Audiovisual Training Effects on L2 Speech Perception and Production

Ying Li[1,*]

[1]School of Foreign Languages, Southwest University of Political Science and Law, China

*Correspondence: School of Foreign Languages, Southwest University of Political Science and Law, China. E-mail: liying_22@163.com

Received: March 7, 2016 Accepted: March 17, 2016 Online Published: April 17, 2016

doi:10.5430/ijelt.v3n2p14 URL: http://dx.doi.org/10.5430/ijelt.v3n2p14

Abstract

This study investigated whether audiovisual perception training can improve learners' auditory perception of L2 speech sounds. 29 subjects (experimental group) who had difficulty in the perception of English /θ/-/s/ and /ð/-/z/ were recruited to attend a 9-session audiovisual perception training programme with identification tasks on the target contrasts. Their perception performance was tested before, during and at the end of the training with an AXB task. A control group was tested with the same AXB task and intervals as that of the experimental group. The results showed that the experimental group's perception accuracy improved substantially during and by the end of the training programme. The control group also showed perception improvement across the pre-test and post-test. Their degree of improvement, however, was significantly lower than that of the experimental group. These results therefore confirm the value of the audiovisual modality in L2 speech perception training.

Keywords: *audiovisual; speech perception; phonetic training; L2*

1. Introduction

1.1 Factors Affecting L2 Speech Perception

The research on second language (L2) speech perception has found that there are many factors affecting the acquisition of L2 speech. The age factor and language learners' first language (L1) experience has usually been given great attention with regard to the learners' ultimate L2 achievement. Specifically, younger L2 learners were shown to have more advantages than older ones. Their advantages show first in the brain system of youth 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). 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, 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). The validity of CPH may be also compromised since there is no consensus on when the "critical period" ends. It was found that the "critical period" may be ended at different ages in different linguistic domains. In speech perception, for example, 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).

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 similarities between learners' L1 and L2 systems facilitate their L2 learning, whereas 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. More supporting findings can be found from early studies on L2 acquisition (i e., Robinett and Schachter, 1983; Banathy, Trager, and Waddle, 1966; Berger, 1952; Lado, 1957). 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. In the first situation, listeners tend to equate an L1 sound to a correlated L2 sound on a phonological level, despite their phonetic difference(s). For example, native English listeners are likely to equate the French /r/ with the English /ɹ/. In the second situation two L2 sounds may be perceived as the same phonetic category, yet one is perceived as a better exemplar than the other. Third, two L2 speech sounds are assimilated into a single L1 phonetic category, which is the most difficult situation for listeners. The last situation occurs when an L2 sound can be perceived without being assimilated to the listener's L1 category, meaning no assimilation occurs. 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, 1991b, 1992a, b, 1995a, 2003) Speech Learning Model (SLM) discusses the influence of language learners' L1 on their perception of L2 speech sounds. SLM proposed an extensive set of assumptions and hypotheses. Here are four of the essential hypotheses that are relevant to the present study. (1) Learners' capacity in speech learning remains intact throughout their life. 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. (2) The amount of language experience plays a significant role in language learners' perception of L2 speech sounds. That is, greater L2 experience serves to enhance language learners' capability for the perception of L2 speech sounds. (3) The perceived similarity or phonetic space between L1 and L2 phonetic categories determines whether a new L2 category can be formed. On this point, 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.

1.2 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. It was found to be able to facilitate language learners' perception of L2 speech sounds (Hirata and Kelly, 2010; Navarra and Soto-Faraco, 2007; Chen, 2001; Walden, Prosek, Montgomery, and Scherr, 1977; Sumby and Pollack, 1954). The McGurk effect (McGurk and MacDonald, 1976) can be seen as one of the embodiments of the influence of articulatory gestures on speech perception. According to McGurk effect, 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). Moreover, successful lip reading training studies on speech perception have further illustrated the facilitating role of visual articulatory information in speech perception (Walden et al., 1977; Walden et al, 1981).

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). About the age factor, 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(Note 1) and whether it is a tone language(Note 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).

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 (Wang et al., 2009). Hazan et al., (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).

The present study aimed to reveal whether audiovisual speech perception training could benefit adult L2 learners' auditory perception of L2 speech sounds. Based on these hypotheses and findings discussed above, it was predicted that the subjects' auditory perception accuracy could be enhanced with audiovisual perception training. 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.

In addition to the factors discussed above, individual variances in gender (Reid, 1987; Powell and Baters, 1985; Kaylani, 1996; Oxford, Nyikos and Ehrman, 1988; Oxford, 1993), motivations (Gardner, 1985; Taylor, 1974; Lenneberg, 1967; MacNamara, 1973), 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 et al., 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 (Bradlow et al., 1997; Grant and Seitz, 1998; Hazan et al., 2005; Bernstein et al., 2013).

1.3 The Relation of Audiovisual Integration, Auditory and visual Skills in Speech Perception

As mentioned above, McGurk effect 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). Grant and Seitz (1998) claims 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. Similar findings can be found from DiStefano (2010), James (2009) and Gariety (2009).

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 et al., (1991) reported that visual codes of articulatory information have an entry in the auditory cortex. Similar evidence is available from Calvert et al. (1997), 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 was further illustrated by Sato et al., (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.

