Spoken English Discrimination (SED) Training with Multilingual Malaysians: Effect of Adaptive Staircase Procedure and Background Babble in High Variability Phonetic Training

Christine Leong Xiang Ru

Thesis submitted to the University of Nottingham

for the degree of Doctor of Philosophy

December 2016

Abstract

High variability phonetic training (HVPT) has been shown to improve non-native speakers' perceptual performance in discriminating difficult second language phonemic contrasts (Bradlow, Akahane-Yamada, Pisoni, & Tohkura, 1999; Bradlow, Pisoni, Akahane-Yamada, & Tohkura, 1997; Lively, Logan, & Pisoni, 1993; Lively, Pisoni, Yamada, Tohkura, & Yamada, 1994; Logan, Lively, & Pisoni, 1991). The perceptual learning can be generalized to novel words (Wang & Munro, 2004), novel speakers (Nishi & Kewley-Port, 2007; Richie & Kewley-Port, 2008) and even to speech production (Bradlow et al., 1997). However, the rigidity of the laboratory training settings has limited applications to real life situations.

The current thesis examined the effectiveness of a new phonetic training program - the Spoken English Discrimination (SED) training. SED training is a computerized individual training program designed to improve non-native speakers' bottom-up perceptual sensitivity to discriminate difficult second language (L2) phonemic contrasts. It combines a number of key training features including 1) natural spoken stimuli, 2) highly variable stimuli spoken by multiple speakers, 3) multi-talker babble as background noise and 4) an adaptive staircase procedure that individualizes the level of background babble.

The first experiment investigated the potential benefits of different versions of the SED training program. The effect of stimulus variability (single speaker vs. multiple speakers) and design of background babble (constant vs. adaptive staircase) were examined using English voiceless-voiced plosives /t/-/d/ phonemic contrast as the training materials. No improvements were found in the identification accuracy on the /t/-/d/ contrast in post-test, but identification improvements were found on the untrained English /ɛ/-/æ/ phonemic contrast.

The effectiveness of SED training was re-examined in Chapter 3 using the English /ɛ/-/æ/ phonemic contrast as the training material. Three experiments were conducted to compare the SED training paradigms that had the background babble implemented either at a constant level (Constant SED) or using the adaptive staircase procedure (Adaptive Staircase SED), and the longevity of the training effects. Results revealed that the Adaptive Staircase SED was the more effective paradigm as it generated greater training benefits and its effect generalized better to the untrained /t/-/d/ phonemic contrast. Training effects from both SED paradigms retained six months after the last training section.

Before examining whether SED training leads to improvements in speech production, Chapter 4 investigated the phonetics perception pattern of L1 Mandarin Malaysian speakers, L1 Malaysian English speakers and native British English speakers. The production intelligibility of the L1 Mandarin speakers was also evaluated by the L1 Malaysian English speakers and native British English speakers. Single category assimilation was observed in both L1 Mandarin and L1 Malaysian English speakers whereby the /ɛ/ and /æ/ phonetic sounds were assimilated to a single/æ/ category (Best, McRoberts, & Goodell, 2001). While the British English speakers showed ceiling performance for all phonetic categories involved, the L1 Malaysian English speakers had difficulty identifying the British English /ɛ/ phoneme and the L1 Mandarin speakers had difficulty identifying the /d/ final, /ɛ/ and /æ/ phonemes. As seen by their perceptual performance, the L1 Mandarin speakers also had difficulty producing distinct /d/ final, /ɛ/ and /æ/ phonemes.

Two experiments in Chapter 5 examined whether the effects of SED training generalizes to speech production. The results showed that L1 Malaysian English speakers and native British English speakers found different SED paradigms to be more effective in inducing the production improvement. Only the production intelligibility of the /æ/ phoneme improved as a result of SED training.

Collectively, the seven experiments in this thesis showed that SED training was effective in improving Malaysian speakers’ perception and production performance of difficult English phonemic contrasts. Further research should be conducted to examine the efficacy of SED training in improving speech perception and production across different training materials and in speakers who come from different language backgrounds.

Acknowledgements

I would like to express my gratitude to all who had contributed to the completion of this thesis.

First and foremost, my deepest and greatest appreciation goes to my supervisor, Dr. Jessica Price for her invaluable guidance, inspiration and unconditional encouragement. She has been very generous in giving praise and suggestions that kept me going throughout the journey. There is nothing more that I can ask from her and it would be impossible to complete this work without her support. For this, I am truly grateful to have the opportunity to work under her supervision.

I would also like to extend my gratitude to my second supervisors, Dr. Walter van Heuven and Dr. Nicola Pitchford, for their insightful ideas and knowledge that had built up the foundations of my studies. I am also thankful for their advice and support during my two visits to the UK campus for data collection and research purpose. The studies could never be completed without their deliberate supervision.

Also, many thanks to the staff and students from the School of Psychology from the University of Nottingham (both Malaysia and UK campus) for providing me research resources; and to my fellow PhD friends in office BB47 (especially Yee Mun, Adila, Cheng, Azlina and Kok Wei) who provided me great encouragements and support. Of course, none of this work would have been possible without participants and friends who willingly volunteered hours of their time. Their participations are sincerely appreciated.

Last but not least, I could never have completed this work without my family and friends. Thank you for their unconditional love and both emotional and financial supports. Although they were a little surprised at how long it took for me to complete this, they have never doubted my effort.

Table of Contents

Chapter 1: Introduction

Chapter 3: Spoken English Discrimination (SED) Training on English Vowel /ɛ/-/æ/ Phonemic Contrast

Chapter 4: Phonetics Perception and Production Complication of Malaysian Speakers

Chapter 5: Effect of SED Training on Speech Production Intelligibility

Chapter 6: General Discussion and Conclusion

List of Figures

List of Tables

List of Abbreviations

2AFC	two alternatives forced-choice
ANCOVA	analysis of covariance
ANOVA	analysis of variance
AoA	age of acquisition
AS	adaptive staircase
CALL	computer assisted language learning
db SPL	decibels of sound pressure level
DT	discrimination task
ELF	English as Lingua Franca
HVPT	High Variability Phonetic Training
ID	identification
IELTS	International English Language Testing System
IPA	International Phonetic Alphabets
L1	first language
L2	second language
LP	letters to phonemes
m	mean
MA	mean age
ME	Malaysian English
MMN	mismatch negativity
MRI	magnetic resonance imaging
N	number
N/A	not available
N/P	not provided
NLM	Native Language Theory
PAM	Perceptual Assimilation Model
RMS	root mean square
s.d.	standard deviation
SED	Spoken English Discrimination
SLM	Speech Learning Model
SNR	signal-to-noise ratio

SR	self-rated

Chapter 1: Introduction

1.1 Linguistic Background of Malaysia

1.1.1 History of English Education in Malaysia

During British colonial expansion through the later 19^{th} and early 20^{th} century, English was once the *de facto* official language used in the administration, judicial system and educational infrastructure on the land of Malaya (now commonly known as Peninsular or West Malaysia; Muniandy, Sekharan Nair, Krishnan, Ahmad, & Mohamed Noor, 2010; Tan, 2013). In fact, the first major English-medium school in Malaya was founded as early as 1816 in Penang. English remained the most important lingua franca for inter-ethnic communication until the mid-20^{th} century. In order to strengthen the new national identity after the independence of the Federation of Malaya in 1957, the Malaysia government gradually introduced Malay as the official national language and it replaced the official status of English in government functions. Malay has also become the new medium of instructions in all schools (Baskaran, 2005; Tan, 2013). All English medium schools were converted into Malay medium schools, from Standard 1 of primary schools, moving all the way up to the tertiary education system. Since then, the Malaysia government has vigorously implemented the teaching of the new national language (now referred to as Bahasa Malaysia), while maintaining the education of English language and other local languages in all vernacular schools (Gaudart, 1987).

Nevertheless, English still preserve as the second most common language spoken in the domain of Malaysia's national education until today. Due to globalization, competence in English is believed to be a prerequisite for the nation's scientific and technological developments, modernization and also in international trade and commerce with the world (Gaudart, 1987; Omar, 2016).

Although the former education systems highlighted the need to preserve national identity through the use of a national language, a language policy shift was observed in the past decades to ensure that the youth of today can speak proficient English and the country is capable of keeping pace in the increasingly competitive global economy (Omar, 2016). Such government efforts can be seen in the PPSMI (Pengajaran dan Pembelajaran Sains dan Matematik dalam Bahasa Inggeris) policy implemented in the year 2003. With the enforcement of the policy, English language became the language of instruction for Mathematics and Science subjects in primary and secondary schools, to produce a new generation who are scientifically and technologically knowledgeable and fluent in English (Yahaya et al., 2009). PPSMI was however, abolished in the year 2012. Omar (2016) attributed the instability of the policy to the inadequate language competence of young non-native students, insufficient trained teachers and the ideological objections rose from the perceived threat to cultural identity.

1.1.2 Role of English in Formal Education

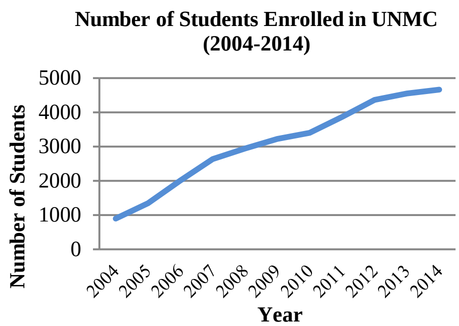
English is seen as the second most important language in Malaysia, and is widely used academically and even among the urbanized population (Crystal, 1997). Following the education policy set up by the Ministry of Education in year 1956 (also known as the Razak Report), Malay and English were made compulsory subjects in all primary and secondary schools (Gaudart, 1987; Tan, 2013). The nation tries to promote a bilingual education system in order to encourage bilingualism at an early age. Children are required to learn at least Malay and English, irrespective to their first language or the medium of instruction used in the schools they attend. The Malaysia education system uses Received Pronunciation as the core linguistic model to teach English in classes.

As for tertiary education today, mastering Malay has become an inevitable requirement for public university admission and to ensure academic success because the once English-medium institution has now changed to a Malay-medium institution. Only the designated courses such as the language course and English literature course are still taught in the related languages. Nevertheless, there is a growth of English being used as a medium of instruction in higher education due to the increasing diversity of student population and the demand from stakeholders who identify English as an important communicative tool for career life and global competition (Omar, 2016).

The emergence of international university branches (e.g. Monash University from Australia, University of Nottingham from UK), twinning programs offered by colleges and private pre-tertiary colleges during the last decades has reinstated the status of English-medium tertiary education in Malaysia (Tan, 2013). With the establishment of a pro-Malay/Bumiputera education policy, this significantly limits the number of public university admissions open to ethnic groups other than the Malay population. Many students of other ethnicities struggle to secure an education spot in the public universities, especially those who fail to meet the Malay language entry requirements. Choosing to study in branch campuses of international universities and private colleges which use English as the medium of instruction has become a popular choice among students who fail to enter public universities or prefer English-medium education but can't afford to study abroad. According to the Ministry of Education Malaysia (Malaysia, 2015), there are a total of 510 registered higher education institutions located throughout the nation, of which 8 of them are branch campuses of international universities. Nevertheless, Malaysian students are still required to score at least a credit (60% and above) for the national language in order to enter the English-medium international universities. Otherwise, students will have to complete a compulsory module to further study Bahasa Malaysia as a precondition for the award of recognized certificates, diplomas and degrees in private education institutions.

Figure 1.1 shows the rapid increase in the number of students who have enrolled in University of Nottingham, Malaysia Campus over the past ten years (Nottingham, 2015); while Figure 1.2

shows the number of students enrolled in public and private higher education institutions from year 2003 to 2013 (Ministry of Education Malaysia, 2015). It can be seen that approximately half of the students in tertiary education are exposed to an environment that uses English throughout as the means of communication. If this trend continues to grow, it is of no surprise that the use of English (as lingua franca) for intra-nation communication will be more prevalent in the community.



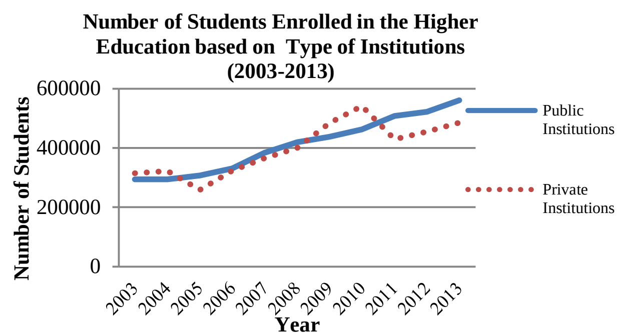


Figure 1.1 (left) & *Figure 1.2* (right): Number of students enrolled in University of Nottingham, Malaysia Campus (UNMC) from 2004-2014; Number of students enrolled in the higher education institution based on the type of institutions (public vs. private) from 2003-2013.

1.1.3 Linguistic background in Malaysia

Malaysia is a multi-ethnic, multilingual and multicultural country, with a complex and rich linguistic history. With the population over 30 million, Malays form the largest ethnic group of Malaysia's demography (50.1%), with the Chinese population coming in second (22.6%), the indigenous group coming in third (11.8%) and the Indian population coming in forth (6.7%; Central Intelligence Agency, 2015).

Malay, being the ancestral language of the largest ethnic group, is today's Malaysia's national language. Being educated in Malay-medium schools, the Malay generation born after Malaysia's independence are usually most proficient in their native language, Malay (the standard Malay taught in the education system or other dialects of Malay such as Kelatan-Pattani Malay, Kedah Malay and Terengganuan Malay which are commonly spoken in different regions of Malaysia); with reasonable fluency in English due to the Malaysia's bilingualism education policy. Some of them also learn some basic Arabic for religious purposes (Quran of the religion Islam is traditionally written in Classical Arabic).

The Chinese community, however, still speaks a wide range of ancestral languages such as Cantonese, Hokkien, Hakka, Teochew and Hainanese (originated from southern China) as their first language, as well as Mandarin, which has become the intra-ethnic lingua franca among the Chinese-educated community. Tan (2013) attributed the successful maintenance of ancestral language of Chinese to three factors: 1) the significant population size, 2) well-established Chinese-medium education and 3) the development of Chinese media in the nation (e.g. newspaper, films, TV channels and etc.). The multilingualism phenomenon is most manifested among the Chinese community as those who are educated in Chinese-medium schools learn Mandarin (as well as Malay & English) in school. On top of those, some of Malaysian Chinese also inherit other ancestral languages from their home environment. They are typically proficient in Mandarin and some ancestral languages depending on their living environment (e.g. Cantonese - most widely-spoken ancestral language in Kuala Lumpur; Hokkien – most widely-spoken ancestral language in Penang); with varying degrees of success in learning Malay and English. Chinese students who attend Malay-medium schools are more likely to be more proficient in English or Malay; but not all of them can speak their ancestral languages. On the other hand, the older generation who was born before 1967 (before the gradual removal of the official status of English language imposed by the *1967 National Language Act*, which was enforced to promote Malay nationalism) and educated in English-medium schools is usually fluent in English as well as their ancestral languages. They are only able to speak simple Malay for daily-life communication, having learnt it through interactions with the community, the media and the government.

Furthermore, the indigenous people from the Peninsula Malaysia and Borneo speak various distinctive native languages such as Iban, Kadazan, Bajau and Melanau; while a portion of the Indians still maintain their ancestral languages such as Tamil, Malayalam and Telegu. All of them have at least minimal proficiency in speaking Malay and English, depending on their learning accomplishment in schools.

Despite the variety of languages spoken by the multilingual community, Malay remains the most widely-used language for inter-ethnic communication in Malaysia, followed by English, then Mandarin (Tan, 2013). Usually, most Malaysians use two or more predominant languages for daily communication needs. Language choice is often made after considering the interlocutors (speakers) involved in the conversation, the discourse context and the communicative function of the discourse.

Code-switching is also common throughout conversations in Malaysia where speakers alternate between two or more languages in a single conversation. Malaysian speakers fluently use words or phrases of different languages in a spoken sentence to promote better comprehension between the interlocutors who come from different cultural backgrounds; especially when the subject described is culturally relevant (e.g. *tanglong* which means lantern in Mandarin, *maggie goreng* which means stir fried instant noodle in Malay). The most common code-switching takes place between Mandarin and English (mainly within the Chinese community) or between Malay and English (in Malays, Indians and those who are Malay/English-medium educated; Tan, 2013).

Although Malaysians from this modern era learn English at a young age, many of them do not speak English as their first language (L1; see section 2.1 for more details on terms usage in this thesis). According to Crystal’s (2003) estimation, almost one –third of the nation’s population speaks English and 95% of them do not speak English as their L1 (as cited in Tan, 2013). These second language (L2) English speakers often encounter inevitable difficulties faced by L2 speakers when they communicate using their L2. Some general disadvantages include: native accent (Flege, Munro, & Mackay, 1995), inaccuracy in pronunciation (Flege et al., 1995; Rajadurai, 2006), inaccuracy in perceiving speech, and poor vocabulary library (some examples of perception and production inaccuracy of Malaysian speakers are better addressed in the experimental chapters).

1.1.4 Malaysian English

English is spoken by the majority of the population for intra-national communication between its users of various ethnic groups, with a minority who speak it as their first language (L1; Crystal, 1997). The widespread use of English in the multiethnic and multilingual context of Malaysia promoted systematic changes in English's language system and contributed to the emergence of a local variety of English, Malaysian English (ME). It reflects the influences of diverse cultures and language backgrounds of its speakers. Tan (2013) proposed that contact between English and other languages such as Malay, Mandarin and other ancestral languages (e.g. Hokkien, Cantonese, Tamil and etc.) has shaped today's Malaysian English. The contact-induced changes in ME become more complex when the conversation is conducted by a group of interlocutors with diverse language backgrounds and when code-switching becomes a common language practice.

Phoon, Abdullah and Maclagan (2013) identified three sub-varieties of Malaysian English: 1) Malay-influenced ME, 2) Chinese-influenced ME and 3) Indian-influenced ME, mainly spoken by the three major ethnic groups; while Baskaran (1994) classified Malaysian English according to its usage contexts: 1) *acrolect* which is used for formal and educational purposes, 2) mesolect as the most common form, is used for semiformal and casual situations and lastly 3) basilect. Basilect, as the patois form of ME, is usually used as the informal communicative tool by speakers who have learnt the language on an ad lib basis. According to Baskaran (2005), it is the mesolect "*that truly represents Malaysian English as a fully-fledged variety of English"*. The Malay and Mandarin languages have been regarded to be most influential to the mesolect form of Malaysian English today because of the languages' interaction among the two largest ethnic groups in Malaysia: the Malays and the Chinese.

