheless, S3's accuracy was 90.27%. He was the only subject whose accuracy was above 90% both in the production of /θ/ and /ð/.

The subjects' production performance further improved in the post-test. S16 was the only subject whose accuracy was below 60% (58.00%). 9 subjects' accuracy was between 60% and 80% in the production of /θ/, vs. 14 subjects who achieved this level in the production of /ð/. There were 13 subjects with accuracy between 80% and 90% in the production of /θ/, vs. 10 subjects in the production of /ð/. Moreover, 4 subjects (S1, S5, S15, S23) achieved an accuracy of between 90% and 100% in the production of /θ/, vs. 2 subjects (S5, S29) in the production of /ð/.

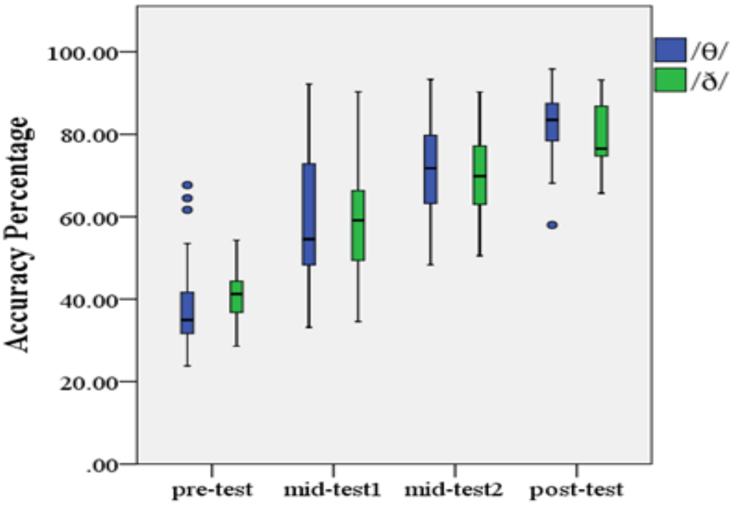


Figure 5.26 Boxplots of the subjects' accuracy in the production of /θ, ð/ in the pre-test, mid-test 1, mid-test 2 and the post-test.

Figure 5.26 above provides us with a visual depiction of the distribution of the experimental group's accuracy in the production of /θ/ and /ð/ in the four tests. Their production performance in the pre-test was fairly poor. Specifically, their accuracy fell in similar ranges in the production of /θ/ and /ð/. Nevertheless, the mean accuracy in the production of /θ/ (38.76%) was lower than for /ð/ (40.88%).

In mid-test 1, the subjects' accuracy ranged from S14's 33.17% to S10's 92.17% in the production of /θ/, and from S19's 34.58% to S10's 90.28% in the production of /ð/. Most of the subjects' accuracy fell in similar ranges in the production of /θ/ and /ð/, which were higher than in the pre-test.

The subjects' accuracy was further improved in mid-test 2. The mean accuracies were 71.13% and 70.14% in the production of /θ/ and /ð/ respectively. The majority of the subjects' accuracy levels in the production of /ð/ fell in a relatively lower percentage range than for /θ/.

In the post-test, their accuracy ranged from 58.00% to 95.83% in the production of /θ/, whereas from 65.69% to S5's 93.19% in the production of /ð/. Similar to in the mid-test 2, most subjects' performance in the production of /ð/ fell in a relatively lower percentage range than for /θ/.

On the whole, most subjects displayed low accuracy (below 60%) in the pre-test in the production of /θ/ and /ð/, whereas more subjects achieved a medium level of accuracy (60%—80%) in the mid-test 1 and the mid-test 2. Their performance further improved in the post-test, in which most subjects showed high accuracy (80%—90%). Moreover, the subjects' production performance rose linearly from the pre-test to the post-test in terms of mean, maximum and minimum accuracy.

	Production of /θ/			**Production of /ð/**		
	(mid-test1—pre-test)%	(mid-test2—pre-test)%	(post-test—pre-test)%	(mid-test1—pre-test)%	(mid-test2—pre-test)%	(post-test—pre-test)%
N Valid	29	28	27	29	28	27
Mean	20.42	32.78	44.19	18.48	29.3	38.46
Median	18	33.33	48	15.56	27.68	35.83
Std. Deviation	10.6	10.42	11.21	11.52	9.58	9.46
Minimum	4.84	17.33	23.5	5.28	10.55	23.6
Maximum	52.5	50.17	63.5	48.06	49.02	55.14

Table 5.31 The subjects' improved accuracy from the pre-test to mid-test 1, mid-test 2 and the post-test in the production of /θ/ and /ð/.

Table 5.31 shows the experimental group's improvement in the production of the target contrasts from the pre-test to the three after-training tests. At the end of the first 3 training sessions, all the subjects showed some production improvement, with the mean improved accuracy 44.19% and 38.46% in the production of /θ/ and /ð/ respectively. The degree of improvement, however, varied across the subjects (see Appendix 10 for individual subjects' production accuracy).

5.6.4.2 Repeated-measures ANOVA analysis

Given that it was the same group of subjects who were repeatedly tested 4 times in the production test, a *repeated-measures ANOVA* was performed to detect which factor(s) displayed a significant effect on the subjects' production of /θ/ and /ð/. Considering that there was not a big difference regarding the subjects' age, years of L2-English learning, etc. (see the factors listed in Table 5.4), only *gender difference* was coded as a between-subjects factor. The within-subjects factors were the target contrasts' *phonetic position* (*initial, medial and final*) and *training* (*non-trained* (*pre-test*)*, trained for 3 sessions (mid-test 1*)*, trained for 6 sessions (mid-test 2*) and *trained for 9 session (post-test*).

factor	F-value	Sig.	Partial Eta Squared
training	F(1, 1212)=325.353	**p<0.001**	η^2=0.605
gender	F(1, 403)=1.015	**P=0.031**	η^2=0.463
gender* training	F(1, 403)=9.316	**P=0.002**	η^2=0.723
phonetic position	F (6, 1260)=5.335	p=0.06	η^2=0.014
phonetic position * training	F (6, 1206)=0.627	p=0.709	η^2=0.019
phonetic position * gender	F (6, 1206)=1.470	p=0.232	η^2=0.011
gender * phonetic position * training	F (6, 1206)=1.708	p=0.116	η^2=0.013

Table 5.32 Statistical analysis results for the subjects' production of /θ/.

factor	df and F-value	Sig.	Partial Eta Squared
training	F(3, 933)=224.864	**p<0.001**	η^2=0.270
gender	F(1, 403)=8.909	**P=0.003**	η^2=0.959
gender* training	F(1, 403)=5.453	**P=0.020**	η^2=0.406
phonetic position	F(2, 1158)=2.605	p=0.074	η^2=0.004
phonetic position * training	F(2, 403)=0.755	p=0.407	η^2=0.018
phonetic position * gender	F(2, 403)=2.589	P=0.059	η^2=0.021
gender * phonetic position * training	F(2, 403)=1.980	P=0.066	η^2=0.015

Table 5.33 Statistical analysis results for the subjects' production of /ð/.

Table 5.32 and Table 5.33 present the results from a *repeated-measures ANOVA* of the subjects' production of /θ/ and /ð/. According to the results, the within-subjects factor *training*, its interaction with *gender*, as well as the between-subjects factor *gender*, showed a significant impact on the subjects' production of /θ/ and /ð/ (*p<0.05*). However, *phonetic position*, as well as its interaction with *training* and/or *phonetic position* and/or *gender* were all found to be non-significant for the subjects' production of /θ/ and /ð/ (*p>0.05*).

1. *Training*

Training was found to play a significant effect on the subjects’ production performance. A *Post Hoc Test* was carried out to further detect whether the subjects’ production improvement during and at the end of the training programme was statistically significant. As displayed in Table 5.34 below, the differences between the subjects’ mean scores in the pre-test and mid-test 1 were 0.481 in the production of /θ/ and 0.855 in the production of /ð/. There were higher mean differences between the pre-test and mid-test 1, which were 1.184 in the production of /θ/, and 1.757 in the production of /ð/. The largest mean differences were found between the pre-test and the post-test, which were 1.843 in the production of /θ/, and 2.299 in the production of /ð/. These differences were all revealed to be statistically significant. Therefore, it seems the more training sessions the subjects went through, the more likely they were to be able to accurately produce the target contrasts. Moreover, the interaction of *training* with *phonetic position* was also found to be statistically non-significant.

(I) test	(J) test	Mean Difference (I-J)		Std. Error		Sig.		95% Confidence Interval			
								Lower Bound		Upper Bound	
		/θ/	/ð/	/θ/	/ð/	/θ/	/ð/	/θ/	/ð/	/θ/	/ð/
Pre-test	Mid-test 1	-0.481	-0.855	0.143	0.154	0.001	0	-0.764	-1.159	-0.198	-0.551
	Mid-test 2	-1.184	-1.757	0.162	0.225	0	0	-1.505	-2.202	-0.863	-1.312
	Post-test	-1.843	-2.299	0.172	0.218	0	0	-2.182	-2.73	-1.504	-1.867
Mid-test 1	Pre-test	0.481	0.855	0.143	0.154	0.001	0	0.198	0.551	0.764	1.159
	Mid-test 2	-0.703	-0.902	0.135	0.18	0	0	-0.97	-1.259	-0.435	-0.545
	Post-test	-1.362	-1.443	0.161	0.211	0	0	-1.681	-1.861	-1.043	-1.026
Mid-test 2	Pre-test	1.184	1.757	0.162	0.225	0	0	0.863	1.312	1.505	2.202
	Mid-test 1	0.703	0.902	0.135	0.18	0	0	0.435	0.545	0.97	1.259
	Post-test	-0.659	-0.541	0.145	0.226	0	0.018	-0.945	-0.988	-0.373	-0.095
Post-test	Pre-test	1.843	2.299	0.172	0.218	0	0	1.504	1.867	2.182	2.73
	Mid-test 1	1.362	1.443	0.161	0.211	0	0	1.043	1.026	1.681	1.861
	Mid-test 2	0.659	0.541	0.145	0.226	0	0.018	0.373	0.095	0.945	0.988

Table 5.34 *Post Hoc Tests* of the subjects’ production results of /θ/ and /ð/.

2. *Gender*

As a between-subjects factor, *gender* displayed a significant effect on the subjects' production of /θ/ and /ð/. According to the *Post Hoc Tests* results shown in Table 5.35 and Table 5.36 below, the mean difference between the male and female subjects was 0.839 in the production of /θ/ and 0.918 in the production of /ð/. Both of the differences were revealed to be statistically significant. Thus, the male subjects outperformed the females in the production of the two speech sounds.

(I) gender	(J) gender	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval for Difference	
					Lower Bound	Upper Bound
male	female	0.839	0.219	0.031	-0.221	0.644
female	male	-0.839	0.219	0.031	-0.644	0.221

Table 5.35 *Post Hoc Tests* of the male and female subjects' production of /θ/.

(I) gender	(J) gender	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval for Difference	
					Lower Bound	Upper Bound
male	female	0.918	0.252	0.003	-0.318	0.812
female	male	-0.918	0.252	0.003	-0.812	0.318

Table 5.36 *Post Hoc Tests* of the male and female subjects' production of /ð/.

3. The interaction between *gender* and *training*

The interaction between *gender* and *training* was also found to display a significant effect on the subjects' production of /θ/ and /ð/. Both in the production of /θ/ and /ð/, as shown in Table 5.37 and Table 5.38 below, the male subjects' mean score was lower than the females in the pre-test. However, after being trained for 3 sessions, the male subjects displayed a higher mean score than the females. They further outperformed the female subjects in mid-test 2 and the post-test in terms of mean score, lower bound and upper bound. On the whole, with the training sessions carried out, the male subjects outperformed the female subjects in the production of the two speech sounds.

gender		Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
male	Pre-test	3.365	0.214	2.941	3.788
	Mid-test 1	6.141	0.279	5.589	6.692
	Mid-test 2	7.443	0.287	6.875	8.01
	Post-test	8.724	0.194	8.339	9.109
female	Pre-test	3.59	0.205	3.186	3.995
	Mid-test 1	5.333	0.267	4.806	5.861
	Mid-test 2	7.105	0.274	6.562	7.647
	Post-test	8.386	0.186	8.018	8.753

Table 5.37 Table 5.37 Scores for the male and female subjects' production of /θ/ in the pre-test, mid-test 1, mid-test 2 and the post-test.

Gender	Test	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
male	Pre-test	5.056	0.212	4.637	5.476
	Mid-test 1	6.005	0.213	5.584	6.426
	Mid-test 2	6.71	0.253	6.209	7.212
	Post-test	6.932	0.273	6.393	7.471
female	Pre-test	4.629	0.204	4.224	5.033
	Mid-test 1	5.39	0.205	4.985	5.796
	Mid-test 2	6.489	0.244	6.006	6.972
	Post-test	7.35	0.263	6.831	7.869

Table 5.38 *Post Hoc Tests* of the male and female subjects' production of /ð/ in the pre-test, mid-test 1, mid-test 2 and the post-test.

4. *Phonetic position*

The within-subjects factor *phonetic position* was found to be non-significant for the subjects' production of /θ/ and /ð/. According to the *Post Hoc Tests* results shown below, the mean differences between the subjects' production of /θ/ and /ð/ in initial and medial positions were 0.208 and 0.077 respectively. In the production of /θ/ and /ð/ in initial and final positions, their mean differences were 0.531 and 0.120 respectively. The mean differences between the subjects' production of /θ/ and /ð/ in medial and final positions were 0.322 and 0.043 respectively. However, all the differences were revealed to be statistically non-significant (*p*>0.05). Furthermore, the interaction among *gender, phonetic position* and *training* did not have a significant effect on the subjects' production of /θ/ and /ð/ (*p*>0.05).

(I) phonetic position	(J) phonetic position	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval for Difference	
					Lower Bound	Upper Bound
Initial	Medial	-0.208	0.207	0.316	-0.618	0.201
	Final	-0.531	0.223	0.19	-0.973	-0.089
medial	Initial	0.208	0.207	0.316	-0.201	0.618
	Final	-0.322	0.221	0.147	-0.759	0.115
final	Initial	0.531	0.223	0.19	0.089	0.973
	Medial	0.322	0.221	0.147	-0.115	0.759

Table 5.39 *Post Hoc Tests* of the male and female subjects' production of /θ/ in different phonetic positions.

(I) phonetic position	(J) phonetic position	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval for Difference	
					Lower Bound	Upper Bound
Initial	Medial	-0.077	0.123	0.53	-0.32	0.165
	Final	-0.12	0.126	0.34	-0.369	0.129
medial	Initial	0.077	0.123	0.53	-0.165	0.32
	Final	-0.043	0.096	0.652	-0.232	0.146
final	Initial	0.12	0.126	0.34	-0.129	0.369
	Medial	0.043	0.096	0.652	-0.146	0.232

Table 5.40 *Post Hoc Tests* of the male and female subjects' production of /ð/ in different phonetic positions.

5.6.5 Production test results of /s/ and /z/

In the pilot study, the subjects were not found to have any difficulty in the pronunciation of /s/ and /z/. After being audiovisually trained, their pronunciation of /s/ and /z/ was assessed to be correct as well.

5.6.6 Individual variances

According to the data presented above, all the subjects in the experimental group achieved significant improvement both in the perception and production of the target contrasts from pre-test to post-test. Their degree of improvement, however, varied across the subjects. There was also considerable individual variation in perception and production accuracy at the end of the training programme. For instance, S10 achieved an accuracy of above 90% both in the perception and production of the target contrasts after being trained for only 3 sessions. S3 achieved similar performance at the end of the 6th training session. In contrast, none of the remaining subjects' accuracy was above

90% both in the perception and production of the target contrasts at the end of the training programme, despite some of them achieving an accuracy of above 90% either in the perception or the production of the contrasts.

A striking individual difference emerged from the comparison between S16 (around 60%) and S9's (above 80%) accuracy in the production of /θ/ and /ð/ in the post-test, since S16 performed slightly better than S9 in the pre-test. Moreover, some subjects displayed a higher degree of improvement from pre-test to post-test than others (see Appendix 11). For instance, S17 and S27's perception accuracy was comparable in the pre-test (S17: 21.29% vs. S18: 23.15%). At the end of the training programme, S17's perception improvement was about 55%, whereas S27 improved by less than 40%. Additionally, from pre-test to post-test, most subjects showed a higher degree of perception improvement than production improvement. Nevertheless, a few of them achieved a higher degree of production improvement than perception improvement (as seen with S9, S24 and S29).

Further insights into individual differences can be gained by examining two of the best and two of the poorest performing subjects' perception and production performance in the main study. S3 and S10 were the only two subjects who achieved an accuracy of above 90% both in the perception and production of the target contrasts before the end of the training programme. According to their answers in the questionnaire (see Appendix 2), both S3 (male) and S10 (female) had been learning English as an L2 for 7 years. However, S3 was the only subject who studied English primarily as a hobby rather than for getting high scores in English exams. Moreover, S3 spent 1-2 hours per day watching English movies, listening to English songs and the BBC news, as well as reading English Newspapers. In comparison, S10's primary motivation for learning English, as for the rest of the subjects, was to get high scores in English exams. She spent 1-2 hours learning English in her spare time by doing English exercises and watching English movies. Although their perception and production performance was comparable in the pre-test (see Appendix 10), S10 seemed to have completely learned the target contrasts after only 3 training sessions, whereas it took 3 more sessions for S3 to learn the contrasts. The *repeated-measures ANOVA* results indicate that the male subjects performed better than the female subjects in the experimental group. S10's performance provides supporting evidence for this finding.

Regarding the subjects who performed the poorest at the end of the training programme, both their perception and production performance was taken into consideration. S14 and S16 were selected. Their accuracy levels in the perception of the two contrasts was below 90% in the post-test. In the production of /θ/ and /ð/, S14's accuracy was 74.70% and 73.30% respectively. S16's production performance was even poorer, with an accuracy of 58.00% in the production of /θ/, and 66.38% in the production of /ð/, which were the lowest accuracies among all the subjects. Both S14 and S16 were female, and had a similar profile to the rest of the subjects. Their performance provided further evidence in support of the statistical significance of *gender* on the subjects' perception and production performance. Although the majority of the subjects reported that they spent some of their spare time on L2-English learning, the way(s) in which they had been learning English may have in-part contributed to their variant performance. For instance, S14 learned English in her spare time through watching English movies, listening to English songs in addition to doing exercises in English textbooks. In contrast, S16 reported that the only method she employed for English learning in her spare time was doing exercises in English textbooks. S14's comparatively better production performance in the post-test, therefore, might be attributable to the different methods which they employed in English learning.

5.7 Answers to the research questions

So far, the research questions can be answered with the findings presented above.

(1) To what extent, if at all, can the subjects' capability in the auditory perception of English contrasts /θ/-/s/, /ð/-/z/ be improved by audiovisual perception training?

According to the results, the capability of all the subjects in the experimental group regarding the auditory perception of English /ð/-/z/ and /θ/-/s/ was improved after the audiovisual perception training. The mean improvement in accuracy was about 50% in in the perception of both contrasts in the post-test compared against the pre-test (see Table 5.1). The degree of improvement varied across the subjects (see Appendix 11). Due to being tested 4 times with the same stimuli, the subjects' (experimental group) improved accuracy in the perception of the target contrasts may be in part attributed to the repeated testing experience. Nevertheless, compared to the control group, the experimental group showed a significantly higher degree of improvement. The experimental group's perception improvement, therefore, can be largely attributed to audiovisual training, rather than to repeated testing experience.

> *(2) To what extent, if at all, can the subjects' capability in the production of English contrasts /θ/-/s/, /ð/-/z/ be improved by audiovisual perception training?*

All the subjects' capability in the production of English /ð/ and /θ/ was improved by the audiovisual perception training. The mean accuracy of improvement was about 44% in the production of /θ/, and about 38% in the production of /ð/ (see Table 5.12). The degree of improvement varied across the subjects (see Appendix 11). Moreover, as in the pilot-study, the subjects' production of English /s, z/ was evaluated to be native-like (a score of 10 in the *10-score Likert scale*).

5. 8 Conclusion

This chapter displayed the main study of the present study, which was the central part of the study. The aim of the present study was to investigate whether audiovisual perception training can facilitate L2 learners' auditory perception and production of the target contrasts. Accordingly, two research questions were formulated and hypotheses devised in view of the theoretical considerations discussed in the literature review. In the methodology part, the preparation of the stimuli (both for training and testing), and the procedure of the study were described in detail. For the presenting of the findings, tables, graphs were employed for the comparison of the subjects' perception and production performance in the four tests, as well as the data collected from the questionnaire. A *repeated-measures ANOVA* was employed to detect the factors that had a significant/non-significant effect on the subjects' perception and/or production performance. Moreover, the research questions were answered with the findings of the main study.

Chapter 6 Discussion and conclusion

6.1. Introduction

The final chapter discusses the findings of the main study. First of all, the hypotheses of CAH, CPH, PAM-L2, SLM, NLM/NL-e and PI, which serve as the main theoretical basis of the present study, are applied to the discussion of the findings. The findings in the main study are compared with those in previous studies as reviewed in the literature review (chapter 2). Moreover, the factors that were revealed to have had a significant effect on the subjects' perception and/or production of the target contrasts are discussed. After that, the present study is briefly summarized. The advantages and limitations of the present study are analysed critically to identify how progress can be made in future studies. Some suggestions for future research are given.

6.2 Discussion of the main study

6.2.1 The effect of audiovisual perception training on the subjects' auditory perception and production performance

One of the major findings of the present study was that compared with in the pre-test, the accuracy of all the subjects in the experimental group in the auditory perception and production of the target contrasts improved to different degrees by the post-test. Their improvement was revealed to be statistically significant. Audiovisual training was found to have had a significant effect on their improvement. Moreover, due to the effect of repeated testing experience, the control group also showed some auditory perception improvement from pre-test to post-test. Nevertheless, their degree of improvement was statistically lower than that of the experimental group. The findings of the present study may shed some light on the hypotheses of the theories/models discussed in the literature review.

1. PAM-L2—Perception Assimilation Model-L2

One of the essential hypotheses of PAM-L2 is that the perception of speech sounds occurs through the discovery of the articulatory gestures of the target speech sounds. Language learners are predicted to assimilate unfamiliar L2 speech sounds to the most articulatorily-similar sounds in their L1 (Best & Taylor, 2007). According to the production test results in the pilot-study, the majority of the subjects realized /θ/ as /s/, and /ð/ as /z/. In terms of the hypothesis of PAM-L2, this finding may be caused by the fact that the articulatory gestures of Mandarin/ CQd /s, z/, which are produced as dental, are similar to, or even the same as that of English /θ, ð/ when produced as dental. Nonetheless, after undergoing the audiovisual training programme, all the subjects in the experimental group achieved significant improvement both in the perception and production of /θ/-/s/, /ð/-/z/. *Training* was detected to be a factor that had significantly affected their improvement in the accurate production of the target contrasts. Given that the training programme involved audiovisual demonstration of the articulatory gestures of /θ/-/s/, /ð/-/z/, it might have been the visible articulatory differences between /θ/ and /s/, /ð/ and /z/ that facilitated the subjects' improvement. It may also be possible that their improved accuracy in the perception and production of the contrasts was because they perceived the acoustic differences between /θ/ and /s/, /ð/ and /z/, since the audiovisual training programme provided them with both visual codes and auditory information. Nonetheless, the investigator personally observed the experimental group's production of /θ/ and /ð/ in the 4 production tests, and found that they did perceive the articulatory differences between /θ/ and /s/, /ð/ and /z/. In the pre-test, none of the subjects raised their tongue to in between the teeth when producing the two sounds. In mid-test 1, some of the subjects began to pronounce the two sounds as interdental. It may be because interdental is a non-native viseme and the subjects were unfamiliar with the use of it that they did not manage to produce every stimulus word that contained /θ/ and /ð/ correctly. However, in mid-test 2, and particularly in the post-test, more subjects produced /θ/ and /ð/ as interdental. Nevertheless, due to a lack of further evidence, it was unclear whether the subjects' improvement in the perception and/or production performance was influenced by the observed articulatory gestures, perceived acoustic differences between /θ/ and /s/, /ð/ and /z/, or both.

Another important hypothesis of PAM/PAM-L2 is that even adult L2 learners can eventually learn L2 speech sounds that they initially have difficulty with. From the pre-test to tests after training, the experimental group's significant improvement in auditory perception and production performance may support this hypothesis. It may be because the training programme was not very long that the experimental group in the perception and production of the target contrasts did not achieve native-like performance (none of them received full scores both in the perception and production of the contrasts in the tests after training). If given further training, the results might be more satisfying.

2. SLM—Speech Learning Model

The experimental group's improved accuracy, both in the perception and production of the target contrasts in the tests after training, may provide supporting evidence for some of the hypotheses of SLM (Flege, 1981, 1987, 1988, 1991a, 1992a, b, 1995a). First of all, contrary to CPH, SLM predicts that language learners' capability remains intact throughout their life. All the subjects of the experimental group were adults. Comparing their perception performance in test 1 with that in test 4, the experimental group's accuracy in the perception of the target contrasts improved significantly. Due to the potential bias caused by using the same stimuli for perception tests, the experimental group's improved accuracy could be partly attributed to the repeated testing effect. However, compared with the control group, the experimental group showed a significantly higher degree of improvement. Thus, the experimental group's substantial improvement at the end of the training programme may indicate that their capability for L2 learning still remains.