1.4 Articulatory and Acoustic Features of English /s, θ, z, ð/, Mandarin /s/, and CQd /s, z/

The subjects of the present study came from Chongqing—a municipality located in the southwest part of China. They speak Mandarin as their L1 and Chongqing dialect (CQd thereafter) as their L1-dialect. CQd is one of the Mandarin dialects (Ramsey, 1987). According to the consonant phonetic inventories of Englissh, Mandarin and CQd, neither /θ/ nor /ð/ exists in Mandarin and CQd. /z/ occurs in the phonetic inventories of English and CQd but not in Mandarin. Only /s/ exists in English, Mandarin and CQd.

Table 1. Articulatory Gestures of /θ, ð, s, z/ in English, Mandarin and CQd

	English	Mandarin	CQd
/θ, ð/	interdental (Prator and Robinet, 1985; Want et al., 2009); dental (Taylor, 1976; Ladeforged, 1996).	N/A	N/A
/s/	alveolar (Toda and Honda, 2003); dental (Taylor, 1979; Ladefoged, 1996)	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)	apical dental (Zhong, 2005; Zhou, 2012)
/z/	alveolar (Toda and Honda, 2003); dental (Taylor, 1979; Ladefoged, 1996)	N/A	apical dental (Zhong, 2005; Zhou, 2012)

Table 2. Acoustic features of /θ, ð, s, z/ in English, Mandarin and CQd (data was adopted from Stevens (1960); From Kent and Read (2002: 182); Pickett (1999: 140); Behrens and Blumstein (1988), Li (unpublished)).

Acoustic cues	English /θ, ð/	/s,z/ in English CQd/Mandarin
Frequency range	1.5-8.5 kHz	above 4 kHz
Strongest frequency range	around 5 kHz	at and avove 4 kHz
Amplitude range	around 54 dB	around 65 dB
Inherent duration	110 ms	125 ms
Vowel transition	downward F2	no vowel transition
Relative intensity	low	high
Relative Effective Spectra Length	relatively long	relatively short
Spectra shape	relatively flat spectrum with no clear dominating peak	well-defined, distinct shape with a primary spectra peak in high frequencies

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). Table 4 displays the articulatory features of /θ, ð, s, z/ in English, Mandarin and CQd.

On acoustic level, as shown in Table 2, 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/.

In order to demonstrate the visible differences between /θ, ð/ and /s, z/ in terms of articulatory gestures, RP (received pronunciation) 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.

2. Methodology

The study included a 9-session audiovisual training programme. The stimuli and training process were designed according to the principles of High Variability Phonetic Training (HVPT thereafter)—"natural" and "variability" (Lively et al., 1993; Logan et al., 1993; Bradlow et al., 1997). HVPT is predicted to be able to direct language learners' attention towards relevant phonetic cues by providing them with stimuli of high-variability in different phonetic contexts.

2.1 Subjects

All the subjects were undergraduates of different majors at a university in China. Their accuracy in the perception of /θ/-/s/ and /ð/-/z/ was below 70%(Note 3). They all spoke Mandarin as L1, CQd as L1-dialect, English as L2. The experimental group included 29 subjects (mean age=20.03; mean OA=13.41; male=14; female=15). The control group were another 20 undergraduates (mean age= 20.50; OA=12.92; male=10; female=10).

2.2 Stimuli and Recording

Stimuli for perception tests: the stimuli were nonsense words that contained /θ/-/s/ and /ð/-/z/ in initial, medial, and final positions of vowel contexts /i, a, u/. The syllable structures were VC, VCV and CV, which were counter-balanced. Each stimulus was repeated 3 times, thus yielding to 108 stimuli for each contrast (interstimulus interval=1,000ms). Moreover, the order of the items was automatically randomized with the Praat program (Boersma and Weenink, 2013), so as to avoid bias. The stimuli were auditorily recorded by a female and a male RP speaker. The recording was carried out in a soundproof booth with Roland-05 recorder (settings: 16-bit mono channel, sampling frequency=44.1KHz).

Stimuli for audiovisual training: 9 sessions of minimal pairs. 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). For instance, in *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 stimuli were auidovisually recorded by 3 RP speakers in a quite room with a digital DVD camera and Roland-09 recorder. 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, so the subjects could observe the speakers' mouth movements. The recordings were then transferred to a computer. To obtain a high quality recording, the sound recorded by the DVD camera was erased. The video channel and the auditory recording were synchronised. The recordings were cut and merged (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 1 and 2 below. Providing the subjects with immediate feedback helps hold and increase their attention during the training process (McGuire, 2010).

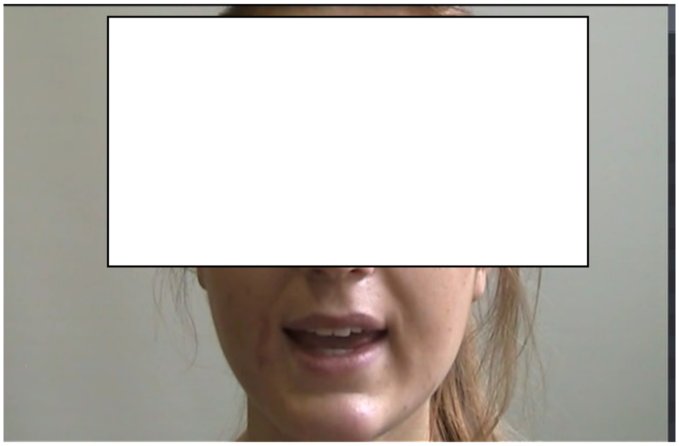


Figure 1. Screenshot of the First-Time Production of the “minimal pair” *A. sink B. think* from RP2 in Training Session 2, 5, 8