Some of the most notable changes in Malaysian English can be seen from the lexical borrowing/ creation and structural nativization phenomena take place during conversation, or even in the media (Baskaran, 2005). Lexical borrowing and lexical creation is seen as the effort of language maintenance (in this case English), to ensure continuing use of a language despite of being in contact with or facing competition from other languages (e.g. Malay, Mandarin and Tamil; Tan, 2013). Lexical borrowing and creation were initiated when intense linguistic interactions were needed between the British sojourners (e.g. civil servants and merchants) and the local communities (e.g. native farmers and fishermen, Chinese and Indian laborers) for commerce and trade, and ruling purposes during late 19th century of British colonization (Tan, 2013). Vocabulary items were introduced into Malaysian English from other languages whenever references to local phenomena became necessary in conversation. For examples, the word *kampong* originated from Malay used to represent "village" and the word *kongsi* originated from Chinese languages to represent "company". These newly introduced and localized "Malaysian English words" are commonly used among the Malaysian community (regardless of their L1) to refer to culturally related subjects. Other than the institutionalized concepts of the borrowed words, lexical indigenization in Malaysian English can also happen when the possible English translation has restriction semantically, culturally or emotionally in the local context (Baskaran, 2005). For instance, the word "dadah" is sometimes preferred over the English word "drugs" in Malaysia media because the Malay term "dadah" is confined to refer to drugs used illicitly.

Structural nativization is another common habit observed from Malaysian English speakers. For instance, utterances that will be considered ungrammatical in standard English varieties such as

"Why you no understand?" and "You go first." are often used and tolerated in mesolect and basilect because the structure of these sentences would be correct in some of the local language systems (i.e. the sentence "Why you no understand?" in Malaysian English would be "为什么你不明白？" in Mandarin and "Kenapa kamu tidak faham?" in Malay). In Malaysian English, the *wh-* questions words can also be placed at the sentence final position, a sentence structure that is possible in Malay and Mandarin (i.e. the sentence "You have *what*?" in Malaysia English in consistent with "你有*什么*?" in Mandarin and "Kamu ada *apa*?" in Malay). This phenomenon can also occur when Malaysian English speakers form the intended sentences by translating the words consecutively and directly from their L1. Sentences with such nativized word order, however, are considered incorrect at the acrolect level.

Malaysian English as a strongly localized variety is impacted by several commonly spoken languages in Malaysia such as Malay, Mandarin and Tamil. It is a colloquial variety, which Tan (2013) further characterized it with its excessive discourse particles (e.g. "lah", "meh" and "lo"), variable word order and localized pronunciation and intonation patterns. It also has phonological features that were influenced by the local languages. Some phonological characteristics of Malaysian English include consonant cluster reduction (e.g. omission of /l/ in pronouncing the word "always", /ɔːweis/ instead of / ɔːlweis/), ambiguous fricatives voicing (e.g. devoicing the voiced fricative /v/ to in pronouncing the word give, /gif/ instead of /giv/), different vowel quality (e.g. the back vowels /ɔ/, /ɔː/, /aː/ and /ɑː/ are pronounced with a more close quality compared to the Received Pronunciation), reduced vowel length distinction (e.g. shortening of long vowels /iː/, /aː/, /ɔː/, /uː/, /əː/ and lengthening of short vowels /i/, /a/, /ɔ/, /u/, /ə/) and reduced dipthongs (Baskaran, 2005). Malaysian English also tends to differ from the standard English with its own system of prefixes, suffixes, redundancies, synonyms, adjectives and appellations (Baskaran, 1994).

1.2 Learning a Second Language

Being bilingual can bring some cognitive advantages; for instance, early studies have shown positive effects of bilingualism in suppressing irrelevant information (Bialystok, 1999; Ellen Bialystok & Martin, 2004), task switching (Prior & Macwhinney, 2010), solving conflicting resources (Costa, Hernández, Costa-Faidella, & Sebastián-Gallés, 2009) and even delaying the onset of dementia (Alladi et al., 2013). In general, bilinguals' advantages in mental flexibility and executive control tasks are attributed to their constant practice in controlling and switching between the languages because the two languages are constantly activated, even though only one language may be spoken at a time.

However, the cognitive benefits of bilingualism have been recently challenged as more replication studies have showed null results on the cognitive effect of bilingualism (e.g. Morton & Harper, 2007; Paap & Greenberg, 2013). Several explanations were offered to explain the decline effect, including regression to the mean, research practices and biases, and publication bias (see de Bruin, Treccani, & Della Sala, 2015 for more information). Although de Bruin and Sala (2015) concluded that the effects of bilingualism on cognitive processing is small and only exists selectively (e.g. in older adult population and for certain executive functions), the benefits of speaking a second language in cross-national communication, broadening life horizons and expanding one's social world is still unquestionable.

1.2.1 General Difficulties faced by Second Language Speakers

It is increasingly common in today's society for individuals to speak more than one language. However, not all bilinguals/multilinguals tend to possess the same level of native proficiency for each non-native language learnt. Non-native speakers may also experience more challenges when conversation is conducted over the telephone (with absence of visual cues and reduced clarity of speech) or in an open public area with the presence of background noise. Studies have shown that background noise can interfere with bilinguals' perceptual performance in perceiving full sentences (as measured by Speech Perception in Noise test in Mayo, Florentine, & Buus, 1997) and in monosyllabic word recognition (Rogers, Lister, Febo, Besing, & Abrams, 2006) to a greater extent as compared to native, monolingual listeners; the perceptual deficit in noise is observed even in early bilinguals (see Lecumberri, Cooke, & Cutler, 2010 for review). In terms of low-level perceptual processes, non-native language speakers may experience greater difficulty in segmenting words or phonemes in running speech, discriminating similar sounding phonemes and reproducing motor articulations for some non-native speech sounds (Iverson et al., 2003).

Differences in language performance between native and non-native speakers may not be apparent in quiet conditions nor in speech production but they manifest in adverse ‘noisy’ conditions (Cutler, Lecumberri, & Cooke, 2008; Rogers et al., 2006); or when the tasks involve higher-level linguistic information (e.g. a word recognition task that depends on the sentence context cues; see Mayo et al., 1997). Rogers et al. (2006) conducted a study comparing monolingual American English speakers and early Spanish-English bilinguals’ monosyllabic English word recognition either with or without noise. Although the monolinguals outperformed the early bilinguals in the noisy condition, no performance difference was observed in the quiet condition. Regardless of how proficient the bilinguals were, recognizing words in noisy environments is often difficult for L2 speakers.

A detrimental effect of noise or adverse conditions in speech perception may be observed in both native and non-native listeners but native listeners’ advantage lies in their capability to recover rapidly and more efficiently in the unfavorable conditions (see review Lecumberri et al., 2010). For instance, the Lombard effect occurs when speakers naturally raise their volume when masking noise exists (Lane & Tranel, 1971) and the “phoneme restoration” phenomenon occurs when listeners’ unconsciously “restore” sound/ words that had been omitted to the “rightful” place (Warren, 1970). These adaptive strategies are commonly practiced by native speakers to overcome the adverse conditions and to fill in the missing information received from the source; the non-native speakers on the other hand, rarely show these practices (Lecumberri et al., 2010).

According to Lecumberri et al. (2010), additive noise sources have been frequently used in studies of L2 speech perception to produce two types of perceptual masking: energetic masking (e.g. white, pink, speech-shaped noise and multi-talker babble) and informative masking (e.g. competing speech). Energetic masking interferes with the perception of target signals at the level of the auditory periphery and causes uncertainty or complete loss of information to the target signals (Lecumberri et al., 2010; e.g. Rogers et al., 2006).

On the other hand, informational masking interferes with decision-making at higher levels of processing whereby audible components from the informational maskers compete for attentional resources with the target signals (Lecumberri & Cooke, 2006; Lecumberri et al., 2010; e.g. Mayo et al., 1997). At a fixed signal-to-noise ratio (SNR), Lecumberri and Cooke (2006) found competing speech (informational masker) being the least effective masker on a consonant identification task compared to stationary speech-shaped noise (energetic masker) and 8-talker babble; with the 8-talker babble being the most challenging due to the combination of both energetic and informational masking.

1.2.2 Factors Affecting Second Language Competence

The success in learning a second language can vary according to a number of common factors, such as: age of acquisition (James Emil Flege & MacKay, 2011; James Emil Flege, Yeni-Komshian, & Liu, 1999), amount and manner of exposure (Kuhl, Tsao, & Liu, 2003), quality of input (Bosch & Ramon-Casas, 2011), L1 influence (Bohn & Best, 2012; Kuhl et al., 2008), similarity between L1 and L2 (Antoniou, Liang, Ettlinger, & Wong, 2014), and years of education (Butler & Hakuta, 2006).

Age of acquisition has been the most frequently used explanation to account for the substantial variability in L2 attainment because an age-related decline is always observed in the performance of different linguistic domains. For instance, Flege et al., (1995) examined the effect of arrival age in Canada on perceived foreign accent (rated by native English speakers). Non-native speakers' age of acquisition turned out to be one of the major factors that determined the strength of their foreign accent. The foreign accent becomes especially evident when the native Italian speakers arrived in Canada and started learning English after the age of 3.

Studies on the effect of age of acquisition always seem to come to the same conclusion: the earlier the L2 acquisition, the higher the chance for non-native speakers to perceive and produce it in a native-like fashion. In fact, functional MRI studies have shown there is a difference in brain activation between the early bilinguals and late bilinguals during L2 speech processing; early bilinguals activate overlapping regions of the brain for the two languages while the late bilinguals activate two distinct regions (see Kuhl, 2000). This neurophysiological finding showed that early bilinguals processed the L1 and L2 more similarly while the late bilinguals tend to process them differently. It further stated that age of L2 acquisition can affect L2 learning in many aspects, starting from the underlying neurological functions.

According to the Critical Period Hypothesis by Lenneberg (1967), there is a maturational constraint on language learning, whereby ability to learn a new language (especially on phonological aspects) to a native-like competence becomes increasingly difficult after a certain point of one's maturation process. Hence, the theory believes that early bilinguals stand a higher chance to perform in a more native-like manner compared to late bilinguals (Flege, 1991; Thornburgh & Ryalls, 1998). Nevertheless, even though the early bilinguals usually out-perform late bilinguals, there are still substantial differences between the early bilinguals and native speakers (e.g. speech perception in noise as discussed above; see Mayo et al., 1997; Rogers et al., 2006).

Lennerberg (1967) believes that language acquisition should take place between the age of two and puberty, along with the lateralization of brain functions. However, more recent neurological research suggests different time frames for the 'critical period' (also known as the sensitivity period) of different language functions (Singleton, 2007), with speech perception starting the earliest (Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992). Infants' perceptual abilities and patterns changed actively and showed preference for native language speech sounds by nine-months-old (Jusczyk & Luce, 1994; Kuhl et al., 1992); with some patterns start to be shaped to resemble that of adult speakers (e.g. perceptual magnet effect; Kuhl, 1991). According to the critical hypothesis, having the right language input during the critical period would be crucial for successful language learning.

While the critical period can begin as early as 1 month after birth, the precise developmental timing of critical period for different linguistic domain however, also varies (Kuhl, 2011). The literature mainly attributes the closure of sensitive period for linguistic abilities to two reasons: 1) the loss of brain plasticity or neurological specification due to maturation (Lenneberg, 1967; Kuhl et al., 2008) and 2) the deep-rooted L1 processing habits (Best, McRoberts, & Goodell, 2001; Flege et al., 1995).

Although the Critical Period Hypothesis emphasizes biological maturation as the limitation for L2 learning, it is not compelling to attribute the function of age-susceptibility to language input to the maturation explanation solely. There are several other environmental factors that can concurrently affect L2 learning in later age.

In the long term and in immersion contexts, early L2 learners start acquisition earlier and tend to have more exposure to the input over the years compared to later learners. Factors, such as the learners' quality experience with L2 are also critical determiners of L2 competence. L2 speakers tend to have less opportunity to learn their L2 in a context full of diverse, native and variable speech inputs as compared to native speakers (Lecumberri et al., 2010). They proposed that late learners miss the opportunity to receive as much native inputs as early learners, which usually can be found through interactions in school or academic settings.

L2 competence can be affected by a wide range of factors that range from the biological limitation to the linguistic environment speakers are exposed to. Another well-discussed factor determining the L2 competence is L1 influence. Section 1.3 below discussed the effect of L1 influence on L2 phonetic perception.

1.3 Phonetic Perception

Phonetic perception is important as the foundation of spoken speech learning because phonemes are regarded as the most basic developed perceptual processing units. The term "phonetic" is usually used to refer to all basic speech sounds produced by human vocal tracts, which may or may not be linguistically meaningful (Catford, 2002); whereas the term phoneme is defined as a "*distinctive sound; that is, for each sound which, being used instead of another, in the same language, can change the meaning of a word*" (IPA, 1999, p. 27). Infants first have to be capable to discriminate phonetic units to respond to distributional properties of sounds and form phonetic categories that correspond to their native language. Phonetic perception ability is also responsible for infants' sensitivity towards the sequential and transitional probabilities[1] between adjacent syllables of a continuous speech (Kuhl, 2004). Infants then use both phonotactic patterns learned and information about transitional probabilities to parse and identify words from running speech (Kuhl et al., 2008; Kuhl, Conboy, Padden, Nelson, & Pruitt, 2005), before forming speech sound categories and making associations to the real world objects. In adulthood, phonetic perception remains important in parsing running speech and recognizing spoken words heard.

[1] This refers to the likelihood of perceived combination of sounds. The probabilities of discrete units in a stream of sequential auditory input assisted listeners to judge whether the combination of sounds are 'legal' in the language (van der Lugt, 2001; Leonard, Bouchard, Tang & Chang., 2015).

Speech production is also closely linked to speech perception behaviorally (Flege, Frieda, & Nozawa, 1997) and neurally as shown by functional magnetic resonance imaging (fMRI) (Pulvermüller et al., 2006; Skipper, van Wassenhove, Nusbaum, & Small, 2007; Wilson & Iacoboni, 2006) . According to Liberman's motor theory (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Liberman & Mattingly, 1985), speech production is believed to develop through vocal imitation, whereby hearing the vocalization of others and oneself guides learners' vocal production. Auditory, visual and motor information are interconnected and co-registered for speech categories. Failure to perceive the critical acoustical properties of language input can lead to formation of inappropriate phonetic categories, which would then be reflected in inaccurate speech production (see section 5.1.2 for more discussion).

1.3.1 Developmental Pattern - Diversion from Non-native Speech Perception

Infants are born with an early universal perceptual ability to discriminate all the phonetic units of the world's languages (Eimas, Siqueland, Jusczyk, & Vigorito, 1971; Kuhl et al., 2008; Werker & Tees, 1984). However, the ability is lost by adulthood and discrimination of some non-native phonetics becomes more difficult (Best et al., 2001; Flege, 1989; Iverson et al., 2003), while discrimination of native language relevant sounds is retained or even refined (e.g. Narayan, Werker, & Beddor, 2010). Reasons given to explain the development transition mechanism varies. Both Eimas' phonetic feature detector account (Eimas, 1975) and Liberman's motor theory (Liberman et al., 1967) theorized that infants' initial ability to detect all phonetics is selectively maintained or lost as a function of early linguistic experience. The difference between the two accounts lies on Liberman's hypothesis that all infants initially detect phonetically relevant gestures from speakers before deciding the phonetic identity (see section 5.1.2.2 for Liberman's Speech Motor Theory).

According to the selective view, infants will eventually lose perceptual sensitivity towards phonetic distinctions which they have not been exposed to during the critical period (Kuhl et al., 2008; Kuhl, 2000). This theory was undermined when studies revealed variability in non-native speakers' ability to discriminate non-native phonetic contrasts. The ability to discriminate some non-native phonetic contrasts can still be found in non-native speakers, although at a reduced level when compared to the native speakers (e.g. Zulu clicks to native English speakers; Best & Mcroberts, 1988). Furthermore, perceptual sensitivity to all the non-native distinctions is also not lost permanently. Phonetic training studies showed that with the right training method, adults' discrimination and identification performance on non-native phonetic contrasts can be improved (see section 1.3.3 for further details). Perhaps sensitivity to non-native phonetics is not lost as claimed; rather the sensitivity for the rarely encountered speech sounds may be reduced across development.

An alternative explanation for the reduced performance on non-native phonetics proposed by Werker and Pegg (1992) suggests a link between the discrimination ability and cognitive control abilities. Inhibitory skills, an ability to disregard irrelevant information to enhance cognitive efficiency was associated with the reduction of sensitivity for non-native phonetics. Infants who have better performance in inhibitory tasks (e.g. detour-reaching tasks) also have greater reduction in non-native discrimination skills (Kuhl et al., 2008).

Attention is regarded as an important factor that determines phonetic perception performance. A study conducted by Iverson and colleagues (2003) showed that the Japanese speakers performed significantly different than the L1 American English speakers in L2 English perception due to their erroneous attention to irrelevant acoustic cues. The L1 American English speakers used the third formant (F3) as acoustical cues to categorize the English /r/-/l/ phonemes whereas the Japanese speakers erroneously relied on the second formant (F2) acoustical cue that does not facilitate the categorization. High sensitivity towards the acoustical cues that are perceptually salient but irrelevant for categorization processes could increase the cognitive load in non-native speakers. Furthermore, the different attentional network of non-native speakers could also cause non-native speakers to form erroneous category representations, an undesired outcome for new language learners.

1.3.2 Linguistic Theories on Non-native Phonetics Perception Difficulty

Second language (L2) speakers often have difficulty discriminating L2 phonetic distinctions that do not exist or comply with their L1 phonetic system. For instance, Chinese speakers have problems discriminating the English voiced-unvoiced stops /t/-/d/ in word-final position because their L1 phonologically contrasts the voiced-and unvoiced stops /t/-/d/ only in word-initial position (Flege, 1989). Another widely-researched example is the perception of /r/-/l/ English contrast by L1 Japanese speakers. This phonemic distinction does not exist in the Japanese language system; therefore, Japanese speakers are long-known to have difficulty discriminating and producing the two phonetic sounds (Bradlow, Akahane-Yamada, Pisoni, & Tohkura, 1999; Bradlow, Pisoni, Akahane-Yamada, & Tohkura, 1997; Flege, 1989; Lively, Logan, & Pisoni, 1993; Lively, Pisoni, Yamada, Tohkura, & Yamada, 1994; Logan, Lively, & Pisoni, 1991). Several speech learning models have been proposed during the past few decades to offer explanations to account for these discriminative difficulties of L2 speakers. Three of the most cited models will be discussed in the section below.