Secondly, SLM predicts that the more dissimilar the L1 and L2 sounds are, the more likely it is that language learners will develop a new phonetic category for the L2 sounds (Flege, 1981, 1987, 1988, 1991a, 1992a, b, 1995a, b, 2002, 2003). As discussed in chapter 3, when produced as interdental and alveolar respectively, /θ/ and /s/ are distinct from each other in terms of articulatory gestures and acoustic characteristics, despite the fact that both of them are voiceless fricatives. The same applies to /ð/ and /z/, though they are both voiced fricatives. It seems the experimental group's large degree of improvement during and at the end of the training programme supports this hypothesis. Nevertheless, the subjects in the control group also achieved significant improvement from pre-test to post-test in the perception of the target contrast, though they were not audiovisually trained. Their improvement was found to be significantly due to repeated testing experience. Accordingly, the experimental group's perception improvement cannot be totally attributed to the training effect. Therefore, it is not clear to what extent the dissimilarity between /θ/ and /s/ and /ð/ and /z/ contributed to the experimental group's improved accuracy in the discrimination of the contrasts.

Thirdly, the subjects' production of English /s/ and /z/ was evaluated as native-like both in the pilot study and the main study, despite the fact that Mandarin /z/ and CQd /s, z/ vary slightly from the English sounds in terms of acoustic properties and articulatory gestures depending on speaker differences. At first glance, this finding seems to be at odds with the “equivalence classification” predicted by SLM, which indicates that if an L2 sound is similar to or identical to the counterpart in their L1, language learners may be able to perceive the acoustic differences, but unable to use the perceived acoustic differences in the production of the sound. Nonetheless, the accuracy of the subjects’ production of the target contrasts was evaluated by native English speakers alone, without being acoustically analysed. It is possible that the subjects’ production of /s, z/ displayed acoustic differences from that of native English speakers, which were not detected by the raters.

Another hypothesis of SLM is that greater L2 experience can help language learners’ perception and production of L2 speech sounds (Flege, 1981, 1987, 1988, 1991a, 1992a, b, 1995a, b, 2002, 2003). Findings from the experimental group may have provided supporting evidence for this hypothesis.

In the perception of the target contrasts, the *Post Hoc Tests* (see Table 5.8) indicate that the experimental group’s improved mean accuracy from the pre-test to mid-test 1, from mid-test 1 to mid-test 2, and from mid-test 2 to the post-test were all statistically significant. That is, the more training sessions the subjects went through, the higher accuracy they achieved in the perception of the target contrasts. Although the control group, who were not audiovisually trained, also achieved significant improvement from pre-test to post-test, the degree of their improvement was significantly lower than that of the experimental group (see Chapter 5, Table 5.48). Therefore, it might be possible to speculate that the experimental group’s substantial improvement is largely due to the L2 experience from the audiovisual training programme.

In the production of /θ/ and /ð/, statistical results from a *repeated-measures ANOVA* confirmed the significant effect of *training* on their production performance. Moreover, the data in Table 5.12 and Appendix 10 indicates that the experimental group’s accuracy in the production of /θ/ and /ð/ improved linearly and stably from mid-test 1 to the post-test, both as a group and individually. Based on these data, we may predict that if given more training sessions, the subjects’ production performance would be further improved.

SLM also predicts that L2 speech perception precedes its production. L2 perception training can eventually lead to improvements in L2 production, because a foreign accent is in part caused by the inaccurate perception of L2 speech sounds (Flege, 1987, 1995, 2003). In the present study, the subjects of the experimental group only received speech perception training. As a result, their accuracy in the production of the target contrasts improved together with perception performance. Although this finding does not serve to prove whether speech perception precedes production, it is in accordance with the view that the accurate perception of L2 speech sounds contributes to the correct production of them.

Additionally, SLM hypothesizes that L2 speech sounds can be produced only as accurately as they are perceived (Flege, 1987, 1995, 2003). According to the findings, although both showed significant improvement at the end of the training programme, the subjects' perception performance surpassed their production performance (see Appendix 11). In the perception tests, at the end of the training programme, except for S29 whose accuracy was 79.63% in the perception of /ð/-/z/, most of the subjects' accuracy was around 90%. In respect of production, however, the subjects' accuracy ranged from around 60% to above 90% in the post-test. Most of the subjects' accuracy was around 70%-80%, which was lower than in the perception test. With these findings, it might be tempting to agree with this hypothesis of SLM. However, as suggested by Bradlow et al. (1997), since no extensive production training was carried out, it is not surprising to have this result. Further production training may result in greater improvement in the production of /θ/ and /ð/. Moreover, repeated testing experience was revealed to have significantly benefited the control group's perception performance. Accordingly, the experimental group's perception improvement may be partly attributed to the repeated testing effect. The experimental group's real perception improvement, therefore, might be less than that was obtained from the AXB tests. In other words, it was unclear whether (or to what extent) the experimental group's perception accuracy was better than their production accuracy at the end of the training programme.

3. NLM/NLL-e— Native Language Magnet theory/Native Language Magnet theory-e

In respect of NLM and NLM-e (Kuhl, 1992, 1994), the central hypothesis of the two models is the constraint of language learners' early language experience (typically, their L1) on their acquisition of L2 speech sounds. NLM-e predicts that adult L2 learners can circumvent the negative influence of their L1 by recapitulating the way in which infants learn L1 speech sounds. That is, by receiving exaggerated L2 input with "multiple instances by many talkers, and massed listening experience" (Kuhl et al., 2008; also see the review from Flege, 2003). During the recording of the training materials, the RP speakers were asked to exaggerate the articulatory gestures in the production of the target contrasts, which mimicked infant-directed speech. Specifically, when producing /θ/ and /ð/, the RP speakers raised their tongue blades to in between the upper and lower teeth, so that the subjects could observe the articulatory differences between /θ/ and /s/, /ð/ and /z/. The positive training results may provide supporting evidence for the effects of exaggerated cues on guiding the language learners' perception and production of L2 speech sounds.

Moreover, according to NLM-e, early language experience constrains learners' future learning of L2 speech sounds, specifically in terms of tuning their language "map" to their L1. However, adult language learners are predicted to be able to acquire L2 speech sounds eventually. In the present study, the subjects initially realized /θ/ as /s/ and /ð/ as /z/ in the production test, and had difficulty in distinguishing /θ/- /s/ and /ð/- /z/ in the perception test. Their substantial improvement in the perception and production of the target contrasts at the end of the training programme is consistent with this prediction.

NLM-e also attaches great importance to social interaction for early language learning on the phonetic level, and this is predicted to be the same for L2 learning (Kuhl, 1992, 1994). In the present study, subjects in the experimental group watched the training video passively without any interaction with the RP speakers during the training programme; in the identification task, they were given immediate feedback concerning the correct answer in each trial. On this point, it seems the successfulness of the training programme in the present study may provide counterevidence to this hypothesis. Nonetheless, as mentioned above, due to the significant influence of repeated testing experience, the experimental group's actual degree of perception improvement may be less than that which was observed. Furthermore, it might be possible that their perception/production performance could be further improved if given opportunities for interaction during the training.

Another important hypothesis of NLM-e is that speech perception precedes speech production, and thus the perception-production link is forged developmentally (Kuhl, 1992, 1994). This is identical to the prediction of SLM. In the present study, it was the perception training that led to the experimental group's improvement in the production of the target contrasts. Thus this may confirm the link between speech perception and production. However, it could be possible that production training would also successfully improve their perception performance. For instance, in Williams and

McReynolds (1975), speech production training led to the subjects' successfulness in the perception of the trained speech sounds.

4. PI— Perception Interference

Similar to NLM/NLM-e, PI predicts that due to the influence of language learners' L1, whether language learners can acquire L2 speech sounds depends on the degree of interference between their L1 and the L2 speech sounds (Iverson et al., 2003). Although it is not clear to what extent the subjects experienced interference upon /θ, ð/ from /s, z/, their improved perception and production performance during and at the end of the training programme may provide supporting evidence for the prediction of PI. That is, even adult language learners can eventually learn L2 speech sounds. However, PI posits that the critical acoustic cues that L2 learners depend on in the perception of L2 speech sounds may be different from those employed by native L2 speakers, but this is beyond the research domain of the present study.

5. CPH—Critical Period Hypothesis

Let us apply the results of the present study to the CPH. According to the CPH, due to loss of the neural plasticity which is relevant to language acquisition, L2 learners are predicted to be unable to achieve a native-like proficiency level if they commence their L2 study after the so called "critical period" (Lenneberg, 1967; Oyama, 1976). In previous studies, AO has been found to have a significant effect on language learners' L2 proficiency level, particularly with regard to accent. L2 learners who commenced their L2 learning after puberty are often found to have a detectable foreign accent, whilst those who started their L2 study before puberty are usually revealed to be accent-free (Tahta et al., 1981; Flege et al., 1995; Flege et al., 1999; Bongaerts et al., 1997). In the meantime, L2 learners with younger AOs of L2 learning are detected to have better perception performance than those with older AOs (Mayo et al., 1997; Shi, 2010).

Among the subjects of the experimental group, 17 of them did not commence their L2-English study until 13 years old. Another 12 subjects' AO of L2-English learning was 14 years old. Given that no consensus has been achieved regarding the exact age when the "critical period" ends, it is difficult to assess whether the subjects' AO is before or after the end of the "critical period". For instance, if the "critical period" is defined as ending at 9 years old (Penfield and Roberts, 1959) or 12 years old (Scovel, 1988), then all the subjects started their L2-English learning after the “critical period”. If it is defined as 11-14 years old (Lenneberg, 1967), however, the subjects’ AO of learning English as an L2 would be at the end of the “critical period”. If the end of the “critical period” is defined as 15 years old (Patkowski, 1990), however, all the subjects commenced their L2-English study before the end of the “critical period”. Therefore, this finding itself can hardly provide either supporting or disproving evidence for the CPH.

6. CAH—Contrastive Analysis Hypothesis

The central hypothesis of CAH is that the differences between language learners’ L1 and L2 pose difficulty for their L2 learning, whereas the similarities between their L1 and L2 facilitate their L2 acquisition. Given that English /θ/ and /ð/ are missing from the phonetic inventories of the subjects L1 and L1-dialect, the subjects’ poor perception and production performance in the pilot-study seems to provide supporting evidence for this hypothesis. Nevertheless, this hypothesis itself lacks specific quantification, both regarding how to determine the degree of “differences/similarities” and the extent to which the difficulty and/or facilitation can influence L2 learning. Moreover, CAH does not predict whether or how L2 learners can eventually overcome the difficulty which results from the differences between their L1 and the L2. Thus, the subjects’ improved accuracy in the perception and production of /θ/-/s/ and /ð/-/s/ in the tests after training might be viewed as irrelevant to the hypothesis of CAH.

On the whole, the experimental group’s improved accuracy in the perception and production of the target contrasts in the tests after training provides supporting evidence for the common hypotheses of PAM/PAM-L2, SLM, NLM/NLM-E and PI. That is, even adult language learners can ultimately learn L2 speech sounds that they initially have difficulty with.

6.2.2 *The effect of articulatory information on the subjects’ perception performance*

Compared with the pre-test, the experimental group’s accuracy in the perception of the target contrasts substantially improved in the tests after training. Although there was some repeated testing effect, as found for the control group, their increase in accuracy was significantly greater than that of the control group. Given that in the training programme, the subjects received an articulatory demonstration of /θ/-s/, /ð/-/z/, it might be possible to speculate that visual cues facilitated their perception performance. This may provide evidence for some hypotheses concerning the critical role of articulatory information in speech perception. Nevertheless, due to lack of comparison between the effects of auditory-only and audiovisual training in the present study, the role of visual cues in facilitating the experimental group's perception and production of the target contrasts may be mitigated.

Moreover, given that the experimental group's perception improvement, to a large extent, can be attributed to the audiovisual training, this could confirm the view that instead of being independent skills, audiovisual, auditory and visual skills in speech perception are integrated with each other (Berstein et al., 2013). This finding is at odds with those presented in Grant and Seitz (1998), James (2009), Gariety (2009), and DiStefano (2010), which suggest that audiovisual integration is independent from auditory and visual skills in speech perception. For instance, in DiStefano (2010), audiovisual training on the perception of bilabial, alveolar and velar contrasts did not improve the subjects' capability in the auditory or visual perception of these sounds. There might be two reasons for the discrepancy. First of all, the stimuli used in DiStefano (2010) only included 8 different words, though with different patterns of combinations. In the present study, however, 60 different minimal pairs of each contrast were created in each training session. Therefore the subjects were exposed to a much wider range of stimuli in the present study than in DiStefano (2010). Consequently, the subjects in the present study may have benefited more than those in DiStefano (2010). Secondly, all the stimuli employed in the present study were naturally produced and not synthesized. In DiStefano (2010), however, degraded stimuli were adopted. As predicted by Logan et al. (1991), synthetic speech may mislead, or provide subjects with incomplete information about the target phonetic category in speech perception. On the whole, the present training mainly followed the HVPT approach, which emphasizes "natural variability" (Logan et al., 1991; Yamada, 1993). In comparison, the approach in DiStefano (2010) seems more like LVT, despite the fact that five different speakers were asked to record the training stimuli.

In addition, the employment of non-native language visual cues in the perception and particularly in the production of the target contrasts by subjects in the experimental group is at odds with the hypothesis that tone language speakers are less likely to use visual information in non-native speech perception/production (Sekiyama, 1997; Sekiyama and Tohkura, 1993). In previous studies, Mandarin speakers showed a relatively lower degree of use of visual information in speech perception than non-tone language speakers (de Gelder and Vroomen, 1992). It has been argued that since Mandarin is a tone language, L1-Mandarin speakers rely more on tones than on visual cues in speech perception (Sekiyama, 1997; Sekiyama and Tohkura, 1993). Moreover, Hazan et al. (2006) indicated that L2 listeners may lose sensitivity to visemes that do not exist in their L1. These predictions would be confirmed if we look at the results in the pre-test, in which the subjects' accuracy in the perception of the target contrasts was pretty low. Before being audiovisually trained, the subjects had been learning English for 6 to 8 years. Although they all reported that they had never been to English-speaking countries, they had definitely watched English movies. Moreover, as far as the author knows, each week they had one class given by native English speakers in the first year of their university study. There might have been other chances to speak to, or observe the speech of other native English speakers. Nonetheless, their perception and production performance in the pre-test indicate that they may have not discovered the non-native language viseme – interdental. However, after being audiovisually trained, the subjects' accuracy in the perception and production of the target contrasts substantially increased from the pre-test to tests after training. Apart from the scores in the perception and production tests, the author personally observed their production, and found that most of them placed their tongue in between the upper and the lower teeth to different degrees when producing /θ, ð/ after being trained, particularly in the post-test. On this point, two conclusions can be reached: (1) as hypothesized by Wang et al. (2009), language learners are able to discover and use non-native visual cues in speech perception and production; (2) audiovisual training may facilitate language learners' correlation of non-native speech sounds with corresponding visual cues (Hazan et al., 2005).

6.2.3 Transferred effect of perception training on speech production.

Given that no exclusive training on production was carried out, the experimental group's improved accuracy in the production of the contrasts in the tests after training is attributable to the audiovisual perception training. In other words, speech perception training displayed a transferred beneficial effect on speech production. This finding is congruent with that in some previous studies (Jamieson and Rvachew, 1992; Bradlow et al., 1997; Lambacher et al., 2005). It further confirmed the hypothesis of Flege's SLM – L2 perception training can eventually lead to production improvement (Flege, 1981, 1987, 1988, 1991a, 1992, 1995a). However, it is at odds with findings in some early experiments, in which perception training only served to improve the subjects' perception performance, rather than benefiting their production ability as a whole. For example, in Williams and McReynolds (1975), although production training enhanced the subjects' perception and production capability, perception training only improved their perception performance. Similar findings are available from studies in Winitz and Bellerose (1963), Guess (1969), Guess and Baer, (1973). The discrepancy between findings in these studies and those of the present study can be explained by the differences in training approaches. In these studies, perception training was conducted with an auditory modality, while an audiovisual modality was employed in the present study. The subjects of the present study may, somehow, have benefited from the visual codes demonstrated by the RP speakers. Nonetheless, due to only perception training being conducted in the present study, the finding could not serve to answer the question of whether production training can benefit speech perception.

Moreover, as mentioned above, the subjects applied the observed visual cue (interdental) in the audiovisual training programme to produce the target contrasts in the tests after training. It is predicted that during the process of language learning, *mirror neurons* enable language learners to learn and imitate on the basis of observation (Fadiga et al., 1995; Strafella and Paus, 2000; Kohler et al., 2002; Rizzolatti and Craighero, 2004). This finding has provided us with further evidence in support of the close relationship between speech perception and production. As observed by the investigator, the subjects' production performance showed that they mimicked the RP speakers' production of /θ/ and /ð/ in terms of articulatory gestures. Consequently, the discovery of the articulatory gestures may have facilitated their perception of /θ/-/s/ and /ð/-/z/. However, this result neither serves to demonstrate that speech perception and production share a common link or a common processing strategy, as proposed by MT (Liberman et al., 1967; Liberman and Mattingly, 1985; 1989; Hawkins, 1999; Liberman and Whalen, 2000), nor does it support the view that instead of being tied together by a mediating link, speech perception and production form an integrated system, as proposed by Fowler (1986).

6.2.4 Factors that significantly affect the experimental group's perception/production performance

According to the statistical analysis results from the *repeated-measures ANOVA*, in the perception tests, except for the *training* effect, the factors *gender*, the interaction between *training* and *gender*, *phonetic position*, and the interaction between *phonetic* *position* and *training* were all revealed to have had a significant effect on the experimental group's perception performance. Meanwhile, *gender* and its interaction with *training* also had a significant impact on the experimental group's production of /θ/ and /ð/.

6.2.5 Gender difference

Let us first have a look at the effect of the gender difference on the subjects' perception and production performance. There were 15 female and 14 male subjects in the main study. Results from the *Post Hoc Tests* indicate that the male subjects in the experimental group showed better perception and production performance than the female subjects (see Chapter 5 for details). Nonetheless, there were individual differences. For instance, S10 (female) achieved an accuracy of above 90% both in the perception and production of the target contrasts after being trained for only 3 sessions.

In some previous studies on speech perception and production, the gender difference of subjects was either not specified as a significant factor for the subjects' perception and/or production performance (Flege and Fletcher, 1992; Elliott, 1995), or revealed to be statistically non-significant for the subjects' perception and/or production of L2 speech sounds (Piske, MacKay, and Flege, 2001). However, in previous studies on SLA, the female language learners were found to out-perform males in terms of being more mature and serious about their studies (Clark, 1995; Clark and Trafford, 1995; Wright, 1999), or made greater use of strategies in learning vocabulary than the male learners (Catlan, 2003), which may consequently have resulted in their greater achievement in L2 learning (Asher and Garcia, 1969). Nevertheless, there are some studies in which the male subjects performed better than the female subjects. For instance, in Fullana and Mora (2008), the male subjects displayed a higher correctness rate than female subjects in the perception of English voicing contrasts in word-final position.

In some studies, gender difference was found to be statistically non-significant for the subjects' perception and/or production performance, it could be attributed to the interaction with other social factors. For instance, Piske et al. (2001) revealed that gender difference did not display a significant effect on the subjects' correctness in L2 pronunciation. They attributed this to the neutralized effect of the interaction among gender, AO of L2 learning and the amount of L2 experience. The neutralization effect, however, may not work in the present study. Subjects in the experimental group were characterized by quite similar ages, AO of L2 learning and amount of L2 experience (years of L2-Englsh learning, and the ways in which they had been learning English). All of them reported that they had been learning English in public school/university. Moreover, they neither had any chance to use English on a daily basis nor had travelled to/lived in English-speaking countries. Except for S3, all the rest of the subjects' primary reason for English learning was to get high scores in exams.

Therefore, the most convincing explanation for the male subjects' better performance than the female subjects may be their greater visual-spatial ability (Bouchard and McGee, 1977; Harris, 1978; Sanders et al., 1982; Goldstein et al., 1990). During the audiovisual training programme, the visible articulators were the RP speakers' tongue tip, teeth and lips. The inside part of the mouth was not visible. The male subjects may have used the visible articulators to form a complete picture of the movements of the articulators, which may have consequently led to their better performance in the perception and production of the target contrasts.

6.2.6 Phonetic position* and its interaction with *training

Both *phonetic position* and its interaction with *training* were detected to be significant for the experimental group's perception performance. They were revealed to perform better in the perception of the target contrasts in initial and medial positions than in final position, despite the perception training stimuli including the target contrasts embedded in all three positions. The subjects' perception performance as a function of *phonetic position* replicates findings from some earlier studies, such as Bradlow et al. (1997) and Gillette (1980), and Lively et al. (1993). In these studies, the subjects' perception performance was found to be significantly different across different phonetic positions. Particularly, it is congruent with the findings in Flege (1989), in which Chinese subjects had difficulty in the perception of English /t/-/d/ in word-final position. This was explained by the fact that word-final /t/ and /d/ do not occur in Chinese. This explanation can also be employed to explain the finding in the present study. That is, the non-occurrence of /s/, /z/ and their replacement /θ/, /ð/ in syllable final position in Mandarin and CQd may have led to the subjects' difficulty in the perception of the target contrasts in word-final position. This point is in accordance with the prediction that L2 learners' syllable-processing strategies vary according to L1 differences (Cutler et al., 1983, 1986; Flege, 1989; Flege and Davidian, 1984; Flege and Wang, 1989), which may further confirm the influence of the L1 on language learners' L2 learning.

Moreover, it is also congruent with the prediction of CAH. That is, the differences on syllable structure between the subjects' L1/L1-dialect and L2 may have posed difficulty for their perception of the L2 contrasts.

However, *phonetic position* and its interaction with *training* did not show a significant effect on the subjects' production of /θ/ and /ð/. Their accuracy in the production of /θ/ and /ð/ varied across initial, medial and final positions without showing a regular pattern. This finding is at odds with the suggestion that language learners learn "syllabically" (Morosan and Jamieson, 1989), since the training stimuli embedded the target contrasts in different phonetic positions. It is also different from findings in some previous studies, such as Bada (2001) and Flege and Davidian (1984), in which the subjects' production of the target speech sounds was significantly affected by phonetic position. It seems the subjects' production of /θ/ and /ð/ was not affected by the syllable rules of their L1 and L1-dialect. Their improvement in the production tests after training revealed that they had successfully discovered the articulatory gesture (interdental) of /θ/ and /ð/. Nonetheless, perhaps due to the lack of exclusive production training, the subjects may either be hesitant or unfamiliar with the use of this non-native articulatory gesture in the production of /θ/ and /ð/. If audiovisually trained for a longer period of time, or given further training on the production of the target contrasts, the subjects may show greater production improvement. In the meantime, *phonetic position* may show some effect on their production performance.

6.2.7 Factors that are non-significant for the experimental group's perception/production performance

6.2.7.1 Vowel context

As discussed in the literature review, the coarticulation effect is attached great importance in speech perception, and it may change some acoustic characteristics of a target speech sound (Chomsky and Halle, 1968; Nittrouer and Studdert-Kennedy, 1987; Pickett, 1999; Kent and Read, 2002). Nonetheless, in the experimental group, it seems the subjects were not affected by the coarticulation effect. Vowel context was revealed to be non-significant for the subjects' perception performance. This finding is in accordance with that in Liberman et al. (1967), in which listeners successfully identified /d/ both in /u/ and /i/ contexts, despite the F2 trajectory varying in the two different vowel contexts. This may be because, as predicted by MT, language listeners are able to perceive the intended articulatory gestures of a target speech sound.

However, this finding seems to be at odds with some findings in previous studies, in which the subjects' responses were significantly affected by vowel contexts. For instance, in Mann and Repp (1980), in the perception of synthetic fricative noises from a /ʃ/-/s/ continuum, the subjects perceived more instances of /s/ in the /u/ context than in the context of /a/. The discrepancy between the finding in the present study and that in Mann and Repp (1980) might be explained by the difference of stimuli used in the tests. The stimuli used in the present study were naturally produced, whereas synthesized stimuli were employed in Mann and Repp (1980). As suggested by Logan et al. (1991), synthetic speech may not provide the subjects with complete information about a target speech sound. Although some studies revealed that the perception of a speech sound may largely depend on the length of its adjacent segment (Miller and Liberman, 1979; Diehh, Souther, and Convis., 1980; Miller, 1987; Summerfield, 1981), the vowel contexts in the present study differ in terms of height, backness and roundness, rather than length. Therefore, the influence of the length of the adjacent vowel could be precluded.

Nonetheless, in some previous studies, naturally produced stimuli were used, yet the subjects' perception performance was revealed to be a function of vowel context. For example, Hardison (2003) reported that there was a significant effect of vowel context on the subjects' perception of English /l/-/r/. The L1-Japanese speakers had serious difficulty in the perception of initial clusters with the vowel contexts /u, o/, whereas L1-Korean speakers' most challenging phonetic environment was the final singleton with /i, ɪ/. The findings were explained by the negative influence of the phonetic inventories of their L1s. The discrepancy between the findings in Hardison (2003) and those in the present study might be because of the variance of training materials. In Hardison (2003), the adjacent vowels of the target contrast were /u, o, ɑɪ, e, ɛ, I, ɪ/. In the present study, however, a larger number of adjacent vowel and consonant contexts were employed in the training materials, which may have better enhanced their capability in the perception of the target contrasts. As a result, the effect of vowel context on the subjects' perception performance was found to be non-significant.