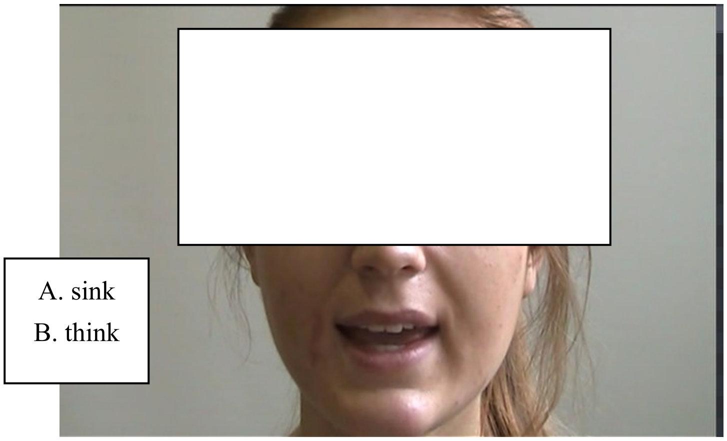


Figure 2. Screenshot of the Second-Time Production of the “Minimal Pair” *A. sink B. think* from RP2 in Training Session 2, 5, 8

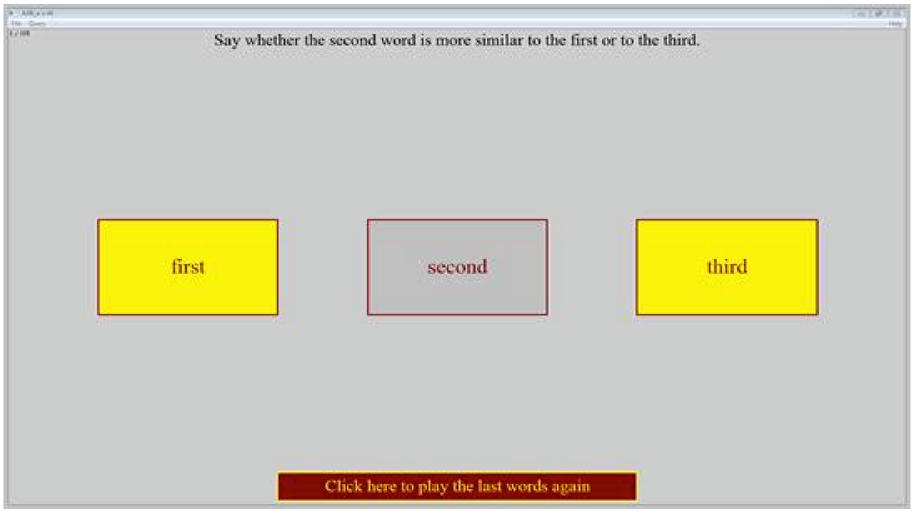
2.3 Procedure

Before training, the subjects were asked to fill in a questionnaire to collect their personal information on L2-English study. Table 3 shows the schedule of the programme. The experimental group experienced the audiovisual training sessions, whereas the control group only experienced the perception tests. The training sessions were carried out in a quite room. 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 via the headphone at a comfortable listening level (65—70dB), and visually presented via the monitor (33*20 cm) in front of them. Instructions were given in Mandarin. All the desktops were connected to and controlled by the “central computer”, which was controlled by the investigator. The subjects were asked to do 2AFC tasks. For each trial, the subjects were asked to circle on an answer sheet concerning 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 2), so that the subjects could check their answers.

Table 3. The Schedule of the Programme

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)

All the subjects' perception performance was tested before, during and after the training programme with an AXB discrimination test. The test was presented with Praat Program (Boersma and Weenink, 2013). They 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 mouse-clicking 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 3 below). After clicking on this button, the trial is played again. The subjects' responses were automatically recorded by the Praat program (Boersma and Weenink, 2013). Any subjects who achieved an accuracy of 90% or above in the perception of the target contrasts in mid-test 1 or mid-test2 were dropped from the following training sessions and test(s), because they were assumed not to need further training. The same principle was applied to mid-test-2.



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

Figure 3. Screenshot of the AXB Test

3. Results

3.1 Overall Results

To avoid potential bias in the subjects' responses, the subjects' perception accuracy was converted to *d-prime*(Note 4) scores (the signal detectability measure *d-prime*) with SPSS and Excel.