1.3.2.1 Flege's Speech Learning Model (SLM)

According to the Speech Learning Model by Flege (1995), new L2 phonological categories will be established only when the L2 sounds are sufficiently dissimilar to the L1 sounds. Flege found differences in the degree of perceived vowel dissimilarity between L1 English and L1 Spanish speakers (L2 English speakers) on English vowel minimal pairs (e.g. /a/-/æ/, /a/-/ɛ/); in which L1 English speakers rated some of the English vowel pairs to be significantly more dissimilar compared to L1 Spanish speakers. L1 English speakers thus were more likely to identify the vowels as realizations of phonetically distinct categories while L1 Spanish speakers would have greater difficulty in establishing new and distinct phonetic categories for the similar sounding

English vowels. When L2 speakers barely distinguish the similar sounding L2 phonetics, the 'equivalence classification' mechanism takes place where L2 sounds are assimilated into a single phonetic category of L1. Such assimilation can be seen from the most widely cited example of Japanese speakers in discriminating the English phonemic contrasts /r/-/l/ because both phonemes are assimilated to the /l/ category (Bradlow et al., 1997; Callan et al., 2003; Handley, Sharples, & Moore, 2009; Hazan, Sennema, Iba, & Faulkner, 2005; Iverson, Hazan, & Bannister, 2005).

1.3.2.2 Best's Perceptual Assimilation Model (PAM)

The Perceptual Assimilation Model (Best et al., 2001) also proposes that the discrimination difficulties can arise when phonemes from the L2 are assimilated inappropriately into the phoneme categories of the L1. Flege's model mainly focuses on acquisition or formation of individual L2 phonemic category. It fails to make explicit predictions for different types of L2 phoneme assimilation. Building on Flege's SLM, Best pointed out that discrimination performance on similar L2 sounds that do not occur in L1 is not always uniform. For example, American English speakers' ability to discriminate the Zulu clicks is still preserved although the distinction does not occur in American English (Best et al., 1988).

To account for the discrimination variation of different types of L2 contrasts, Best et al. (2001) introduced three main types of phonemic assimilation; when two L2 phonemes are phonetically similar to two distinct L1 phonemes, the two category assimilation occurs; when they assimilated equally well or poorly to a single L1 phoneme, the single category assimilation occurs; and lastly, when both L2 phonemes are assimilated into a single L1 phoneme but with different degree of fitness, the assimilation with category goodness difference occurs. PAM posits that L1 phonology interferes with L2 discrimination the most when two L2 phonemes are assimilated to the same L1 phoneme (i.e. single category assimilation); however, it may facilitate the discrimination when the two L2 phonemes are also separated by the L1 phonological boundary (i.e. two category assimilation).

The single category assimilation phenomenon has been widely researched to examine the reason behind L2 speakers' difficulty in perceiving sound difference that seems to appear quite prominently in the native population. This type of research also shows the influence of L1 on L2 perception sensitivity; for example, Mandarin speakers with the voiced-unvoiced plosives /t/-/d/ at word-final position (Aliaga-García & Mora, 2009; Flege, 1989) and Spanish speakers with the English vowels /æ/-/ʌ/ (Aliaga-García & Mora, 2009). These phonemic pairs are typically assimilated into one category by the non-native speakers.

1.3.2.3 Kuhl's Native Language Magnet Theory (NLM)

Kuhl's Native Language Magnet Model (NLM) was initially constructed based on the observation of perceptual magnet effect (see section 1.3.2.3.1 for more details). NLM proposes three phases of development (Kuhl et al., 2008) in the initial state (phase 1), infants are born with innate neural specification for all possible phonetic units. The capability to differentiate all the sounds from human speech is not influenced by the speech-specific mechanism yet, but derived from the general auditory mechanisms. By about 6-months-old (phase 2), infants' experience with the distributional properties of ambient language alters their perception of the phonetic units of speech and causes physical changes in neural tissue and circuitry that reflect the statistical and prosodic features of the language input (Kuhl, 2010). The learning and changes are believed to continue until stability is achieved. According to Kuhl et al., (2005), stability is reached when the language input has created a distribution that can predict new instances and that the new occurrence of speech sound no longer significantly alters the underlying distribution. In other words, the maturational process may start the mechanisms of phonetic learning, but it is the language experience that closes the learning. In this phase, phonetic representations (prototypes) are also formed and they function as the perceptual magnets. These magnets serve to increase the perceived similarity between members of the same phonetic category.

In the latest phase (phase 3), the distortion of perception intervenes with infants' language acquisition. The native language neural commitment produces a bi-directional effect: it facilitates native language learning that capitalizes on learned phonetic pattern (Kuhl et al., 2006), while reducing phonetic sensitivity to alternate phonetic schemes (e.g., foreign language phonetic system that does not conform to the learned pattern; Kuhl et al., 2008). Kuhl and colleagues (2005) conducted a study to support the facilitation and inhibition assumption of native language neural commitment. In the study, infants' performance in discriminating non-native phonetic contrasts was assumed to reflect the extent to which the brain remains open or uncommitted to their native language. At 7.5-months-old, infants who performed well in native phonetic skills (neural commitment to native language pattern) showed great advancement in their native language acquisition. Native phonetic perception skill turned out to be a good predictor of future language ability, which was measured in terms of word understanding and word production. In contrast, non-native phonetic perception skill turned out to be a strong predictor of slower native language growth. Both native and non-native phonetic perception skills were shown to be inversely correlated, and they predicted later language ability in opposite directions.

1.3.2.3.1 Perceptual Magnet Effect

Kuhl (1991) first discovered that not all /i/ instances from the phonetic category were graded equally in terms of their perceived category goodness. In the study, raters were presented with 64 synthetic variants of vowel /i/ and were asked to rate the 'typicality/ goodness' of the phonetic sound heard. Interestingly, the best instance or the prototype of /i/ vowels was consistently identified across the raters. Furthermore, the perceived goodness of /i/ vowels declined systematically as they were moving further away from the chosen prototypic /i/ vowel (see Figure 1.3). In other words, /i/ instances with acoustic properties dissimilar to the chosen prototypic /i/ vowel would be rated as bad instances.

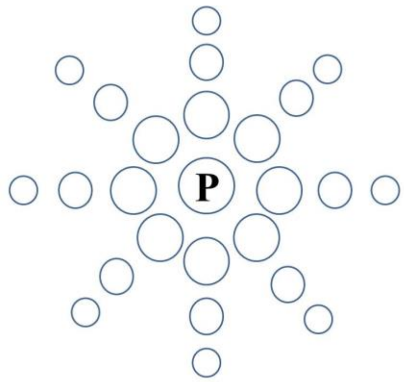


Figure 1.3 The prototype ‘P’ located at the center of each phonetic category, with phonetic variants surrounding it in systematic steps. Adapted from Kuhl (1991).

Another important finding of that study was the discovery of perceptual magnet effect. According to Kuhl, the prototype of the /i/ vowel category functions like a perceptual magnet when it serves as the referent vowel, pulling and assimilating neighboring stimuli to a greater extent compared to the condition where other non-prototypic /i/ vowel instances served as the referent vowel. The phenomenon was observed even when the psychophysical distance between the referent vowels and the surrounding variants were controlled (Kuhl, 1991). Kuhl found that perceptual discrimination became more difficult in the prototype condition than the non-prototypic condition because the /i/ vowel instances in prototype condition sounded more alike to each other. This human-specific phenomenon was then known as the perceptual magnet effect and it can be observed in early infancy as young as six-months-old (Kuhl, 1991).

Iverson and Kuhl (2000) proposed that the perceptual magnet effect can arise due to two reasons: 1) auditory processing- ambient language experience produces distortions in auditory processing due to the formation of neural maps in the auditory system (Guenther & Gjaja, 1996) and 2) phonemic encoding – stimuli input are always referred to the category prototypes and encoded based on their distance from the prototypes, with variation of goodness.

The language auditory system of humans is believed to evolve with one important objective, which is to enhance speech intelligibility. Speech perception sensitivity is never uniform towards speech sounds although their acoustical properties differ in uniform steps. Increased perceptual sensitivity can be observed at phonetic positions that signify distinctiveness of different phonemic categories, i.e. the category boundary (Iverson & Kuhl, 2000). The peak sensitivity at category boundary observed was named as the phoneme boundary effect (Iverson & Kuhl, 2000). In contrast, the perceptual magnet effect may serve to strengthen within-category cohesiveness, by minimizing the perceptual difference of stimuli near phonemic category centers (Kuhl, 1991). Both instinctive cognitive strategies facilitate the process of phonemic categorization during speech perception, at the cost of less behaviorally relevant within-category discrimination (Guenther & Gjaja, 1996).

When the context variance (variety of stimuli heard within each testing) is reduced and the task becomes less dependent on memory, performance on the task eventually relies more on the sensory resolution. Under this circumstance, phoneme boundary effect diminishes; perceptual magnet effect however, was not affected by the manipulation of context variance (Iverson & Kuhl, 2000). The presence of perceptual magnet effect in such conditions that rely mainly on sensory information indicates that the effect is more to do with general auditory processing, instead of cognitive processing.

1.3.2.4 Differences between Linguistic Theories

All three language learning models posit that L1 works as a “sieve” on L2 and influences L2 phonetics processing. According to Flege’s SLM and Best’s PAM, phonetics perception changes in adult L2 speakers are due to the higher-level linguistic processes or more explicitly, the cognitive representation in memory (e.g. phonological encoding). What distinguishes PAM from SLM is that PAM offers a more extensive account on the assimilation patterns of L2, and the articulatory gesture (e.g. movement of lips or vocal tracts) is taken into account. PAM proposes that discriminative difficulty of infants on non-ambient contrasts could partially be induced by the articulatory similarity between the ambient and non-ambient phonetics production (Best & McRoberts, 2003; Best, 1994; Best, 1995). SLM however, provides a more elaborative description on the developmental process of L2 phonological categories in long term memory (Flege, 1995).

Unlike SLM and PAM which focus on cognitive explanations, Kuhl's NLM attributed L2 speakers' difficulty in L2 phonetic perception to the changes of earliest sensorimotor sensitivity starting from infanthood, thus targets more on the lower-level linguistic processes. The model provides a comprehensive account of the phonetics perception development trajectory and uses neurological and language input factors to explain the sensitivity diversion from L2 phonetics perception.

1.3.3 Phonetics Perception Training

Previous literature has shown that perceptual training can re-shape speakers' L2 phonetic perception by shifting their attention to the meaningful acoustic cues that can help distinguish the phonetic sounds phonologically (e.g. Flege, 1989; Logan et al., 1991; Bradlow et al., 1997; 1999). A large number of training methodologies have been developed over the last thirty years to improve the ability to discriminate between non-native phonemes.

1.3.3.1 Training Method: Discrimination vs. Identification Task

The training tasks can generally be categorized into two designs. The identification task (ID task) was the most commonly used design. In the task, participants are first presented with an auditory stimulus (e.g. *rock*) before matching the auditory stimulus with its orthographic form by choosing from at least two response alternatives (e.g. *rock* or *lock*; see Lengeris & Hazan, 2010; Lively et al., 1994; Wang, 2013). Few studies have included more than two categories as the response alternatives. For example, Pisoni et al. (1982) used a three-category ID task where three response options were provided for every identification trial; while more recently, Nishi and Kewley-Port (2007) used a nine-category ID task where participants identified the vowel in real or nonsense words by choosing from nine response alternatives presented in Consonant-Vowel-Consonant form.

Other studies have used a discrimination tasks (DT) which can be further divided into three main designs: the two-alternative forced-choice (2AFC) discrimination task (also known as AX discrimination task by Strange and Dittmann, 1984), the oddity discrimination task, or the categorical discrimination task (e.g. Flege, 1995). Participants in the Strange and Dittman (1984) study listened to stimulus pairs before indicating whether the second auditory stimulus was identical to the first stimulus previously heard. Similarly, Kraus et al., (1995) used a training paradigm consisting of a same-different two-alternative forced-choice discrimination task in which participants heard two consecutive auditory stimuli and had to indicate whether the stimuli were the same or different.

In an oddity discrimination task however, participants discriminate several auditory stimuli and point out the odd stimulus out of others (e.g. Handley et al., 2009). Lastly, the categorical discrimination task (often referred to as XAB discrimination task or ABX discrimination task) requires participants to discriminate two stimuli out of three that share the same phonemic category membership, by matching either A or B stimuli to the appointed stimulus X or vice versa (e.g. Moore, Rosenberg, & Coleman, 2005).

Both identification and discrimination training tasks involve different cognitive mechanisms. Identification training engages top-down processing of speech signals in which participants have to respond based on their categorized or phonetic representations in memory. On the other hand, discrimination training influences primarily the bottom up processing of speech signals. The training engages a lower level and sensory-based information in speech signals by improving speakers' sensitivity to detect minor differences between the similar sounding stimuli (Kraus et al., 1995).

To assess training effects, an identification task is often used to compare phonetic identification accuracy between pre- and post-test performance. It is sometimes combined with a discrimination task to assess discrimination sensitivity (e.g. Jamieson & Morosan, 1986; Jamieson & Morosan, 1989; McCandliss, Fiez, Protopapas, Conway, & McClelland, 2002) or other test batteries to assess generalization of training effect towards other aspects of performance (e.g. reading for production accuracy as examined in Iverson, Pinet, & Evans, 2012; word recognition and imitation for production accuracy as examined in Richie & Kewley-Port, 2008; Phonological Assessment Battery for receptive phonological skill as examined in Moore et al., 2005).

1.3.3.2 Synthetic Training

Training studies conducted pre-1990 commonly used synthetic speech materials (e.g. Pisoni et al., 1982; Strange & Dittman, 1984; Jamieson & Morosan, 1986; Jamieson & Morosan, 1989; Flege, 1989). In these studies, stimuli were acoustically synthetized by manipulating criterial acoustical cues that signifies the phonological identity of speech sounds. Some of the most common manipulated features include: the voice onset time (Pisoni et al., 1982), formant onset frequency (Strange & Dittman, 1984), formant frequency (Golestani, Paus, & Zatorre, 2002; Iverson et al., 2005; Zhang et al., 2009) or duration of frication (Jamieson & Morosan, 1986; Jamieson & Morosan, 1989). Through the manipulation of acoustical properties, many speech items that vary in small, systematic steps can be created. This manipulation introduces acoustical context variation to the training stimuli.

The main aim of the synthetic trainings is to increase participants' perceptual sensitivity to detect small differences between similar sounding non-native speech sounds. When participants can perceive the similar non-native speech sounds as two distinctive sounds, new phonological categories can be formed to help future identification. Table 1.1 summarizes 17 articles that used synthetic stimuli for phonetic training found in the literature from the year 1982 to 2013. The most common training materials consisted of the English consonant phonemic contrasts - /r/-/l/ (26%), /t/-/d/ (11%), /θ/-/ ð/ (11%), the English vowels contrasts (21%), and other consonant contrasts such as the Hindi /ta/-/da/ and the Mandarin pitch contours (31%). More than half of the studies trained on English (63%), with the remaining studies trained on languages such as the Mandarin, Hindi and Quebec French.

Training effects were found by all the studies; however the study by Brosseau-Lapré, Rvachew, Clayards, and Dickson (2011) reported training effects only for certain training conditions. Adults' perception of the L2 phonetic contrasts were only enhanced when the training stimuli were produced by multiple speakers (i.e., when stimulus variability was introduced). The learning was also greater when the exemplars distributed near the phonetic category prototypes were used as training materials compared to the exemplars distributed near the boundaries of phonetic categories. They proposed that adults learn phonetic category identities by referring to the perceptual characteristics of individual categories, instead of the location of the boundary between categories. Hence, the representative acoustical information given by prototypes facilitates the category learning.

One concern is whether the generalization effect of synthetic training materials is able to transfer to natural speech perception. Findings were consistent across the studies so far, showing that training effect observed in synthetic stimuli generalized to the natural speech stimuli (Giannakopoulou, Uther, & Ylinen, 2013; Jamieson & Morosan, 1986; Jamieson & Morosan, 1989; Nobre-Oliveira, 2008; Strange & Dittmann, 1984; Zhang et al., 2009).

Although repetitive stimuli with artificially modulated sound features leads generally to a greater training effect size on trained materials, some of the studies reported non-uniform generalizability effect towards novel speech materials (Logan et al., 1991; e.g. Strange & Dittman, 1984; McCandliss et al., 2002; Brosseau-Lapre et al., 2011). For instance, Strange and Dittman (1984) and McCandliss et al. (2012) only reported generalization effect to another novel synthetic continuum (e.g. from *road-load* to *rock-lock* in McCandliss et al., 2012), Iverson et al. (2005) reported non-uniform generalization effect to novel syllable positions and Brosseau-Lapre et al. (2011) reported limited generalization effect to stimuli produced by novel talkers.

In order to promote generalizability of training effect, stimulus variability seems to play a crucial role in synthetic training. Variability of synthetic training can be improved by including a wide range of training stimuli that vary in acoustical cues and phonetic context (e.g. Iverson et al., 2005). Referring to Table 1.1, some synthetic training studies reported retention of training effect up to 3 months (i.e. Wang & Munro, 2004 and etc.); while others did not look at long term benefits. The training duration across the trainings also varied greatly, ranging from less than an hour to 24 hours.

Table 1.1 Overview of perceptual phonetic training studies that used synthetic speech stimuli, detailing participants' L1, language trained, phonemic contrast trained, training method/ task, training duration (in hours), training effect/improvement size (in percentage), the retention of training effects and generalization of training effects.

Authors (Year)	**L1**	**Training language**	**Contrast**	**Task (Special Features)**	**Training Duration (h)**	**General Improvement (average %)**	**Training Effect Retention**	**Generalization Effect**
Pisoni et al. (1982)	English	N/P	/b/, /p/, /ph/	ID & DT	4	N/P	N/A	To the full stimulus continuum
Strange & Dittman (1984)	Japanese	English	/r/-/l/	DT	5	N/P	N/A	Limited transfer to another synthetic stimulus series (from rock-lock to rake-lake) To natural speech stimuli
Jamieson & Morosan (1986)	French	English	/θ/-/ ð/	ID (Perceptual Fading Method)	1.5	30 (synthetic stimuli) 11 (natural stimuli)	N/A	To natural speech stimuli
Jamieson & Morosan (1989)	French	English	/θ/-/ ð/	ID (with prototypes)	0.67	N/P	N/A	To natural speech stimuli
Flege (1989a)	Mandarin/ Taiwanese/ Shanghainese/ Hakka	English	/t/-/d/	ID	0.67	9	N/A	To novel words with different vowels

(1989b)	Mandarin/ Taiwanese/ Southern Chinese Dialect	English	/t/-/d/	ID	0.67	11	N/A	To novel words with different vowels
Kraus et al. (1995)	N/P	N/P	variants of /da1/ or /da2/	DT	6	11	1 month (12% Improvement)	N/P
Golestani et al. (2002)	English	Hindi	retroflex /ta/ - dental /da/	ID (Perceptual Fading Method)	1.5	N/P	N/A	N/P
McCandliss et al. (2002)	Japanese	English	/r/-/l/	ID	1	N/P	N/A	To another untrained synthetic continuum (e.g. from *road-load*, to *rock-lock* or vice versa) when feedback was provided.
Golestani & Zatorre (2004)	English	Hindi	retroflex /ta/- dental /da/	ID (Perceptual Fading Method)	5	N/P	N/A	N/P
Wang & Munro (2004)	Mandarin/ Cantonese	English	/i/-/ɪ/, /ɛ/-/æ/, /ʊ/-/u/	ID (Perceptual Fading Method)	24	21	3 months (16% improvement)	To novel words To novel talkers
Iverson et al. (2005)	Japanese	English	/r/-/l/	ID (All Enhancement, Perceptual Fading, Secondary Cue Variability)	5	18	N/A	To novel words To novel talkers. Failed to generalize to novel syllable positions.