6.2.7.2 General factors

When doing the statistical analysis, due to the subjects having the same/very similar answers to most of the questions in the questionnaire, these factors were not employed as between-subjects factors in the analysis of their perception and production performance. Nonetheless, these factors may provide us with further evidence regarding the subjects' perception and production performance.

1. Motivation

Let us first have a look at the subjects' *motivation*. Except for S3 who had been learning English primarily as a hobby, all the rest of the subjects reported that they had been learning English for the purpose of getting high scores in English exams. This is due to the educational system in China. Nowadays, most universities in China require the students to pass the CET-4 exam (College English Test – level 4). It is one of the prerequisites for getting their Bachelor's Degree. Moreover, passing all the English exams at university is also compulsory. Therefore, most students who are not English majors are still studying English largely for this reason.

It may be hard to evaluate which purpose motivates the subjects more in L2-English learning. Nonetheless, Gardner (1985) suggests, self-motivated learners may desire to interact with the target language group. They may also have positive attitudes toward the learning of the target language, and thus desire to learn the language. Findings in some previous studies confirm this point of view, such as MacNamara (1973), Flege (1987), Suter, 1976, Purcell and Suter (1980), and Elliott (1995). In the present study, it may be because S3 had been learning English as a hobby that the ways he employed in English learning in his spare time were different from other subjects. He reported that he had been learning English through reading English newspapers; watching English movies, listening to English songs and the BBC news. In comparison, most of the rest of the subjects preferred to do exercises in English textbooks, because the items in English exams are mostly from the exercises. S3 achieved an accuracy of above 90% both in the perception and production of the target contrasts in mid-test 2, and so was dropped from the following 3 training sessions. He was one of the two subjects who achieved satisfactory perception and production performance before the end of the whole training process. According to this finding, it seems learning an L2 as a hobby may be better able to motivate a learner compared with learning it for exams.

Nevertheless, all the rest of the subjects showed significant improvement in the perception and production of the target contrasts, despite their primary motivation for L2-English learning being to get high scores in exams. On this point, it seems even without a strong motivation, such as learning an L2 as a personal hobby, language learners can also acquire non-native speech sounds if given sufficient L2 input. This is congruent with findings in Oyama (1976) and Thompson (1991), in which motivation was suggested to be non-significant for language learners' degree of L2 proficiency, specifically concerning foreign accent.

2. Amount of L2-English learning experience

The amount of experience in language learning is predicted to be significant for language learners' achievement in L2 learning (Flege, 1981, 1987, 1988, 1991a, 1992a, b, 1995a, b, 2002, 2003; Cumming, 1994; Carroll, 1969; Riney and Flege, 1998; Purcell and Suter, 1980). In the present study, the subjects' answers to the question of how many years they had been learning English and the amount of time they spent on English learning in their spare time may shed some light on this issue.

Among the subjects in the experimental group, the majority of them had been learning English as an L2 for 6 or 7 years; only S13 had been learning English for 8 years. Yet the comparatively longer time of English learning seemed not to benefit her much in the perception and production of the target contrasts. S13's accuracy in the pre-test was around 40% in the perception of /θ/-/s/ and /ð/-/z/, and about 30% in the production of /θ/ and /ð/. At the end of the training programme, she achieved an accuracy of about 89% in the perception of /θ/-/s/ and /ð/-/z/, and about 75% in the production of /θ/ and /ð/, which was a medium level performance compared with other subjects. This may be explained by the English educational system of China. In China, English teaching attaches great importance to grammar and comprehensive reading, which is also embodied in English exams. As a result, although S13 had been learning English for one or more years more than other subjects, she may not have benefited from English classes.

Moreover, the majority of the subjects reported that they learned English in their spare time (n=25). Among the 25 subjects, the amount of time they spent on English learning, and the ways they learned English in their spare time were similar to each other (see Appendix 2). That is, most of them reported that they read articles, did exercises, and recited vocabulary items in their English textbooks. These results most likely have something to do with their primary *motivation* in English learning. The methods that most of the subjects employed in English study in their spare time could hardly help their perception and/or production of L2 English sounds, because they could not receive native English speakers' input in these ways. Although a few of them also watched

English movies or listened to English songs, the amount of time they spent in doing so was very limited.

3. Age

All the subjects of the experimental group were young adults (19-22 years old), thus there should not be a big difference in the cognitive ability among individual subjects, despite biological aging being predicted to start from 20 years old (Birdsong, 2005, 2006, 2007). The experimental group's significant improvement in the perception and production of the target contrasts further confirmed the common hypothesis of PAM/PAM-L2, SLM, NLM/NLM-e and PI. That is, with sufficient L2 input, L2 learners can eventually acquire L2 speech sounds regardless of their age. Furthermore, it is predicted that adults are more likely to be influenced by visual information in distinguishing consonants than children (Massaro et al., 1986; Sekiyama et al., 2003). Since the present study involved an audiovisual training programme, the subjects being adults may have facilitated their perception and production performance.

6.2.8 Statistical analysis results for the experimental group's perception of /θ/-/s/ and voiced /ð/-/z/, as well as production of voiceless /θ/ and /ð/.

It seems that the subjects in the experimental group performed better in the perception of voiceless /θ/-/s/ than voiced /ð/-/z/ (see Appendix 10), and in the production of voiceless /θ/ than voiced /ð/. However, *Post Hoc Tests* results indicate that this difference was statistically non-significant. Considering the experimental group's perception performance, the *d-prime* scores, which preclude bias, show little observable difference between the subjects' perception of /θ/-/s/ and /ð/-/z/ than that shown by accuracy (see Table 5.3). With respect to their production performance, before being audiovisually trained, the subjects realized voiceless /θ/ as voiceless /s/ and voiced /ð/ as voiced /z/. Thus they were not likely to have difficulty with the pronunciation of voicing. This may explain the finding that the experimental group did not perform significantly different in the perception of voiceless /θ/-/s/ compared with voiced /ð/-/z/, as well as in the production of voiceless /θ/ and voiced /ð/.

6.2.9 Factors of significant and non-significant effect on the control group's perception performance.

The factors *repeated testing experience* (*experience* thereafter) and *phonetic position* of the target contrasts were revealed to have had a significant effect on the control group's perception performance. The remaining factors, which included *gender* difference, *vowel context*, and their interaction with each other and/or with *experience* were all found to be non-significant for the control group's perception performance. Moreover, the interaction between *phonetic position* and *gender* as well as between *experience* and *phonetic position* were also revealed not to have a significant impact on their perception of the target contrasts. The interaction between/among *experience* and/or *gender* and/or *phonetic position* and/or *vowel context* may have neutralized their effect on the subjects' perception of the target contrasts, and thus led to the non-significant effect. Here we will only have a look at the significant/non-significant effect of *repeated testing experience*, *gender* difference and *phonetic environments* on the control group's perception performance.

1. *Repeated testing experience*

It was found that *experience*, as a within-subjects factor, displayed a significant effect on the control group's perception performance. Given that subjects of the control group had a similar profile to that of the experimental group (i.e. same age, years of English learning, etc.), and the same testing materials were employed, it was quite possible that the subjects of the experimental group had also benefited from the repeated training experience in the perception tests. Thus the experimental group's perception improvement, to some extent, would be attributed to the influence of repeated testing experience. Nevertheless, the experimental group's perception improvement was statistically higher than the control group, and thus the beneficial effect of audiovisual training would have played a critical role.

The comparison between the experimental group and the control group's perception performance in the present study, however, does not seem to be consistent with that in some previous studies. For example, the subjects of the control group in Lively et al. (1994), who only participated in a pre-test and a post-test without being trained, did not show a statistically significant difference in the two perception tests. Similar findings are available from other relevant studies, such as Hardison (2003), Moradi et al. (2013), Lidestam et al. (2014). The discrepancy might be caused by the time of repeated testing.

The control group's perception performance in the many of the previous studies was only tested twice (pre-test and post-test). In the present study, however, the subjects were repeatedly tested (4 times) with the same stimulus materials, although the order was rearranged each time. Thus it is reasonable to find that the control group significantly benefited from the repeated testing experience.

Nonetheless, the *Post Hoc Tests* results indicate that in the perception of both /θ/-/s/, the subjects did not benefit significantly from the repeated testing experience until reaching the post-test. While in the perception of /ð/-/z/, it was from mid-test 2 and the post-test that the subjects significantly benefitted from the repeated testing experience. Thus, it might be possible to speculate that if the subjects were only tested twice with the same stimuli, no/little beneficial effect from repeated testing experience would be found. However, in Bradlow et al. (1999), subjects of the control group who did not undergo phonetic training displayed no significant changes in perception identification accuracy from pre-test to post-test, as well as a 3-month follow-up test, though the same testing materials were used in the three tests. This might be explained by the fact that, in Bradlow et al. (1999), the interval between the post-test and the follow-up test was much longer than that between each test in the present study.

2. *Gender*

Gender and its interaction with *training* were found to have had a significant effect on the experimental group's perception and production performance. In the control group, however, neither *gender* nor its interaction with *experience* was revealed to be significant for the subjects' perception performance. As discussed above, in the experimental group, the male subjects' better performance may be attributed to their greater visual-spatial ability (Bouchard and McGee, 1977; Harris, 1978; Sanders et al., 1982; Goldstein et al., 1990), which may have facilitated their perception of the target contrasts through the audiovisual training. In the control group, the female subjects' perception performance was comparable with that of the male subjects. This result is consistent with that obtained by Piske, MacKay, and Flege (2001), in which the gender difference was non-significant for the subjects' perception of L2 speech sounds. Given the fact that the control group did not receive audiovisual training, the male subjects were not likely to have benefited from their comparatively greater visual-spatial ability. Accordingly, their perception performance was not likely to be significantly better than the female subjects in the mid-tests and the post-test.

3. *Phonetic environments*

As in the experimental group, it was revealed that the control group's perception performance was significantly affected by the *phonetic position* of the target contrasts, whereas it was not significantly affected by *vowel contexts*.

Regarding *phonetic position*, the control group performed better when the target contrasts were embedded in initial and medial positions than in final position. This finding is identical to that for the experimental group. Considering that the control group shared the same L1 and L1-dialect as the experimental group, their perception difference as a function of phonetic position may also be explained by the influence of their L1 and/or L1-dialect. That is, /θ, ð/ do not exist in their L1/L1-dialect. Their "replacement" /s, z/ do not occur in the word-final position in their L1 and L1-dialect. Similar findings are available from Lively et al. (1993), Lively et al. (1991), and Logan et al. (1991).

4. *Vowel context*

In respect of *vowel context*, it was expected that the vowels /i/ and /u/ generally contribute more difficult contexts for perception accuracy than lower unrounded vowels, such as /ɑ/ (Hagiwara, 1995). Nonetheless, it was found that the control group's perception performance was not significantly affected by the *vowel contexts*. The experimental group showed the same result. In the discussion of the experimental group's perception performance, this finding was explained by the fact that they were trained with stimuli that included different vowel and consonant environments. For the control group, however, it was unclear what reason led to this result, since they did not receive audiovisual training.

6.2.10 Individual variances in the perception/production performance

One of the interesting findings is the wide range of individual differences in the perception and/or production performance. In the experimental group, as presented in the sections containing test results, the subjects achieved different accuracies and degrees of improvement both/either in the perception and/or production of the target contrasts. This is consistent with some previous findings in L2/non-native speech perception/production studies (e.g., Yamada et al., 1994; MacKain et al., 1981; Gordon et al., 2001). Given that independent variables, such as the subjects' age, AO and motivation were similar to each other, it was unclear which specific factor(s) determined individual performance. Nevertheless, the variance of individual capability or skills in lip-reading (Demorest, Bernstein, and DeHaven, 1996), sensitivity to visual cues (Sennema et al., 2003), and in the integration of auditory and visual information in speech perception/production (Grant and Seitz, 1998) may, in part, contribute to the explanation. Moreover, in addition to *gender differences*, which were found to have a significant effect on the experimental group's performance, the investigation of S14 and S16's method(s) of learning English in their spare time may provide us with further explanation. It may also have something to do with individual differences in learning strategies, intelligibilities, or cognitive abilities in language learning, and so on (Ellis, 1985; Skehan, 1998; Munro and Derwing, 1995).

Individual variances are also evident in the control group. Some of the subjects seem to have benefited much more than others from the repeated testing experience across the 4 tests. For instance, in the perception of /θ/-/s/, S36's accuracy increased from 54.63% in pre-test to 64.81% in post-test. In comparison, S30's accuracy only increased about 2% from pre-test to post-test. Moreover, some of the subjects' accuracy increased linearly from pre-test to post test in the perception of the target contrasts, whereas that of others did not. S42's accuracy in the perception of /θ/-/s/, for example, increased from 50.93% in pre-test to 54.63% in mid-test 1, yet decreased to 52.78% in mid-test 2. S40 and S44's accuracy maintained the same level in mid-test 1 and mid-test 2. The reason for the individual variances in the control group's perception performance was unclear, though individual intelligibility, cognitive ability and/or other general factors may contribute to the explanation.

6.3 Brief summary of the study

This study endeavoured to explore whether audiovisual training on speech perception can lead to adult language learners' improvement in auditory perception of the L2 speech sounds which they initially have difficulty with, and whether the training can benefit their production of the L2 speech sounds as a transferred beneficial effect. The motivation was to provide further evidence in support of the significance of articulatory information in speech perception. In addition, the present study may shed some light on the controversial issue of whether speech perception training can improve language learners' capability in speech production.

To accomplish the aims of the present study, a pilot study was carried out for the purpose of selecting suitable subjects and target speech sounds (English consonants only) for the following main study. 42 university level students were recruited from Chongqing, China. They were L1-Mandrain speakers of L2-English. Their L1-dialect was CQd. Their production of all the English consonants was tested first. Given that speech perception and production could be closely connected (Williams and McReynolds, 1975; Jamieson and Rvachew, 1992; Watkins, Strafella and Paus, 2003), or innately linked to each other (Liberman et al., 1967; Liberman, 1985), the subjects' incorrectly produced consonants were tested in perception tests. According to the results, 29 subjects who were found to have had serious difficulty in the perception and production of /θ/-/s/, /ð/-/z/ were selected to participate in the main study as the experimental group. Another 20 subjects, who had similar profiles as the 29 subjects, were recruited to be the control group.

In the main study, the experimental group received an audiovisual perception training programme. Their perception and production performance was repeatedly tested (4 times: before, during and after the training programme). Considering that the stimuli used in the 4 perception tests were the same, though with different orders, the control group's perception of the target contrasts was also tested 4 times with the same testing intervals. The purpose was to detect whether there was a repeated testing effect in the perception tests.

The key findings of the main study were: (1) the experimental group's accuracy in the perception and production of the target contrasts increased linearly and significantly from pre-test to post-test as a function of training effect. (2) The male subjects performed significantly better than the female subjects both in the perception and production of the target contrasts in the experimental group. Yet, the factor *gender difference* was found to be non-significant for the control group's perception performance. (3) Both the experimental group and the control group showed better perception performance when the target contrasts were embedded in initial and medial positions than in final position, while vowel context was revealed to be non-significant for their perception performance. (4) Repeated testing experience was found to be statistically significantly for the control group's perception performance. Accordingly, the experimental group may have benefited from the repeated testing experience in the tests after training. Nonetheless, the experimental group's perception accuracy was significantly higher than that of the control group. Therefore, audiovisual training appears to have played a more significant role in facilitating the experimental group's perception performance than repeated testing experience.

Overall, the results from the main study demonstrate that the audiovisual perception training facilitated the subjects' capability in auditory perception and production of English /θ/-/s/ and /ð/-/z/.

6.4 Implications for L2 learning

Findings from the present study have implications for L2 learning, specifically concerning the perception and production of L2 speech sounds. Firstly, given that audiovisual perception training was revealed to be effective for the subjects' auditory perception and production of the target L2-English sounds, it would be useful to adopt audiovisual techniques in the teaching of L2 speech sounds. For instance, the teachers can demonstrate the articulatory gestures of L2 speech sounds by (1) producing them exaggeratedly to facilitate the learners' observation of the articulatory gestures; (2) for some speech sounds, the articulatory gestures of which are not visible due to the fact that they are produced at the back of the vocal tract, the movements of articulators could be demonstrated with the help of pictures, videos or other available techniques.

Secondly, during the training programme, the experimental group was provided with native English speakers' input. Their perception and production of the target contrasts was both improved as a result of the training programme. Therefore, it is predicted that providing language learners with native speakers' input would contribute to their acquisition of L2 speech sounds. There are many ways available for doing so, such as watching TV programmes or movies with the L2 as the target language, or having classes given by native L2 speakers.

Moreover, the view that L2 learners whose L1 is a tone language may be less likely to employ visual information in L2 perception and production was supported (Gelder and Vroomen, 1992). In particular, this view is illustrated by findings in the pilot study. Although the subjects had been learning English for 6-8 years, they did not manage to perceive or produce the visible "interdental" cue, which is non-native. It could be helpful if L2 learners' sensitivity to visual cues, particularly non-native cues, could be enhanced at the beginning of L2 learning. Specifically, L2 teachers can direct the learners' attention to the articulatory gestures of the L2 speech sounds at the beginning of L2 sounds teaching.

In addition, to a large extent, the training programme followed the principles of HVPT – "Natural variability" (Logan et al., 1993). Specifically, the experimental group was provided with naturally produced stimuli from multiple speakers. As such, the subjects were exposed to a large number of "minimal pairs" of the target contrasts. The experimental group's improved accuracy during and at the end of the training programme indicates that this approach is useful in teaching L2 speech sounds. Therefore, it could be employed in L2 teaching. It would be beneficial if L2 learners are exposed to different native L2 speakers' input with a wide range of input content.

6.5 Critique of the study

The present study is different from previous audiovisual training studies in two respects. Firstly, compared with other relevant studies (e.g., Lively et al., 1993), the training materials include a larger number of stimulus words, which contained a wide range of phonetic environments concerning vowel contexts and phonetic positions. Secondly, instead of testing their performance before and after the training programme, the accuracy of subjects in the experimental group in the perception and production of the target contrasts was tested before, during and at the end of the training programme, so that the degree of their improvement during the training programme was revealed.

Moreover, qualitative data was collected for the purpose of further examining the findings in the main study. For instance, although the majority of the subjects reported that they spent some time on L2-English learning in their spare time, further investigation of the amount of time, and the ways in which they had been learning English provides us with valuable insight on this issue.

In addition, the validity and reliability of the present study was enhanced by the careful preparation of the training and testing stimuli, recruitment of the subjects, testing procedure, assessment process, and the choice of method in the analysis of collected data.

However, the study also bears some limitations. First of all, it lacks a generalization test as many previous studies have done (e.g., Hardison, 2003; Hazan et al., 2005), which served to detect whether the experimental group could generalize to the perception of the target contrasts in new words. In order to minimize this limitation, the stimuli employed in the perception tests were nonsense words which did not occur in the training materials. Nevertheless, due to the subjects being repeatedly tested (4 times) with the same testing materials (though with different orders), their perception improvement may, in part, be attributed to the repeated testing experience. This is another limitation of the study. In fact, it was confirmed by the improvement of the control group's perception performance across the 4 tests. Nevertheless, the experimental group's perception improvement was significantly greater than that of the control group. Thus the significant effect of audiovisual training on the experimental group's perception and production improvement is clear.

Secondly, the training programme lasted for 9 sessions, with about 35 minutes per session, which was a medium duration compared with previous studies (e.g., Hazan et al., 2005; Iverson and Evans, 2009). At the end of the training programme, the subjects' performance was not good enough. Further improvement might be observed if they were given more training sessions; in particular, room for improvement remained in their accuracy in the production of /θ/ and /ð/.

Thirdly, due to the fact that the whole study was carried out in China while the author was studying in the UK, it was not convenient to carry out a long-term retention test. Therefore it was unclear whether, or how long the audiovisual training effect would last. However, evidence from previous studies indicates that the effect of audiovisual training can last for a long time, such as that in Lively et al. (1994).

Moreover, the training materials were produced by 3 RP speakers. According to HVPT, exposing L2 learners to various speakers' input of the L2 benefits their perception/production of target speech sounds. Although the training programme was successful, the subjects may have achieved a higher degree of improvement if they were exposed to more RP speakers' production of the stimuli.

Another limitation of the present study was that, in the pre-test, the control group and the experimental group's accuracy in the perception of the target contrasts was not comparable. As shown in Table 5.9 and Table 5.18, in the pre-test, the subjects of the control group were better able to perceive the target contrasts than the experimental group. The control group's test results were adopted because the purpose of including the control group was to detect whether there was repeated testing effect. Different degrees of repeated testing effect might be observed if a control group of comparable perception accuracy was selected.

In addition, in order to facilitate the subjects' observation of the articulatory differences of the target contrasts, the RP speakers were asked to exaggerate their production by producing /θ/ and /ð/ as interdental, and /s/ and /z/ as alveolar. Doing this aimed to help the subjects' differentiation of the target contrasts with visible articulatory gestures. Nonetheless, it reduced the *variability* sought by the HVPT approach.

Furthermore, the present study lacks comparative conditions regarding training modalities, such as a comparison between audiovisual training and auditory-only training (e.g., Hazan et al., 2005). The critical role of articulatory gestures in speech perception and production may be better demonstrated if comparative training conditions were employed.

6.6 Suggestions for further research

The present study provided supporting evidence for the critical role of articulatory gestures in language learners' perception of L2 speech sounds, and the transferred beneficial effect of speech perception training on the production of L2 speech sounds. However, there are still some domains in the research of L2 speech perception and production which remain to be studied in the future.

First of all, although speech perception and production are typically viewed as closely connected to each other (Williams and McReynolds, 1975; Jamieson and Rvachew, 1992; Watkins, Strafella and Paus, 2003), it is still an open debate concerning whether training of one benefits the other. The present study provides supporting evidence for the view that perception training benefits speech production. The reverse scenario was not explored in this study. Future research, therefore, can examine whether speech production training can help language learners' perception of L2 speech sounds.

Another prospective area for research could be the exploration of novel methods in audiovisual training. In recent years, most audiovisual training of speech perception and production are with the help of software, such as the CSLU toolkit used in Hazan et al. (2005), or the more recently used software such as TP. These techniques could be useful in speech perception training. Although the software can provide the subjects with immediate feedback regarding the correctness of their responses, they may be compromised concerning the lack of interaction with the subjects during the training process. Therefore, communicative approaches for speech perception and/or production training are another potential area for future research, which may lead to even better training results.

Moreover, NLM/NLM-e, PAM-L2, and PI all predict that given sufficient L2 input, even adult L2 learners can eventually acquire L2 speech sounds. It would be of interest to investigate whether language learners who commence their L2 learning in adulthood, thus far beyond the "critical period", can also be trained to perceive and produce the L2 speech sounds which they initially have difficulty with. Suppose they can be successfully trained in the perception and production of L2 speech sounds, future research may compare the amount of time that they need in the learning of these sounds with those who started L2 learning in early childhood.

In addition, according to PI, even though language learners can manage to perceive L2 speech sounds. The critical acoustic cues they employ in the perception of the sounds, however, may vary from those used by native L2 speakers (Iverson et al., 2003). Therefore, future studies can also investigate whether the cues that L2 learners employ in the perception of L2 speech sounds are the same as those used by native L2 speakers.

6.7 Conclusion

This chapter discussed the findings of the main study with the support of relevant theories/models and previous studies reviewed in chapter 2. Then, it briefly summarized the pilot study and main study, and listed the main findings of the present study. The implications and limitations of the study were analysed. Moreover, possible topics for future research concerning audiovisual training on L2 speech perception and production were suggested.

Appendix 1

QuickPlacementTest

Part 1

Question 1 – 5

v Where can you see these notices?

v For questions **1** to **5**, mark one letter **A**,**B** or **C** on your **Answer Sheet**.

1. YOU CAN LOOK, BUT DON'T TOUCH THE			**A**	**B**	**C**
A▶ in an office	**B▶** in a cinema	**C▶** in a museum			
2. PLEASE GIVE THE RIGHT MONEY TO THE			**A**	**B**	**C**
A▶ in a bank	**B▶** on a bus	**C▶** in a cinema			
3. NO PARKING PLEASE			**A**	**B**	**C**
A▶ in a street	**B▶** on a book	**C▶** on a table			
4. CROSS BRIDGE FOR TRAINS TO EDINBURGH			**A**	**B**	**C**
A▶ in a bank	**B▶** in a garage	**C▶** in a station			
5. KEEP IN A COLD PLACE			**A**	**B**	**C**
A▶ on clothes	**B▶** on furniture	**C▶** on food			

Question 6 –10

v In this section you must choose the word which best fits each space in the text below.

v For questions **6** to **10**, mark **one** letter **A**, **B**, or **C** on your Answer Sheet

THE STARS

There are millions of stars in the sky. If you look **(6)...............**the sky on a clear night, it is possible to se about 3000 stars. They look small, but they are really **(7).............**big hot balls of burning gas. Some of them are huge, but others are much smaller, like our planet Earth. The biggest stars are very bright, but they only live for a short time. Every day new stars **(8)..........**born and old stars die. All the stars are very far away. The light from the nearest star takes more **(9)..........**four years to reach Earth. Hundreds of years ago, people **(10)............**stars, like the North Star, to know which direction to travel in. Today you can still see that star.