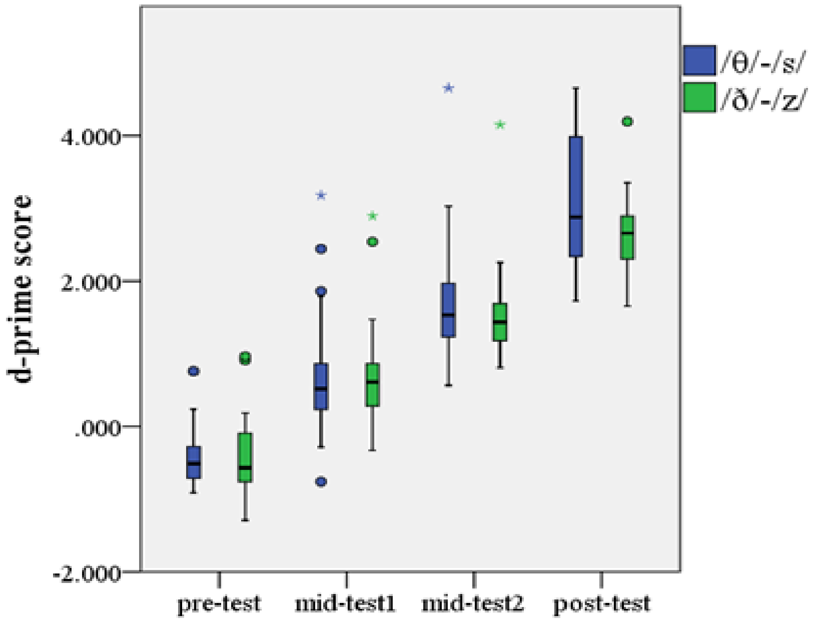


Figure 4. Boxplots of the Experimental Group's d' Scores in the Pre-Test, Mid-Test 1, Mid-Test 2 and the Post-Test

Table 4. The Experimental Group's Calculated *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
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

As shown in Table 4 and Figure 4, the experimental group showed linear perception improvement from pre-test to post-test. The majority of the subjects' *d'* scores were negative in the pre-test. In mid-test 1, however, except for S1's (S=subject thereafter) accuracy in the perception of /θ/-/s/ and S15's accuracy in the perception of the two contrasts, the remaining subjects' *d'* scores were all above 0. S10 achieved the highest accuracy in the perception of /θ/-/s/ (94.44%, *d'*=3.18) and /ð/-/z/ (92.59%, *d'*=2.19). Therefore, S10 was dropped from the following training and tests.

After the 6th training session, the perception performance of the remaining 28 subjects' accuracy was further improved in mid-test 2. S3 achieved the highest accuracy among the remaining subjects: 98.15% (*d'*=4.65) in the perception of /θ/-/s/ and 100% (*d'*=4.15) in the perception of /ð/-/z/. Therefore, S3 was dropped from the following training and test. At the end of the training programme, the remaining subjects received further improvement in post-test. Specifically, compared with that in mid-test2, they displayed comparatively higher minimum, maximum, and mean scores as well as the range of the majority of the subjects' accuracy in post-test.

A *repeated-measures ANOVA* analysis revealed that the experimental group's 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*). A further *Post Hoc Test* revealed that 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.

Table 5. *Post Hoc Tests* for the Experimental Group's Perception Performance in the Four AXB Tests

(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-test 2	Pre-test	2.221	2.04	0.202	0.17	0	0	1.82	1.7	2.622	2.38
	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

Due to the same stimuli were used in the AXB tests for 4 times, though with different orders, there might be repeated testing effect in the experimental group's perceptual progress. To detect this issue, subjects in the control group were tested 4 times with the same intervals as that of the experimental group, yet without being trained.

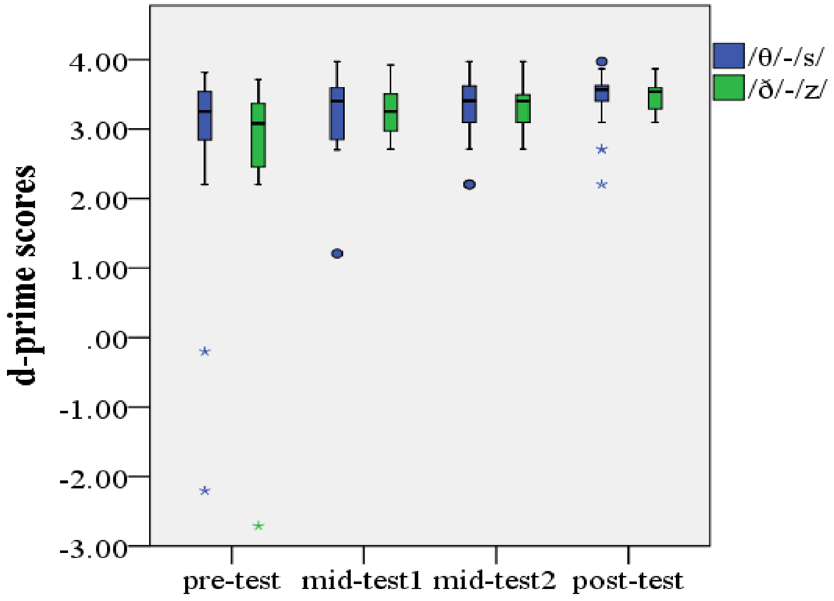


Figure 5. Boxplots of the Control Group's *d'* Scores in the 4 Tests

Table 6. The Control Group's *d'* Scores in the 4 Tests

		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

As displayed in Figure 5 and Table 6, similar to that of the experimental group, the control group showed some degree of improvement from pre-test to post-test. However, detailed examination of individual subjects' *d'* scores revealed that some subjects' accuracy did not improve linearly from pre-test to post-test. For instance, S33, S37, S40 and S41's *d'* scores decreased in mid-test 1. Moreover, A *repeated-measures ANOVA* analysis indicated that 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.

3.2 Factors Had Significant Effect on the Experimental Group's Performance

Table 7. Data Collected from the Questionnaire (Experimental Group)

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%

Data collected from the questionnaire indicated that 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.

A *repeated-measures ANOVA* was conducted to detect which factor(s) may have had a significant impact on their perception performance. The subjects *d'* scores were coded as the dependent variable. *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. 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)*.