Moore et al. (2005)	English (4 speak unspecified L1)	English	/b/-/d/, /d/-/g/, /ɛ/-/a/, /er/-/or/, /i/-/e/, /l/-/r/, /m/-/n/, /s/-/sh/, /s/-/th/, /v/-/w/, /a/-/ʌ/	DT (Adaptive Staircase Procedure)	6	N/P	5-6 weeks	N/P
Nobre Oliveira (2008)	Brazilian Portuguese	English	/i-ɪ/, /ɛ-æ/, /ʊ-u/	ID	4.5	14 (synthetic stimuli) 11 (natural stimuli)	N/A	To natural speech stimuli To novel syllable structure
Zhang et al. (2009)	Japanese	English	/r/-/l/	ID (Adaptive Staircase Procedure; with visual articulation cues)	12	22	N/A	To natural speech stimuli with different vowels To novel talkers
Brosseau-Lapré et al. (2011)	English	Quebec French	/ə/-/ø/	ID	1	N/P	N/A	To novel words Failed to generalize to novel talkers
Perrachione et al. (2011a)	American English	Mandarin (Pseudowords with pitch contours)	3 pitch contours (level, rising and falling)	ID	8	N/P	N/A	To novel talkers

(2011b)	American English	Mandarin (Pseudowords with pitch contours)	3 pitch contours (level, rising and falling)	ID	8 or 32	N/P	N/A	To novel talkers
Giannakopoulou et al. (2013)	Modern Greek	English	/i/-/ɪ/	ID	5	N/P	N/A	To the untrained modality of stimuli (from synthetic to natural or vice versa) To novel talkers

Remarks: ID – Identification Task, DT – Discrimination Task, N/P – Not Provided, N/A – Not Applicable

1.3.3.3 Natural Speech Training (High Variability Phonetic Training, HVPT)

The phonetic cues available in synthetic tokens may be considered somewhat impoverished because of the absence of acoustical variability that can only be found in natural speech (e.g. variability induced by differences in speakers due to the different shapes of articulators, pronunciation habits, speaking rate and etc.). Therefore, a great number of studies conducted post-90s started to examine the effect of phonetic training using natural speech stimuli (see Table 1.2).

In the early 90s, Lively and colleagues emphasized the effectiveness of using natural speech stimuli in phonetic training and developed a highly effective phonetic training procedure, known as High Variability Phonetic Training (HVPT) (Logan et al., 1991; Lively et al., 1993; Lively et al.,1994). Using the HVPT paradigm, they trained native Japanese speakers (experiment 1) on the English /r/-/l/ contrast with highly variable speech stimuli using an ID task that provided feedback during training after each response. The high variability speech stimuli used in HVPT consisted of speech recordings naturally produced by multiple speakers. In experiment 2, native Japanese speakers (Lively et al., 1993) were trained with an identical training regime, but this time with training stimuli produced by a single speaker (low variability). The Japanese speakers trained with high variability training materials showed significant training and generalization effect but speakers trained with low variability training materials did not. Speakers from the low variability training regime were found to develop highly detailed representations of trained stimuli when identification improvement observed in trained stimuli did not generalize to novel stimuli and speakers. The term “novel stimuli” will be used to refer to stimuli or words that share the same phonological category membership as the training materials, but are not used during the training; while the term “novel speakers” refer to the speakers whose productions are not included as the training materials.

Phonetic training related studies published post-1994 often used training procedure modelled after Lively and colleagues' HVPT and focused more on naturalistic training. Table 1.2 presents an overview of 17 articles found in the literature from 1991 to 2013, which used an adapted HVPT paradigm and natural speech stimuli produced by multiple speakers (number of speakers range: 2-10, mean 5.65) in their perceptual phonetic training. The studies vary in a number of aspects, such as in the training methodology (e.g., stimuli, task and training time), participants' language background and how the results are reported. The variation renders difficulty in comparing the effectiveness across training methods. Almost all studies from Table 1.1 (94%) trained L2 English speakers on English phonemic contrasts. The most common training materials consisted of English /r/-/l/ contrast (42%), vowels contrasts (42%), and other consonant contrasts such as /p/-/b/ and /t/-/d/ (10%). The study by Wang (2013) however, trained naive participants on Mandarin tones; while Werker and Tees (1984) trained L1 English speakers on Hindi and Thompson consonant contrasts.

Half of the training studies targeted L2 English Japanese speakers who reported to have poor speaking and listening proficiency with the target language. The remaining studies trained on other L1 populations (e.g. Mandarin, Greek or French) who reported to be intermediate to advanced L2 speakers of the target language.

From Table 1.2, a change in training trend can be seen in the last decade where research has focused more on training English vowels perception instead of to the consonants. The change might also be induced by the change of subject interest across the years. In the early years, a heavy interest was observed in improving Japanese speakers' perception of the English /r/-/l/ contrast. Starting from the mid-2000s, the research interest has moved to study the effect of phonetic training on the English vowels perception of L2 English speakers. Those L2 English speakers were reported to have a higher English proficiency level compared to the Japanese speakers.

Training benefits in perceptual phonetic training were reported by all studies in Table 1.2. Findings from these studies suggests that high-variability training using natural speech stimuli encourages a long-term modification on L2 speakers' phonetic perception that can be generalized to novel stimuli produced by novel speakers. Logan et al. (1991) proposed that stimulus variability promotes attention re-allocation to criterial acoustic cues across a range of speech exemplars of the same phonological category, thus the generalization of training effect observed in the high variability training.

Two studies reported that training benefits last up to six months (Lively et al., 1994; Wang & Munro, 2004). Training benefits were also found to be generalized to novel stimuli (e.g., Lively et al., 1994; Hazan et al., 2005), produced by novel speakers whose voice were not heard during the training (e.g., Iverson, Hazan & Bannister, 2005; Nishi & Kewley-Port, 2007). Lively et al. (1993) failed to find generalization effects with novel stimuli from participants who were trained with single speaker stimuli (low variability). Therefore, Lively et al. (1993, 1994) concluded that high variability training materials are essential to promote generalization of training benefits. Although the training benefits of phonetic training appears to be well-established, the training procedure often took days if not weeks to be completed (mean training time 10 hours for studies reported in Table 1.2) before improvements in phonetic identification or discrimination tasks could be observed.

Table 1.2 Overview of perceptual phonetic training studies that used natural speech stimuli, detailing participants' L1, language trained, phonemic contrast trained, training method/ task, training duration (in hours), training effect/improvement size (in percentage), the retention of training effects and generalization of training effects.

Authors (Year)	**L1**	**Training Language**	**Contrast**	**Task (Special Features)**	**Training Duration (h)**	**General Improvement (average %)**	**Training Effect Retention**	**Generalization Effect**
Werker & Tees (1984)	English	Hindi & Thompson	Hindi /ṭa/–/ta/ & Thompson /`ki/-/`qi/	DT	N/P	N/P	N/A	N/P
Logan et al. (1991)	Japanese	English	/r/-/l/	ID	10	8	N/A	To novel words To novel talkers
Lively et al. (1993a)	Japanese	English	/r/-/l/	ID	7.5	6	N/A	To untrained phonetic environment or position To novel words To novel talkers
(1993b)	Japanese	English	/r/-/l/	ID	10	N/P (Training effect found only in some phonetic environments)	N/A	Failed to generalize to novel words and speakers
Lively et al. (1994)	Japanese	English	/r/-/l/	ID	10	12	6 months (5% Improvement)	To novel words To novel talkers (but old talkers > novel talkers)
Bradlow et al. (1997; 1999)	Japanese	English	/r/-/l/	ID	22.5	16	3 months	To novel words To novel talkers

Callan et al. (2003)	Japanese	English	/r/-/l/	ID	22.5	18	N/A	To novel words To another phonemic contrast (/b-v/)
Wang & Munro (2004)	Mandarin/ Cantonese	English	/i/-/ɪ/, /ɛ/-/æ/, /ʊ/-/u/	ID (Perceptual Fading Method)	24	21	3 months	To novel words To novel talkers
Hazan et al. (2005a)	Japanese	English	/v/-/b/-/p/	ID (Auditory Training vs. Visual Training vs. Audiovisual Training)	6.7	17 (Audio and Audio-visual training conditions only)	N/A	To nonsense words that contain trained phonemes, from real words (training materials)
(2005b)	Japanese	English	/r/-/l/	ID (Auditory Training vs. Audiovisual Training with natural face vs. Audiovisual Training with synthetic face)	6.7	18	N/A	To nonsense words that contain trained phonemes, from real words (training materials)
Iverson et al. (2005)	Japanese	English	/r/-/l/	ID	5	18	N/A	To novel words To novel talkers Failed to generalize to novel syllable positions
Nishi & Kewley-Port (2007)	Japanese	American English	/iː/, /ɪ/, /ɛ/, /æː/, /ɑː/, /ʌ/, /ɔː/, /ʊ/, /uː/ or /ɑː/, /ʌ/, /ʊ/	ID	13.5	26 (Group that trained on full set of stimuli only)	3 months (24% Improvement)	To novel words and novel talkers (but only in the group trained on the full set of stimuli)

Nobre Oliveira (2008)	Brazilian Portuguese	English	/i/-/ɪ/, /ɛ/-/æ/, /ʊ/-/u/	ID	4.5	14 (Synthetic stimuli) 11 (Natural stimuli)	N/A	To novel talkers
Richie & Kewley-Port (2008)	American English	English	/i/-/ɪ/-/e/-/ɛ/-/æ/-/ɑ/-/ʌ/-/o/-/ʊ/-/u/	ID	6	29	N/A	To novel talkers
Aliaga-Garcia & Mora (2009)	Catalan/ Spanish	English	/p/-/b/, /t/-/d/, /æ/-/^/, /i/-/I/	ID & DT Phonetic transcription Exposure to Native Speakers' sounds Production Training (Perceptual Fading Method)	12	9 (Perception of vowels only)	N/A	N/P
Handley, Sharples & Moore (2009)	Mandarin	English	/r/-/l/	ID or Oddity DT	N/P	N/P	N/A	To novel words To novel talkers
Lengeris & Hazan (2010)	Greek	English	/ɛ/-/ɑ/-/a/-/ʌ/, /i/-/ɪ/-/aɪ/-/eɪ/, /ɒ/-/əʊ/-/ɔ/, /u/-/aʊ/-/ɜ/	ID (Adaptive Staircase Procedure)	3.75	17	N/A	To novel talkers To speech-in-noise condition (training in quiet laboratory)

Iverson et al. (2012)	French	English	/ɛ/-/ɑ/-/a/-/ʌ/, /i/-/ɪ/-/aɪ/-/eɪ/, /ɒ/-/əʊ/-/ɔ/, /u/-/aʊ/-/ɜ/	ID (Adaptive Staircase Procedure)	6	21 (Inexperienced speakers) 17 (Experienced speakers)	N/A	N/P
Giannakopoulou et al. (2013)	Modern Greek	English	/i/-/ɪ/	ID	5	N/P	N/A	To the untrained modality of stimuli (from natural to synthetic) To novel talkers
Wang X. (2013)*	Hmong/ English/ Japanese/ Spanish/ Khmer/ Tagalog	Mandarin	Mandarin tone contours	ID Imitation (with visual feedback)	6	21 (Auditory training only) 11 (Auditory training with visual feedback on production)	N/A	To novel words To novel talker

Remarks: ID – Identification Task, DT – Discrimination Task, N/P – Not Provided, N/A – Not Applicable

1.3.3.4 Synthetic Training vs. Natural Speech Training

Overall, significant training effects are reported by both synthetic training and natural speech training. The generalization effect to novel tokens and novel speakers is usually observed as long as high variability training materials are used in the training. In natural speech training, variability of training materials can be easily attained by including training materials produced by different speakers of different genders. No speaker will produce identical speech sounds due to the uniqueness of vocal tract shapes and articulation habits. Having natural speech sounds produced by multiple speakers introduces sufficient variability to the training materials. However, manipulating acoustical properties of the speech sounds in synthetic training would limit the variability. To compensate the loss of acoustical manipulation and to enrich the variability of training materials, synthetic training can involve more stimuli that vary in acoustical properties and phonetic context (i.e., different phonetic position in words; e.g., phoneme /t/ at word final position - sa*t* and word initial position - *t*ame); which in turn results in more intensive training that requires longer sessions to be completed.

The generalization effect seems to be more uniform when using natural speech materials while two studies of synthetic training reported a lack of generalization to novel syllable positions (Iverson et al., 2005) and to novel talkers (Brosseau-Lapré et al., 2011). The difference might be partially attributed to the limitation of stimuli variability. For instance, Brosseau-Lapré et al. (2011) used only four minimal pairs (8 single words) in their relatively short 1 hour training; while Iverson et al. (2005) used only words contrasting /r/ and /l/ phonemes at word initial position in their training. Up to now, studies that compared the efficiency of synthetic and natural speech trainings found both methodologies to be equally effective (Iverson et al., 2005; Nobre-Oliveira, 2008).

If synthetic training does not produce additive gains in performance after training, natural speech training might be the most efficient method because signal processing techniques used in synthetic training require intensive labor efforts. Giannakopoulou et al.'s (2013) study however showed an advantage of synthetic training. Participants who were trained with synthetic materials identified and discriminated synthetic materials better than participants who were trained with natural speech materials; while both groups performed similarly for the natural speech materials. The advantage with synthetic materials nevertheless, may not be relevant outside of laboratory settings as the encounter of synthetic speech sounds is rare in the daily life communication context.

1.3.3.5 Effects of Training on Neural Processing

Phonetic training does not affect behavioral performance solely, but it also results in neurological changes in post-test performance. Substantial neural plasticity was observed along with L2 phonetics learning whereby relative brain areas showed different neurophysiological activation for the perception of newly learnt L2 contrast (e.g. Zhang et al., 2009). For example, Kraus et al.'s (1995) study found greater mismatch negativity event-related potential in post-test compared to pre-test, when different stimuli were presented in a pair. The changes in neurophysiologic responses suggested that participants learnt to discriminate the two similar sounds after training.

Callan et al.'s study (2003) also showed greater activation in the right hemisphere after successful phonetic training. This brain area is usually associated with native language acquisition. Golestani and Zatorre's (2004) study showed suppression of the left middle temporal region and the angular gyri after L2 phonetics learning. The left middle temporal region usually shows activation during language-related tasks that involve cognitive recognition while the angular gyri show activity in a perceptual noise baseline. In other words, suppression in the particular areas may indicate a reduction in inappropriate cognitive interference (e.g., deep-rooted L1 language system) and in the perceived noise level from the L2 speech input after phonetic training.

1.3.3.6 Factors Affecting Training Effect

The effectiveness of phonetic training does not seem to rely on the amount and period of exposure to L2 sounds, but rather on the training's effectiveness in directing listeners' attention to the relevant phonetic cues native speakers attend to. For instance, studies such as the Flege (1989), Aliaga-Garcia and Mora (2009), and Perrachione et al., (2011) showed that intensive training with more training trials did not generate greater improvement in discriminating the non-native phonetic sounds. In fact, Perrachione et al.'s (2011) study implied that the design of phonetic training rather than the amount of exposure is more crucial in generating significant learning in the target language. In their studies, the low aptitude learners did not benefit from the high variability phonetic training but showed impairments in learning after being trained, which is unusual. When Perrachione et al. (2001) reduced the variability of training materials within each experimental block, the low aptitude learners improved in their perceptual performance after training. Simply multiplying the number of training trials did not result in significant improvement in the low aptitude learners' perception performance. As such, Perrachione and colleagues concluded that the efficiency of training paradigms would also be determined by the aptitude of the training population towards the training language.

Previous studies have demonstrated that the laboratory training was most effective when training stimuli with high variability were involved; the stimuli can be presented in various phonetic environments and produced by multiple speakers (e.g. Logan et al., 1991; Bradlow et al., 1997; Bradlow et al., 1999). Training effects generated by this kind of training design were more likely to be generalized to novel words produced by novel speakers. The importance of stimulus variability in phonetics studies was even showed in a synthetic training study conducted by Brosseau-Lapré and colleagues (2011). In their study, adult speakers of English were trained to discriminate the synthesized French /ə/ and /ø/ vowel contrast in three different stimuli conditions. Participants were either trained with a single prototypical exemplar from each vowel category (prototype condition - low variability), multiple tokens closed to the prototypes to maximize the differences between categories (close condition – high variability) or multiple tokens closed to the category boundary to minimize difference between categories (far condition – high variability). Significant perceptual improvement was found only when the training stimuli were highly variable (i.e. in close and far conditions and only when stimuli were productions by multiple speakers). The training effect however did not generalize to stimuli spoken by novel speakers.

1.3.4 Implications for Perception Training

Most of the training studies target English as the training language due to its global lingua franca status. It serves as a common means of communication between speakers of different L1 backgrounds and is widely spoken all around the world (Jenkins, Cogo, & Dewey, 2011). As the most common second language, research on enhancing the learning experience of English would be of the highest interest. Other than the core language trained, the training materials (i.e. phonemic contrasts) used are also quite finite. Up to now, the most widely researched contrasts are the English consonant pair /r/-/l/ and the English vowel pairs.

All the existing training studies aimed to improve competence of L2 speakers in their perception ability. None of them applied the training to examine its effect in improving accented language perception. With English functions as a lingua franca, conversations often take place among speakers who speak different varieties of English (e.g., British English, Malaysian English and Indian English), especially in occasions where people of different nationalities gather for a common purpose (e.g., academic or business conferences; Jenkins et al., 2011; see section 6.4.2 for further discussion). Even among proficient speakers, unfamiliar accents sometimes render comprehension difficulty due to the different manners of intonation and pronunciation. Since it is rather impractical to unify and have only one language variety, phonetic training could be useful in familiarizing speakers with the perception of new unfamiliar accents.

The existing training studies have trained L2 speakers with speech materials which they have difficulty discriminating, but to the best of our knowledge, none of the studies conducted training in difficult environments or background noise settings. One of the most prominent common weaknesses of L2 speakers is the ability to conduct conversation fluently with the presence of noise. Although numerous studies have been conducted to investigate the effect of adverse condition on L2 speakers' perceptual performance (see Lecumberri et al., 2010 for review), it is notable that there have been few studies examining the effect of simulated adverse conditions on phonetic training.