6.			**A**	**B**	**C**
A▶ at	**B▶** up	**C▶** on			
7.			**A**	**B**	**C**
A▶ very	**B▶** too	**C▶** much			
8.			**A**	**B**	**C**
A▶ is	**B▶** be	**C▶** are			
9.			**A**	**B**	**C**
A▶ that	**B▶** of	**C▶** than			
10.			**A**	**B**	**C**
A▶ use	**B▶** used	**C▶** using			

Question 11 - 15

v In this section you must choose the word which best fits each space in the texts.

v For questions **11** to **20**, mark one letter **A**, **B**, **C** or **D** on your Answer Sheet.

Good smilies ahead for young teeth

Older Britons are the worst in Europe when it comes to keeping their teeth. But

British youngsters **(11)............**more to smile about because **(12).............**teeth are among the best. Almost 80% of Britons over 65 have lost all ore some **(13).............**their teeth according to a World Health Organisation survey. Eating too **(14)............**sugar is part of the problem. Among **(15)............**, 12-year-olds have on average only three missing, decayed or filled teeth.

11.				**A**	**B**	**C**	**D**
A▶ getting	**B▶** got	**C▶** have	D▶ having				
12.				**A**	**B**	**C**	**D**
A▶ their	**B▶** his	**C▶** them	**D▶** theirs				
13.				**A**	**B**	**C**	**D**
A▶ from	**B▶** of	**C▶** among	**D▶**between				
14.				**A**	**B**	**C**	**D**
A▶ much	**B▶** lot	**C▶** many	**D▶**deal				
15.				**A**	**B**	**C**	**D**
A▶ person	**B▶** people	**C▶** children	**D▶**family				

Question 16 - 20

Christopher Columbus and the New World

On August 3, 1492, Christopher Columbus set sail from Spain to find a new route to

India, China and Japan. At this time most people thought you would fall off the edge of the world if you sailed too far. Yet sailors such as Columbus had seen how a ship appeared to get lower and lower on the horizon as it sailed away. For Columbus this **(16)**..........that the world was round. He **(17)**..........to his men about the distance travelled each day. He did not want them to think that he did not **(18)**...........exactly where they were going. **(19)**............., on October 12, 1492, Columbus and his men landed on a small island he named San Salvador.

Columbus believed he was in Asia, **(20)**............he was actually in the Caribbean.

16.				**A**	**B**	**C**	**D**
A▶ made	**B▶** pointed	**C▶** was	**D▶** proved				
17.				**A**	**B**	**C**	**D**
A▶ lied	**B▶** told	**C▶** cheated	**D▶** asked				
18.				**A**	**B**	**C**	**D**
A▶ find	**B▶** know	**C▶** think	**D▶**expect				
19.				**A**	**B**	**C**	**D**
A▶ Next	**B▶** Secoundly	**C▶** Finally	**D▶**Once				
20.				**A**	**B**	**C**	**D**
A▶ as	**B▶** but	**C▶** because	**D▶**if				

Question 21 - 30

- In this section you must choose the word or phrase which best completes each sentence.
- For questions **21** to **40**, mark one letter **A**, **B**, **C** or **D** on your Answer Sheet.

21. The children won ´t go to sleep.......we leave a light on outside their bedroom.				A	B	C	D
A▶ except	**B**▶ otherwise	**C**▶ unless	**D**▶ but				
22. I´ll give you my spare keys in case you.........home before me.				A	B	C	D
A▶ would get	**B**▶ got	**C**▶ will get	**D**▶ get				
23. My holiday in Paris gave me a great..........to improve my French accent				A	B	C	D
A▶ occasion	**B**▶ chance	**C**▶ hope	**D**▶ possibility				
24. The singer ended the concert...........her most popular song				A	B	C	D
A▶ by	**B**▶ with	**C**▶ in	**D**▶ as				
25. Because it had not rained for several months, there was a............of water.				A	B	C	D
A▶ shortage	**B**▶ drop	**C**▶ scare	**D**▶ waste				
26. I ´ve always.............you as my best friend.				A	B	C	D
A▶ regarded	**B**▶ thought	**C**▶ meant	**D**▶ supposed				
27. She came to live her............a month ago.				A	B	C	D
A▶ quite	**B**▶ beyond	**C**▶ already	**D**▶ almost				
28. Don´t make such a..........! The dentist is only going to look at your teeth.				A	B	C	D
A▶ fuss	**B**▶ trouble	**C**▶ worry	**D**▶ reaction				
29. He spent a long time looking for a tie which..........with his				A	B	C	D
A▶ fixed	**B**▶ made	**C**▶ went	**D**▶ wore				
30. Fortunately,.........from a bump on the head, she suffered no serious injuries from her fall				A	B	C	D
A▶ other	**B**▶ except	**C**▶ besides	**D**▶ apart				

Question 31 – 40

31. She had changed so much that.........anyone recognised her.				**A**	**B**	**C**	**D**
A▶ almost	**B▶** hardly	**C▶** not	**D▶** nearly				
32. ....teaching English, she also writes children´s books.				**A**	**B**	**C**	**D**
A▶ Moreover	**B▶** As well as	**C▶** In addition	**D▶** Apart				
33. It was clear that the young couple were.........of taking charge of the restaurant.				**A**	**B**	**C**	**D**
A▶ responsible	**B▶** reliable	**C▶** capable	**D▶**able				
34. The book.........of ten chapters, each one covering a different topic				**A**	**B**	**C**	**D**
A▶ comprises	**B▶** includes	**C▶** consists	**D▶**contains				
35. Mary was disappointed with her new shirt as the colour...........very quickly.				**A**	**B**	**C**	**D**
A▶ bleached	**B▶** died	**C▶** vanished	**D▶**faded				
36. National leaders from all over the world are expected o attend the......meeting.				**A**	**B**	**C**	**D**
A▶ peak	**B▶** summit	**C▶** top	**D▶** apex				
37. Jane remained calm when she won the lottery and......about her business as if nothing had				**A**	**B**	**C**	**D**
A▶ came	**B▶** brought	**C▶** went	**D▶**moved				
38. I suggest we.........outside the stadium tomorrow at 8.30.				**A**	**B**	**C**	**D**
A▶ meeting	**B▶** meet	**C▶** met	**D▶**will meet				
39. My remarks were..........as a joke, but she was offended by them				**A**	**B**	**C**	**D**
A▶ pretended	**B▶** thought	**C▶** meant	**D▶**supposed				
40. You ought to take up swimming for the..........of your health				**A**	**B**	**C**	**D**
A▶ concern	**B▶** relief	**C▶** sake	**D▶**cause				

Part 2

Do not start this part unless told to do so by your test supervisor

Questions 41 – 45

v In this section you must choose the word which best fits each
. space in the texts.
v For questions **41** to **45**, mark one letter **A**, **B**, **C** or **D** on your Answer Sheet.

CLOCKS

The clock was the first complex mechanical machinery to enter the home, **(41)**………..it was too expensive for the **(42)**………person until the 19th century, when **(43)**………production techniques lowered the price. Watches were also developed, but they **(44)**………luxury items until 1868, When the first cheap pocket watch was designed in Switzerland. Watches later became **(45)**………available, and Switzerland became the world´ s leading watch manufacturing centre for the next 100 years.

41.				**A**	**B**	**C**	**D**
A▶ despite	**B▶** although	**C▶** otherwise	D▶ average				
42.				**A**	**B**	**C**	**D**
A▶ average	**B▶** medium	**C▶** general	**D▶** common				
43.				**A**	**B**	**C**	**D**
A▶ vast	**B▶** large	**C▶** wide	**D▶** mass				
44.				**A**	**B**	**C**	**D**
A▶ lasted	**B▶** endured	**C▶** kept	**D▶** remained				
45.				**A**	**B**	**C**	**D**
A▶ mostly	**B▶** chiefly	**C▶** greatly	**D▶** widely				

Questions 46 - 50

f

Dublin City Walks

What better way of getting to know a new city than by walking around it? Whether you choose the Medieval Walk, which will **(46)**………you to the

1000 years ago, find out about the more **(47)**………history of the city on the Eighteenth

Century Walk, or meet the ghosts of Dublin´s many writers on

The Literary Walk, we know you will enjoy the experience.

46.				**A**	**B**	**C**	**D**
A▶ introduce	**B▶** present	**C▶** move	D▶ show				
47.				**A**	**B**	**C**	**D**
A▶ near	**B▶** late	**C▶** recent	**D▶** close				
48.				**A**	**B**	**C**	**D**
A▶ take place	**B▶** occur	**C▶** work	**D▶** function				
49.				**A**	**B**	**C**	**D**
A▶ paying	**B▶** reserving	**C▶** warning	**D▶** booking				
50.				**A**	**B**	**C**	**D**
A▶ funds	**B▶** costs	**C▶** fees	**D▶** rates				

Question 51– 60

v In this section you must choose the word or phrase which best completes each sentence.

v For questions **51** to **60**, mark one letter **A**, **B**, **C** or **D** on your Answer Sheet.

51. If you´re not too tired we could have a……..of tennis after lunch.				**A**	**B**	**C**	**D**
A▶ match	**B▶** play	**C▶** game	**D▶** party				
52. Don´t you get tired………watching TV every nigh?				**A**	**B**	**C**	**D**
A▶ with	**B▶** by	**C▶** of	**D▶** at				
53. Go on, finish the dessert. It needs………up because it won´t stay fresh until.				**A**	**B**	**C**	**D**
A▶ eat	**B▶** eating	**C▶** to eat	**D▶** eaten				
54. We´re not used to……….invited to very formal occasions				**A**	**B**	**C**	**D**
A▶ be	**B▶** have	**C▶** being	**D▶** having				
55. I´d rather we……….meet this evening, because I´m very tired				**A**	**B**	**C**	**D**
A▶ wouldn´t	**B▶** shouldn´t	**C▶** hadn´t	**D▶**didn´t				
56. She obviously didn´t want to discuss the matter so I didn´t……..the point.				**A**	**B**	**C**	**D**
A▶ maintain	**B▶** chase	**C▶** follow	**D▶** pursue				
57. Anyone………after the start of the play is not allowed in until the interval.				**A**	**B**	**C**	**D**
A▶ arrives	**B▶** has arrived	**C▶** arriving	**D▶** arrived				
58. This new magazine is ………..with interesting stories and useful information.				**A**	**B**	**C**	**D**
A▶ full	**B▶** packed	**C▶** thick	**D▶** compiled				
59. The restaurant was far too noisy to be………to relaxed conversation				**A**	**B**	**C**	**D**
A▶ conducive	**B▶** suitable	**C▶** practical	**D▶** fruitful				
60. In this branch of medicine, it is vital to ……….open to new ideas				**A**	**B**	**C**	**D**
A▶ stand	**B▶** continue	**C▶** hold	**D▶** remain				

Appendix 2

Subjects of the experimental group's information (All the subjects' English proficiency level was intermediate).

Subjects in pilot study	Rearranged number of selected subjects in the main study	age	gender	English learning duration (in year)
S1	S1	19	Male	6
S2		20	Female	6.5
S3	S2	20	Male	6.5
S4		22	Female	7
S5	S3	21	Male	7
S6		23	Female	7.5
S7	S4	19	Male	6
S8		20	Female	6.5
S9	S5	19	Male	6
S10	S6	19	Male	6
S11		21	Female	7
S12	S7	20	Male	6.5
S13	S8	19	Female	6
S14	S9	23	Male	7.5
S15	S10	20	Female	6.5
S16		22	Male	7
S17		21	Male	7
S18	S11	20	Female	6.5
S19		19	Male	6
S20	S12	20	Female	6.5
S21	S13	22	Female	7.5

S22		20	Male	6.5
S23	S14	19	Female	6
S24	S15	19	Female	6
S25	S16	21	Female	7
S26	S17	20	Female	6.5
S27	S18	19	Female	6
S28	S19	20	Female	6.5
S29		23	Male	7
S30		20	Male	6.5
S31	S20	21	Female	7
S32		22	Male	7
S33	S21	20	Female	6
S34	S22	19	Male	6
S35		22	Female	7
S36	S23	22	Male	7
S37	S24	20	Male	6.5
S38	S25	20	Male	6
S39	S26	20	Male	6
S40	S27	21	Female	6.5
S41	S28	22	Female	7
S42	S29	20	Female	6
average		20.50		6.55

subject	AO (onset age of learning English as a L2)	gender	age	Length of English study (in year)
1	13	male	19	6
2	13	male	20	7
3	14	male	21	7
4	13	male	19	6
5	13	male	19	6
6	13	male	19	6
7	13	male	20	7
8	13	female	19	6
9	14	male	21	7
10	13	female	20	7
11	13	female	20	7
12	13	female	20	7
13	14	female	22	8
14	13	female	19	6
15	13	female	19	6
16	14	female	21	7
17	13	female	20	7
18	13	female	19	6
19	13	female	20	7
20	14	female	21	7
21	14	female	20	6
22	13	male	19	6
23	14	male	21	7
24	13	male	20	7

25	14	male	20	6
26	14	male	20	6
27	14	female	21	7
28	14	male	22	8
29	14	female	20	6

subject	Primary motivation of English learning	Learn English in other institute?	Lear English in spare time?	In which ways/ about how many hours per day?	Have you ever traveled or lived in English speaking countries?
1	score[8]	no	no		no
2	score	no	yes	about 1 hours per day/ watch English movies; listen to English songs;	no
3	hobby	no	yes	about 1-2 hours per day/ read Englis news paper; watch English movies; listen to English songs and BBC news.	no
4	score	no	yes	half an hour/ read English text book in the morning	no
5	score	no	no		no
6	score	no	yes	less than 1 hour per day in week days/ read articles on English text book	no

[8] 'score' means the subject's primary motivation of English learning was to get high scores in English exams.

7	score	no	yes	about 1 hour per day, 2-3days a week/ pre-read articles in English text book and do exercises in the text book	no
8	score	no	yes	1-2 hours per day in week days/ do English exercises in English text book; read articles on English text book	no
9	score	no	no		no
10	score	no	yes	1-2 hours per day, 3-4 days per week/ do exercises in English text book; watch English movies	no
11	score	no	yes	about half an hour per day, 5 days a week/ read articles in English text book	no
12	score	no	yes	about 1 hour per day in week days/ do exercises in English text book; read artiles in the text book	no
13	score	no	yes	less than 1 hour per day/ read English text book; do exerices in text book; listen to English songs	no
14	score	no	yes	about 3 hours on every Saturday and Sunday/ Watch English movies	no
15	score	no	yes	1 hour per day in week days/ do exercieses in English text book and relevant books for the preparation of English exams	no
16	score	no	yes	1-1.5 hours per day in week days/ do exercises for the preperation of English exams	no
17	score	no	yes	about half an hour per day/ read articles on text book in the morning	no
18	score	no	yes	2 hours perday/ do exercises and read articles on text book; listen	no

				to English songs	
19	score	no	yes	less than 1 hour per day, 5 days per week/ do exercises on text book and relevant exercises for the preperation of English exams	no
20	score	no	yes	about 3 hours in weekend/ watch English movies; read articles on English text book	no
21	score	no	yes	2 hours per day, 4 days per week/ do exercises on text book; recite English vocabulary on text book	no
22	score	no	yes	1 hour per day in week days/ read articles on text book; recite English vocabularies on text book	no
23	score	no	yes	about 1-2 hours, 3-4 days per week/ read articles on text book; do exercises on text book;	no
24	score	no	yes	about 2 hours per day, 4-5 days per week/ do English exercises	no
25	score	no	yes	1 hour per day,5-6 days per week during term time/ read articles on text book; do exercises on text book	no
26	score	no	yes	about 1 hour per day/ do exercises on English text book	no
27	score	no	yes	less than 1 hour per day, about 4 days per week/recite English vocabularies on text book	no
28	score	no	no		no
29	score	no	yes	1-2 hours per day in week days/ recite English vocabulary on English text book; do English exercises	no

Appendix 3

Stimuli for production test in pilot study (revised from *Comma Gets a Cure*, McCullough, Somerville, & Honorof, 2000)

Sarah once dreamed to be a lawyer, and dwell in UK. Yet, she became a nurse who had been working at a zoo in Asia, so she was very happy to start a new job at a private practice ahead north square near the Tower. That area was much nearer for her and more to her liking. She took a shower. Then she put on a plain beige dress, picked up her kit and headed off for work.

There was a woman, Mary, with a goose waiting for her. It could be suffering from a form of mouth disease, which normally happens to a dog.

That goose began to “scream” like a child. Mary called twice, "Comma, Comma," which Sarah thought was strange. Comma was huge, so they didn’t wish to trap her easily. Sarah tried gently stroking the goose's lower back, then singing to her, which worked. Then Sarah managed well to bathe the goose, and gave it back go Mary.

Stimulus words contained in the reading text: put, trap, been, job, tower, child, much, huge, kit, work, goose, dog, from, off, very, gave, thought, north, there, bathe, so, zoo, was, shower, wish, Asia, beige, had, ahead, singing, liking, like, well, right, Sarah, dwell, yet, lawyer

Appendix 4

Frequency of occurrence (the 38 stimulus words in the reading text of Appendix 3)

stimulus words	frequency of occurrence
put	57050
trap	1630
been	256779
job	21904
tower	3255
child	23486
much	89035
huge	7516
kit	1772
work	88643
goose	497
dog	7746
from	419502
off	66938
very	118490
gave	21708
thought	53213
north	21044
there	316871
bathe	142
so	236850
zoo	748
was	872575
shower	1502
wish	11330
Asia	2810
beige	226
had	415001
ahead	8446
singing	2676
liking	1432
like	145993
well	141308
right	89822
Sarah	3204
dwell	355
yet	33498
lawyer	2098

Assessed from http://corpus.byu.edu/bnc/ (on 15/07/2013)

Appendix 5

Nonsense words for perception test

/zi/ /ði/ /θi/ /si/

/za/ /ða/ /θa/ /sa/

/zu/ /ðu/ /θu/ /su/

/izi/ /iði/ /iθi/ /isi/

/aza/ /aða/ /aθa/ /asa/

/uzu/ /uðu/ /usu/ /uθu/

/iz/ /ið/ /iθ/ /is/

/az/ /að/ /aθ] /as/

/uz/ /uz/ /uθ/ /us/

/si/ /sa/ /su/

/isi/ /asa/ /uθu/

/is/ /as/ /us/

/θi/ /θa/ /θu/

/iθi/ /aθa/ /usu/

/iθ/ /aθ/ /uθ/

Appendix 6

Stimuli for production test of the main study

1. I think Cathy likes going to that zoo instead of this one to see animals, though it's a thousand miles away.
2. Don't sleep, or ask anything about them in class.
3. Although he was an athlete, he can dance with rhythm.
4. Hold your breath when he is cleaning your teeth with cloth.
5. He claimed throne by a cruel method.
6. Three theatres and a fourth museum will be built in a wealthy state soon.
7. Father told brother to bathe himself before putting on a new clothe with a zipper on it.
8. Neither of us took a bath before visiting her.
9. A Master student designed a Wreathe with a badge of scythe on it.
10. The zip code includes one zero.
11. A so called "user" doesn't exist.
12. It's easy to zoom in and out.

The table below shows the stimulus words, the pronunciation of which were selected to be judged to detect the subjects' production performance.

Target sound	Initial position	Medial position	Final position
/ð/	that, this, though, them, the	although, rhythm, father, brother, neither	with(occurred 4 times), bathe, clothe, scythe, wreathe
/z/	zoo, zero, zip, zipper, zoom	designed, easy, visiting, user, exist	miles, is, was, animals, theatres
/θ/	think, thousand, throne, three, theatres	Cathy, anything, athlete, method, wealthy	breath, teeth, cloth, fourth, bath
/s/	see, sleep, state, soon, student	instead, ask, himself, Master, exist	likes, this, class, dance, us

Appendix 7

Questionnaire

Name: Age:
Gender:

(姓名) (年龄） (性别)

1. How old were you when you began English learning? （你几岁开始学英语的？）

2. How many years have you been learning English? （你学习英语几年了？）

3. What is your primary motivation(s) of learning English? (You can choose more than one choices. If your answer is F, please give specific answer(s) in the following bracket. （你学习英语的动机是什么？你可以选择多项。如果你的答案是 F，请把具体的答案写到选项后的括号里面。）

A. hobby （兴趣）

B. the need of work （工作的需要）

C. the need of getting high scores in English exams （为了考试得高分而学）

D. cater to parents’ wish （为父母的期望而学）

E.travel to foreign countries （为了方便到其他国家旅游）

F. others ( ) （其它）

4. Apart from the study at school/university, do you study English in other institute? （你在什么机构里学习英语？）

A Yes (please give detailed infromation)

B.No.

5. Do you study English in your spare time? (If the answer is *A. Yes*, please specify the amount of time, and the ways you sued in English learning in your spare time.)

（在课余时间，你会用其它方式学习英语吗？如果你的答案是 A，即，有，请将你所用的英语学习方式写到选项后的括号里。）

A. Yes (such as:) （有。比如：）

B. No （没有）

6. Have you ever travelled or lived in English speaking countries? A. Yes 有 B. No 没有

8. Do you have any chance to use English on a daily basis? (日常生活中你有使用英语的机会吗?)

A. Yes (有) B. No (没有)

Appendix 8

Stimuli used in audiovisual training

Session 1, 4, 7

/θ/-/s/

1. A. [sɪk] sick	B.[θɪk] thick
2. A. [sɔːt] sought	B.[θɔːt] thought
3. A. [sɒŋ] song	B.[θɒŋ] thong
4. A. [ˈsɔːɹɪə] soria	B.[ˈθɔːɹɪə] thoria
5. A. [siːf] safe	B.[θiːf] thief
6. A. ['esɪk] esic	B.['eθɪk] ethic
7. A. ['liːs(ə)l] lisal	B.['liːθ(ə)l] lithal
8. A. ['tesə] teaser	B.['teθə] teather
9. A. ['ɔːsə] ausor	B.['ɔːθə] author
10. A. [hels] hels	B.[helθ] health
11. A. [fɪfs] fifs	B.[fɪfθ] fifth
12. A. [des] dess	B.[deθ] death
13. A. [əʊs] oas	B.[əʊθ] oath
14. A. [bɜːs] birs	B.[bɜːθ] birth
15. A [‘bɒsɪ] bossy	B[‘bɒsɪ] bothey

/ð/-/z/

1. A. [ðem] them	B.[zem] zem
2. A. [ðiːz] these	B.[ziːz] zese
3. A. [ðeə] there	B. [zeə] zere
4. A. [ðen] then	B.[zen] zen
5. A. [ðiː] thee	B. [ziː] zee
6. A. ['kɹeɪðɪ] crathey	B. ['kɹeɪzɪ] crazy
7. A. ['fɜːzə] furzer	B.['fɜːðə] further
8. A. ['leðə] leather	B.['lezə] leazer
9. A. [dʒæð] jathe	B.[dʒæz] jazz
10. A.['nɔːz(ə)n] norzern	B.['nɔːð(ə)n] northern
11. A. [ˈpɔɪz(ə)n] poisin	B.[ˈpɔɪð (ə)n] poithin
12. A. [saɪð] sithe	B.[saɪz] size
13. A. [luːð] loothe	B.[luːz] lose
14. A. [leɪð] lathe	B. [leɪz] laze
15. A. [kwɪð] quithe	B.[kwɪz] quize

Session 2, 5, 8

θ/-/s/

1. A. [sɹeɪs] srais	B.[θɹeɪs] thrais
2. A. [ˈsɪs(ə)l] sistle	B.[ˈθɪs(ə)l] thistle
3. A. [sɔːn] sorn	B.[θɔːn] thorn
4. A. [sɹɪft] srift	B.[θɹɪft] thrift
5. A. ['sɪmpəsi] sympasy	B.['sɪmpəθi] sympathy
6. A. [ɔː'sɒɹɪtɪ] ausority	B.[ɔː'θɒɹɪtɪ] authority
7. A. ['gɔsik] gosic	B.['gɔθik] gothic
8. A. [əu'seləu] oselo	B.[əu'θeləu] othelo
9. A. ['nesɪ] nesy	B.['neθɪ] nethy
10.A. ['i:sən] Esan	B.['i:θən] Ethan
11.A. [pas] pass	B.[paθ] path
12.A. [æs] ass	B.[æθ] ath
13.A. [mʌs] mars	B.[mʌθ] marth
14.A. [sus] soos	B.[suθ] sooth
15.A. [zɪs] zis	B.[zɪθ] zith