Table 8. Factors of Significant Effect on the Experimental Group's Perception of /θ/-/s/ in the 4 Tests

factor	df and F-value	Sig.	Partial Eta Squared
training	F(1, 25)=127.262	p<0.001	η^2=0.853
gender	F(1, 25)=154.289	p<0.001	η^2=0.861
training * gender	F(1, 25)=5.266	p=0.030	η^2=0.714
phonetic position	F(2, 50)=6.911	p=0.002	η^2=0.855
phonetic position * training	F(6, 156)=2.339	p=0.034	η^2=0.083

Table 9. Factors of Significant Effect on the Experimental Group's Perception of /ð/-/z/ in the 4 Tests

factor	df and F-value	Sig.	Partial Eta Squared
training	F(3, 75)=90.317	p<0.001	η^2=0.749
gender	F(1, 25)=233.281	p<0.001	η^2=0.903
training * gender	F(1, 25)=3.458	p=0.029	η^2=0.657
phonetic position	F(2, 50)=34.346	p<0.001	η^2=0.579
phonetic position * training	F(6, 156)=3.477	p=0.003	η^2=0.122

According to the analysis results, *training* was found to have displayed a significant effect on the subjects' perception of the target contrasts. A further *Post Hoc Test* results indicated that the more training sessions the subjects received, the higher *d'* scores they received in the perception of the target contrasts (see Table 10).

Table 10. *Post Hoc Tests* for the Experimental Group's Perception Performance in the Four AXB Tests

(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-test 2	Pre-test	2.221	2.04	0.202	0.17	0	0	1.82	1.7	2.622	2.38
	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

Moreover, *gender and its* interaction with *training* were also shown to have a significant effect on their perception performance. As show in Table 11, there was not a significant difference between the experimental and control group's accuracy in pre-test. With the training sessions carried out, the male subjects performed better than the female subjects.

Table 11. The Female and Male Subjects of the Experimental Group's Mean *d'* Scores in the Perception of /θ/-/s/ and /ð/-/z/

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

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 12), 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 13), the mean difference between initial and medial positions was revealed to be statistically 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.

Table 12. *Post Hoc Tests* of the Experimental Group's Perception of /θ/-/s/ in Different Phonetic Positions

(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 13. *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

As shown in Table 14, the rest of the factors were found to have no significant effect on the experimental group's perception performance in the 4 tests.

Table 14. Factors Which Were Statistically **Non-significant** for the Experimental Group's Perception of /θ/-/s/ and /ð/-/z/

factor	df and F-value		Sig.		Partial Eta Squared	
Vowel context	/θ/-/s/ F(2, 50)=0.003	/ð/-/z/ F(2, 50)=0.528	/θ/-/s/ p=0.997	/ð/-/z/ p=0.593	/θ/-/s/ η^2=0.003	/ð/-/z/ 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

3.3 Factors Had Significant Effect on the Control Group's Performance

Table 15. Data Collected from the Questionnaire (Control Group)

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%

In the selection of control group, it was designed to select subjects of a similar profile to that of the experimental group. The control group, therefore, has similar or the same age, OA of L2-English learning, etc. as that of the experimental group. Thus, only *gender difference* was adopted as a between-subjects factor. *Repeated testing experience* (*experience*, hereafter) was coded as a within-subjects factor to detect whether there was a repeated testing effect. Given that each subject was tested 4 times with equal intervals, 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)*.

According to the results, *experience* was revealed to have had a significant effect on the control group's perception of /θ/-/s/ (F(3, 54)=3.884, p=0.014, η^2=0.177) and /ð/-/z/ (F(3, 54)=2.872, p=0.045, η^2=0.138). In other words, there was a repeated testing effect on the subjects' perception performance. Further *Post Hoc Tests* results indicated that the repeated testing effect was not significant until mid-test 2 (see Table 16). That is, repeated testing effect had significantly facilitated the subjects' perception improvement in mid-test 2 and post-test.

Table 16. *Post Hoc Tests* of the Control Group's Perception Performance in the Fours AXB Tests

(I) experience	(J)experience	Mean Difference (I-J)	Std. Error	Sig.
no experience	experience 1	-0.516	0.269	0.071
	experience 2	-0.564	0.261	**0.045**
	experience 3	-0.732	0.3	**0.025**
experience 1	no experience	0.516	0.269	0.071
	experience 2	-0.048	0.063	0.456
	experience 3	-0.216	0.055	**0.001**
experience 2	no experience	0.564	0.261	0.045
	experience 1	0.048	0.063	0.456
	experience 3	-0.168	0.064	**0.017**
experience 3	no experience	0.732	0.3	0.025
	experience 1	0.216	0.055	0.001
	experience 2	0.168	0.064	0.017

Moreover, *phonetic position* was found to be highly significant for the control group's perception performance. According to the *Post Hoc Tests* results, the control group did not display a significant difference in the perception of the target contrasts in initial and medial positions ($p>0.05$). However, 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 ($p<0.05$). Specifically, they were more likely to have accurate perception when the contrasts were embedded in initial and medial positions than in final position.

The rest of the factors and their interaction with one another were revealed to be non-significant for the control group's perception performance. ($p>0.05$). Furthermore, the control group's difference between the perception of the voiced and voiceless contrast was statistically non-significant ($F(1, 38)=0.049, p=0.826, \eta^2=0.001$).

4. Discussion

4.1 Audiovisual Training Effect on the Subjects' Auditory Perception 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 of the target contrasts was significantly improved by the post-test. 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 introduction section.