Some studies did present stimuli under adverse conditions but not as a way to improve training effectiveness. For instance, Richie and Kewley-Port (2008) used a speech-shaped stationary noise with spectrum modulated to simulate hearing loss in listeners with normal hearing, whereas Lengeris and Hazan (2010) incorporated background noise only in pre- and post-tests but not during training to examine the effect of training-in-quiet had on speech-in-noise L2 perception. The type of background noise used in their study was not fully described. As far as we know, only Jamieson and Morosan (1986) used background "cafeteria" noise to increase perceptual difficulty during the latter phase of training, although no further attention was given to the implication of using background noise.

Training under difficult conditions might better tackle and improve L2 speakers' weakness in L2 communication. Hence, the purpose of this thesis was to examine the effect of phonetic training under adverse condition (i.e. with background babble), using a newly designed training paradigm known as Spoken English Discrimination (SED) training.

1.4 Research Questions

The present thesis aims to address the following questions. Firstly, the effects of background babble implementation in phonetic training (i.e. SED) was examined on the perceptual accuracy of listeners under adverse condition (Chapters 2 & 3). The thesis also examined whether manipulating the background babble using adaptive staircase procedure would further enhance SED training effect. The longevity of SED training effect and its generalizability to novel words and speakers were examined in chapter 3 and 5.

Perception pattern of speakers from different linguistic background was examined on different varieties of English (Chapter 4). L1 Malaysian English and native British English speakers' perceptual accuracy on British English spoken words and Malaysian English spoken words were compared. The speakers were expected to show better identification accuracy on spoken words of English variety they familiar with. The effects of SED training was later examined on speech production (Chapter 5). Improvement in perceptual sensitivity was expected to generalize to production domain.

Chapter 2: Spoken English Discrimination (SED) Training on Voiced and Voiceless /t/-/d/ Phonemic Contrast Perception

2.1 Introduction

From this point onwards, the term first language (L1) refers to the language one first acquired since birth and speakers are assumed to possess a native-like competence in their L1 as the terms used in the literature can be inconsistent across studies. Speakers could speak a variety of languages since birth at home (e.g., Hokkien, Cantonese and Tamil) but speak other formal languages more dominantly and more proficiently in the later stage of life (e.g., Mandarin, English and Malay). For instance, speakers can claim to have native-like competence in languages (e.g. Mandarin) other than their L1 which is not commonly spoken in the community (e.g. Hokkien and Hakka). In that case, L1 is used to refer to the self-reported most proficient language. On the other hand, the term second language (L2) will be used to describe any other languages acquired after L1. According to Lecumberri et al. (2010), the term L2 is usually applied to the language learnt after the L1 that has been fully established or learnt post-puberty. To be qualified as a L2, it also needs to be used widely among the community when the person acquires it. Malaysians start learning their L2s latest by the age of 7 due to the nation's education system; henceforth their age of acquisition for the L2s would be much earlier than the L2 learning time suggested by Lecumberri and colleagues. Because the usage of the term L2 in this thesis does not fit with the L2 description by Lecumberri et al. (2010), speakers' self-rated proficiency on each language spoken is also used to assist their L1 and L2 identification.

2.1.1 Phonetic Training

Previous research has shown that perceptual training can improve speakers’ ability to discriminate non-native phonetic sounds. The training re-shapes speakers’ L2 phonetics perception by shifting their attention to the meaningful acoustic cues that can help distinguish between the phonetic sounds (e.g., Flege, 1989; Logan et al., 1991; Bradlow et al., 1997; Bradlow et al., 1999). Numerous neurophysiological studies have also shown learning-induced enhancement in neural sensitivity corresponding to the formation of new phonemic categories in L2 speakers after phonetic training (Callan et al., 2003; Golestani et al., 2002; Golestani & Zatorre, 2004; Kraus et al., 1995; Y. Zhang et al., 2009).

Werker and Tees (1984) claims that the language experience effects from phonetic training are only able to modify higher-level cognitive processes (e.g. phonological encoding or memory retention) that are malleable, but not lower-level sensorimotor sensitivity that is shaped earlier in life. This claim was refuted when studies showed that training tasks that require minimum memory loads (rely less on cognitive processes and more on the underlying auditory sensitivity; e.g., Kraus et al., 1995; Golestani et al., 2002; Golestani & Zatorre, 2004) can also improve listeners' discriminative ability. These studies trained listeners with synthetic non-linguistic speech sounds so that listeners did not need to refer to the phonetics representation in memory and performed based on the acoustical properties perceived from the stimuli. The non-linguistic speech stimuli are not real words in any language and often just comprise of one consonant and one vowel (i.e., CV).

2.1.1.1 Effects of Phonetic Training

Previous studies investigating phonetic training have observed that training can result in perceptual improvements that generalized to novel words that have not been used in the training (e.g., Lively et al., 1993) and to novel talkers whose productions are only used in pre- and/or post-tests, but not in the training (e.g., Iverson et al., 2005). Some studies even found generalization effects across speech sounds of different modality; for instance, training effect of synthetic speech training generated to naturally produced speech sounds (e.g., Zhang et al., 2009) or training effect on naturally produced speech sounds generalized to nonsense words that consist of the training phonemes (e.g. Hazan et al., 2005). The improvement in phonetics perception also generalized to speech production in some studies (e.g., Bradlow et al, 1999; Iverson et al., 2012).

Although the training benefits of phonetic training are prominent across the literature, the training does not benefit all language groups equally. The effectiveness of training depends on the similarity between speakers' native language and the training language used. In a study by Flege (1989), phonetic training was selectively more effective to the speakers whose native language shares a more similar structure or phonological rules with the training language, when compared to those whose native language's structure is very different to the training language. Hence, Flege indicated that even after training, native influences or more specifically, syllable-processing strategies established during native language acquisition could not be diminished completely. The deep-rooted native influence will continue to affect later L2 learning.

The effectiveness of phonetic training also varies across the training materials. Numerous studies reported significant improvement only in some of the phonetic contexts. For instance, small yet significant improvement was found in post-test by Handley, Sharples and Moore (2009) when the English /r/-/l/ contrast is located at word final position as singleton phoneme (e.g., *war-wall*), but not as clusters (e.g., *cord-called*). Surprisingly, their findings were reversed when the training effect was examined on novel words; significant improvement was found when English /r/-/l/ contrast is located at words final position as clusters, but not as singletons. They attributed the absence of improvements in some parts of the training to the relatively restricted set of phonetic contexts (i.e. /r/-/l/ contrast appear as the final clusters and singletons that are preceded by the vowel /ɔː/) which Mandarin speakers have difficulty discriminating. In a related study by Iverson, Hazan and Bannister (2005), a greater discriminative improvement was found when the /r/ and /l/ contrast is at word initial position rather than when it is at word final position. Aliaga-Garcia and Mora (2009) on the other hand, found inconsistency in training generalization towards the production domain; improvements in production were found in consonants but not in vowels.

The effectiveness of phonetic training seems to rely on the phonetics dimension trained and the difficulty nature and level the speakers experienced. Hence, choosing the right training materials for the target language population would be critical in generating perception improvement because listeners' perceptual limitation can override the benefits of the training (Perrachione et al., 2011).

2.1.2 Difficulties in Speech Perception under Adverse Conditions

Differences in speech perception ability between L1 speakers and proficient L2 speakers has been found to be absent in quiet environments (without or limited background noise), but they become more prominent in adverse conditions where a high level of background noise is present (Mayo et al., 1997; Lecumberri & Cooke, 2006; for review see Lecumberri et al., 2010). In the present chapter, we examined the effects of background noise on perceptual phonetic training. We implemented multi-talker babble in the training to create a naturalistic yet moderately challenging training context.

2.1.3 Rationale of Present Study – Introducing the Spoken English Discrimination Training

This thesis focuses on improving speech perception because perception is believed to be the most fundamental stage that builds up language learning, including speech production. Lecumberri and colleagues (2010) claim that phonetics perception, the most basic segment of a spoken word, plays a crucial role in word perception as it rules out other unmatched words and determines the right word heard. Hence, the ability to detect differences between phonemes plays a vital role when comparing native and non-native listeners' ability in perceiving L2 speech. Almost all infants show a universal phonetic capacity since birth, where they can discriminate all types of speech sounds regardless of their functions in different languages (Kuhl et al., 2006). However, the phonetics perception competence attenuates over time when some phonetics is no longer phonologically distinctive in the language learning environment of the infants.

The present study introduces a new phonetic training program: Spoken English Discrimination (SED) training that combines a number of key training features to create a single computerized individual training program. SED training is a variant of High Variability Phonetic Training (HVPT) with newly added features: 1) the use of natural recordings of the target phonemic contrasts, 2) stimuli spoken by multiple speakers, 3) background noise in the form of multi-talker babble and 4) an adaptive staircase procedure in which the background noise level for training is set by manipulating the background noise volume to yield 79% correct auditory discrimination of participants.

Realistic background noise (i.e., multi-talker babble) was included due to the expectation of greater training benefits when listeners are trained under circumstances which they find difficult. The adaptive staircase procedure introduced in SED training is used to measure a person's sensory capabilities by finding out his/her threshold or limits to detect and discriminate similar and confusable physical stimuli (Leek, 2001). The adaptive staircase procedure adjusts the level of background babble by decreasing the babble's volume with each incorrect response and increasing the volume with each correct response. The adjustment stops after a few reversals (changing from ascent-to-descent or descent-to-ascent) and a threshold level will be calculated by averaging the babble's volume from the last consecutive reversals. This method allows us to identify each participant's discrimination threshold to perform at a desired percent correct discrimination accuracy level. The threshold level will then be used during the subsequent training phase.

This experiment was designed based on the study by Lively et al. (1993). We examined the importance of variability in phonetic training. In SED training, listeners were either trained with single-speaker or multiple-speaker training materials. In accordance to Lively and colleagues, listeners who are trained with multi-speaker training materials should show greater training benefits that could be generalized to novel words and novel talkers.

A discrimination task (DT) was used in the training sessions. The perceptual changes in participants were then measured using identification task (ID task) in pre-test and post-tests.

2.1.3.1 Training Materials

The present study examined the effects of SED training on L1 Mandarin Malaysian Chinese speakers who speak Malaysian English as L2. The English /t/-/d/ phonemic contrast was chosen as SED training materials in this study due to its discriminative difficulty among the L1 Mandarin speakers. Previous research has shown that discriminating the phonemic contrast /t/-/d/ at the word final position is difficult for Chinese speakers because this contrast does not exist as the word final obstruent in Chinese languages including Mandarin (see section 4.1.1.2 for more details; Flege, 1989; Aliaga-Garcia & Mora, 2009). Similarly, words from the Malay language (another L2 spoken by the training population) also do not end with voiced stop consonants (including /d/), except for borrowed words from other languages such as the Arabic (Hassan, 2005). Malay words ending with the letter "d" are often devoiced or neutralized. As their L1 and one of the L2s do not encourage /t/-/d/ word final discrimination, L1 Mandarin speakers targeted in this study were expected to have reduced perceptual sensitivity towards the word final contrast compared to the L1 English speakers.

The English /ɛ/-/æ/ phonemic contrast is also included during pre-test and post-test to examine generalization effect between contrasts. The contrast is difficult to discriminate by the multilingual Malaysian English speakers because of its ambiguous categorical representation in Mandarin (L. M. Huang, 1992), and Malay (Yap, Liow, Jalil, Faizal, & Louis, 2010; see Figure 4.1, Figure 4.2 and Figure 4.3 for the vowel charts of Malay, Mandarin and Received Pronunciation English respectively) as well as Malaysian English (Pillai, Don, Knowles, & Tang, 2010). A previous study by (Yap, Wong, & Aziz, 2010) had also found that L1 Malay Malaysian speakers unconsciously assimilate the phoneme /ɛ/ (e.g., bet) into the phoneme /æ/ (e.g., bat). Learning from the phonetic training would reconstruct the underlying phonological representation of the vowels by shifting the existing category boundaries or forming new and more accurate phonological category for the newly learnt vowel phonemes.

2.1.3.2 Variability of Training Materials

Conversation flow between English as Lingua Franca (ELF) speakers can be disrupted when speakers fail to pick up spoken words of another party. The confusion arises when speakers fail to discriminate and identify spoken words that are unfamiliarly pronounced. Therefore in this study, we trained moderately to highly proficient Malaysian English speaking participants to discriminate similar phonemic sounds of an unfamiliar variety of spoken English, British English. Participants heard spoken English words in British English accent in pre- and post-tests, but trained with words in Southern Irish English, Irish English and American English accents consecutively across three training sessions. SED training was expected to improve participants' sensitivity and ability in picking up relevant acoustic cues spontaneously for future phonemic words identification, even from a different variety of spoken English (i.e. British English) which was not used as the training material. SED training was designed to improve listeners' general auditory sensitivity (especially around the phonological space that was either lost or restrained due to the influence of L1 phonological system), rather than forming new and intractable phonemic categories for a single language variety learnt from training. This type of perceptual learning has greater training benefits in terms of its generalizability to general speech perception.

An ideal phonetic training paradigm should be tailored to benefit language speakers with different levels of proficiency and language backgrounds. Furthermore, the training should be able to effectively improve the learners' perceptual sensitivity towards a wide range of speech varieties (generalizability), within a minimum amount of time. Considering the wide varieties of English spoken all around the world, SED training's primary purpose is to improve communication between speakers of different varieties of English, at different levels of proficiency.

2.1.3.3 Adaptive Staircase Procedure

The adaptive staircase procedure is a psychometric method used to measure a person's sensory capabilities, by finding out his/her threshold or limits to detect and discriminate similar and confusable physical stimuli (Leek, 2001). The adaptive staircase procedure has been previously implemented in a few phonetic training studies using synthetic stimuli (McCandliss et al., 2002; Moore et al., 2005). The acoustic properties of stimuli were carefully manipulated to produce speech stimuli that systematically differ with other members in a consistent step size. Difficulty level or the distinctiveness of the stimuli pairs can then be manipulated by pairing stimuli from different step sizes. Moore et al. (2005) and McCandliss et al. (2002) used the adaptive staircase procedure to determine the magnitude of distinction between stimuli presented based on participants' preceding performance. In order to ensure that perceptually distinguishable stimuli were presented across the course of learning, participants of their studies were initially trained with highly distinguishable stimuli before proceeding with acoustically closer stimuli.

None of the previous natural speech training studies used an adaptive staircase procedure as it would require a systematic manipulation of the speech stimuli which would affect the naturalistic and variability of the training materials. In SED training, the adaptive staircase procedure was used to set the background babble volume to increase or decrease phonemic discrimination difficulty, without affecting the naturally produced speech stimuli.

The aim of the study is to: 1) investigate if SED training enhances auditory discrimination skills, 2) to examine whether implementing background babble using an adaptive staircase procedure in SED training would generate greater training effects 3) to examine effects of training variability in phonetic training using SED training and lastly 4) to examine SED training effect on L1 Malaysian English speakers and L1 Mandarin speakers.

2.2 Experiment 1 (SED Training) - Method

2.2.1 Design and Materials

The stimuli consisted of three groups of English minimal pairs (see appendix 1 for full stimuli list): Sixteen /t/-/d/ minimal pairs of which eight differ at the initial position (e.g., *tame–dame*) and eight at the final position (e.g., *sat–sad*), and sixteen /ɛ/-/æ/ minimal pairs (e.g., *leg–lag*). The sixty-four English words consisted of nouns and verbs and the minimal pairs across the three groups were similar in word frequency (/t/-/d/ initial: 117.72, /t/-/d/ final: 101.32, /ɛ/-/æ/: 172.34 occurrences per million based on SUBTLEX-US, $F < 1$; Brysbaert, New, & Keuleers, 2012) and number of syllables (/t/-/d/ initial: 1.0, /t/-/d/ final: 1.0, /ɛ/-/æ/: 1.3; $F < 1$). The /t/-/d/ final minimal pairs were shorter than the other minimal pairs in terms of the number of letters (/t/-/d/: 3.75, /t/-/d/: 3.50, /ɛ/-/æ/: 4.09).

During the training sessions, the /t/-/d/phonemic contrasts were split into two groups. Half of the participants were presented with half of the words from the /t/-/d/ initial and final minimal pairs, whereas the other half of the participants were presented with the other group of /t/-/d/ initial and final minimal pairs. The words of the minimal pairs were spoken by four speakers (2 females) with different English accents (female British English, male Southern Irish English, female American English and male Irish English). Their word productions were recorded in an Anechoic chamber using a microphone (AKG Perception 400) and a 24-bit/96k FireWire recording interface (Presonus FireBox) connected to an Apple Macbook Pro. Speech was recorded at 44.1.hKz (sampling depth: 16 bit) using Amadeus Pro (version 2). Recordings were edited using Amadeus Pro and the volume of each word recording was equated using Amadeus Pro by amplifying the sound recordings of each speaker to an average root mean square (RMS) power of -25 dB (200 ms window).

The background noise consisted of 6 talker babble and was created by combining the audio recording of 6 native English speakers (3 females) taken from six BBC Radio 4 interviews in which they talked about their life and work. The interviewer's voice was edited out and the volume of each speaker was equated using Amadeus Pro by amplifying the sound to an average root mean square (RMS) power of -25 dB (200 ms window). The resulting 6 audio files were combined into a single 6 talker babble mono audio file (44.1 kHz, 16 bit) of 6 minutes.

In pre-test and post-test, the multi-talker babble was played simultaneously with the stimuli words, at half of the stimuli level and an approximate signal-to-noise ratio[2] (SNR) of 3 db (SNR range from 3.78 to 6.84 db depend on the stimuli word, mean SNR 2.63 db). For the training paradigms that implement background babble at constant level, multi-talker babble during the training phase was played at a lower level than the pre- and post-tests, which is 10 percent of the stimuli level with an approximate SNR of – 37 db. For paradigms that implemented background babble using adaptive staircase procedure, the background babble level could be reduced to 0 (SNR ~ -68 db) or increased up to 100 percent (SNR ~ 9 db) of the stimuli level, depending on the participants' performance in the pre-training task.

Two features of SED training were manipulated in this study to identify the most effective paradigm for phonetic training. The participants were either trained with single speaker or multiple speakers, with background babble implemented either at a constant level or with adaptive staircase procedure. The manipulation resulted in four different training paradigms (see Table 2.1). Identification accuracy of participants in pre- and post-tests was then compared to examine (any) training effects.

[2] Signal-to-noise ratio (SNR) was calculated by subtracting the mean of stimuli intensity (db SPL) from the mean of multi-talker babble intensity (SNR = Mean Intensity of Noise – Mean Intensity of Stimuli). The intensity level of stimuli and multi-talker babble were measured using Praat software (Boersma & Weenink, 2007).

Table 2.1 The four training paradigms.