/ð/-/z/

1. A. [ðəʊz] those	B.[zəʊz] zose
2. A. [ðəʊ] though	B.[zəʊ] zough
3. A. [ðaɪ] thy	B.[zaɪ] zy
4. A. [ðæn] than	B.[zæn] zan
5. A. [ðʌs] thus	B.[zʌs] zus
6. A. ['gæzə] gazer	B.['gæðə] gather
7. A. ['fezə] feazer	B.['feðə] feather
8. A. [ˈmʌzə] mozer	B.[ˈmʌðə] mother
9. A. [ˈbɹʌzə] brozer	B.[ˈbɹʌzə] brozer
10. A. [təˈgɛzə] togezer	B.[təˈgɛðə] together
11. A. [bɹiːð] breathe	B.[bɹiːz] breaze
12. A. [bʌgð] bugthe	B. [bʌgz] bugs
13. A. [tʃiːð] cheethe	B.[tʃiːz] cheese
14. A. [kʌbð] cubthe	B.[kʌbz] cubz
15. A. [biː ð] beethe	B.[biː z] bees

Session 3, 6, 9

/θ/-/s/

1. A. [sɛsp] sesp	B.[θɛsp] thesp
2. A. [ˈsɪlə] siller	B.[ˈθɪlə] thiller
3. A. [sæŋk] sank	B.[θæŋk] thank
4. A. [sraɪs] srais	B.[θraɪs]thrais
5. A. [ˈsɪk(ə)n] sicken	B.[ˈθɪk(ə)n] thicken
6. A.[sʌs] sars	B. [sʌθ] sarth
7. A. ['dentɪs] dentis	B.['dentɪθ] dentith
8. A. [tes] tess	B.[teθ] teth
9. A. [dɹes] dress	B.[dɹeθ] dreth
10. A. [tæks] tax	B.[tækθ] tacth
11. A. ['besəl] bessal	B.['beθəl] bethal
12. A. [ɑːsk] ask	B.[ɑːθk] athk
13. A. ['ɹɪsk] risk	B.['ɹɪθk] rithk
14. A. ['fæsɪk] fathic	B.['fæθɪk] fasic
15. A. ['glɔsi] glossy	B.['glɔθi] glothy

/ð/-/z/

1. A. [legð] legthe	B.[legz] legs
2. A. [niːð] kneethe	B.[niːz] knees
3. A. [seð] sethe	B.[sez] says
4. A.[dʒiːnð] jeanthe	B.[dʒiːnz] jeans
5. A. [gɹəʊð] growthe	B.[gɹəʊz] grows
6. A. [ˈpænzi] panzy	B.[ˈpænði] panthy
7. A. [ˈʌzə] ozer	B.[ˈʌðə] other
8. A. [ˈʌðə] other	B. [ˈʌzə] ozer
9. A. ['smʌzə] smozer	B.['smʌðə] smother
10. A. ['rɑːðə] rather	B. ['rɑːzə] raser
11. A. [ˈðændə] thander	B.[ˈzændə] zander
12. A. [ðɪf] thiff	B.[zɪf] ziff
13. A. [ˈðəʊɪk] thoic	B.[ˈzəʊɪk] zoic
14. A. [ðeɪl] they'll	B.[zeɪl] zey'll
15. A. [ðeɪd] they'd	B.[zeɪd] zey'd

Appendix 9

Mandarin initials and finals

Mandarin initials

		Bilabial		Labiodental	Alveolar		Retroflex		Alveolo-palatal	Velar
		Voiceless	Voiced	Voiceless	Voiceless	Voiced	Voiceless	Voiced	Voiceless	Voiceless
Nasal			m [m]			**n** [n]				
Plosive	Unaspirated	**b** [p]			**d** [t]					**g** [k]
	Aspirated	**p** [pʰ]			**t** [tʰ]					**k** [kʰ]
Affricate	Unaspirated				**z** [ts]		**zh** [ʈʂ]		**j** [tɕ]	
	Aspirated				**c** [tsʰ]		**ch** [ʈʂʰ]		**q** [tɕʰ]	
Fricative				**f** [f]	**s** [s]		**sh** [ʂ]	**r** [ʐ~ɻ]1	**x** [ɕ]	**h** [x]
Lateral						**l** [l]				
Approximant		**y**[3] [j]/[ɥ][2] and **w**[3] [w]								

[1] /r/ may phonetically be [ʐ] (a voiced retroflex fricative) or [ɻ] (a retroflex approximant). This pronunciation varies among different speakers, and is not two different phonemes.

[2] /y/ is pronounced [ɥ] (a labial-palatal approximant) before /u/.

[3] the letters *w* and *y* are not included in the table of initials in the official pinyin system. They are an orthographic convention for the medials /i, u, ü/ when no initial is present. When /i, u/ or /ü/ are finals and no initial is present, they are spelled [yi], [wu], and [yu], respectively.

Initials of Mandarin (the bold letters indicate pinyin and the brackets enclose the symbol in the International Phonetic AlphabetAssessed from http://en.wikipedia.org/wiki/Mandarin_pinyin 22/07/2013) Check Norman, 1988

Mandarin initials and finals

Assessed from (http://en.wikibooks.org/wiki/Chinese_(Mandarin)/Table_of_Initial-Final_Combinations 22/07/2013)

Pinyin table		**Initials**																						**Pinyin table**	
		(no initial)	**b**	**p**	**m**	**f**	**d**	**t**	**n**	**l**	**g**	**k**	**h**	**j**	**q**	**x**	**zh**	**ch**	**sh**	**r**	**z**	**c**	**s**		
Group a Finals	**(no final)**																zhi	chi	shi	ri	zi	ci	si	**(no final)**	**Group a Finals**
	a	a	ba	pa	ma	fa	da	ta	na	la	ga	ka	ha				zha	cha	sha		za	ca	sa	**a**	
	o	o	bo	po	mo	fo																		**o**	
	e	e			me		de	te	ne	le	ge	ke	he				zhe	che	she	re	ze	ce	se	**e**	
	ê																							**ê**	
	ai	ai	bai	pai	mai		dai	tai	nai	lai	gai	kai	hai				zhai	chai	shai		zai	cai	sai	**ai**	

	ei	ei	bei	pei	mei	fei	dei		nei	lei	gei		hei				zhei		shei		zei			**ei**	
	ao	ao	bao	pao	mao		dao	tao	nao	lao	gao	kao	hao				zhao	chao	shao	rao	zao	cao	sao	**ao**	
	ou	ou		pou	mou	fou	dou	tou	nou	lou	gou	kou	hou				zhou	chou	shou	rou	zou	cou	sou	**ou**	
	an	an	ban	pan	man	fan	dan	tan	nan	lan	gan	kan	han				zhan	chan	shan	ran	zan	can	san	**an**	
	en	en	ben	pen	men	fen			nen		gen	ken	hen				zhen	chen	shen	ren	zen	cen	sen	**en**	
	ang	ang	bang	pang	mang	fang	dang	tang	nang	lang	gang	kang	hang				zhang	chang	shang	rang	zang	cang	sang	**ang**	
	eng	eng	beng	peng	meng	feng	deng	teng	neng	leng	geng	keng	heng				zheng	cheng	sheng	reng	zeng	ceng	seng	**eng**	
	er	er																						**er**	
	i	yi	bi	pi	mi		di	ti	ni	li				ji	qi	xi								**i**	
	ia	ya								lia				jia	qia	xia								**ia**	**G**

Group i Finals	**io**	yo																						**io**	roup i Finals
	ie	ye	bie	pie	mie		die	tie	nie	lie				jie	qie	xie								**ie**	
	iai	yai																						**iai**	
	iao	yao	biao	piao	miao		diao	tiao	niao	liao				jiao	qiao	xiao								**iao**	
	iu	you			miu		diu		niu	liu				jiu	qiu	xiu								**iu**	
	ian	yan	bian	pian	mian		dian	tian	nian	lian				jian	qian	xian								**ian**	
	in	yin	bin	pin	min				nin	lin				jin	qin	xin								**in**	
	iang	yang							niang	liang				jiang	qiang	xiang								**iang**	
	ing	ying	bing	ping	ming		ding	ting	ning	ling				jing	qing	xing								**ing**	
Group u	**u**	wu	bu	pu	mu	fu	du	tu	nu	lu	gu	ku	hu				zhu	chu	shu	ru	zu	cu	su	**u**	**Group u**
	ua	wa									gu	ku	hu				zhu	chu	shu					**ua**	

Finals											a	a	a				a	a	a					**Finals**
	uo	wo					duo	tuo	nuo	luo	guo	kuo	huo				zhuo	chuo	shuo	ruo	zuo	cuo	suo	**uo**
	uai	wai									guai	kuai	huai				zhuai	chuai	shuai					**uai**
	ui	wei					dui	tui			gui	kui	hui				zhui	chui	shui	rui	zui	cui	sui	**ui**
	uan	wan					duan	tuan	nuan	luan	guan	kuan	huan				zhuan	chuan	shuan	ruan	zuan	cuan	suan	**uan**
	un	wen					dun	tun		lun	gun	kun	hun				zhun	chun	shun	run	zun	cun	sun	**un**
	uang	wang									guang	kuang	huang				zhuang	chuang	shuang					**uang**
	ong	weng					dong	tong	nong	long	gong	kong	hong				zhong	chong		rong	zong	cong	song	**ong**
Grou	**ü**	yu							nü	lü				ju	qu	xu								**ü**	**Grou**
	ü	yu							nü	lü				ju	qu	xu								**ü**	

püFinals	e	e							e	e				e	e	e								e	püFinals
	üan	yuan								lüan				juan	quan	xuan								üan	
	ün	yun								lün				jun	qun	xun								ün	
	iong	yong												jiong	qiong	xiong								iong	
Pinyin table		**(no initial)**	**b**	**p**	**m**	**f**	**d**	**t**	**n**	**l**	**g**	**k**	**h**	**j**	**q**	**x**	**zh**	**ch**	**sh**	**r**	**z**	**c**	**s**	**Pinyin table**	
		Initials																							

Appendix 10

The experimental group's perception and production tests results (main study)

Perception test results

subject	pre-test		mid-test1		mid-test2		post-test	
	/θ/-/s/	/ð/-/z/	/θ/-/s/	/ð/-/z/	/θ/-/s/	/ð/-/z/	/θ/-/s/	/ð/-/z/
S1	35.18%	37.04%	63.89%	66.67%	84.26%	78.70%	95.37%	91.67%
S2	39.81%	38.89%	70.37%	63.89%	84.26%	77.78%	98.15%	92.59%
S3	64.81%	67.59%	88.89%	89.82%	100.00%	98.15%	dropped	dropped
S4	39.81%	33.33%	60.19%	51.85%	76.85%	65.74%	95.37%	80.56%
S5	42.59%	25.93%	56.67%	41.67%	79.63%	62.04%	98.15%	85.18%
S6	32.41%	45.37%	53.70%	65.74%	70.37%	78.70%	87.04%	92.59%
S7	32.41%	53.71%	57.41%	65.74%	76.85%	80.55%	98.15%	92.59%
S8	46.30%	48.15%	76.85%	62.96%	91.67%	76.85%	99.07%	89.81%
S9	34.26%	38.89%	48.15%	55.56%	71.30%	67.59%	87.96%	89.81%
S10	55.56%	68.52%	94.44%	92.59%	dropped	dropped	dropped	dropped
S11	40.74%	45.37%	63.89%	66.67%	80.56%	75.93%	94.44%	92.59%
S12	39.81%	48.15%	49.07%	70.37%	69.45%	87.04%	87.96%	92.59%
S13	48.15%	39.82%	61.11%	63.89%	77.78%	71.30%	89.81%	88.89%
S14	34.26%	37.04%	50.93%	50.00%	73.15%	75.00%	80.56%	87.04%
S15	32.41%	31.48%	44.44%	43.52%	61.11%	67.59%	87.96%	80.55%
S16	36.11%	32.41%	50.00%	56.48%	67.59%	73.15%	86.11%	85.18%
S17	33.33%	33.33%	54.63%	52.78%	67.59%	76.85%	88.89%	87.96%
S18	43.52%	45.37%	65.74%	60.19%	79.63%	73.15%	87.59%	89.81%
S19	40.74%	27.78%	59.26%	57.41%	80.56%	75.00%	97.22%	92.59%
S20	36.11%	39.82%	64.81%	55.56%	83.33%	67.59%	96.30%	90.74%

S21	54.63%	48.15%	82.41%	71.30%	93.52%	84.26%	98.15%	94.44%
S22	53.70%	35.19%	81.48%	51.85%	93.52%	75.00%	100.00%	81.48%
S23	33.33%	38.89%	53.70%	62.04%	77.78%	74.07%	93.52%	93.52%
S24	51.85%	51.85%	78.70%	76.85%	87.96%	87.04%	99.07%	94.44%
S25	37.04%	37.04%	59.26%	58.33%	78.70%	82.41%	90.74%	95.37%
S26	40.74%	48.15%	62.04%	68.52%	73.15%	78.70%	81.48%	94.44%
S27	37.96%	50.93%	61.11%	75.00%	76.85%	87.04%	87.04%	90.74%
S28	46.30%	28.70%	66.67%	50.00%	75.00%	71.30%	88.89%	82.41%
S29	44.44%	35.19%	63.89%	53.70%	73.15%	71.30%	92.59%	79.63%

Production test Results

Subject	Pre-test		Mid-test1		Mid-test2		Post-test	
	/θ/	/ð/	/θ/	/ð/	/θ/	/ð/	/θ/	/ð/
S1	29.17%	37.36%	50.00%	62.92%	64.50%	74.03%	92.67%	89.86%
S2	31.17%	38.61%	57.00%	63.89%	76.33%	70.14%	86.50%	81.11%
S3	45.33%	41.25%	77.66%	81.11%	93.33%	90.27%	dropped	dropped
S4	33.17%	39.58%	51.00%	47.50%	71.50%	71.25%	81.17%	78.19%
S5	41.67%	38.47%	64.50%	75.69%	79.50%	82.22%	90.83%	93.19%
S6	64.50%	36.11%	79.00%	52.92%	81.83%	65.28%	89.67%	82.36%
S7	38.67%	48.89%	54.17%	66.39%	72.00%	70.42%	89.83%	80.42%
S8	40.67%	40.14%	56.50%	49.44%	67.33%	63.61%	80.00%	75.83%
S9	23.83%	32.92%	48.17%	47.36%	62.83%	70.69%	83.50%	88.06%
S10	49.83%	42.22%	92.17%	90.28%	dropped	dropped	dropped	dropped
S11	53.50%	41.67%	74.67%	66.25%	72.17%	71.94%	81.00%	75.14%
S12	34.33%	41.39%	61.83%	47.78%	74.83%	63.19%	79.67%	65.69%
S13	37.00%	32.64%	55.33%	45.56%	63.55%	69.58%	77.83%	75.69%

S14	26.50%	42.08%	33.17%	51.25%	58.50%	62.85%	74.67%	73.33%
S15	67.67%	54.31%	83.67%	74.86%	87.00%	87.22%	95.83%	89.31%
S16	28.67%	42.78%	42.50%	49.86%	48.33%	53.33%	58.00%	66.38%
S17	31.00%	34.58%	48.33%	49.44%	57.67%	58.89%	79.50%	74.44%
S18	31.67%	33.06%	49.67%	43.19%	60.33%	50.56%	68.17%	68.06%
S19	37.00%	28.61%	44.50%	34.58%	70.33%	61.25%	77.17%	75.00%
S20	61.67%	47.78%	72.83%	53.06%	82.33%	67.78%	85.17%	74.17%
S21	33.00%	43.47%	54.56%	59.13%	67.00%	69.31%	77.17%	75.14%
S22	30.67%	35.69%	43.83%	51.25%	77.17%	60.31%	86.50%	76.53%
S23	41.67%	38.33%	78.67%	72.50%	88.00%	82.08%	93.33%	87.78%
S24	31.83%	36.81%	84.33%	76.25%	82.00%	82.36%	85.50%	87.22%
S25	36.83%	45.28%	41.67%	56.84%	57.67%	62.64%	88.00%	75.00%
S26	32.17%	47.78%	50.83%	59.83%	80.00%	68.06%	85.00%	73.06%
S27	41.83%	45.14%	53.83%	61.25%	65.00%	71.32%	69.17%	80.97%
S28	34.00%	54.31%	47.33%	64.86%	55.67%	80.28%	79.00%	86.39%
S29	34.96%	44.38%	64.50%	66.32%	75.00%	82.92%	87.00%	92.22%

Appendix 11

Subjects of the experimental group’s improved perception and production accuracy.

Improved accuracy in perception tests

subject	Improvement in mid-test1 /θ/-/s/	Improvement in mid-test2/θ/-/s/	Improvement in post-test/θ/-/s/	Improvement in mid-test1/ð/-/z/	Improvement in mid-test2/ð/-/z/	Improvement in post-test/ð/-/z/
S1	28.71%	49.08%	60.19%	29.63%	41.67%	54.63%
S2	30.56%	44.45%	58.33%	25.00%	38.89%	53.70%
S3	24.08%	35.19%	dropped	22.22%	30.56%	dropped
S4	20.37%	37.04%	55.56%	18.52%	32.41%	47.22%
S5	14.08%	37.04%	55.56%	15.74%	36.11%	59.26%
S6	21.30%	37.96%	54.63%	20.37%	33.33%	47.22%
S7	25.00%	44.45%	65.74%	12.04%	26.85%	38.89%
S8	30.55%	45.37%	52.78%	14.81%	28.70%	41.67%
S9	13.89%	37.04%	53.70%	16.67%	28.70%	50.93%
S10	38.89%	dropped	dropped	24.07%	dropped	dropped
S11	23.15%	39.81%	53.70%	21.30%	30.56%	47.22%
S12	9.26%	29.63%	48.15%	22.22%	38.89%	44.44%
S13	12.96%	29.63%	41.67%	24.07%	31.48%	49.07%
S14	16.67%	38.89%	46.30%	12.96%	37.96%	50.00%
S15	12.04%	28.71%	55.55%	12.04%	36.11%	49.07%
S16	13.89%	31.48%	50.00%	24.07%	40.74%	52.78%
S17	21.29%	34.26%	55.55%	19.45%	43.52%	54.63%
S18	22.22%	36.11%	44.07%	14.82%	27.78%	44.44%
S19	18.52%	39.81%	56.48%	29.63%	47.22%	64.81

						%
S20	28.70%	47.22%	60.18%	15.74%	27.78%	50.93 %
S21	27.78%	38.89%	43.52%	23.15%	36.11%	46.30 %
S22	27.78%	39.81%	46.30%	16.67%	39.81%	46.30 %
S23	20.37%	44.44%	60.19%	23.15%	35.18%	54.63 %
S24	26.85%	36.11%	47.22%	25.00%	35.19%	42.59 %
S25	22.22%	41.67%	53.70%	21.30%	45.37%	58.33 %
S26	21.30%	32.41%	40.74%	20.37%	30.56%	46.30 %
S27	23.15%	38.89%	49.07%	24.07%	36.11%	39.82 %
S28	20.37%	28.71%	42.59%	21.30%	42.59%	53.70 %
S29	19.45%	28.71%	48.15%	18.52%	36.11%	44.44 %
average	21.91%	37.10%	50.47%	20.31%	34.63%	47.44 %
max	38.80%	49.08%	65.74%	29.63%	47.22%	64.81 %
min	9.26%	28.71%	40.74%	12.04%	26.85%	38.89 %

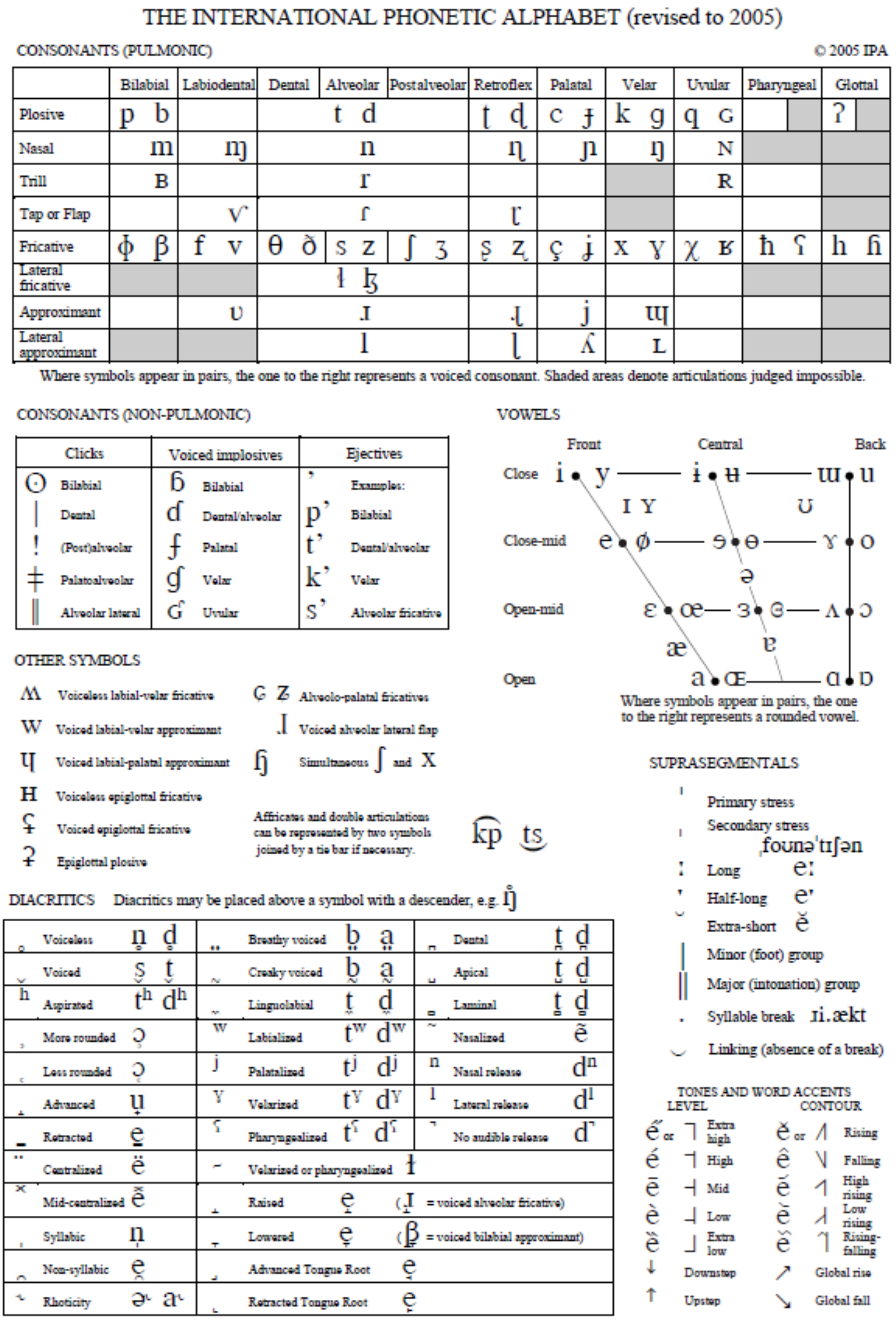
Improved accuracy in production tests

Subject	From pre-test to mid-test1		From pre-test to mid-test2		From pre-test to post-test	
	/θ/	/ð/	/θ/	/ð/	/θ/	/ð/
S1	20.83%	25.56%	35.33%	36.67%	63.50%	52.50%
S2	25.83%	25.28%	45.16%	31.53%	55.33%	42.50%
S3	32.33%	39.86%	48.00%	49.02%	dropped	dropped
S4	17.83%	7.92%	38.33%	31.67%	48.00%	38.61%
S5	22.83%	37.22%	37.83%	43.75%	49.16%	54.72%
S6	14.50%	16.81%	17.33%	29.17%	25.17%	46.25%

S7	15.50%	17.50%	33.33%	21.53%	51.16%	31.53%
S8	15.83%	9.30%	26.66%	23.47%	39.33%	35.69%
S9	24.34%	14.44%	39.00%	37.77%	59.67%	55.14%
S10	42.34%	48.06%	dropped	dropped	dropped	dropped
S11	21.17%	24.58%	18.67%	30.27%	27.50%	33.47%
S12	27.50%	6.39%	40.50%	21.80%	45.34%	24.30%
S13	18.33%	12.92%	26.55%	36.94%	40.83%	43.05%
S14	6.67%	9.17%	32.00%	20.77%	48.17%	31.25%
S15	16.00%	20.55%	19.33%	32.91%	28.16%	35.00%
S16	13.83%	7.08%	19.66%	10.55%	29.33%	23.60%
S17	17.33%	14.86%	26.67%	24.31%	48.50%	39.86%
S18	18.00%	10.13%	28.66%	17.50%	36.50%	35.00%
S19	7.50%	5.97%	33.33%	32.64%	40.17%	46.39%
S20	11.16%	5.28%	20.66%	20.00%	23.50%	26.39%
S21	21.56%	15.66%	34.00%	25.84%	44.17%	31.67%
S22	13.16%	15.56%	46.50%	24.62%	55.83%	40.84%
S23	37.00%	34.17%	46.33%	43.75%	51.66%	49.45%
S24	52.50%	39.44%	50.17%	45.55%	53.67%	50.41%
S25	4.84%	11.56%	20.84%	17.36%	51.17%	29.72%
S26	18.66%	12.05%	47.83%	20.28%	52.83%	25.28%
S27	12.00%	16.11%	23.17%	26.18%	27.34%	35.83%
S28	13.33%	10.55%	21.67%	25.97%	45.00%	32.08%
S29	29.54%	21.94%	40.04%	38.54%	52.04%	47.84%

Appendix 12

IAP Chart



THE INTERNATIONAL PHONETIC ALPHABET (revised to 2005)

CONSONANTS (PULMONIC) © 2005 IPA

	Bilabial	Labiodental	Dental	Alveolar	Postalveolar	Retroflex	Palatal	Velar	Uvular	Pharyngeal	Glottal
Plosive	p b			t d		ʈ ɖ	c ɟ	k ɡ	q ɢ	▨	ʔ ▨
Nasal	m	ɱ		n		ɳ	ɲ	ŋ	ɴ	▨	▨
Trill	ʙ			r				▨	ʀ		▨
Tap or Flap		ⱱ		ɾ		ɽ		▨			▨
Fricative	ɸ β	f v	θ ð	s z	ʃ ʒ	ʂ ʐ	ç ʝ	x ɣ	χ ʁ	ħ ʕ	h ɦ
Lateral fricative	▨	▨		ɬ ɮ						▨	▨
Approximant		ʋ		ɹ		ɻ	j	ɰ			▨
Lateral approximant	▨	▨		l		ɭ	ʎ	ʟ		▨	▨

Where symbols appear in pairs, the one to the right represents a voiced consonant. Shaded areas denote articulations judged impossible.