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 pre-test results, the majority of the subjects had serious difficulty in the discrimination of /θ/ and /s/, /ð/ and /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 in the perception of /θ/-/s/, /ð/-/z/. *Training* was detected to be a factor that had significantly affected their perception improvement.

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 perception accuracy was because they perceived the acoustic differences between /θ/ and /s/, /ð/ and /z/. 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 performance may support this hypothesis. It may be because the training programme was not very long that the experimental group in the perception of the target contrasts did not achieve native-like performance (none of them received full scores in the after-training tests). If given further training, the results might be more satisfying.

2. SLM—Speech Learning Model

The experimental group's improved accuracy after training may provide supporting evidence for some of the hypotheses of SLM (Flege, 1981, 1987, 1988, 1991a, b, 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 pre-test with that in post-test, the experimental group's perception accuracy 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 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 section 1.4, 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, due to there was repeated testing effects, the experimental group's perception improvement cannot be totally attributed to the training effect. Therefore, it was 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.

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* indicated 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 there was repeated testing effects, the degree of the control group's improvement was significantly lower than that of the experimental group. Thus, it might be possible to speculate that the experimental group's substantial improvement is largely attributed to the L2 experience from the audiovisual training programme.

3. NLM/NLL-e, and PI

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. Similarly, 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). In the present study, the subjects initially had difficulty in the discrimination of /θ/-/s/ and /ð/- /z/. Their substantial improvement at the end of the training programme is consistent with this prediction

4. CPH—Critical Period Hypothesis

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, 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.

5. 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 performance in the pre-test 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 perception accuracy 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 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.

4.2 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, despite there was some repeated testing effects. 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. 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 may be mitigated.

Moreover, given that the experimental group's perception improvement, to a large extent, can be attributed to the audiovisual training. It 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 of the target contrasts by 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. However, after being audiovisually trained, the subjects' accuracy in the perception of the target contrasts substantially increased from the pre-test to tests after training. 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).

4.3 Factors That Significantly Affect the Experimental Group's Perception Performance

According to the statistical analysis results, 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.

Regarding gender difference, it was found that the male subjects in the experimental group showed better perception performance than the female subjects in after trained tests. 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). 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, 1958; 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 of the target contrasts.

As for *phonetic position*, the experimental group were found had better perception performance when 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. 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. To apply to the present study, 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.

5. Conclusion

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. The motivation was to provide further evidence in support of the significance of articulatory information in speech perception. Findings of the present study confirmed the facilitating role of audiovisual training in auditory speech perception.

References

Ausubel, D.P. (1964). Adults versus children in second-language learning: Psychological considerations. *The Modern Language Journal*, *48*(7), 420-424. http://dx.doi.org/10.1111/j.1540-4781.1964.tb04523.x

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., & Blumstein, S. E. (1988). Acoustic characteristics of English voiceless fricatives: A descriptive analysis. *Journal of Phonetics*, *16*, 295-298. http://dx.doi.org/10.3389/fnins.2013.00034

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

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

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. (1995b). 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., & 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., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In Bohn, O. S, & Munro, M. J. (Ed.), *Language experience in second language speech learning: In honor of James Emil Flege*. John Benjamins Publishing, 13-34. http://dx.doi.org/10.1075/lllt.17.07bes

Best, C. T. (1995a). 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.

Bialystok, E., & 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.

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/

Bouchard Jr, T. J., & 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. http://dx.doi.org/10.1080/19485565.1977.9988304

Bradlow, A., Pisoni, D., AkahansYamada, R., & Tohkura, Y. I. (1997). Training Japanese listeners to identify English/r/and /l/: IV. Some effects of perceptual learning on speech production. *Journal of the Acoustical Society of America*, *101*(4), 2299-23. http://dx.doi.org/10.1121/1.418276

Calvert, G. A., Bullmore, E. T., Brammer, M. J., Campbell, R., Williams, S. C., McGuire, P. K., & David, A. S. (1997). Activation of auditory cortex during silent lipreading. *Science*, *276*(5312), 593-596. http://dx.doi.org/10.1126/science.276.5312.593

Chen, T. (2001). Audiovisual speech processing. *IEEE Signal Processing Magazine*, 9–31. http://dx.doi.org/10.1109/79.911195

DiStefano, S. (2010). *Can audio-visual integration improve with training?* Senior Honors Theses, The Ohio State University.

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. http://dx.doi.org/10.1111/j.1540-4781.1995.tb01112.x

Ellis, R. (1985). *Understanding second language acquisition*. Oxford: Oxford U. P.

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. (1981). The phonological basis of foreign accent: A hypotheses. *TESOL Quarterly, 15*(4), 443-455. http://dx.doi.org/10.2307/3586485

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. 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. http://dx.doi.org/10.1121/1.398599

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. http://dx.doi.org/10.1075/lald.2.15fle

Flege, J. E. (1991b). The interlingual identification of Spanish and English vowels: Orthographic evidence. *Quarterly Journal of Experimental Psychology*, *43*(3), 701-731. http://dx.doi.org/10.1080/14640749108400993

Flege, J. E. (1992a). Speech learning in a second language. In Ferguson, C., Menn, L., & 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. (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. (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. http://dx.doi.org/10.1515/9783110895094.319

Flege, J. E., & 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. http://dx.doi.org/10.1121/1.402780

Fullana, N., & 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.