Training Features	Constant Background Babble	Adaptive Staircase Background Babble
Single Speaker	Paradigm 1	Paradigm 2
Multiple Speakers	Paradigm 3	Paradigm 4

2.2.2 Participants

68 Malaysian Chinese were recruited from the University of Nottingham Malaysia campus but only data from 56 participants was included in data analysis due to failure in meeting the criteria of this study (e.g. have a near perfect score in pre-test, incomplete data set etc.). Participants were divided into two groups based on their L1. 17 participants were classified as L1 Malaysian English (ME) speakers while 39 participants were classified as L2 ME speakers who speak L1 Mandarin or other Chinese languages (i.e., 2 Hokkien, 2 Cantonese and 1 Teochew). 14 participants were alternatively assigned to each experimental paradigm with the number of L1 and L2 ME speakers matched across the groups. Therefore, each paradigm consisted of 10 L2 ME speakers and 4 L1 ME speakers; except for paradigm 2 that consisted of 9 L2 ME speakers and 5 L1 ME speakers. All participants reported to have normal eye sight or corrected to normal vision, and no history of any hearing, speech and reading disorder.

The language background of all participants was assessed using a language background questionnaire (see Appendix 2). Table 2.2 presents an overview of their subjective English proficiency, the age they first acquired English (AoA), and their International English Language Testing System (IELTS) standard scores[3] according to their paradigm group.

[3] International English Language Testing System (IELTS) standard scores (some were converted from Malaysia UCLES 1119 Syllabus English scores based on the standardized scores provided by the University of Nottingham; see https://www.nottingham.ac.uk/ugstudy/applying/entryrequirements.aspx)

Table 2.2 Summary of language background for each paradigm group.

SED Paradigm	N	Mean Age	Gender	Self-Rated English Proficiency	AoA	IELTS
Paradigm 1	14	21.29	7 females, 7 males	6.11	2.64	7.21
Paradigm 2	14	21.86	9 females, 5 males	5.45	3.21	7.07
Paradigm 3	14	21.07	6 females, 8 males	5.36	4.36	6.82
Paradigm 4	14	20.64	9 female, 5 males	4.55	4.00	6.82

* Average self-rated English proficiency scores were used to represent participants' overall self-rated English proficiency. The score was calculated by averaging reading, writing, speaking and listening proficiency given by the participants.

Independent-samples t-test revealed a higher self-rated English proficiency for the L1 ME speakers (*m* 5.91, s.d. 0.94) compared to L2 ME speakers (*m* 5.13, s.d. 1.10), $t(54) = 2.55$, $p = .01$, but a similar IELTS score between the two groups (L1 ME speakers: *m* 7.18, S.D. 0.53; L2 ME speakers: *m* 6.90, s.d. 0.60), $t(54) = 1.66$, $p = .10$. The Pearson's correlation coefficient conducted however showed a moderate positive relationship between self-rated English proficiency and IELTS scores, $r(56) = 0.40$, $p < 0.01$.

2.2.3 Procedure

The participants completed 5 sessions spread across 5 days. On the first day, they completed the pre-test, followed by three training sessions spread across 3 consecutive days and then on the final day participants completed the post-test. The SED training was constructed with an educational version of LiveCode from RunRev and the tasks were run on a 14 inch screen HP EliteBook 8460p laptop, with a HP USB optical mouse to make responses.

2.2.3.1 Pre- and Post-tests

The pre- and post-tests involved a 2 alternative forced-choice (2AFC) identification task. The stimuli (32 minimal pairs: 16 /ɛ/-/æ/ and 16 /t/-/d/) in this task were repeated four times resulting in a total of 128 experimental trials. In each trial, two words of a minimal pair were displayed side by side at center of the computer screen. At the same time one of the words was presented auditorily. Participants were asked to indicate which word on the computer screen matched with the word they had heard. Auditory stimuli were presented at a comfortable listening level set by each participant through a pair of Sony Noise Cancelling Headphones, Model MDR-NC8/WHI. Background noise (6-talker babble) was present at half of the auditory stimuli level throughout the pre-test and post-test tasks. Participants completed eight practice trials first in order to familiarize themselves with the task. Presentation of the minimal pairs was randomized. No feedback was given after each trial. Only at the end of the practice trials and after each block of experimental trials was the total correct percentage presented. The pre- and post-tests each took approximately 15 minutes. Participants heard stimuli spoken by a female British English speaker in the pre- and post-tests.

2.2.3.2 Training

The training procedure differed according to the paradigm to which participants were assigned. Participants from Paradigm 1 and 2 were trained with stimuli produced by the male Southern Irish English speaker while participants from Paradigm 3 and 4 were trained with three different speakers across three consecutive days (two males - Southern Irish and Irish and one female - American English). Each training session took approximately 20 minutes to complete.

2.2.3.2.1 SED Training with Constant Background Babble (Paradigm 1 and 3)

Participants performed a Same-Different word discrimination task with background babble presented at 10 percent of stimuli level. Participants heard pairs of words and had to decide whether the words were the same or different by clicking on one of the two buttons presented on the computer screen using a computer mouse. The second word was played 1000ms after the first word. Participants received feedback for each response. After a correct response, the response button with the correct answer turned green and then the next trial started. For an incorrect response, the button with the incorrect answer turned red and the correct answer turned green. The word pair was played again and no response was needed. After this, the next trial was presented.

Each training session lasted approximately 20 minutes. Instructions were given verbally before each session. Participants were instructed to focus on the auditory words they heard and try to ignore the babble played in the background; emphasis was placed on accuracy rather than the speed of their response.

2.2.3.2.2 SED Training with Adaptive Staircase Background Babble (Paradigm 2 and 4)

In order to determine the level of noise, a threshold task was implemented before the training sessions. Participants performed a Same-Different word discrimination task in which the level of background babble was manipulated. Participants heard pairs of words from the /t/-/d/ phonemic contrast and had to decide whether the words were the same or different by clicking on one of the two buttons presented on the computer screen using a computer mouse. The second word was played 1000ms after the first word. The goal of the pre-training task was to determine the optimal volume of the background babble for the subsequent training. An adaptive staircase procedure using a three-down, one-up algorithm was implemented to obtain about 79% correct responses (Leek, 2001). The procedure adjusted the volume of the background babble, which had a range from 0 – 100 relative to the volume of the stimuli word (volume level 100). In the first trial, the volume of the babble was set to zero and the amount of change (step size) was set initially at 64. For every reversal made, the step size was divided by 2 so that the amount of change in volume decreased during the task. A reversal is defined as a correct response followed by an incorrect response or vice versa. If the response in a trial was correct, the volume was increased by 1 x step size, if it was incorrect, it was decreased by 3 x step size. For instance, the babble volume increased from 0 to 64 after the first correct trial. In the second trial, the volume was increased by another 64, from 64 to 100 (maximum volume as 100) when the response was correct again. On the other hand, if the response was incorrect for the second trial, the volume was decreased by 96 (3 x 32; step size was reduced by half from 64 to 32 due to the first reversal which took place when a correct response was made in first trial followed by an incorrect response made in second trial); the volume reduced from 64 to 0 (minimum volume as 0). The minimum step size was 1. The staircase procedure stopped after 12 reversals were achieved. At the end of the staircase procedure the final babble level or the threshold level was determined by taking the average of the babble levels from the last 4 reversals.

The subsequent training session was identical to the training task described in section 2.2.3.2.1., except for the background babble which was implemented at the individualized threshold level indicated by the threshold task. The babble level remained the same throughout that training session.

2.3 Results

The mean percentage of correct identification was calculated for every phonemic contrast in pre- and post-tests respectively and was used for the rest of the data analysis. Data from the /t/-/d/ initial phonemic contrast was excluded from the data analysis due to the near ceiling identification performance in pre-test for both L1 groups (mean above 98%).

2.3.1 Pre-test

Due to the smaller sample size of the L1 ME speakers compared to the L2 ME speakers, Levene's test of equality was conducted and results showed that the homogeneity of variance between the two language groups was equal (p = .92).

A between-subjects analysis of covariance (ANCOVA) was conducted on the mean percentage of correct identification made in pre-test for stimuli from both /t/-/d/ final and /ɛ/-/æ/ contrasts; with IELTS and self-rated English proficiency scores as covariates, and Language Group (L1 ME vs. L2 ME) as the between-subjects factor. The analysis was conducted to examine the potential perception differences on British English spoken words between the L1 and L2 ME speakers when the English proficiency scores were controlled. A significant effect of language was still found on the identification accuracy, $F(1,52) = 6.11$, $p = 0.02$, $\eta^2 = 0.11$ whereby the L1 ME speakers scored a higher identification accuracy than the L2 ME speakers, after IELTS scores and self-rated English proficiency scores were controlled. The IELTS scores and self-rated English proficiency scores however, did not have a significant effect on the participants' pre-test identification accuracy, $F(1,52) = .09$, $p = .76$, $\eta^2 = 0.06$ and $F(1,52) = 3.29$, $p = .08$, $\eta^2 < .01$ respectively.

2.3.2 Training Effect on /t/-/d/ Final Phonemic Contrast

A 2 x 2 x 2 x 2 mixed-design analysis of variance (ANOVA) was conducted on the mean percentage of correct identification on /t/-/d/ final phonemic contrast; with speaker variability (single vs multiple speakers), background babble (constant vs. adaptive staircase) and Language Group (L1 vs. L2 ME speakers) as the between-subjects factors, and lastly time of test (pre- vs. post-test) as the within-subjects factor (see Figure 2.1). No main effect of time of test was found, $F(1,48) = 2.29$, $p = .14$. However, there was a main effect of language group whereby L1 ME speakers were more accurate compared to the L2 ME speakers, $F(1,48) = 23.25$, $p < .001$. The participants who were trained with the single speaker scored an overall higher identification accuracy compared to those who were trained with multiple speakers, $F(1,48) = 6.76$, $p = .01$ but the speaker variability factor did not interact with time of test, $F(1,48) = .19$, $p = .67$.

A significant interaction was also found between the time of test and background babble, $F(1,48) = 4.36$, $p = .04$. Post-hoc t-tests were conducted to examine the interaction observed; the paired-samples t-tests comparing post-test identification accuracy to pre-test showed an approaching significant difference between the pre- and post-tests identification accuracy for participants from constant SED training, $t(27) = -1.86$, $p = .07$. No difference was found for participants from adaptive staircase SED training, $t(27) = .88$, $p = .39$. In addition, the between-subjects t-tests comparing identification accuracy between the participants from constant SED training and adaptive staircase SED training in pre- and post-tests respectively showed a similar performance between the two training paradigms in pre-test, $t(54) = .78$, $p = .45$, but a marginally significant higher accuracy for the constant training paradigm compared to adaptive staircase paradigm in post-test, $t(54) = 1.99$, $p = .05$.

There was also a marginally significant interaction between time of test and language, $F(1,48) = 3.87$, $p = .06$. Post-hoc t-tests were conducted to examine the interaction; between-subjects t-tests showed a higher identification accuracy of L1 ME speakers compared to the L2 ME speakers in both pre-test, $t(54) = 3.81$, $p < .001$ and post-test, $t(54) = 4.40$, $p < .001$. Interestingly, the paired-samples t-tests showed that L1 ME speakers scored a higher identification accuracy in post-test (94%) than pre-test (88%) by 6%, $t(16) = 2.77$, $p = .01$ but L2 ME speakers identified the stimuli words similarly between pre-test (77%) and post-test (76%), $t(38) = .37$, $p = .72$.

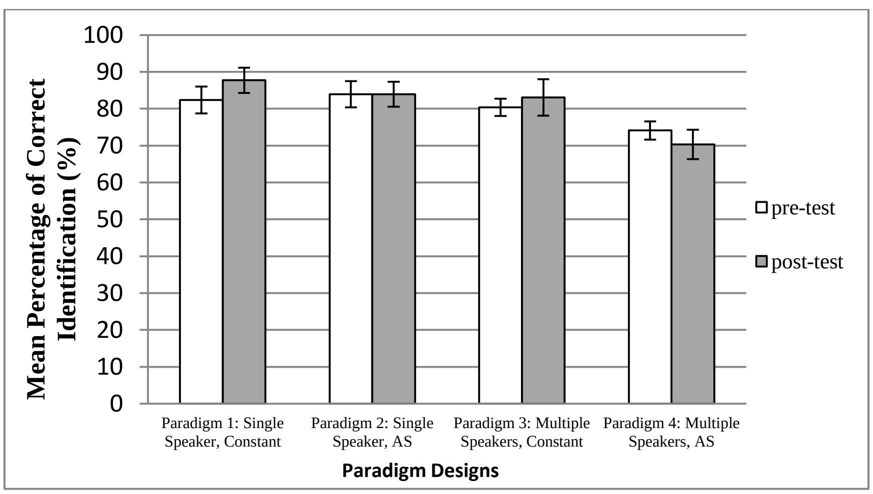


Figure 2.1 Mean Percentage of Correct Identification Accuracy on /t/-/d/ final phonemic words in pre- and post-tests made by each paradigm group (with error bars as standard error).

*Note: AS – Adaptive Staircase

2.3.3 Generalization of Training Effect to Novel Words

This analysis was conducted to examine whether identification accuracy in pre- and post-tests was affected by the training when the stimuli were separated as the trained and untrained (novel) stimuli. A 2 x 2 x 2 x 2 x 2 mixed-design ANOVA was conducted with stimuli set (trained vs. untrained /t/-/d/) and time of test (pre- vs. post-test) as the within-subjects factors, and speaker variability (single vs multiple speakers), background babble (constant vs. adaptive staircase) and language group (L1 vs. L2 ME speakers) as the between-subjects factors. The post-test identification accuracy was marginally higher than the pre-test. Importantly, no interaction was found between training status and time of test, $F(1,48) = .76$, $p = .39$; suggesting the identification accuracy across time of test was not affected by the stimuli words' training status. There were no other effects or interactions found, *F*s < 2.76, *p*s > .05.

2.3.4 Generalization of Training Effect across Untrained Phonemic Contrast /ɛ/-/æ/

In order to examine the potential identification changes on the untrained phonemic contrast /ɛ/-/æ/ , a 2 x 2 x 2 x 2 mixed-design analysis of variance (ANOVA) was conducted on the mean percentage of correct identification on /ɛ/-/æ/ phonemic contrast; with speaker variability (single vs multiple speakers), background babble (constant vs. adaptive staircase) and language group (L1 vs. L2 ME speakers) as the between-subjects factors, and lastly time of test (pre- vs. post-test) as the within-subjects factor (See Figure 2.2). The results showed that participants improved in their identification accuracy in the post-test compared to the pre-test, $F(1,48) = 11.64$, $p = .001$. The main effect of language group was also found; L1 ME speakers scored a higher identification accuracy on the /ɛ/-/æ/ phonemic words compared to the L2 ME speakers, $F(1,48) = 5.81$, $p = .02$. No other effects or interactions were found, *F*s < 2.40, *p*s > .05.

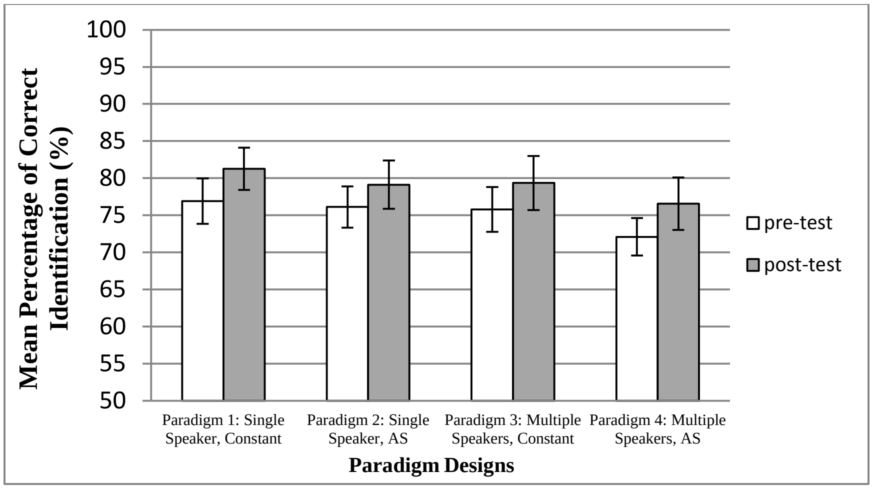


Figure 2.2 Mean Percentage of correct identification accuracy on /ɛ/-/æ/ phonemic words in pre- and post-tests made by each paradigm group (with error bars as standard error).

*Note: AS – Adaptive Staircase

2.4 Discussion

After SED training, identification accuracy on the /t/-/d/ final phonemic words did not show strong improvement in performance. It was worth noting that the SED training using constant background babble showed a trend in generating identification improvement although the results were not significant. In addition, participants who were trained with low variability speech materials (i.e. produced by single speaker) overall had higher identification accuracy than those who were trained with high variability speech materials (i.e. produced by multiple speakers). The speaker variability however did not interact with the time of test factor. Hence, the variability of training materials did not affect the effectiveness of SED training in generating perceptual improvement on the /t/-/d/ final phonemic words.

There was also no difference in identification accuracy across time of test between the trained /t/-/d/ and novel /t/-/d/ final phonemic words. This finding is contrary with Lively and colleagues' (1993; 1994) finding, in which they proposed the importance of training variability in promoting better training effects; the current study failed to find a similar pattern of results.

When participants' IELTS standard scores and self-rated scores were controlled, their pre-test identification differed according to their L1. Both groups of speakers generally identified the phonemic words well in pre-test. However, the L1 ME speakers had higher identification accuracy for the British English /t/-/d/ final phonemic words (88%) compared to the L2 ME speakers (77%). Interestingly, the statistical results also showed that L1 ME speakers improved significantly in identifying the /t/-/d/ final phonemic words in post-test but the L2 ME speakers did not. This finding contradicts our prediction because the L2 ME speakers who had lower identification accuracy in pre-test should have a greater capacity for potential improvement.

The L1 ME speakers did not have much difficulty discriminating and identifying the word final contrast. The failure to achieve ceiling performance in pre-test could be due to the unfamiliar pronunciation pattern of the British English spoken words. Assuming speakers' L1 ME language system does not reduce their perception sensitivity at word final position, L1 ME speakers should still be sensitive and more capable to pick up the critical phonetic cues that can assist them in identifying the British English phonetic sounds after being trained.

There is the possibility that L2 ME speakers have reduced perception plasticity. The L2 ME speakers from this study speak Mandarin as their L1 or the most dominant language. Their deep-rooted L1 language system that does not promote word final consonants discrimination may impose a certain degree of perceptual limitations within the phonetic environment. The results suggested that phonetic training alone was not sufficient to re-shape their perception sensitivity to resemble those of the L1 British English speakers or the L1 ME speakers.