CONSONANTS (NON-PULMONIC)

Clicks		Voiced implosives		Ejectives	
ʘ	Bilabial	ɓ	Bilabial	ʼ	Examples:
ǀ	Dental	ɗ	Dental/alveolar	pʼ	Bilabial
ǃ	(Post)alveolar	ʄ	Palatal	tʼ	Dental/alveolar
ǂ	Palatoalveolar	ɠ	Velar	kʼ	Velar
ǁ	Alveolar lateral	ʛ	Uvular	sʼ	Alveolar fricative

VOWELS

Front Central Back

Close i • y — ɨ • ʉ — ɯ • u

ɪ ʏ ʊ

Close-mid e • ø — ɘ • ɵ — ɤ • o

ə

Open-mid ɛ • œ — ɜ • ɞ — ʌ • ɔ

æ ɐ

Open a • ɶ — ɑ • ɒ

Where symbols appear in pairs, the one to the right represents a rounded vowel.

OTHER SYMBOLS

ʍ	Voiceless labial-velar fricative	ɕ ʑ	Alveolo-palatal fricatives
w	Voiced labial-velar approximant	ɺ	Voiced alveolar lateral flap
ɥ	Voiced labial-palatal approximant	ɧ	Simultaneous ʃ and x
ʜ	Voiceless epiglottal fricative		
ʢ	Voiced epiglottal fricative		
ʡ	Epiglottal plosive		

Affricates and double articulations can be represented by two symbols joined by a tie bar if necessary. k͡p t͜s

SUPRASEGMENTALS

ˈ	Primary stress	
ˌ	Secondary stress	ˌfoʊnəˈtɪʃən
ː	Long	eː
ˑ	Half-long	eˑ
̆	Extra-short	ĕ
|	Minor (foot) group	
‖	Major (intonation) group	
.	Syllable break	ɹi.ækt
‿	Linking (absence of a break)	

DIACRITICS Diacritics may be placed above a symbol with a descender, e.g. ŋ̊

◌̥	Voiceless	n̥ d̥	◌̤	Breathy voiced	b̤ a̤	◌̪	Dental	t̪ d̪
◌̬	Voiced	s̬ t̬	◌̰	Creaky voiced	b̰ a̰	◌̺	Apical	t̺ d̺
ʰ	Aspirated	tʰ dʰ	◌̼	Linguolabial	t̼ d̼	◌̻	Laminal	t̻ d̻
◌̹	More rounded	ɔ̹	ʷ	Labialized	tʷ dʷ	◌̃	Nasalized	ẽ
◌̜	Less rounded	ɔ̜	ʲ	Palatalized	tʲ dʲ	ⁿ	Nasal release	dⁿ
◌̟	Advanced	u̟	ˠ	Velarized	tˠ dˠ	ˡ	Lateral release	dˡ
◌̠	Retracted	e̠	ˤ	Pharyngealized	tˤ dˤ	̚	No audible release	d̚
◌̈	Centralized	ë	◌̴	Velarized or pharyngealized	ɫ			
◌̽	Mid-centralized	e̽	◌̝	Raised	e̝		(ɹ̝ = voiced alveolar fricative)	
◌̩	Syllabic	n̩	◌̞	Lowered	e̞		(β̞ = voiced bilabial approximant)	
◌̯	Non-syllabic	e̯	◌̘	Advanced Tongue Root	e̘			
˞	Rhoticity	ɚ a˞	◌̙	Retracted Tongue Root	e̙			

TONES AND WORD ACCENTS

LEVEL			CONTOUR		
e̋ or	˥	Extra high	ě or	˩˥	Rising
é	˦	High	ê	˥˩	Falling
ē	˧	Mid	e᷄	˦˥	High rising
è	˨	Low	e᷅	˩˨	Low rising
ȅ	˩	Extra low	e᷈	˧˦˧	Rising-falling
↓		Downstep	↗		Global rise
↑		Upstep	↘		Global fall

Appendix 13

Carrier sentences used in the analysis of CQd /s, z/

/su/, /zu/, /so/, /zo/, /sɚ/, /zɚ/, /se/, /ze/,/sʅ/, /zʅ/, /sa/

1. /pa/ **/su/** /tu/ /ts^{h}eu/ /laɪ/
2. /pa/ **/zu/** /tu/ /ts^{h}eu/ /laɪ/
3. /pa/ **/so/** /tu/ /ts^{h}eu/ /laɪ/
4. /pa/ **/zo/** /tu/ /ts^{h}eu/ /laɪ/
5. /pa/ **/sɚ/** /tu/ /ts^{h}eu/ /laɪ/
6. /pa/ **/zɚ /** /tu/ /ts^{h}eu/ /laɪ/
7. /pa/ **/se/** /tu/ /ts^{h}eu/ /laɪ/
8. /pa/ **/ze/** /tu/ /ts^{h}eu/ /laɪ/
9. /pa/ **/sʅ/** /tu/ /ts^{h}eu/ /laɪ/
10. /pa/ **/zʅ /** /tu/ /ts^{h}eu/ /laɪ/
11. /pa/ **/sa/** /tu/ /ts^{h}eu/ /laɪ/

Appendix 14

Control group’s perception test results

Control group’s perception of /θ/-/s/ (in %)

Subject	Pre-test	Mid-test1	Mid-test2	Post-test
S30	48.15	50.93	50.00	50.93
S31	59.26	62.04	66.67	66.67
S32	52.78	55.56	61.11	60.19
S33	64.81	62.96	69.44	67.59
S34	61.11	64.81	65.74	70.37
S35	50.93	52.78	55.56	58.33
S36	54.63	59.26	63.89	64.81
S37	62.04	63.89	66.67	64.81
S38	55.56	61.11	59.26	62.04
S39	49.07	54.63	59.26	65.74
S40	56.48	59.26	59.26	63.89
S41	65.74	67.59	70.37	74.07
S42	50.93	54.63	52.78	57.41
S43	61.11	62.04	64.81	65.74
S44	60.19	62.96	62.96	64.81
S45	58.33	64.81	67.59	70.37
S46	64.81	68.52	69.44	67.59
S47	59.26	63.89	67.59	70.37
S48	56.48	61.11	60.19	65.74
S49	55.56	60.19	63.89	62.96

Control group's perception of /ð/-/z/ (in %)

subject	Pre-test	Mid-tset1	Mid-test2	Post-test
S30	51.85	54.63	50.93	56.48
S31	55.56	59.26	61.11	65.74
S32	58.33	62.96	62.04	65.74
S33	61.11	61.11	64.81	68.52
S34	63.89	64.81	65.74	64.81
S35	49.07	52.78	54.63	60.19
S36	60.19	59.26	62.96	66.67
S37	56.48	61.11	62.04	65.74
S38	50.00	56.48	61.11	62.04
S39	52.78	55.56	60.19	61.11
S40	57.41	60.19	58.33	62.04
S41	59.26	60.19	64.81	65.74
S42	55.56	54.63	59.26	59.26
S43	60.19	61.11	65.74	66.67
S44	57.41	60.19	60.19	62.96
S45	63.89	62.04	66.67	69.44
S46	54.63	58.33	62.04	62.04
S47	62.04	62.96	66.67	68.52
S48	55.56	57.41	60.19	63.89
S49	54.63	56.48	61.11	64.81

Appendix 15

The control group's profile

Subject	OA (onset age of learning English as a L2)	Gender	Age	Years of English study
S30	14	female	21	7
S31	14	female	20	6
S32	14	female	20	6
S33	13	female	19	6
S34	14	female	21	7
S35	13	female	20	7
S36	14	female	21	7
S37	13	female	19	6
S38	14	female	20	6
S39	14	female	21	7
S40	14	male	21	7
S41	13	male	20	7
S42	14	male	21	7
S43	14	male	20	6
S44	14	male	20	6
S45	14	male	21	7
S46	13	male	20	7
S47	14	male	20	6
S48	13	male	19	6
S49	14	male	20	6

(All the subjects of the control group reported that they had been learning English in publish schools and university; they spent about 1 hour per day in weekdays in English learning mainly by doing English exercises in English textbooks; none of them had travelled to/lived in English-speaking countries, or had any chance to use English on a daily basis.)

References

Abramson, A. S., and Lisker, L., 1970. Discriminability along the voicing continuum: Cross-language tests. In *Proceedings of the sixth international congress of phonetic sciences*. Prague, Czechoslovakia: Academia, 569-573.

Aliaga-García, C., and Mora, J. C., 2009. Assessing the effects of phonetic training on L2 sound perception and production. In Watkins, M. A. and Rauber, A. S. (Ed.), *Recent Research in Second Language Phonetics/Phonology: Perception and Production.* Newcastle upon Tyne: Cambridge Scholars Publishing, 2-31.

Asher, J. J., and Garcia, R., 1969. The optimal age to learn a foreign language. *The Modern Language Journal*, 53(5), 334-341.

Ashby, M., and Maidment, J., 2005. *Introducing phonetic science*. Cambridge University Press.

Ausubel, D.P., 1964. Adults versus children in second-language learning: Psychological considerations. *The Modern Language Journal*, 48(7), 420-424.

Bada, E., 2001. Native language influence on the production of English sounds by Japanese learners. *Reading Matrix: An International Online Journal*, 1(2).

Barkat-Defradas, M., Al-Tamimi, J. E., and Benkirane, T., 2003. Phonetic variation in production and perception of speech: a comparative study of two Arabic dialects. Solé, Recasens, Romero, Proceedings. 15th International Congress of Phonetic Science, Barcelona, 857-860.

Baker, W., and Trofimovich, P., 2006. Perceptual paths to accurate production of L2 vowels: The role of individual differences. *IRAL–International Review of Applied Linguistics in Language Teaching*, 44(3), 231-250.

Banathy, B., Trager, E. C., Waddle, C. D., 1966. The Use of Contrastive Data in Foreign Language Course Development. In Valdman, A. (Ed.), *Trends in Language Teaching*. New York: McGraw Hill, 27-56.

Behrens, S. J., and Blumstein, S. E., 1988a. Acoustic characteristics of English voiceless fricatives: A descriptive analysis. *Journal of Phonetics*, 16, 295-298.

Behrens, S., and Blumstein, S.E., 1988b. On the role of the amplitude of the fricative noise in the perception of place of articulation in voiceless fricative consonants. *Journal of the Acoustical Society of America*, 84(3), 861-867.

Bell-Berti, F., Raphael, L. J., Pisoni, D. B., and Sawusch, J. R., 1979. Some relationships between speech production and perception. *Phonetica*, 36, 373-383.

Bernstein-Ratner, N., 1984. Patterns of vowel modification in mother-child speech. *Journal of Child Language.* 11(3), 557-578.

Bernstein, L. E., Auer Jr, E. T., Eberhardt, S. P., and Jiang, J., 2013. Auditory perceptual learning for speech perception can be enhanced by audiovisual training. *Frontiers in neuroscience*, 7(34).

Best, C. T. and Strange, W., 1992. Effects of phonological and phonetic factors on cross-language perception of approximants. *Journal of Phonetics,* 20(3), 305-330.

Best, C. T., 1994. The emergence of native-language phonological influences in infants: A perceptual assimilation model. In Nusbaum, H. C. (Ed.), *The development of speech perception: the transition from speech sounds to spoken words*. MIT Press, 167-224.

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

Best, C. T., 1995 b. Learning to perceive the sound pattern of English. In Rovee-Collier, C., and Lipsitt, L. (Ed.), *Advances in infancy research*. Hillsdale, NJ: Ablex. 217-304.

Best, C. T., McRoberts, G. W., and Goodell, E., 2001. Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener’s native phonological system. *The Journal of the Acoustical Society of America*, 109(2), 775.

Best, C. C., and McRoberts, G. W., 2003. Infant perception of non-native consonant contrasts that adults assimilate in different ways. *Language and speech*, *46*(2-3), 183-216.

Best, C. T., and Tyler, M. D., 2007. Nonnative and second-language speech perception: Commonalities and complementarities. In Bohn, O. S, and Munro, M. J. (Ed.), *Language experience in second language speech learning: In honor of James Emil Flege*. John Benjamins Publishing, 13-34.

Berger, M. D., 1952. *The American English pronunciation of Russian immigrants*. Doctoral dissertation, Columbia University.

Bialystok, E., 1997. The structure of age: In search of barriers to second language acquisition. *Second Language Research*, 13(2), 116-137.

Bialystok, E. and Hakuta, K., 1999. Confounded age: Linguistic and cognitive factors in age differences for second language acquisition. In Birdsong, D., (Ed.), *Second Language Acquisition and the Critical Period Hypotheses*. Mahwah, NJ: Erlbaum, 162-181.

Birdsong, D., 2005. Interpreting age effects in second language acquisition. In Kroll, J. F., and De Groot, A., (Ed.), *Handbook of bilingualism: Psycholinguistic approaches*. New York: Oxford University Press, 109-127.

Birdsong, D., 2006. Age and second language acquisition and processing: A selective overview. *Language Learning*, *56*(s1), 9-49.

Birdsong, D., 2007. Native like pronunciation among late learners of French as a second language. In Bohn, O. S, and Munro, M. J. (Ed.), *Language experience in second language speech learning: In honor of James Emil Flege.* John Benjamins Publishing, 99-116.

Bitar, N., 1993. Strident feature extraction in English fricatives. paper presented at the *125th meeting of the Acoustical Society of America*. Ottawa, Canada.

Boersma, P. & Weenink, D. J. M. (2013). Praat: doing phonetics by computer (Version 5.3.64). Amsterdam: Institute of Phonetic Sciences of the University of Amsterdam. [Computer program]. Retrieved from http://www.praat.org/

Bohn, O.S., and Flege, J. E., 1990. Interlingual identification and the role of foreign language experience in L2 vowel perception. *Applied Psycholinguistics*, 11(3), 303–328.

Bohn, O. S., and Flege, J. E., 1992. The production of new and similar vowels by adult German learners of English. *Studies in Second Language Acquisition*, *14*(2), 131-158.

Bohn, O. S., 1995. Cross language speech perception in adults: First language transfer doesn't tell it All. In Strange, W., (Ed.), *Speech perception and linguistic experience: Theoretical methodological Issues.* Timonium, MD: York Press, 279–304.

Bongaerts, T., van Summeren, C., Planken, B. and Schils, E., 1997. Age and Ultimate Attainment in the Pronunciation of a Foreign Language. *Studies in Second Language Acquisition,* 19(4), 447-465.

Bouchard Jr, T. J., and McGee, M. G., 1977. Sex differences in human spatial ability: Not an X‐linked recessive gene effect. *Biodemography and Social Biology*, *24*(4), 332-335.

Bradlow, A., 1995. A comparative study of English and Spanish vowels. *Journal of the Acoustical Society of America*, 97(3), 1916–1924.

Bradlow, A., Pisoni, D., AkahansYamada, R., and Tohkura, Y. I., 1997. Training Japanese listeners to identify English/r/and /l/: IV. Some effects of perceptual learning on speech production. *JourmI of the Acoustical Society of America*, *101*(4), 2299-23.

Bradlow, A. R., Kraus, N., Nicol, T. G., McGee, T. J., Cunningham, J., Zecker, S. G., and Carrell, T. D., 1999. Effects of lengthened formant transition duration on discrimination and neural representation of synthetic CV syllables by normal and learning-disabled children. *The Journal of the Acoustical Society of America*, *106*(4), 2086-2096.

Bradlow, A. R., 2008. Training non-native language sound patterns. In Edwards, J. G. H., and Zapini, M. L. (Ed.), *Phonology and second language acquisition*. John Benjamins Publishing, 287-308.

Breeuwer, M., and Plomp, R., 1984. Speech reading supplemented with frequency‐selective sound‐pressure information. *The Journal of the Acoustical Society of America*, 76(3), 686.

Broen, P. A., Strange, W., Doyle, S. S., and Heller, J. H., 1983. Perception and production of approximant consonants by normal and articulation-delayed preschool children. *Journal of Speech, Language and Hearing Research*, 26(4), 601.

Burnham, D., Kitamura, C., and Vollmer-Conna, U., 2002. What's new, pussycat? On talking to babies and animals. *Science*, 296(5572), 1435-1435.

Calvert, G. A., Bullmore, E. T., Brammer, M. J., Campbell, R., Williams, S. C., McGuire, P. K., and David, A. S., 1997. Activation of auditory cortex during silent lipreading. *Science*, 276(5312), 593-596.

Cao A. Y., 2002. Analysis of acoustic cues for identifying the consonant /ð/ in continuous speech. Doctoral dissertation, Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science.

Catalan, R. M. J., 2003. Sex differences in L2 vocabulary learning strategies. *International Journal of Applied Linguistics*, 13(1), 54-77.

Carney, A.E., Widin, G.P., and Viemeister, N.F., 1977. Non-categorical perception of stop consonants differing in VOT. *Journal of the Acoustical Society of America*. 62(4), 961-970.

Carroll, J.B., 1969. Psychological and educational research into second language teaching to young children. In Stern, H. H., and Carroll, J. B. (Ed.), *Language and the young school child*. London: Oxford University Press, 56-68.

Carter, R. J., *An Approach to a Theory of Phonetic Difficulties in Second-Language Learning*. Bolt Beranek and Newman Inc. Report No. 1575.

Chan, M. K.M., 1987. Post-stopped Nasals in Chinese: An Areal Study. *UCLA Working Papers in Phonetics,* 68, 73-119.

Chang, C. B., Haynes, E. F., Yao, Y., and Rhodes, R., 2009. A tale of five fricatives: Consonantal contrast in heritage speakers of Mandarin. *University of Pennsylvania Working Papers in Linguistics*, *15*(1), 6.

Chang, C. B., Yao, Y., Haynes, E. F., and Rhodes, R. 2011. Production of phonetic and phonological contrast by heritage speakers of Mandarin. *The Journal of the Acoustical Society of America*, 129(6), 3964-3980.

Chao, Y. R., 1948. *Mandarin primer.* Massachusetts: Harvard University Press.

Chao, Y. R., 1968. *A Grammar of Spoken Chinese*. Berkeley: University of California Press.

Chen, T., 2001. Audiovisual speech processing. *IEEE Signal Processing Magazine*, 9–31.

Cheng, R.L., 1966. Mandarin Phonological Structure. *Journal of Linguistics,* 2(2), 135-262.

Cheng, R. L., 1968. Tone Sandhi in Taiwanese. *Linguistics*, *6*(41), 19-42.

Chomsky, N., and Halle, M., 1968. *The sound pattern of English.* New York: Harper and Row.

Clark, A., 1995. Boys into modern languages: An investigation of the discrepancy in attitudes and performance between boys and girls in modern languages. *Gender and Education*, *7*(3), 315-326.

Clark, A. and Trafford, J., 1995. Boys into modern languages: an investigation of the discrepancy in attitudes and performance between boys and girls in modern languages. *Gender and Education*, 7(3), 315-325.

Coblin, W. S., 2000. A brief history of Mandarin, *Journal of the American Oriental Society*, 120 (4), 537–552.

Cooper, F.S., Delattre, P.C., Liberman, A.M., Borst, J.M., and Gerstman, L.J., 1952. Some experiments on the perception of synthetic speech sounds. *Journal of the Acoustical Society of America*, 24(6), 597-606.

Cooper, W., 1979. *Speech perception and production: Studies in selective adaptation*. Ablex Publishing Corporation.

Crowder, R. G., 1982. Decay of auditory memory in vowel discrimination. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *8*(2), 153.

Cumming, A., 1994. Writing Expertise and Second-Language Learning Proficiency. In Cumming, A. H. (Ed.) *Bilingual Performance in Reading and Writing*. John Benjamins North America, 821 Bethlehem Pike, Philadelphia, PA 19118, 173-221.

Cutler, A., Mehler, J., Norris, D., and Segui, J., 1983. A language-specific comprehension strategy. *Nature*, *304*(5922), 159-160.

Cutler, A., Mehler, J., Norris, D., and Segui, J., 1986. The syllable's differing role in the segmentation of French and English. *Journal of memory and language*, *25*(4), 385-400.

Dalston, R. M., 1975. Acoustic characteristics of English/w, r, l/spoken correctly by young children and adults. *The Journal of the Acoustical Society of America*, *57*(2), 462-469.

Dart, S. N., 1991. *Articulatory and acoustic properties of apical and laminal articulations*. Ph.D theses, University of California, Los Angeles, CA.

Davenport, M., and Hannahs, S. J., 2010. *Introducing phonetics and phonology*. Routledge.

De Boer, B., and Kuhl, P. K., 2003. Investigating the role of infant-directed speech with a computer model. *Acoustics Research Letters Online*, 4(4), 129-134.

De Boysson-Bardies, B., 1993. Ontogeny of language specific syllabic productions. In Boysson-Bardies, B., Schonen, S., Jusczyk, P., McNeilage, P., and Morton, J., (Ed.), *Developmental neurocognition: Speech and face processing during the first year of life*. Drdrecht: Kluwer, 353-365.

Demorest, M. E., Bernstein, L. E., and DeHaven, G. P., 1996. Generalizability of speech reading performance on nonsense syllables, words, and sentences: Subjects with normal hearing. *Journal of speech and hearing research*, 39(4), 697.

Desai, S., Stickney, G., and Zeng, F. G., 2008. Auditory-visual speech perception in normal-hearing and cochlear-implant listenersa). *The Journal of the Acoustical Society of America*, 123(1), 428-440.

Diehl, R. L., Souther, A. F., and Convis, C. L., 1980. Conditions on rate normalization in speech perception. *Attention, Perception, and Psychophysics*, 27(5), 435-443.

Diehl, R.L.and Kluender, K.R., 1989. On the objects of speech perception. *Ecological Psychology*, 1(2), 121-144.

Diehl, R, L., Lotto, A. J. and Holt, L, L., 2004. Speech perception. *Annual Review of Psychology*, 55, 149-179.

Di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., and Rizzolatti, G., 1992. Understanding motor events: A neurophysiological study. *Experimental Brain Research*, 91(1), 176-180.

Dissosway-Huff, P., Port, R., and Pisoni, D. B.,1982. Context effects in the perception of/r/ and /1/ by Japanese, *Research on Speech PerceptionP rogress Report* No. 8 (Speech Research Laboratory, Indiana University, Bloomington, IN).

DiStefano, S., 2010. *Can audio-visual integration improve with training?* Senior Honors Theses, The Ohio State University.

Dooling, R. J., Okanoya, K., and Brown, S. D., 1989. Speech perception by budgerigars (Melopsittacus undulatus): The voiced-voiceless distinction. *Attention, Perception, and Psychophysics*, 46(1), 65-71.

Edwards, H. T., 1992. *Applied Phonetics: The sounds of American English.* San Diego: Singular Publishing Group.

Elliott, A. R., 1995. Field independence/dependence, hemispheric specialization, and attitude in relation to pronunciation accuracy in Spanish as a foreign language. *The Modern Language Journal*, 79(3), 356-371.

Ellis, R., 1985. *Understanding second language acquisition*. Oxford: Oxford U. P.

Elman, J. L., Diehl, R. L., and Buchwald, S. E., 1977. Perceptual switching in bilinguals. *The Journal of the acoustical Society of America*, 62(4), 971.

Escudero, P., and Boersma, P., 2004. Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition*, 26(4), 551-585.

Escudero, P., 2005. *The Attainment of Optimal Perception in Second-Language Acquisition*. Ph.D dissertation, University of Utrecht: Landelijke Onderzoeksschool Taawetenschap.

Escure, G., 1997. *Creole and dialect continua: Standard acquisition processes in Belize and China (PRC)* (18). John Benjamins Publishing.

Fadiga, L., Craighero, L., Buccino, G., and Rizzolatti, G., 2002. Speech listening specifically modulates the excitability of tongue muscles: A TMS study. *European Journal of Neuroscience*, 15(2), 399-402.

Fadiga, L., Fogassi, L., Pavesi, G., and Rizzolatti, G., 1995. Motor facilitation during action observation: A magnetic stimulation study. *Journal of Neurophysiology*, 73(6), 2608-2611.