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., & Vroomen, J. (1992). Auditory and visual speech perception in alphabetic and non-alphabetic Chinese-Dutch bilinguals. *Advances in psychology*, *83*, 413-426. http://dx.doi.org/10.1016/S0166-4115(08)61508-3

Ghazanfar, A. A., & Schroeder, C. E. (2006). Is neocortex essentially multisensory? *Trends in cognitive sciences*, *10*(6), 278-285. http://dx.doi.org/10.1016/j.tics.2006.04.008

Goldstein, D., Haldane, D., & Mitchell, C. (1990). Sex differences in visual-spatial ability: The role of performance factors. *Memory and Cognition*, *18*(5), 546-550. http://dx.doi.org/10.3758/BF03198487

Grant, K.W., & 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. http://dx.doi.org/10.1121/1.423751

Hardison, D. M. (2003). Acquisition of second-language speech: Effects of visual cues, context, and talker variability. *Applied Psycholinguistics*, *24*(4), 495-522. http://dx.doi.org/10.1017/S0142716403000250

Hardison, D. M. (2005a). Second-language spoken word identification: Effects of perceptual training, visual cues, and phonetic environment. *Applied Psycholinguistics*, *26*(4), 579-596. http://dx.doi.org/10.1017/S0142716405050319

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. http://dx.doi.org/10.1016/j.specom.2005.02.002

Harris, K. S. (1958). Cues for the discrimination of American English fricatives in spoken syllables. *Language and speech*, *1*(1), 1-7.

Hazan, V., Sennema, A., Faulkner, A., Ortega-Llebaria, M., Iba, M., & Chung, H. (2006). The use of visual cues in the perception of non-native consonant contrasts. *The Journal of the Acoustical Society of America*, *119*, 1740-1751. http://dx.doi.org/10.1121/1.2166611

Hazan, V., Sennema, A., Iba, M., & Faulkner, A. (2005). Effect of audiovisual perceptual training on the perception and production of consonants by Japanese learners of English. *Speech communication*, *47*(3), 360-378. http://dx.doi.org/10.1016/j.specom.2005.04.007

Hirata, Y., & 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. http://dx.doi.org/10.1044/1092-4388(2009/08-0243)

Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y. I., Kettermann, A., & Siebert, C. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition*, *87*(1), 47-57. http://dx.doi.org/10.1016/S0010-0277(02)00198-1

James, K. (2009). *The effects of training on intelligibility of reduced information speech stimuli*. Senior Honors Theses, The Ohio State University.

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., & Read, C. (2002). *The Acoustic Analysis of Speech.* San Diego, Calif: Singular Pub. Group.

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. http://dx.doi.org/10.1037/10119-012

Kuhl, P. K. (1994). Learning and representation in speech and language. *Current option in neurobiology*. *4*(6), 812-822. http://dx.doi.org/10.1016/0959-4388(94)90128-7

Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., & 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. http://dx.doi.org/10.1098/rstb.2007.2154

Kuhl, P. K., Tsao, F. M., & Liu, H. M. (2003). Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning. *Proceedings of the National Academy of Sciences*, *100*(15), 9096-9101. http://dx.doi.org/10.1073/pnas.1532872100

Lado, R. (1957). *Linguistics Across Cultures.* Ann Arbor: University of Michigan Press.

Lenneberg, E. H. (1967). *Biological Foundations of Language*. New York: Wiley and sons.

Lively, S. E., Logan, J. S., & Pisoni, D. B. (1993). Training Japanese listeners to identify English/r/and/l/. II: The role of phonetic environment and talker variability in learning new perceptual categories. *The Journal of the Acoustical Society of America*, *94*, 1242. http://dx.doi.org/10.1121/1.408177

Logan, J. S., Lively, S. E., & 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. http://dx.doi.org/10.1121/1.1894649

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.

Massaro, D. W., Thompson, L. A., Barron,B., & Laren, E. (1986). Developmental changes in visual and auditory contributions to speech perception. *Journal of Experimental Child Psychology*, *41*(1), 93-113. http://dx.doi.org/10.1016/0022-0965(86)90053-6

Mayo, C., & 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. http://dx.doi.org/10.1121/1.1738838

Mayo, L. H., Florentine, M., & 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. http://dx.doi.org/10.1044/jslhr.4003.686

McGuire, G. (2010). A brief primer on experimental designs for speech perception research. *Laboratory Report*, *77*.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*, 746-748. http://dx.doi.org/10.1038/264746a0

Munro, M. J., & Derwing, T. M. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, *45*(1), 73-97. http://dx.doi.org/10.1111/j.1467-1770.1995.tb00963.x

Nath, A. R., & Beauchamp, M. S. (2011). A neural basis for interindividual differences in the McGurk effect, a multisensory speech illusion. *NeuroImage, 59*(1), 781-787. http://dx.doi.org/10.1016/j.neuroimage.2011.07.024

Navarra, J., & 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. http://dx.doi.org/10.1007/s00426-005-0031-5

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.

Oxford, R., Nyikos, M., & Ehrman, M. (1988). Vive la difference? Reflections on sex differences in use of language learning strategies. *Foreign Language Annals*, *21*(4), 321-329. http://dx.doi.org/10.1111/j.1944-9720.1988.tb01076.x

Oyama, S. (1976). A Sensitive Period for the Acquisition of a Nonnative phonological System. *Journal of Psycholinguistic Research*, *5*(3), 261-285. http://dx.doi.org/10.1007/BF01067377

Patkowski, M. S. (1990). Age and accent in a second language: A reply to James Emil Flege. *Applied linguistics*, *11*(1), 73-89. http://dx.doi.org/10.1093/applin/11.1.73

Penfield, W., & Roberts, L. (1959). *Speech and Brain Mechanisms.* Princeton, NJ: Princeton University Press.