Unlike most literature on phonetic training and HVPT, SED training failed to show significant training effect on the training materials. Surprisingly, the identification accuracy on the untrained phonemic contrast /ɛ/-/æ/ improved in post-test, merely after a brief exposure in the pre-test. A similar identification improvement was also found by McCandliss et al. (2002) in their control group. Their Japanese participants from the control group identified the /r/-/d/ distinctions at word initial position better after a 3-days test-retest delay interval, with only exposure to the stimuli on the first testing day. These findings suggest that improvements in identification could be achieved through brief exposure to the speech materials. Other phonetic training studies such as the Bradlow et al. (1997; 1999), Wang and Munro (2004), and Lengeris and Hazan (2010) however, did not find any perception changes from the control groups in which no training was administered. An effective phonetic training paradigm should depend on its capability in shifting listeners' attention to relevant acoustic cues that assists phonetic sounds discrimination; speech exposure alone should be insufficient to promote the attention shifts.

L2 ME speakers in this study, however, did improve in /ɛ/-/æ/ identification without any additional training sessions. This pattern of results could be simply due to practice effects or participants' capability to learn from the listening experience. It is important to know whether phonetic training would improve the perception more than mere exposure. Looking at participants' identification accuracy, the /ɛ/-/æ/ phonemic contrast might be a better training target as it also imposed greater identification difficulty to the Malaysian population (identification accuracy in pre-test: L1 ME speakers 80%, L2 ME speakers 73%) when compared to the /t/-/d/ final phonemic contrast.

The /t/ and /d/ phonetic sounds at word final position can be identified by realizing the existence of released burst (/t/) or the final closure voicing (/d/) cues. A study by Flege (1989) however, trained Chinese speakers on the /t/-/d/ final contrast with the final closure voicing and released burst cues removed, so that participants can learn the discrimination through vowel quality. The acoustic information retained was sufficient to cue voicing judgment and the Chinese speakers learnt to identify the final stops /t/-/d/ more accurately in post-test.

According to Flege, the reason that Chinese speakers performed differently than the native English speakers in identifying the voiced and voiceless /t/-/d/ final contrast lies in Chinese speakers' lack of use of the acoustic information (e.g. spectral quality[4], duration and onset/ offset of formants[5]) from the preceding vowel, rather than the reduced sensitivity in discriminating the released burst and closure voicing between /t/-/d/. On the other hand, the native English speakers were not affected as much by the removal of the critical closure voicing and released burst because of their capability in using the vowel acoustic cues to assist the identification. Because it is common in conversational speech for the speakers to drop the final release burst /t/ and devoice the closure voiced /d/, native English speakers were trained through experience to use the vowel acoustical information as secondary cues. The non-native English Chinese speakers however would not be able to adopt this strategy as easily due to differences in speech environment as well as less experience doing so.

As proposed by Flege (1989), we assume some of the naturally produced stimuli used in this study did not have perfect release bursts and closure voicing /t/-/d/. As the L2 ME speakers involved in the present study spoke a similar language (i.e. L1 Mandarin) as participants in Flege's study, we also assumed that discriminating the release burst and closure voicing was not their major difficulty. Thus, training that promotes the use of vowel acoustic cues might be more effective.

The lack of training effect on the /t/-/d/ final phonemic contrast may also be limited because the level of improvement room is reduced in the Malaysian population. If so, the effectiveness of SED training based on their identification changes in post-test might not reveal the training paradigm's full potential. Judging on the level of improvement room (poor pre-test performance),

---

[4] Image or distribution of sound components that are arranged according to their frequency, charge and energy.

[5] Formants are bundles of high-amplitude harmornics that can be used to determine the identity of vowels.

the /ɛ/-/æ/ phonemic contrast should be more of a priority because the identification of the pair was more difficult than the /t/-/d/ phonemic contrast.

Generally, the lack of improvement on /t/-/d/ final contrast and the improvement observed on /ɛ/-/æ/ phonemic contrast also suggested that resetting sensitivity to phonetic details which has been turned off by the formation of L1 may be more challenging than learning a new vowel category or shifting the existing category boundaries of vowels.

To summarize, the contextual constraint against the voiceless-voiced stops /t/-/d/ in word final position is difficult to override with phonetic training. Therefore, the next study was conducted to examine how the SED training could affect the perceptual performance of the /ɛ/-/æ/ phonemic contrast.

Chapter 3: Spoken English Discrimination (SED) Training on English Vowel /ɛ/-/æ/ Phonemic Contrast

3.1 Introduction

Previously, we failed to show the effectiveness of SED training in improving L2 Malaysian English speakers' identification accuracy on the voiced-voiceless plosives /t/-/d/ contrast at word final position. In this chapter, we trained Malaysian participants on the English /ɛ/-/æ/ phonemic contrast because the results from Experiment 1 showed a higher degree of plasticity in /ɛ/-/æ/ perception. The English /ɛ/-/æ/ phonemic contrast is difficult to discriminate by the multilingual Malaysian English speakers because of its ambiguous classification in Mandarin (Huang, 1992), Malay (Yap et al., 2010) as well as Malaysian English (Pillai et al., 2010). Yap et al (2010) found that L1 Malay speakers unconsciously assimilate the phoneme /ɛ/ (e.g., bet) into the phoneme /æ/ (e.g., bat) (see section 4.1.3.2 for more details). The phonemic words from /t/-/d/ phonemic contrast were still included in pre-test and post-test to examine the generalization of training effect across untrained phonemic contrasts.

Another important aspect to examine is the longevity of SED training effect. Any robust training effect should be retained after a certain period of time. Several phonetic training studies have reported a long term benefit of the training (see Table 1.1 and 1.2). In this chapter, the retention of SED training effect was examined after a resting period of six months in Experiment 4.

3.2 Experiment 2 (Constant SED Training)

The design of Experiment 1 could have been too complicated to show the potential of SED training. Therefore, the features of SED training were broken down and investigated separately for this chapter. In Experiment 2, we first examined whether implementation of a constant level of background babble (Constant Spoken English Discrimination (SED) training) in High Variability Phonetic Training (HVPT) improved L1 Mandarin speakers' (L2 ME speakers) identification accuracy. We also examined the effect of task difficulty on the effectiveness of phonetic training. We trained two groups of participants with either a low level or high level background babble. The difficulty level of the high babble condition was set to resemble pre- and post-tests, in which the background babble was played on top of the training stimuli at half of the training stimuli level throughout the training. The difficulty level of the low babble condition was relatively low, with the background babble played at one-tenth of the training stimuli level. The background babble was implemented in the SED training because we believe effective phonetic training should be conducted in a difficult context that resembles the real-life situation. Therefore, we also hypothesized that Constant SED with high babble condition (higher difficulty level) would generate greater training effects in the L1 Mandarin speakers compared to the low level condition.

As reported in previous HVPT studies, we expected L1 Mandarin speakers to improve significantly in identification accuracy in post-test, with training effects being generalized to novel stimuli (spoken words from the same phonemic category as the training materials but not included in the training sessions), produced by novel speakers (whose productions were not included as training materials).

3.2.1 Method

3.2.1.1 Design and Materials

The stimuli were identical to the stimuli used in Experiment 1. In pre-test, post-test and high babble condition training, the multi-talker babble was played simultaneously with the stimulus words, at half of the stimuli level and an approximate signal-to-noise ratio (SNR) of 3 db (SNR range 2.60 to 2.84 db depends on the speakers, mean SNR 2.72 db). In low babble condition training, the multi-talker babble was played at 10 percent of the stimuli level, and at an approximate SNR of -37 db SPL (SNR range -37.16 to -37.39 db depends on the speakers, mean SNR -37.28 db).

The /t/-/d/ minimal pairs were used in pre-test and post-test only. For the training sessions, the /ɛ/-/æ/phonemic contrasts were split into two groups. Half of the participants were presented with half of the words from the /ɛ/-/æ/ minimal pairs, whereas the other half of the participants were presented with the other group of /ɛ/-/æ/ minimal pairs.

3.2.1.2 Participants

28 (age range: 17-22; 16 females, low babble group mean age 18.93 years, high babble group mean age 18.57 years) participants who spoke Mandarin as their L1 were recruited from University of Nottingham, Malaysia Campus and were paid for their participation. All participants spoke ME and Malay, and had learnt both languages from a young age (age range: 0-7, mean age 3.55 years). All participants reported to have normal eyesight or corrected to normal vision, and had no history of any hearing, speech and reading disorder. Participants completed a questionnaire, providing information about their language background and formal education.

Participants were alternatively assigned to the low babble and high babble condition groups. The mean self-rated English and Mandarin proficiency did not differ across the two groups, $t < 1.30$, $p > .05$ (English proficiency: low babble 4.71, high babble = 4.75; Mandarin proficiency: low babble 5.66, high babble = 5.21). Participants also did not differ in terms of their age of English acquisition (AoA) and IELTS standardized test scores, $t < 1$, $p > .05$ (AoA: low babble 2.36, high babble 2.36; IELTS: low babble 7.07, high babble = 6.96).

3.2.1.3 Procedure

The procedure was identical to the procedure in Experiment 1.

3.2.1.3.1 Pre- and Post-tests

The procedure was identical to Experiment 1.

3.2.1.3.2 Training

Participants were allocated to either the low babble (set at 10 percent of stimuli level), or high babble condition (set at 50 percent of stimuli level) and were matched based on their results from the language history questionnaire in order to ensure they did not differ. Participants performed the Same-Different word discrimination task used in Experiment 1, with the background babble set at either high or low level.

3.2.2 Results

The mean percentage of correct identification was calculated for each of the three phonemic contrasts in the pre- and post-test of the two groups of participants (see Table 3.1). Data from the /t/-/d/ phonemic contrast at word initial position was excluded from further analyses because of ceiling performance in each group (mean above 95%).

3.2.2.1 Training Effect of Constant SED

The training effect of Constant SED with low and high background babble level was examined in a 2 X 2 X 2 mixed-design ANOVA with babble level (low vs. high) as between-subject factor, time of test (pre-test vs. post-test) and phonemic contrasts (/t/-/d/ final vs. /ɛ/-/æ/) as within-subject factors (Table 3.1). Although the improvement scale was small, participants trained with Constant SED performed significantly better in post-test (76%) than pre-test (73%), $F(1,26) = 10.41$, $p < .01$. There was no interaction between babble level and time of test, $F(1,26) = 0.09$, $p = .77$, demonstrating that both babble level conditions yielded similar levels of improvement in identification accuracy. No other main effects and interactions were found, Fs < 2.01, ps > .05.

Two separate mixed-design ANOVAs were conducted to investigate the generalization of training benefit yielded by Constant SED, to untrained /ɛ/-/æ/ stimuli and untrained phonemic contrast, /t/-/d/ final (section 3.2.2.2 and section 3.2.2.3).

Table 3.1 Mean percentage of correct identification in pre- and post-tests for each phonemic contrast and babble level condition (with standard error in brackets).

	/t/-/d/ Initial		**/t/-/d/ Final**		**/ɛ/-/æ/**	
Babble Level	**Pre**	**Post**	**Pre**	**Post**	**Pre**	**Post**
Low	98.4 (0.6)	98.7 (0.8)	73.7 (3.9)	75.0 (3.1)	73.1 (2.1)	77.2 (3.3)
High	98.4 (0.5)	98.4 (0.5)	72.3 (3.6)	72.8 (2.9)	72.1 (3.2)	78.2 (3.6)
Total	98.4 (0.4)	98.5 (0.5)	73.0 (2.6)	73.8(2.1)	72.6 (1.9)	77.7 (2.4)

3.2.2.2 Training Generalizations of Constant SED to Novel Words

We first compared identification accuracy of untrained /ɛ/-/æ/ stimuli to trained /ɛ/-/æ/ stimuli to examine the generalization effect within the same phonemic contrast in 2x2x2 ANOVA, with stimulus set (trained /ɛ/-/æ/ vs. untrained /ɛ/-/æ/) and time of test (pre-test vs. post-test) as the within-subject factors, and babble level (low vs. high) as the between-subject factor. The analysis showed that identification accuracy in post-test was significantly higher than pre-test, $F(1,26) = 6.99$, $p = .01$. Importantly, there was no interaction between stimulus set and time of test, $F(1,26) = 1.79$, $p = .19$. No other main effects or interactions were significant, *F*s < 1.58, *p*s > .05.

3.2.2.3 Training Generalizations of Constant SED across Untrained Phonemic Contrast /t/-/d/ Final

We then conducted a 2x2x2 mixed-design ANOVA was conducted by comparing identification accuracy of the untrained /ɛ/-/æ/ stimuli to untrained phonemic contrast, /t/-/d/ final to examine the generalization effect across different phonemic contrast, with untrained contrast (untrained /ɛ/-/æ/ vs. /t/-/d/ final) and time of test (pre-test vs. post-test) as within-subject factors, and babble level (low vs. high) as between-subject factor. The analysis yielded a marginally significant effect of time of test, $F(1,26) = 3.46$, $p = .07$. Identification accuracy on the untrained /ɛ/-/æ/ stimuli was shown to improve significantly in post-test. In order to examine whether it was the performance on the /t/-/d/ final stimuli that rendered the training effect, post-hoc paired-samples t-tests were conducted and results showed that participants' identification accuracy on the untrained /ɛ/-/æ/ stimuli improved in post-test , $t(27) = 3.02$, $p < .01$, but untrained /t/-/d/ final phonemic contrast did not, $t(27) = 0.50$, $p > .05$. No other effects or interactions were significant, Fs < 1, ps > .05.

3.2.3 Discussion

The Constant SED successfully improved participants' identification accuracy on the /ɛ/-/æ/ phonemic words, which is in line with the literature findings so far. Although participants from the present experiment initially identified the /ɛ/-/æ/ phonemic words at a considerably high accuracy (mean 73% in pre-test), the Constant SED was still capable of improving their perceptual performance even though the level of improvement was small – performance improved by 5%.

The training effect generalized to untrained /ɛ/-/æ/ stimuli set, but it did not generalize as effectively to the untrained /t/-/d/ contrast. There was no other HVPT study that reported generalization of training effect across phonemic contrasts except the study by Callan et al. (2003). They trained native Japanese speakers to identify /r/-/l/ contrast using HVPT. Surprisingly, the results also revealed a significant improvement in identification accuracy of the /b/-/v/ contrast along with the trained /r/-/l/ contrast after training. This present experiment did not replicate the same finding; this may be due to the differences in the training materials. In Callan et al.'s study, both training (/r/-/l/) and generalization testing materials (/b/-/v/) were consonant pairs; while in this present experiment, we trained on a vowel phonemic contrast, /ɛ/-/æ/ and examined its generalization effect towards a consonant phonemic contrast, /t/-/d/. The difference between the contrasts examined may explain why an across phonemic contrast generalization effect was not observed. Furthermore, the sensitivity towards the word final phonetic details is difficult to modify, as evidenced in Experiment 1.

Participants' performance in this study also improved similarly regardless of the difficulty level of Constant SED (high babble level vs. low babble level). This suggests that merely increasing the task difficulty level will not affect the training outcome. It is important to note that the typical HVPT paradigm and the present Constant SED does not account for individual differences.

3.3 Experiment 3 (Adaptive Staircase SED Training)

In this experiment, we examined the effect of implementing background babble using an adaptive staircase procedure in SED (named Adaptive Staircase SED) on a wider population of differently proficient ME speakers: L1 ME speakers, and L1 Malay and L1 Mandarin speakers who speak ME as L2. This study examined whether the individual discrimination threshold level set for Adaptive Staircase SED stimulates greater training effect and generalization benefits compared to the Constant SED.

By combining the strengths of existing training methods and an individualized level of background noise that was pre-determined before each training session, SED was expected to be more effective (in terms of training effect size and generalizability) in improving phonetics identification accuracy of differently proficient participants, when compared to the constant SED in Experiment 1.

3.3.1 Method

3.3.1.1 Design and Materials

The materials were identical to Experiment 1. The design was similar except that a pre-training session was implemented in Adaptive Staircase SED, whereby each participant's discrimination threshold was measured by the adaptive staircase procedure. The threshold was then applied to the subsequent training phase.

3.3.1.2 Participants

40 multilingual Malaysian participants (age range: 16-23 years, mean age 18.18 years; 23 females) were recruited from the University of Nottingham Malaysia Campus and were paid for their participation. Participants were divided into three groups based on their first language (L1). 14 participants were classified as L1 Mandarin speakers, 15 as L1 Malay speakers, and 11 as L1 ME speakers. All participants spoke ME and Malay, and had learnt both languages from a young age (range: 0-7 years, mean age 3.95 years). One L1 Malay speaker was also able to speak Mandarin, but proficiency was very poor. All participants had normal or corrected to normal vision, and they had no history of any hearing, speech or reading disability.

Language background of all participants was assessed using the same language background questionnaire used in Experiment 1. Table 3.2 presents an overview of their subjective proficiency in each language, the age they first acquired English (AoA), and their International English Language Testing System (IELTS) standard scores (see footnote 3).

Table 3.2 Participants' language background.

First Language (L1)	N	MA	SR. Eng Prof	SR. Man Prof	SR. BM Prof	AoA	IELTS
Mandarin	14	18.64	4.96	5.87	4.87	3.86	7.04
Malay	15	18.00	5.18	1.50	6.72	5.00	6.83
Malaysian English	11	17.81	6.23	3.34	5.89	2.18	7.32

* N-Number, MA-Mean Age, SR. Language Prof- Self-rated Language Proficiency (from scale 1 = very poor to scale 7 = native like), AoA- Age of Acquisition, IELTS- International English Language Testing Scores

* Average self-rated English proficiency scores were used to represent participants' overall self-rated English proficiency. The score was calculated by averaging reading, writing, speaking and listening proficiency, rated by the participants. When examining the relationship between self-rated English proficiency and IELTS score, there was a positive relationship but it was not statistically significant, $r(40) = 0.17$, $p > .05$.

A one-way between-subject ANOVA was conducted comparing the language history information across the three language groups. The three groups differed significantly in terms of their overall self-rated English proficiency, $F(2,39) = 9.64$, $p < .001$, AoA, $F(2,39) = 5.44$, $p < .01$ and IELTS scores, $F(2,39) = 3.58$, $p < .05$. Pairwise comparisons (Bonferroni) showed that L1 Malay and L1 Mandarin participants did not differ in terms of their self-rated English proficiency scores ($p = .91$). However, both L1 Malay and L1 Mandarin participants reported their self-rated English proficiency lower than the L1 ME participants ($p <.01$). Only L1 Malay participants were found to report significantly lower IELTS standard scores compared to L1 ME participants ($p < .05$), and acquired English later than L1 ME participants ($p < .01$).

3.3.1.3 Procedure

The pre-training task was identical to the one described in Experiment 1. Before the training session, participants performed a Same-Different word discrimination task in which the level of background babble was manipulated by a three-down one-up algorithm adaptive staircase procedure. The procedure estimated a volume level where participants identified the stimuli at about 79% correct accuracy (Leek, 2001). Details of the adjustment procedure can be found in Experiment 1.