Fang, X., and Ping-an, H., 1992. Articulation disorders among speakers of Mandarin Chinese. *American Journal of Speech-Language Pathology*, 1(4), 15-16.

Fant, G., 1973. *Speech Sounds and Features*. MIT Press; Cambridge, MA

Farnetani, E., 1997. Coarticulation and connected speech processes. In Hardcastle, W. J. and Laver, J., and Gibbon, F. E. (Ed.), *The handbook of phonetic sciences*. Wiley-Blackwell, 371-404.

Fisher, C.G., 1968. Confusions among visually perceived consonants. *Journal of Speech and Hearing Research*, 11(4), 796–804.

Flanagan, J. L., 1972. *Speech analysis: Syntheses and perception*. Springer, New York.

Flege, J. E., 1981. The phonological basis of foreign accent: A hypotheses. *TESOL Quarterly,* 15(4), 443-455.

Flege, J. E., and Davidian, R. D., 1984. Transfer and developmental processes in adult foreign language speech production. *Applied Psycholinguistics*, 5(04), 323-347.

Flege, J. E., and Hillenbrand, J., 1986. Differential use of temporal cues to the /s/–/z/ contrast by native and non‐native speakers of English. *The Journal of the Acoustical Society of America*, 79(2), 508-517.

Flege, J. E., 1987. The production of "new" and "similar" phonemes in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics,* 15(1), 47-65.

Flege, J., 1988. The production and perception of speech sounds in a foreign languages. In: Winitz, H. (Ed.), *Human Communication and Its Disorders, A Review*. Norwood, N.J.: Ablex, 224-401.

Flege, J. E., 1989. Chinese subjects' perception of the word‐final English /t/–/d/ contrast: Performance before and after training. *The Journal of the Acoustical Society of America*, 86(5), 1684.

Flege, J. E., and Wang. C., 1989. Native-language phonotactic constraints affect how well Chinese subjects perceive the word-final English/t/-/d/contrast. *Journal of phonetics*. 17, 299-315.

Flege, J. E., 1991a. Perception and production: The relevance of phonetic input to L2 phonological learning. In Heubner, T. and Ferguson, C., (Ed.), *Crosscurrents in second language acquisition and linguistic theory*. Philadelphia: John Benjamins, 249-284.

Flege, J. E., 1991b. The interlingual identification of Spanish and English vowels: Orthographic evidence. *Quarterly Journal of Experimental Psychology*, 43(3), 701-731.

Flege, J. E., 1991c. Age of learning affects the authenticity of voice‐onset time (VOT) in stop consonants produced in a second language. *The Journal of the Acoustical Society of America*, 89(1), 395.

Flege, J. E., 1992a. Speech learning in a second language. In Ferguson, C., Menn, L., and Stoel-Gammon, C. (Ed.), *Phonological development: Models, research, and application*. Timonium, MD: York Press, 565-604.

Flege, J. E., 1992b. The intelligibility of English vowels spoken by British and Dutch talkers. *Intelligibility in speech disorders: Theory, measurement, and management*, 1, 157-232.

Flege, J. E., and Fletcher, K. L., 1992. Talker and listener effects on degree of perceived foreign accent. *The Journal of the Acoustical Society of America*, 91(1), 370.

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

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

Flege, J. E., Munro, M. J., and MacKay, I. R., 1995. Effects of age of second-language learning on the production of English consonants. *Speech Communication*, 16(1), 1-26.

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

Flege, J. E., Bohn, O. S., and Jang, S., 1997. Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics*, 25(4), 437–470.

Flege, J. E., 1999. Age of learning and second language speech. In Birdsong, D. (Ed.), *Second language acquisition and the critical period hypotheses*, 101-131.

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

Flege, J. E., 2003. Assessing constraints on second-language segmental production and perception. In Schiller, N. O., and Meyer, A. S. (Ed.), *Phonetics and phonology in language comprehension and production, differences and similarities*, 319-355.

Flege, J. E., Yeni-Komshian, G. H., and Liu, S., 1999. Age constraints on second-language acquisition. *Journal of memory and language*. 41(1), 78-104.

Fowler, C. A., 1981. Production and perception of coarticulation among stressed and unstressed vowels. *Journal of Speech Hearing Research.* 24(1), 127–139.

Fowler, C. A., 1984. Segmentation of coarticulated speech in perception. *Perception & Psychophysics*, 36(4), 359-368.

Fowler, C. A., 1986. An event approach to the study of speech perception from a direct-realistic perspective. *Journal of Phonetic,* 14(1), 3-28.

Fowler, C. A., 1989. Real objects of speech perception: a commentary on Diehl and Kluender. *Ecological Psychology*, 1(2), 145-160.

Fowler, C. A., 1994a. Speech perception: direct realist theory. In Asher, R. E., and James M. Y. (Ed.), *The Encyclopedia of Language and Linguistics 8*. Oxford: Pergamon, 4199–4203.

Fowler C. A., 1994b. Invariants, specifiers, cues: An investigation of locus equations as information for place of articulation. *Perception & psychophysics*, 55(6), 597-610.

Fowler, C. A., 1996. Listeners do hear sounds, not tongues. *The Journal of the Acoustical Society of America*, *99*(3), 1730-1741.

Fowler, C. A., Brown, J. M., Sabadini, L., and Weihing, J., 2003. Rapid access to speech gestures in perception: Evidence from choice and simple response time tasks. *Journal of Memory and Language*, 49(3), 396-413.

Fox, R. A., Flege, J. E., and Munro, M. J., 1995. The perception of English and Spanish vowels by native English and Spanish listeners: A multidimensional scaling analysis. *Journal of the Acoustical Society of America*, 97(4), 2540–2550.

Fry, D. B., 1966. Mode de perception des sons du langage. A. Moles and B. Vallancien, Phonétique et phonation, 191-206. (cited by Barkat-Defradas, Al-Tamimi, and Benkirane, 2003).

Fujimura, O., and Erickson, D., 1997. Acoustic phonetics. In Hardcastle, W. J. and Laver, J., and Gibbon, F. E. (Ed.), *The handbook of phonetic sciences*. Wiley-Blackwell, 65-115.

Fullana, N., and Mora, J. C., 2008. Production and perception of voicing contrasts in English word-final obstruents: Assessing the effects of experience and starting age. In *New Sounds 2007: Proc 5th International Symposium on the Acquisition of Second Language Speech*, 207-221.

Galantucci, B., Fowler, C. A., and Turvey, M. T., 2006. The motor theory of speech perception reviewed. *Psychonomic Bulletin & Review*, 13(3), 361-377.

García Mayo, M. and García Lecumberri, M. (Ed.), 2003. *Age and the acquisition of English as a foreign language.* Clevedon: Multilingual Matters.

Gardner, R. C., 1985. *Social psychology and second language learning: The role of attitudes and motivation*. London: Edward Arnold.

Gariety, M., 2009. *Effects of training on intelligibility and integration of sine-wave speech*. Senior Honors Thesis, The Ohio State University.

Gelder, B. D., and Vroomen, J., 1992. Auditory and visual speech perception in alphabetic and non-alphabetic Chinese-Dutch bilinguals. *Advances in psychology*, *83*, 413-426.

Gesi, A. T., Massaro, D. W., and Cohen, M. M., 1992. Discovery and expository methods in teaching visual consonant and word identification. *Journal of Speech, Language and Hearing Research*, *35*(5), 1180.

Ghazanfar, A. A., and Schroeder, C. E., 2006. Is neocortex essentially multisensory? *Trends in cognitive sciences*, *10*(6), 278-285.

Giannakopoulou, A., 2012. Plasticity in second language (L2) learning: perception of L2 phonemes by native Greek speakers of English.

Gillette, S., 1980. Contextual variation in the perception of L and R by Japanese and Korean speakers. *Minnesota Papers in Linguistics and the Philosophy of Language*, 6, 59-72.

Goldstein, D., Haldane, D., and Mitchell, C., 1990. Sex differences in visual-spatial ability: The role of performance factors. *Memory and Cognition*, *18*(5), 546-550.

Goto, H., 1971. Auditory perception by normal Japanese adults of the sounds /l/ and /r/. *Neuropsychologia*, *9*(3), 317-323.

Gottfried, T., and Beddor, P., 1988. Perception of temporal and spectral information in French vowels. *Language and Speech*, *31*, 57–75.

Grant, K.W. and Seitz, P.F., 1998. Measures of auditory-visual integration in nonsense syllables and sentences. *The Journal of the Acoustical Society of America, 104 (4)*, 2438-2450.

Guess, D., 1969. Functional analysis of receptive language and productive speech: Acquisition of the plural morpheme. *Journal of applied behavior analysis, 2(1),* 55-64.

Guess, D., and Baer, D. M., 1973. An analysis of individual differences in generalization between receptive and productive language in retarded children. Journal of applied behavior analysis, 6(2), 311-329.

Hagiwara, R., 1995. *Acoustic realizations of American/r/as produced by women and men*.Vol. 90. Phonetics Laboratory, Dept. of Linguistics, UCLA.

Hakuta, K., Bialystok, E., and Wiley, E., 2003. Critical evidence a test of the critical-period hypothesis for second-language acquisition. *Psychological Science*, *14*(1), 31-38.

Handley, Z., Sharples, M., and Moore, D., 2009. Training novel phonemic contrasts: A comparison of identification and oddity discrimination training. In: *Speech and Lanugae Technology in Education (SLaTE)* 2009, 3-5 Sep. 2009, Wroxall, England.

Hardison, D. M., 2003. Acquisition of second-language speech: Effects of visual cues, context, and talker variability. *Applied Psycholinguistics*, *24*(04), 495-522.

Hardison, D. M., 2005a. Second-language spoken word identification: Effects of perceptual training, visual cues, and phonetic environment. *Applied Psycholinguistics*, *26*(4), 579-596.

Hardison, D. M., 2005b. Variability in bimodal spoken language processing by native and nonnative speakers of English: A closer look at effects of speech style. *Speech communication*, *46*(1), 73-93.

Harnad, S. R. (Ed.)., 1990. *Categorical perception: The groundwork of cognition*. Cambridge University Press.

Harrelson, A., 1969. *Effects of productive speech training on receptive language*. Master's theses, University of Kansas.

Harris, K. S., 1958. Cues for the discrimination of American English fricatives in spoken syllables. *Language and speech*, *1*(1), 1-7.

Harris, J., 1969. *Spanish Phonology*. Cambridge MA: The MIT Press.

Harris, L.J., 1978. Sex differences in spatial ability: Possible environmental, genetic, and neurological factors. In Kinsbourne, M. E. (Ed.). *Assymetrical function of the brain*. New York; Cambridge University Press, 405-522.

Hawkins, S., 1999. Looking for invariant correlates of linguistic units: two classical theories of speech perception. In *Pickett, J. M.* (Ed.) *The Acoustics of Speech Communication. Fundamentals, Speech Perception Theory and Technology.* Needham Heights, MA: Allyn and Bacon, 198-232.

Hazan, V., Sennema, A., Iba, M., and 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.

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

Heinz, J. M., and Stevens, K. N., 1961. On the properties of voiceless fricative consonants. *The Journal of the Acoustical Society of America*, *33*(5), 589-596.

Hillenbrand, J., Getty, L. A., Clark, M. J., and Wheeler, K., 1995. Acoustic characteristics of American English vowels. *The Journal of the Acoustical society of America*, *97*(5), 3099-3111.

Hirata, Y., and Kelly, S. D., 2010. Effects of lips and hands on auditory learning of second-language speech sounds. *Journal of Speech, Language, and Hearing Research*, *53*(2), 298-310.

Hoffman, P. R., Daniloff, R. G., Bengoa, D., and Schuckers, G. H., 1985. Misarticulating and normally articulating children's identification and discrimination of synthetic [r] and [w]. *Journal of Speech and Hearing Disorders*, *50*(1), 46.

Holt, L.L, Lotto, A.J, and Kluender, K.R., 2000. Neighboring spectral content influences vowel identification. *The Journal of the Acoustic Society of America,* 108(2), 710–722.

Hu, F., 2008. The three sibilants in Standard Chinese. *Proc. 8th International Seminar onSpeech Production, Strasbourg, France*.

Hughes, G. W., and Halle, M., 1956. Spectral properties of fricative consonants. *The journal of the acoustical society of America*, *28*(2), 303-310.

Hurford, J. R., 1991. The evolution of the critical period for language acquisition. *Cognition*, *40*(3), 159-201.

Iverson, P., and Evans, B. G., 2009. Learning English vowels with different first-language vowel systems II: Auditory training for native Spanish and German speakers. *The Journal of the Acoustical Society of America*, *126*, 866.

Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y. I., Kettermann, A., and Siebert, C., 2003. A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition*, *87*(1), B47-B57.

Jackson, P.L., 1988. The theoretical minimal unit for visual speech perception: Visemes and coarticulation. *The Volta Review, 90(5),* 99-114.

James, A., 1988. *The Acquisition of Second Language Phonology*. Tubingen: Gunter Narr.

James, K., 2009. *The effects of training on intelligibility of reduced information speech stimuli*. Senior Honors Theses, The Ohio State University.

Jamieson, D. G., and Morosan, D. E., 1986. Training non-native speech contrasts in adults: Acquisition of the English /ð/-/θ/ contrast by francophones. *Attention, Perception, and Psychophysics*, *40*(4), 205-215.

Jamieson, D. G. and Morosan, D. E., 1989. Training new, nonnative speech contrasts: A comparison of the prototype and perceptual fading techniques. *Canadian Journal of Psychology,* 43, 88–96.

Jamieson, D. G., and Rvachew, S., 1992. Remediating speech production errors with sound identification training. *Journal of Speech-Language Pathology and Audiology*, 16(3), 201-210.

Jongman, A., 1989. Duration of frication noise required for identification of English fricatives. *Journal of the Acoustical Society of America*, 85, 1718-1725.

Jongman, A., Wayland, R., andWong, S., 2000. Acoustic characteristics of English fricatives. The Journal of the Acoustical Society of America, 108(3), 1252–1263.

Jongman, A., Wang, Y., and Kim, B. H., 2003. Contributions of semantic and facial information to perception of nonsibilant fricatives. *Journal of speech, language, and hearing research*, *46*(6), 1367.

Kaylani, C., 1996. The influence of gender and motivation on EFL learning strategy use in Jorda. In Oxford, R. L. (Ed.), *Language learning strategies around the world: cross-cultural perspectives*. University of hawai'i at Manoa: Second Language Teaching and Curriculum Centre.

Kent, R.D., and Read, C., 2002. *The Acoustic Analysis of Speech.* San Diego, Calif : Singular Pub. Group.

Klatt, D.H., 1974. Duration of [s] in English words. *Journal of Speech and Hearing Research*, 17, 41-50.

Klatt, D. H., 1989. Review of selected models of speech perception.

Kohler, E., Keysers, C., Umiltà, M. A., Fogassi, L., Gallese, V., and Rizzolatti, G., 2002. Hearing sounds, understanding actions: Action representation in mirror neurons. *Science*, 297(5582), 846-848.

Kuhl, P. K., and Miller, J. D., 1975. Speech perception by the chinchilla: Voiced-voiceless distinction in alveolar plosive consonants. *Science*, *190*(4209), 69-72.

Kuhl, P. K., 1991. Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Attention, Perception, and Psychophysics*, *50*(2), 93-107.

Kuhl, P. K., 1992. Psychoacoustics and speech perception: Internal standards, perceptual anchors, and Prototypes. In Werner, L. A., and Rubel, E. W. (Ed.), Developmental Psychoacoustics. APA science volumes Washington, DC, US: *American Psychological Association*, 293-332.

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., and Lindblom, B., 1992. Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, 255(5044), 606-608.

Kuhl, P. K., 1993, Early Linguistic Experience and Phonetic Perception: Implications For Theories of Developmental Speech Production. *Journal ofPhonetics,* 21, 125-139.

Kuhl, P. K., 1994. Learning and representation in speech and language. *Current option in neurobiology*. 4(6), 812-822.

Kuhl, P. K., Andruski, J. E., Chistovich, I. A., Chistovich, L. A., Kozhevnikova, E. V., Ryskina, V. L., and Lacerda, F., 1997. Cross-language analysis of phonetic units in language addressed to infants. *Science*, *277*(5326), 684-686.

Kuhl, P. K., 2000a. A new view of language acquisition. *Proceedings of the National Academy of Sciences*, *97*(22), 11850-11857.

Kuhl, P. K., 2000b. Language, mind, and brain: Experience alters perception. *The new cognitive neurosciences*, *2*, 99-115.

Kuhl, P. K., Tsao, F. M., and Liu, H. M., 2003. Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning. *Proceedings of the National Academy of Sciences*, *100*(15), 9096-9101.

Kuhl, P. K., 2004. Early language acquisition: cracking the speech code. *Nature reviews neuroscience*, *5*(11), 831-843.

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

Kurpaska, M., 2010. *Chinese Language(s): A Look Through the Prism of The Great Dictionary of Modern Chinese Dialects*, Walter de Gruyter,

Ladefoged, P., and Maddieson, I., 1986. *Some of the sounds of the world's languages*. Phonetics Laboratory, Department of Linguistics, UCLA.

Ladefoged, P., 1996. *Elements of Acoustic Phonetics. Chicago* (2nd edition). University of Chicago Press.

Ladefoged, P., and Maddieson, I., 1996. *The sounds of the world's languages*. Oxford: Blackwell.

Ladefoged, P., 2006. A course in phonetics (5th edition). Boston: Thomson Wadsworth.

Ladefoged, P., 2007. Articulatory features for describing lexical distinctions. *Language*, 161-180.

Ladefoged, P., 1993. *A Course in Phonetics (3rd edition)*. Fort Worth: Harcourt, Brace, Jovanovich.

Lado, R., 1957. *Linguistics Across Cultures.* Ann Arbor: University of Michigan Press.

Lambacher, S. G., Martens, W. L., Kakehi, K., Marasinghe, C. A., and Molholt, G., 2005. The effects of identification training on the identification and production of American English vowels by native speakers of Japanese. *Applied Psycholinguistics*, 26(02), 227-247.

Lee, W. S., 1999. An articulatory and acoustic analysis of the syllable-initial sibilants and approximants in Beijing Mandarin. *The International Congress on Phonetic Sciences*. San Francisco, 413-416.

Lee, C. Y., 2003. An acoustic study of strident fricatives in Mandarin Chinese. *The Journal of the Acoustical Society of America*, *113*, 2329.

Lee, S. I., 2011. Spectral analysis of Mandarin Chinese sibilant fricatives. In *The 17th International Congress of Phonetic Sciences (ICPhS XVII)*, 1178-1181).

Lehiste, I., 1964. *Acoustic characteristics of selected English consonants*. Bloomington: Indiana University Research Center in Anthropology. Folklore and Linguistics.

Lidestam, B., Moradi, S., Pettersson, R., and Ricklefs, T. 2014. Audiovisual training is better than auditory-only training for auditory-only speech-in-noise identification. *The Journal of the Acoustical Society of America*, *136*(2), EL142-EL147.

Lenneberg, E. H., 1967. *Biological Foundations of Language*. New York: Wiley and sons.

Li, F., Edwards, J., and Beckman, M., 2007. Spectral measures for sibilant fricatives of English, Japanese, and Mandarin Chinese. In *Proceedings of the XVIth International Congress of Phonetic Sciences*, 4, 917-920.

Li, Y., 2010. *The Influence of L1 and L1 Dialect on the Pronunciation of English Segmentals by Chinese Adult and Child Learners*. Unpublished master's theses, University of Stirling, Stirling, U.K.

Liberman, A. M., Delattre, P.C., and Cooper, F. S., 1952. The role of selected stimulus-variables in the perception of the unvoiced stop consonants. *American Journal of Psychology, 65,* 497-516.

Liberman, A. M., 1957. Some results of research on speech perception. *Journal of the Acoustical Society of America*, 29, 117-123.

Liberman, A. M., Harris, K. S., Hoffman, H. S. and Griffith, B. C., 1957. The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology* 54 (5): 358–368.

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

Liberman, A. M., Harris, K. S., Kinney J. A., and Lane, H., 1961b. The discrimination of relative onset-time of the components of certain speech and nonspeech patterns. *Journal of experimental psychology,* 61(5), 379-388.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., and Studdert-Kennedy, M., 1967. Perception of the speech code. *Psychological review*, 74(6), 431.

Liberman, A. M., and Whalen, D. H., 2000. On the relation of speech to

Language. *Trends in Cognitive Sciences*, 4, 187–196.

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

Liberman, A. M., and Mattingly, I. G., 1989. A specialization for speech perception. *Science* 243(4890), 489–494.

Liberman, A. M., Whalen, D. H., 2000. On the relation of speech to language. *Trends in cognitive sciences*, 4 (5), 187–196.

Lindblom, B. E. F, Studdert-Kennedy, M., 1967. On the role of formant transitions in vowel recognition. *The Journal of the Acoustic Society of America*, 42(4), 830–843.

Lisker, L., and Abramson, A. S., 1967. The voicing dimension: Some experiments in comparative phonetics. In *Proceedings of the Sixth International Congress of Phonetic Sciences*. Prague: Academia, 6-7.

Liu, H. M., Kuhl, P. K., and Tsao, F. M., 2003. An association between mothers' speech clarity and infants' speech discrimination skills. *Developmental Science*, *6*(3), F1-F10.

Lively, S. E., Logan, J. S., and 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. *The Journal of the Acoustical Society of America*, *94*, 1242.

Locke, J. L., 1988. Variation in human biology and child phonology: A response to Goad and Ingram. *Journal of Child Language*, *15*(3), 663-668.

Löfqvist, A., 2010. Theories and models of speech production. In Hardcastle, W. J. and Laver, J., and Gibbon, F. E. (Ed.), *The handbook of phonetic sciences*. Wiley-Blackwell, 406-426.

Logan, J. S., Lively, S. E., and Pisoni, D. B., 1991. Training Japanese listeners to identify English /r/ and /1/. *The Journal of the Acoustical Society of America,* 89(2), 874-886.

Long, M. H., 1990. Maturational constraints on language development. *Studies in Second Language Acquisition* 12(03), 251-285.

Lotto, A. J., Sato, M. and Diehl, R., 2004. Mapping the task for the second language learner: the case of Japanese acquisition of /r/ and /l/. *From sound to sense*, 50(2004), C381-C386.

MacKain, K.S., Best, C.T., and Strange, W., 1981. Categorical perception of English /r/ and /l/ by Japanese bilinguals, *Applied Psycholinguistics*, 2(4), 369-390.

Macmillan, N. A., 1987. Beyond the categorical/continuous distinction: A psychophysical approach to processing modes. In Harnad, S. R. (Ed.), *Categorical perception: The groundwork of cognition*. Cambridge University Press, 53-85.

MacNamara, J., 1973. Attitudes and learning a second language. In Shuy, R., and Fasold, R (Ed.), *Language attitudes: Current Trends and Prospects*. Georgetown University Press, Washington, DC.

Marinova-Todd, S.H., Marshall, D.B., Snow, C.E., 2000. Three misconceptions about age and L2 learning. *TESOL Quarterly*, 34(1), 9-34.

Mann, R. A., and Baer, D. M., 1971. The effects of receptive language training on articulation. *Journal of applied behavior analysis*, 4 (4), 291-298.

Mann, V. A., and Repp, B. H., 1980. Influence of vocalic context on perception of the [ʃ]-[s] distinction. *Attention, Perception, and Psychophysics*, *28*(3), 213-228

Mann, V. A, Repp, B. H., 1981. Influence of preceding fricative on stop consonant perception. *The Journal of the Acoustical Society of America*, *69*(2), 548-558.

Mann, V. A., 1986. Distinguishing universal and language-dependent levels of speech perception: evidence from Japanese listeners' perception of English "l" and "r". *Cognition*, *24*(3), 169-196.

Manrique, A. M. B., and Massone, M. I., 1981. Acoustic analysis and perception of Spanish fricative consonants. *The Journal of the Acoustical Society of America*, 69(4), 1145.

Massaro, D. W., 1984. Children's perception of visual and auditory speech. *Child Development*, 55, 1777-1788.

Massaro, D. W., Thompson, L. A., Barron,B. , and Laren, E., 1986. Developmental changes in visual and auditory contributions to speech perception. *Journal of Experimental Child Psychology*, 41(1), 93-113.

Massaro, D. W., 1987. *Speech perception by ear and eye: A paradigm for psychological inquiry*. Psychology Press.

Massaro, D. W., Cohen, M. M., and Gesi, A. T., 1993. Long-term training, transfer, and retention in learning to lipread. *Attention, Perception, and Psychophysics*, 53(5), 549-562.

Massaro, D. W., Cohen, M. M., and Gesi, A. T., Heredia, R., Tsuzaki, M., 1993. Bimodal speech perception: an examination across languages. *Journal of Phonetics*, 21, 445-478.

Maye, J., Werker, J. F., and Gerken, L., 2002. Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, 82(3), B101-B111.

Mayo, L. H., Florentine, M., and Buus, S., 1997. Age of second-language acquisition and perception of speech in noise. *Journal of Speech, Language and Hearing Research*, 40(3), 686-693.

Mayo, C., and Turk, A., 2004. Adult–child differences in acoustic cue weighting are influenced by segmental context: Children are not always perceptually biased toward transitions. *The Journal of the Acoustical Society of America*, 115(6), 3184-3194.