Pickett, J. M. (1999). *The acoustics of speech communication*. Boston: Allyn and Bacon.

Piske, T., MacKay, I. R., & Flege, J. E. (2001). Factors affecting degree of foreign accent in an L2: A review. *Journal of phonetics*, *29*(2), 191-215. http://dx.doi.org/10.1006/jpho.2001.0134

Pisoni, D. B. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Attention, Perception, and Psychophysics*, *13*(2), 253-260. http://dx.doi.org/10.3758/BF03214136

Powell, R.C., & Baters, J. D. (1985). Pupils' perceptions of foreign language learning at 12+: some gender difference. *Educational Studies*, *11*(1), 11-23. http://dx.doi.org/10.1080/0305569850110102

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.

Reid, J. M. (1987). The learning style preferences of ESL students. *TESOL quarterly*, *21*(1), 87-111. http://dx.doi.org/10.2307/3586356

Robinett, B. W., & Schachter, J. (Ed.) (1983). *Second language learning: Contrastive analysis, error analysis, and related aspects*. Ann Arbor, MI: University of Michigan Press.

Sams, M., Aulanko, R., Hämäläinen, M., Hari, R., Lounasmaa, O. V., Lu, S. T., & Simola, J. (1991). Seeing speech: visual information from lip movements modifies activity in the human auditory cortex. *Neuroscience letters*, *127*(1), 141-145. http://dx.doi.org/10.1016/0304-3940(91)90914-F

Sanders, B., Soares, M. P., & D'Aquila, J. M. (1982). The sex difference on one test of spatial visualization: A nontrivial difference. *Child Development*, 1106-1110. http://dx.doi.org/10.2307/1129153

Sato, M., Troille, E., Ménard, L., Cathiard, M. A., & Gracco, V. (2013). Silent articulation modulates auditory and audiovisual speech perception. *Experimental Brain Research*, *227*(2), 275-288. http://dx.doi.org/10.1007/s00221-013-3510-8

Schwartz, J. L., Basirat, A., Ménard, L., & 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. http://dx.doi.org/10.1016/j.jneuroling.2009.12.004

Scovel, T. (1988). *A Time to Speak. A Spycholinguistic Inquiry into the Critical Period for Human Speech.* Rowley, MA: Newbury House.

Sekiyama, K. (1997). Cultural and linguistic factors in audiovisual speech processing: The McGurk effect in Chinese subjects. *Perception and Psychophysics, 59*(1), 73–80. http://dx.doi.org/10.3758/BF03206849

Sekiyama, K., & Tohkura, Y. (1993). Inter-language differences in the influence of visual cues in speech perception. *Journal of Phonetics, 21*(4), 427–444.

Sennema, A., Hazan, V., & Faulkner, A. (2003). The role of visual cues in L2 consonant perception. In Proc. *15th ICPhS*, Barcelona, Spain, 135-138.

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. http://dx.doi.org/10.1044/1092-4388(2010/09-0081)

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

Stevens, P. (1960). Spectra of fricative noise in human speech. *Language and Speech*, *3*(1), 32-49.

Sumby, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America*, *26*(2), 212-215. http://dx.doi.org/10.1121/1.1907309

Taylor, B.P. (1974). Toward a theory of language acquisition. *Language Learning*, *24*(1), 23-35. http://dx.doi.org/10.1111/j.1467-1770.1974.tb00233.x

Toda, M., & 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.

Walden, B. E., Erdman, S. A., Montgomery, A. A., Schwartz, D. M., & 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. http://dx.doi.org/10.1044/jshr.2402.207

Walden, B. E., Prosek, R. A., Montgomery, A. A., Scherr, C. K., & Jones, C. J. (1977). Effects of training on the visual recognition of consonants. *Journal of Speech, Language and Hearing Research*, *20*(1), 130. http://dx.doi.org/10.1044/jshr.2001.130

Wang, Y., Behne, D. M., & Jiang, H. (2009). Influence of native language phonetic system on audio-visual speech perception. *Journal of Phonetics*, *37*(3), 344-356. http://dx.doi.org/10.1016/j.wocn.2009.04.002

Wardhaugh, R. (1970). The contrastive analysis hypotheses. *TESOL Quarterly, 4*(2), 123-30. http://dx.doi.org/10.2307/3586182

Weinreich, U. (1953). *Language in Contact: Findings and Problems*. New York: Linguistic Circle of New York.

Werker, J. F., & Logan, J. S. (1985). Cross-language evidence for three factors in speech perception. *Attention, Perception, and Psychophysics*, *37*(1), 35-44. http://dx.doi.org/10.3758/BF03207136

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.

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. http://dx.doi.org/10.1121/1.406052

Notes

Note 1. The number of visemes in their phonetic inventory refers to the number of identifiable ‘visual categories’ which are pronounced with visual movements.

Note 2. Tone information is not visually observable.

Note 3. A pilot study of perception test was carried out to select subjects who had difficulty with the perception of /θ/-/s/ and /ð/-/z/. The testing task, stimuli, and procedure were the same as that used in the perception test of the present study.

Note 4. 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.

Appendix

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/