McAllister, R., Flege, J. E., and Piske, T., 2002. The influence of L1 on the acquisition of Swedish quantity by native speakers of Spanish, English and Estonian. *Journal of phonetics*, 30(2), 229-258.

McAllister, R, 2007. Strategies for realization of L2-categories. In Bohn, O. S, and Munro, M. J. (Ed.), *Language experience in second language speech learning: In honor of James Emil Flege*. John Benjamins Publishing, 17, 153-166.

McCandliss, B. D., Fiez, J. A., Protopapas, A., Conway, M., and 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, and Behavioral Neuroscience*, 2(2), 89-108.

McCullough, J., Somerville, B., and Honorof, D. N., 2000. *Comma gets a cure*. A diagnostic passage for accent study.

McGuire, G., 2010. A brief primer on experimental designs for speech perception research. *Laboratory Report*, 77.

McGurk, H., and MacDonald, J., 1976. Hearing lips and seeing voices. *Nature*. 264, 746-748.

McMurray, B. and Aslin, R.N., 2005. Infants are sensitive to within-category variation in speech perception. *Cognition*, 95(2), B15-B26.

Miller, J. L, Liberman, A. M., 1979. Some effects of later-occurring information on the perception of stop consonant and semivowel. *Perception & Psychophysics,* 25(6), 457–465.

Miller, J. L., 1987. Rate-dependent processing in speech perception. In Ellis, A. W. (Ed.), *Progress in the Psychology of Language*. Hillsdale, NJ, England: Lawrence Erlbaum Associates, 3, 119-157.

Miller, J. L., 1994. On the internal structure of phonetic categories: a progress report. *Cognition*, 50(1), 271-285.

Moradi, S., Lidestam, B., and Rönnberg, J. 2013. Gated audiovisual speech identification in silence vs.noise: Effects on time and accuracy, *Frontiers in psychology* . 4, 359.

Morgan, R. A., 1984. Auditory discrimination in speech-impaired and normal children as related to age. *British Journal of Disorders of Communication*, 19(1), 89-96.

Morosan, D. E., and Jamieson, D. G., 1989. Evaluation of a Technique for Training New Speech Contrasts: Generalization Across Voices, but not Word-Position or Task. *Journal of Speech, Language and Hearing Research*, 32(3), 501.

Mortreux, S., 2008. English Coronal Consonants Produced by L2 French Learners-An articulatory and Acoustic Study. In *8th International Seminar on Speech Production*, 145-148.

Murata, A., Fadiga, L., Fogassi, L., Gallese, V., Raos, V., and Rizzolatti, G., 1997. Object representation in the ventral premotor cortex (area F5) of the monkey. *Journal of Neurophysiology*, 78(4), 2226-2230.

Munro, M. J., 1993. Productions of English vowels by native speakers of Arabic: Acoustic measurements and accentedness ratings. *Language and Speech*, 36(1), 39-66.

Munro, M. J., and Derwing, T. M., 1995. Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, 45(1), 73-97.

Nath, A. R. and Beauchamp, M. S., 2011. A neural basis for interindividual differences in the McGurk effect, a multisensory speech illusion. *NeuroImage,* 59(1), 781-787.

Navarra, J., and Soto-Faraco, S., 2007. Hearing lips in a second language: visual articulatory information enables the perception of second language sounds. *Psychological research*, 71(1), 4-12.

Nearey, T. M., 1989. Static, dynamic, and relational properties in vowel perception. *The Journal of the Acoustical Society of America*, 85(5), 2088-2113

Nittrouer, S., and Studdert-Kennedy, M., 1987. The role of coarticulatory effects in the perception of fricatives by children and adults. *Journal of Speech, Language and Hearing Research*, 30(3), 319-29.

Norman, J., 1988, *Chinese*, Cambridge University Press.

Olive J. P., Greenwood A., and Coleman J., 1993. *Acoustics of American English Speech: A Dynamic Approach*. Springer-Verlag, New York 166–175.

Ortega-Llebaria, M., Faulkner, A., and Hazan, V., 2001. Auditory-visual L2 speech perception: Effects of visual cues and acoustic-phonetic context for Spanish learners of English. In *AVSP 2001-International Conference on Auditory-Visual Speech Processing*, 149-154.

Owens, E., and Blazek, B., 1985. Visemes observed by hearing-impaired and normal-hearing adult viewers. *Journal of Speech, Language and Hearing Research*, 28(3), 381.

Oxford, R., Nyikos, M., and Ehrman, M., 1988. Vive la difference? Reflections on sex differences in use of language learning strategies. *Foreign Language Annals*, 21(4), 321-329.

Oxford, R. L., 1993. Instructional Implications of Gender difference in Second/Foreign Language (L2) Learning Styles and Strategies. *Applied language learning*, 4(1), 65-94.

Oyama, S., 1976. A Sensitive Period for the Acquisition of a Nonnative phonological System*. Journal of Psycholinguistic Research*, 5(3), 261-285.

Patkowski, M. S., 1980. The Sensitive Period for the Acquisition of Syntax in a Second. *Language learning*, 30(2), 449-468.

Patkowski, M. S., 1990. Age and accent in a second language: A reply to James Emil Flege. *Applied linguistics*, 11(1), 73-89.

Penfield,W. and Roberts, L., 1959. *Speech and Brain Mechanisms.* Princeton, NJ: Princeton University Press.

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

Peterson, G. E., and Lehiste, I., 1960. Duration of syllable nuclei in English. *Journal of the Acoustical Society of America*, 32(6), 693-703.

Perkell, J. S., 1990. Testing theories of speech production: Implications of some detailed analyses of variable articulatory data. In Hardcastle, W. J., and Marchal, A. (Ed.), *Speech production and speech modelling*. Springer Netherlands 263-288.

Picard, M., 2002. The differential substitution of English/eth/in French: The case against underspecification in L2 phonology. *Lingvisticae Investigationes*, 25(1), 87-96.

Pickett, J. M., 1999. *The acoustics of speech communication*. Boston : Allyn and Bacon.

Piske, T., MacKay, I. R., and Flege, J. E., 2001. Factors affecting degree of foreign accent in an L2: A review. *Journal of phonetics*, 29(2), 191-215.

Pisoni, D. B., 1973. Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Attention, Perception, and Psychophysics*, 13(2), 253-260.

Pisoni, D. B., and Lazarus, J. H., 1974. Categorical and noncategorical modes of speech perception along the voicing continuum. *The Journal of the Acoustical Society of America*, 55(2), 328-333.

Pisoni. D.B. and Tash, J., 1974. Reaction times to comparisons within and across phonetic categories. *Perception and Psychophysics*, 15(2), 285-290.

Pisoni, D. B., Aslin, R. N., Perey, A. J., and Hennessy, B. L., 1982. Some effects of laboratory training on identification and discrimination of voicing contrasts in stop consonants. *Journal of Experimental Psychology: Human Perception and Performance*, 8(2), 297.

Pisoni, D. B., and Lively, S. E., 1995. Variability and invariance in speech perception: a new look at some old problems in perceptual learning. In Strange, W. (Ed.), *Speech perception and linguistic experience: issues in cross-language research*, 429-455.

Prator, C. H., and Robinett, B. W., 1985. *Manual of American English pronunciation*. Harcourt College Pub.

Pulvermüller, F., Huss, M., Kheri, F., Moscoso del Prado Martin, F., Hauk, O., and Shtyrov, Y., 2006. Motor cortex maps articulatory features of speech sounds. *Proceedings of the National Academy of Sciences*, 103(20), 7865-7870.

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

Powell, R.C.and Baters, J. D., 1985. Pupils' perceptions of foreign language learning at 12+: some gender difference. *Educational Studies*, 11(1), 11-23.

Qian, Y., Liang, H., and Soong, F.K., 2009. A cross-language state sharing and mapping approach to bilingual (Mandarin-English) TTS. A*udio, Speech, and Language Processing, IEEE Transaction* on, 17(6), 1231-1239.

Raaymakers, E., and Crul, T., 1988. Perception and production of the final /s-ts/ contrast in Dutch by misarticulating children. *Journal of Speech and Hearing Disorders*, 53(3), 262-270.

Ramsey, S.R., 1987. *The Languages of China.* Princeton, N.J.: Princeton University Press.

Ranta, A., 2010. *How Does Feedback Impact Training in Audio-Visual Speech Perception?* Senior Honors Thesis. The Ohio State University.

Raphael, L. J., 2005. Acoustic cues to the perception of segmental phonemes. In Pisoni, D. B, and Remez, R. E. (Ed.), *The handbook of speech perception*. Blackwell Publishing, 182-206.

Reid, J. M., 1987. The learning style preferences of ESL students. *TESOL quarterly*, 21(1), 87-111.

Remez, R. E., 2008. Perceptual Organization of Speech. In Pisoni, D. B, and Remez, R. E. (Ed.), *The handbook of speech perception*. Blackwell Publishing, 28-50.

Riney T. and Flege, J., 1998. Changes over time in global foreign accent and liquid identifiability and accuracy. *Studies in Second Language Acquisition,* 20(2), 213-243.

Ritchie, W. C., 1968. On the explanation of phonic interference. *Language Learning*, 18(3‐4), 183-197.

Rizzolatti, G., Fadiga, L., Fogassi, L., and Gallese, V., 1997. The space around us. *Science*, 277(5323), 190-191.

Rizzolatti, G., and Craighero, L., 2004. The mirror-neuron system. *Annual Review of Neuroscience*, 27, 169-192.

Robinett, B. W., and Schachter, J. (Ed.), 1983. *Second language learning: Contrastive analysis, error analysis, and related aspects*. Ann Arbor, MI: University of Michigan Press.

Rosenblum, L. D., Schmuckler, M. A., and Johnson, J. A., 1997. The McGurk effect in infants. *Attention, Perception, and Psychophysics*, 59(3), 347-357.

Rvachew, S., 1994. Speech perceptual training can facilitate sound production learning. *Journal of Speech, Language and Hearing Research*, 37(2), 347.

Sams, M., Aulanko, R., Hämäläinen, M., Hari, R., Lounasmaa, O. V., Lu, S. T., and Simola, J., 1991. Seeing speech: visual information from lip movements modifies activity in the human auditory cortex. *Neuroscience letters*, 127(1), 141-145.

Samuel, A. G., 1977. The effect of discrimination training on speech perception: Noncategorical perception. *Perception and Psychophysics*, 22(4), 321-330

Sanders, B., Soares, M. P., and D'Aquila, J. M., 1982. The sex difference on one test of spatial visualization: A nontrivial difference. *Child Development*, 1106-1110.

Sato, M., Troille, E., Ménard, L., Cathiard, M. A., and Gracco, V., 2013. Silent articulation modulates auditory and audiovisual speech perception. *Experimental Brain Research*, 227(2), 275-288.

Schwartz, R. G., and Leonard, L. B., 1982. Do children pick and choose? An examination of phonological selection and avoidance in early lexical acquisition. *Journal of Child Language*, 9(02), 319-336.

Schwartz, J. L., Basirat, A., Ménard, L., and Sato, M., 2012. The Perception-for-Action-Control Theory (PACT): A perceptuo-motor theory of speech perception. J*ournal of Neurolinguistics,* 25(5), 336-354.

Scovel, T., 1969. Foreign accents, language acquisition, and cerebral dominance, *Language learning,* 19(3-4), 245-253.

Scovel, T., 1988. *A Time to Speak. A Spycholinguistic Inquiry into the Critical Period for Human Speech.* Rowley, MA: Newbury House.

Sebastián-Gallés, N., 2005. 22 Cross-Language Speech Perception. In Pisoni, D. B, and Remez, R. E. (Ed.), *The handbook of speech perception*. Blackwell Publishing, 546-566.

Sekiyama, K., and Tohkura, Y., 1993. Inter-language differences in the influence of visual cues in speech perception. *Journal of Phonetics,* 21(4), 427–444.

Sekiyama, K., 1997. Cultural and linguistic factors in audiovisual speech processing: The McGurk effect in Chinese subjects. *Perception and Psychophysics,* 59(1), 73–80.

Sekiyama, K., Burnham, D., Tam, H., and Erdener, D., 2003. Auditory-visual speech perception development in Japanese and English speakers. In *AVSP 2003-International Conference on Audio-Visual Speech Processing*.

Sennema, A. Hazan, V., Faulkner, A., 2003. The role of visual cues in L2 consonant perception. In Proc. *15th ICPhS*, Barcelona, Spain, 135-138.

Shadle, C. H., 1985. The acoustics of fricative consonants. *RLE Technical Report 506,* Cambridge: Massachusetts Institute of Technology.

Shadle, C.H., 1990. Articulatory-acoustic relationships in fricative consonants. In Hardcastle, W. J., and Marchal, A. (Ed.), *Speech production and speech modelling.* Dordrecht, Netherlands: Kluwer, 187-209.

Shadle, C. H., Badin, P., and Moulinier, A., 1991. Towards the spectral characteristics of fricative consonants. *Proc. 12th Int. Congress of Phonetic Sciences* , 42-45.

Shadle, C. H., Mair, S.J. and Carter, J.N., 1996. Acoustic characteristics of the front fricatives [f, v, θ, ð]. Proc. *1st ESCA Tut. Res. Workshop on Speech Production Modeling*, ESCA, Austrans, 193-196.

Shadle, C. H., 2012. 2 The Aerodynamics of Speech. In Hardcastle, W. J. and Laver, J., and Gibbon, F. E. (Ed.), *The handbook of phonetic sciences*. Wiley-Blackwell, 79, 39.

Shadle, C. H., 1990 *Articulatory-acoustic relationships in fricative consonants*. Speech production and speech modelling. Springer Netherlands, 187-209.

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

Shi, L. F., 2010. Perception of acoustically degraded sentences in bilingual listeners who differ in age of English acquisition. *Journal of Speech, Language and Hearing Research*, 53(4), 821.

Shih, Y. T., and Kong, E., 2011. Perception of Mandarin fricatives by native speakers of Taiwan Mandarin and Taiwanese. In *The 23nd North American Conference on Chinese Linguistics* (NACCL-23), 110-119.

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

Skipper, J. I., Nusbaum, H. C., and Small, S. L., 2005. Listening to talking faces: motor cortical activation during speech perception. *Neuroimage*, 25(1), 76-89.

Skipper, J. I., Nusbaum, H. C., and Small, S. L., 2006. Lending a helping hand to hearing: another motor theory of speech perception. In Arbib, M. A. (Ed.), *Action to language via the mirror neuron system*, 250-285.

Skipper, J. I., Van Wassenhove, V., Nusbaum, H. C., and Small, S. L., 2007. Hearing lips and seeing voices: how cortical areas supporting speech production mediate audiovisual speech perception. *Cerebral Cortex*, *17*(10), 2387-2399.

Smith, L.C., 2001. L2 acquisition of English liquids: Evidence for production independent from perception. In Bonch-Bruevich, X. (Ed.), *The Past, Present, and Future of Second Language Research: selected proceedings of the 2000 Second Language Research Forum,* 3-22.

Soli, S. D., 1981. Second formants in fricatives: Acoustic consequences of fricative-vowel coarticulation. *The Journal of the Acoustical Society of America*, 70(4), 976-984.

Stevens, P., 1960. Spectra of fricative noise in human speech. *Language and Speech*, 3(1), 32-49.

Stevens, K. N., 1972. The quantal nature of speech: Evidence from articulatory-acoustic data. In David, E.E., Jr. and Denes P.B. (Ed.), *Human Communication: A unified view, New* York: McGraw-Hill, 51-66.

Stevens, K.N., and Blumstein, S.E., 1981. The search for invariant acoustic correlates of phonetic features. In Eimas, P. D. and Miller, J. L. (Ed.), *Perspectives on the Study of Speech.* Psychology Press, 1-38.

Stevens, K. N., Keyser, S. J., and Kawasaki, H., 1986. Toward a phonetic and phonological theory of redundant features. In Perkell, J. S., and Klatt, D. H. (Ed.), *Invariance and Variability in Speech Processes*. Psychology Press, 426–449.

Stevens, K. N., 1998. *Acoustic Phonetics.* MIT press, 398– 483.

Stockwell, R.P. and Bowen, J.D., 1983. Sound Systems in Conflict: A Hierarchy of Difficulty. In Robinett, B. W., and Schachter, J. (Ed.), *Second language learning: Contrastive analysis, error analysis and Related Aspects*. University of Michigan Press, 20-31.

Strafella, A. P., and Paus, T., 2000. Modulation of cortical excitability during action observation: A transcranial magnetic stimulation study. *NeuroReport*, 11(10), 2289-2292.

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

Suen, C. Y., 1982. Computational analysis of Mandarin sounds with reference to the English language. In *Proceedings of the 9th conference on Computational linguistics*. Academia Praha, 1, 371-376.

Sumby, W. H., and Pollack, I., 1954. Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America*, 26(2), 212-215.

Summerfield, A. Q., 1979. Use of visual information for phonetic perception. *Phonetica*, 36(4-5), 314-331.

Summerfield, Q., 1981. Articulatory rate and perceptual constancy in phonetic perception. *Journal of Experimental Psychology: Human Perception and Performance*, 7(5), 1074-1095.

Summerfield, A. Q., 1983. Audio-visual speech perception. In Lutman, M. E., and Haggard, M. P. (Ed.), *Hearing science and hearing disorders*. London: Academic Press, 131-182.

Suter, R., 1976. Predicators of Pronunciation accuracy in second language learning. *Language Learning*, 26(2), 233-53

Sweet, H., 1877. *A handbook of phonetics* (Vol. 2). Oxford: Clarendon. (Cited by Wood, S. A. J., 1993. Crosslinguistic cineradiographic studies of the temporal coordination of speech gestures. *Working Paper*, 40. Lund University, 251-63.)

Tahta S., Wood, M., and Loewenthal, K., 1981. Foreign accents: Factors relating to transfer of accent from the first language to a second language. *Language and Speech*, 24(3), 265-272.

Taylor, B.P., 1974. Toward a theory of language acquisition, *Language Learning*, 24(1), 23-35.

Taylor, I., 1976. *Introduction to psycholinguistics*. New York: Holt, Rinehart and Winston.

Tercanlioglu, L., 2005. Pre-service EFL teachers' beliefs about foreign language learning and how they relate to gender. *Electronic Journal of Research in Educational Psychology*, 3(5), 145-162.

Thompson, I., 1991. Foreign accents revised: The English pronunciation of Russian immigrants. *Language Learning.* 41(2). 174-204.

Toda, M., and Honda, K., 2003. An MRI-based cross-linguistic study of sibilant fricatives. In *Proceedings of the 6th International Seminar on Speech Production, Sydney,* 290-295.

Tsai, M. Y., and Lee, L. S., 2003. Pronunciation variation analysis based on acoustic and phonemic distance measures with application examples on Mandarin Chinese. In *Automatic Speech Recognition and Understanding, ASRU'03. 2003 IEEE Workshop*, 117-122

Tsao, F. M., Liu, H. M., and Kuhl, P. K., 2004. Speech perception in infancy predicts language development in the second year of life: a longitudinal study. *Child development*, 75(4), 1067-1084.

Tutatchikova, O. P., 1995. *Acquisition of Mandarin Chinese pronunciation by foreign learners: The role of memory in learning and teaching*. Doctoral dissertation, The Ohio State University.

Van Riper, C., 1963. *Speech correction: Principles and methods*. Englewood-Cliffs, NJ: Prentice-Hall.

Viswanathan, N., Magnuson, J. S., and Fowler, C. A., 2010. Compensation for coarticulation: Disentangling auditory and gestural theories of perception of coarticulatory effects in speech. J*ournal of experimental psychology. Human perception and performance,* 36(4), 1005-1015.

Walden, B. E., Prosek, R. A., Montgomery, A. A., Scherr, C. K., and Jones, C. J., 1977. Effects of training on the visual recognition of consonants. *Journal of Speech, Language and Hearing Research*, 20(1), 130.

Walden, B. E., Erdman, S. A., Montgomery, A. A., Schwartz, D. M., and Prosek, R. A., 1981. Some effects of training on speech recognition by hearing-impaired adults. *Journal of Speech, Language and Hearing Research*, 24(2), 207.

Wang, H. M., and Lee, L, S., 1994. Mandarin syllable recognition in continuous speech under limited training data with sub-syllabic acoustic modelling. *Journal of Computer Processing of Chinese Oriental Langauge*, 8, 1-16.

Wang, Y., Behne, D. M., and Jiang, H., 2009. Influence of native language phonetic system on audio-visual speech perception. *Journal of Phonetics*, 37(3), 344-356.

Wardhaugh, R., 1970. The contrastive analysis hypotheses. *TESOL Quarterly,* 4(2), 123-30.

Watkins, K. E., Strafella, A. P., and Paus, T., 2003. Seeing and hearing speech excites the motor system involved in speech production. *Neuropsychologia*, 41(8), 989-994.

Weinreich, U., 1953. *Language in Contact: Findings and Problems*. New York: Linguistic Circle of New York.

Weike, Zhong., 2005. 重庆方言音系研究. (The research on Chongqing dialect). *Journal of Society and Science, Chongqing*, 6.

Werker, J. F., and Tees, R. C., 1984. Phonemic and phonetic factors in adult cross-language speech perception. *The Journal of the Acoustical Society of America*, 75(6), 1866-1878.

Werker, J. F., and Logan, J. S., 1985. Cross-language evidence for three factors in speech perception. *Attention, Perception, and Psychophysics*, 37(1), 35-44.

Wieden, W., 1990. Some remarks on developing phonological representations. In Leather, J., and James, A. (Ed.), *New Sounds 90, Proceedings of the 1990 Amsterdam Symposium on the Acquisition of Second-language Speech*, Amsterdam: University of Amsterdam.

Wilde, L.F. (1995). *Analysis and Syntheses of Fricative Consonants*. Doctoral dissertation, Massachusetts Institute of Technology.

Williams, G.C., and McReynolds, L.C., 1975. The relationship between discrimination and articulation training in children with misarticulations. *Journal of Speech and Hearing Research*, 18(3), 401-412.

Williams, L., 1977. The perception of stop consonant voicing by Spanish-English bilinguals. *Attention, Perception, and Psychophysics*, 21(4), 289-297.

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

Winitz, H., and Bellerose, B., 1962. Sound discrimination as a function of pretraining conditions. *Journal of Speech, Language, and Hearing Research,* 5(4), 340-348.

Winitz, H., and Bellerose, E, B., 1963. Effects of pretraining on sound discrimination learning. *Journal of Speech Hearing Research,* 6(2), 171-180.

Winitz, H.,and Priesler, L., 1965. Discrimination pretraining and sound learning. *Perceptual and motor skills,* 20(3), 905-916.

Winitz, H., and Bellerose, B., 1967. Relation between sound discrimination and sound learning. *Journal of Communication Disorders,* 1(3), 215-235.

Winitz, H., 1969. *Articulatory acquisition and behavior*. New York: Appleton-Century-Crofts.

Winitz, H., 1985. Auditory considerations in articulation treatment. In P.W. Newman, P. W., Creaghead, N. A., and Secord, W. (Ed.), *Assessment and remediation of articulatory and phonological disorders.* Columbus, OH: CE. Merril Publishing Co. 249-358

Wode, H., 1996. Speech perception and L2 phonological acquisition. In Peter, J., and Josine, L., (Ed.), *Investigating Second Language Acquisition*. Berlin: Mouton de Gruyter, 321-353.

Wright, M., 1999. Influences on learner attitudes towards foreign language and culture. *Educational Research*, 41(2), 197-208.

Yamada, R. A., and Tohkura, Y. I., 1992. The effects of experimental variables on the perception of American English /r/ and /l/ by Japanese listeners. *Perception and psychophysics*, 52(4), 376-392.

Yamada, R.A., 1993. Effects of extended training on /r/ and /l/ identification by native speakers of Japanese. *The Journal of the Acoustical Society of America*, 93(4), 2391-2391.

Yamada, R.A., Strange, W., Magnuson, J.S., Pruitt, J.S., and Clarke , W. D., 1994. The intelligibility of Japanese speakers' production of American English /r/, /l/ and /w/, as evaluated by native speakers of American English. *Proceedings of the International Conference of Spoken Language Processing* (Acoustical Society of Japan, Yokohama), 2023-2026.

Yamada, R. A., 1995. Age and acquisition of second language speech sounds: Perception of American English /r/ and /l/ by native speakers of Japanese. In Stange, W. (Ed.), *Speech perception and linguistic experience: Issues in cross-language research*. York Press, 305-320.

Zampini, M. L., and Green, K. P., 2001. The voicing contrast in English and Spanish: The relationship between perception and production. In Nicol, J. (Ed.), *One mind, two languages*, 23-48.

Zhou, L., 2012. (On the negative transfer of Chongqing dialect to English pronunciation). *Chinese Journal of Chongqing Institute of Technology (Social Science)*, 26(12), 103-105

Geographical location of Chongqing Province [map] (2013): retrieved from http://www.google.co.uk/search?q=chongqing+map+chinaandtbm=ischandtbo=uandsource=univandsa=Xandei=Uf7sUZyxBaqf0QWum4D4DQandsqi=2andved=0CC0QsAQandbiw=1466andbih=833, assessed on 22/07/2013.