3.3.2 Results

Data from the /t/-/d/ phonemic contrast at word initial position was excluded from further analyses because of ceiling performance in each group (mean above 95%).

3.3.2.1 Phonetics Perception of Malaysian Speakers on British English

Only the mean percentage of correct identification from pre-test was analyzed in this section to examine phonetics perception of Malaysians before training (see Table 3.3).

A 2 X 3 mixed-design ANOVA was conducted to examine the effect of L1 on Malaysians' phonetics perception; with the phonemic contrast (/ɛ/-/æ/ vs. /t/-/d/ final) as the within-subjects factor and the L1 (L1 ME vs. L1 Malay vs. L1 Mandarin) as the between-subject factor. A main effect of L1 was found , $F(1,37) = 6.07$, $p < .01$ in which the L1 ME speakers were better in identifying the British English spoken words compared to the L1 Mandarin speakers. There was no effect of phonemic contrast, $F(1,37) = .06$, $p = .82$.

There was also an interaction between L1 and phonemic contrast, $F(2,37) = 4.22$, $p < .05$. One-way between-subjects ANOVAs were conducted separately for /t/-/d/ final and /ɛ/-/æ/ contrasts to compare identification accuracy across different L1 groups. Main effects of L1 were found in both analyses, $F(2,39) = 6.54$, $p < .01$ and $F(2,39)= 4.27$, $p < .05$ respectively. For /t/-/d/ final contrast, pairwise comparisons (Bonferroni) revealed lower identification accuracy for the L1 Malay and L1 Mandarin speakers when compared to the L1 ME speakers, $p < .05$. The L1 Malay and L1 Mandarin speakers however performed similarly in this contrast. In addition, pairwise comparisons conducted for the /ɛ/-/æ/ contrast showed that L1 Malay and L1 ME performed similarly. The L1 Mandarin speakers, on the other hand, showed significantly lower identification accuracy compared to the L1 Malay speakers, $p < .05$ and the L1 ME speakers, $p = .05$.

Post-hoc paired-samples t-tests were conducted to examine whether the L1 groups had equal difficulty in discriminating the /ɛ/-/æ/ contrast compared to discriminating the /t/-/d/ final contrast. Results showed that the L1 Mandarin and L1 ME speakers had similar identification accuracy between the two phonemic contrasts, ts < 1.72, ps > .05. However, the L1 Malay group had more difficulty in discriminating the /t/-/d/ final word pairs compared to discriminating the /ɛ/-/æ/ word pairs, $t(14) = 2.64$, $p = .02$.

3.3.2.2 Training Effect of Adaptive Staircase SED

The training effect of Adaptive Staircase SED was examined in a 3 X 2 X 2 mixed-design Analysis of Variance (ANOVA) with time of test (pre- vs. post-test) and phonemic contrast (/t/-

/d/ final vs. /ɛ/-/æ/) as within-subject factors and L1 (Mandarin vs. Malay vs. ME) as the between-subject factor (Table 3.3).

Table 3.3 Mean percentage of correct identification in pre-test and post-test for each phonemic contrast and first language (L1) group (with standard error in brackets).

	/t/-/d/ Initial		**/t/-/d/ Final**		**/ɛ/-/æ/**	
First Language	**Pre**	**Post**	**Pre**	**Post**	**Pre**	**Post**
Mandarin	97.3 (1.0)	99.1 (0.4)	71.7 (3.3)	80.4 (3.5)	70.9 (2.8)	81.2 (2.0)
Malay	95.2 (1.3)	97.7 (0.9)	72.9 (3.3)	80.6 (3.6)	80.7 (2.1)	88.8 (1.7)
Malaysian English	98.6 (0.5)	98.6 (0.6)	87.2 (2.7)	88.6 (2.8)	81.5 (4.0)	86.4 (3.7)

Overall, the identification accuracy of the participants improved significantly in post-test (by 7%, from pre-test 77% to post-test 84%), $F(1,37) = 43.32$, $p < .01$. A marginally significant interaction was found between time of test and L1, $F(2,37) = 3.13$, $p = .06$. Post hoc t-tests were conducted to examine the marginally significant interaction. Paired-sample t-tests revealed that all three language groups improved significantly in post-test: L1 Mandarin group, $t(13) = 4.56$, $p = .001$, L1 Malay group, $t(14) = 4.68$, $p < .001$ and L1 ME group, $t(10) = 2.42$, $p < .05$ respectively. The interaction observed might be attributed to the different training effect size across the language groups. Three between-subject t-tests conducted revealed that correct identification improvement (difference between pre- and post-test) of the L1 Malay (8%) and L1 Mandarin groups (10%) were significantly larger than the improvement of the L1 ME group (3%), $t(24) = 2.34$, $p < .05$ and $t(23) = 2.75$, $p = .01$ respectively. No difference was found between L1 Malay and L1 Mandarin groups, $t(27) = .44$, $p = .66$. Accuracy for the /t/-/d/ final and /ɛ/-/æ/ stimuli improved similarly in the post-test and no interaction was found between phonemic contrast and time of test, $F < 1$. There was also no three-way interaction between time of test, L1 and phonemic contrast, $F(2,37) = 0.21$, $p = .81$. The analysis that did not examine training effects (i.e. time of test factor) can be found in Appendix 3.

3.3.2.3 Training Generalizations of Adaptive Staircase SED to Novel Words and across Untrained Phonemic Contrast /t/-/d/ Final

To examine whether training effects generalized to novel words, a 3 X 2 X 2 mixed-design ANOVA was conducted. Mean percentage of correct identification was analyzed with L1 (Mandarin vs. Malay vs. ME) as a between-subject factor, and time of test (pre- vs. post-test) and stimulus set (trained vs. untrained /ɛ/-/æ/) as within-subject factors. The analysis revealed a higher identification accuracy in post-test compared to pre-test, $F(1,37) = 44.54$, $p < .01$. As before, a main effect of L1 was observed, $F(2,37) = 3.92$, $p < .05$. Post-hoc t-tests revealed that overall accuracy of the L1 Mandarin group was lower than the L1 Malay group, $t(27) = 3.13$, $p < .01$, but only marginally different from the L1 ME group, $t(23) = 1.91$, $p = .07$. Importantly, there was no interaction between stimulus set and time of test, $F(1,37) < 1$. No other main effects or interactions were found, *F*s < 1.83, *p*s > .05.

To examine whether the training effects generalized across phonemic contrast, we compared identification accuracy of /t/-/d/ final contrasts with untrained /ɛ/-/æ/ stimuli. A 3 X 2 X 2 mixed-design ANOVA was conducted with L1 (Mandarin vs. Malay vs. ME) as between-subject factor, and time of test (pre- vs. post-test) and contrast (untrained /ɛ/-/æ/ vs. /t/-/d/ final) as within-subject factors. Across all types of untrained stimuli, accuracy was higher in post-test, $F(1,37) = 35.82$, $p < .01$. Importantly, no interaction was found between contrast and time of test, $F < 1$, and no other main effects or interactions were found, *F*s < 3.13, *p*s > .05.

3.3.2.4 Comparing Training Effect between Constant SED and Adaptive Staircase SED

Identification performance of the L1 Mandarin participants from Experiments 1 and 2 were analyzed together in this section to compare the effectiveness of the Constant SED and Adaptive Staircase SED. Data from the low and high babble conditions from Experiment 1 was collapsed because there was no difference in training benefits between the two conditions. To ensure that training effects of both Constant and Adaptive Staircase SEDs were comparable, the linguistic background and initial perceptual performance of participants from Experiments 1 and 2 were compared. Between-subject t-tests confirmed that participants of both experiments did not differ in terms of their self-rated English proficiency, $t(40) = .96$, $p = .34$, self-rated Mandarin proficiency, $t(40) = .61$, $p = .55$, IELTS standard scores, $t(40) = .10$, $p = .92$, AOA, $t(40) = 1.88$, $p = .07$, and most importantly, their pre-test identification accuracy (including /t/-/d/ final and /ɛ/-/æ/ contrasts), $t(40) = .47$, $p = .64$.

The mean percentage of correct identification from both Experiment 1 and 2 was examined in a 2 x 2 x 2 mixed-design ANOVA; with SED design (adaptive staircase vs. constant) as the between-subject factor, time of test (pre-test vs. post-test) and phonemic contrast (/t/-/d/ final vs. /ɛ/-/æ/) as the within-subject factors. The analysis revealed that L1 Mandarin participants' identification accuracy in post-test was higher than pre-test, $F(1,40) = 41.07$, $p < .01$. An interaction was also found between time of test and SED design, $F(1,40) = 11.11$, $p < .01$. A Post-hoc independent-sample t-test was conducted to compare the training effect size (difference between pre-test and post-test) between both SED paradigms. Results showed that Adaptive Staircase SED yielded greater training effect size than Constant SED, $t(18.17) = 2.86$, $p < .01$. No other significant main effects or interactions were found, Fs < 1.65, ps > .05.

3.3.3 Discussion

The Adaptive Staircase SED did improve participants' identification accuracy irrespective of their L1 groups. The training effect generalized to novel stimuli and novel speakers; even across phonemic contrasts, to /t/-/d/ contrast at the word final position which has never been trained during the training phase. This generalization effect was not found from Constant SED in Experiment 1.

Both the Adaptive Staircase and Constant SED improved identification accuracy. However, when we compared across the SED designs, Adaptive Staircase SED yielded greater training effect size compared to the Constant SED. Language background and identification accuracy in the pre-tests of participants from Experiment 1 and 2 were matched; therefore, the difference in training effect size cannot possibly be attributed to the unequal pre-training criteria. It appears that the adaptive staircase procedure enhances the effectiveness of phonetic training, in terms of its training effect size and its generalization effect.

Experiment 1 showed that manipulating the difficulty level of SED training did not affect participants' identification improvement, which could imply that the effectiveness of phonetic training cannot be attributed solely to SED's difficult training conditions. Rather, training participants at their individualized discrimination threshold seems to account for the greater effectiveness of Adaptive Staircase SED compared to the Constant SED.

To further strengthen the benefits of perceptual phonetic training, previous studies also found long-term training benefits that were retained up to six months (e.g. Lively et al., 1994; Bradlow et al., 1999; Nishi and Kewley-Port, 2007). Experiment 2 and 3 successfully showed the effectiveness of Constant SED and Adaptive Staircase SED. The longevity of training effect however still remains unknown. Experiment 4 was conducted to re-examine participants' identification performance after a delay of six months from the last testing session.

3.4 Experiment 4 (Retention of SED Training Effect)

In this study, we examined the longevity of Adaptive Staircase SED's training effect by re-testing participants' identification accuracy six months after their last training session. Retention of training effect was expected in accordance with the results from previous HVPT studies (e.g., Lively et al., 1994; Bradlow et al., 1999).

3.4.1 Method

3.4.1.1 Design and Materials

The materials and design used for follow-up test are identical to the pre- or post-tests from Experiments 2 and 3.

3.4.1.2 Participants

Thirteen (7 from low babble, 6 from high babble, mean age 18.69) out of twenty-eight L1 Mandarin speaking participants from Experiment 2 (aged 18-22; 4 males, 9 females) and all fourteen L1 Mandarin speaking participants from Experiment 3 (aged 18-23; 7 males, 7 females, mean age 18.64) returned to complete the follow-up testing. Because there were no significant differences between the low and high babble conditions, we collapsed across conditions to form a Constant SED group. The mean self-rated English and Mandarin proficiency did not differ across the two groups, $t < .32$, $p > .05$ (Adaptive Staircase SED: mean English proficiency 4.87, mean Mandarin proficiency 5.87; Constant SED: mean English proficiency 4.56, mean Mandarin proficiency 5.54). Participants also did not differ in IELTS standardized test scores, $t(25) = .15$, $p > .05$ (Adaptive Staircase SED IELTS mean score 7.04, Constant SED IELTS mean score 7.31). Participants from Adaptive Staircase SED training however had a significant later age of English acquisition (mean 3.93) compared to the participants from the Constant SED training (mean 2.00), $t(25) = 2.17$, $p < .05$. The AoA difference is probably due to the different understanding of the term "age of acquisition" by participants from Experiment 2 and 3[6]. The participants also reported a similar level of self-rated English and Mandarin proficiency.

[6] When filling up the AoA information in language history questionnaire, participants from Experiment 2 were reminded of the language exposure through media when they were young but participants from Experiment 1 were not.

3.4.2 Results

3.4.2.1 Retention of Training Effect

Preliminary analysis was conducted to ensure that the training data from the returning participants still replicated the previous findings from Experiment 2 and 3. A 2 X 2 mixed-design ANOVA that compared pre-test and post-test identification accuracy was conducted with the time of test (pre-test vs. post-test) as the within-subject factor and the SED design (Constant vs. Adaptive Staircase) as the between-subjects factor. The analysis produced similar result patterns as previous findings. Identification accuracy in post-test was higher than in pre-test, $F(1,25) = 29.30$, $p < .01$. Interaction between time of test and SED design also confirmed that overall improvement size of participants from Adaptive Staircase SED was still greater than Constant SED, $F(1,25) = 5.31$, $p < .05$. There was no other significant effect or interaction, $F < 1.33$, $p > .05$.

Table 3.4 Mean percentage of correct identification in pre-, post- and six months post-tests for each phonemic contrast and SED design (with standard error between brackets)

	/t/-/d/ Final			**/ɛ/-/æ/**		
SED Design	**Pre**	**Post**	**6 Months Post**	**Pre**	**Post**	**6 Months Post**
Adaptive Staircase	71.6 (3.3)	80.4 (3.5)	77.0 (3.1)	70.9 (2.8)	81.2 (2.0)	81.7 (2.8)
Constant	74.3 (3.9)	76.2 (3.1)	76.2 (4.5)	77.4 (2.8)	83.2 (3.6)	82.3 (3.8)

To examine the retention of SED training effect, two mixed-design ANOVAs were conducted to analyze the six months post-test data with participants’ pre-test and post-test data. The two ANOVAs were conducted to examine whether participants performed as well as post-test, and still better than their pre-test. Participants’ identification accuracy in post-test and six months post-test was compared using a 2 X 2 X 2 mixed-design ANOVA with time of test (post-test vs. six months post-test) and phonemic contrast (/t/-/d/ final vs. /ɛ/-/æ/) as the within-subject factors, and SED designs (Adaptive Staircase vs. Constant) as the between-subject factor. The analysis yielded no significant effect nor interaction, $F < 1$, $p > .05$, except for the marginally significant effect of phonemic contrast, $F(1,25) = 3.80$, $p = .06$ in which the identification accuracy on the /ɛ/-/æ/ contrast was higher than the /t/-/d/ final contrast.

Participants' identification accuracy in pre-test and six months post-test was examined in another 2 X 2 X 2 mixed-design ANOVA with time of test (pre-test vs. six months post-test) and phonemic contrasts (/t/-/d/ final vs. /ɛ/-/æ/) as within-subject factors and SED design (Adaptive Staircase vs. Constant) as between-subject factor. The results revealed a higher identification accuracy in the six months post-test compared to pre-test, $F(1,25) = 11.30$, $p < .01$. There was a close to significant interaction between time of test and phonemic contrast, $F(1,25)$, $p = .08$. The potential interaction was driven by the difference in training effect retention between the /t/-/d/ final and /ɛ/-/æ/ phonemic contrasts. Paired-sample t-tests showed no difference in identification performance on untrained /t/-/d/ final phonemic contrast between pre-test and six months post-test, $t(26) = 1.52$, $p > .05$. However, participants still identified /ɛ/-/æ/ phonemic contrast better in six months post-test than pre-test, $t(26) = 4.76$, $p < .01$. No other effect or interaction was found, $F < 2.57$, $p > .05$.

3.4.2.2 Training Generalizations to Novel Words over Six Months

To examine the generalization of training effect to untrained stimuli set and untrained phonemic contrast in six months post-test, two mixed-design ANOVAs were conducted. The generalization of training effect to untrained stimuli set was examined with a 2 X 2 X 2 mixed-design ANOVA, with stimulus set (trained vs. untrained /ɛ/-/æ/) and time of test (pre-test vs. six months post-test) as the within-subject factors and SED design (Adaptive Staircase vs. Constant) as the between-subject factor (Table 3.5).

Table 3.5 Mean percentage of correct identification in pre-, post- and six months post-tests for each stimulus set and training design (with standard error between brackets)

	Trained /ɛ/-/æ/			**Untrained /ɛ/-/æ/**		
SED Design	**Pre**	**Post**	**6 Months Post**	**Pre**	**Post**	**6 Months Post**
Adaptive Staircase	70.8 (2.6)	82.1 (2.3)	79.9 (2.9)	71.0 (4.1)	80.4 (2.6)	83.5 (3.2)
Constant	76.9 (3.1)	83.9 (3.6)	84.6 (4.3)	77.9 (3.3)	82.5 (4.1)	80.0 (4.0)

Identification accuracy on the /ɛ/-/æ/ stimuli words in six months post-test was higher than pre-test, $F(1,25) = 24.04$, $p < .001$. Post-hoc paired-samples t-tests revealed higher identification accuracy in six months post-test compared to pre-test for both Adaptive Staircase and Constant SEDs, $t(13) = 4.50$, $p = .001$ and $t(12) = 2.35$, $p = .04$ respectively. The post-hoc independent t-test however showed that the improvement size (difference between pre-test and six months post-test) generated by Adaptive Staircase SED was marginally greater than Constant SED, $t(25) = 1.84$, $p = .08$. The analysis revealed no other effects or interactions, Fs < 1.92, ps > .05.

3.4.2.3 Training Generalizations across Untrained Phonemic Contrast over Six Months

The generalization of training effect across untrained phonemic contrast was examined with another 2 X 2 X 2 mixed-design ANOVA, with contrast (/t/-/d/ final vs. untrained /ɛ/-/æ/) and time of test (pre-test vs. six months post-test) as the within-subject factors and SED design (Adaptive Staircase vs. Constant) as the between-subject factor.

Identification accuracy in six months post-test was greater than pre-test, $F(1,25) = 7.84$, $p = .01$. Importantly, there was a marginal significant interaction between time of test and SED design when trained /ɛ/-/æ/ stimuli were not included in the analysis, $F(1,25) = 3.09$, $p = .09$. Post-hoc paired-sample t-tests on the approaching significant interaction revealed that participants from Adaptive Staircase SED identified the untrained materials in six months post-test better than pre-test, $t(13) = 3.27$, $p < .01$; while participants from Constant SED showed similar identification accuracy between the pre-test and six months post-test, $t(12) = 0.73$, $p = .48$. No other effect or interaction was significant, $F < 2.27$, $p > .15$.

3.4.3 Discussion

As expected, the training effects generated by the Adaptive Staircase and Constant SED were retained even after 6 months of resting. Participants' identification accuracy in six months post-test was still better than their pre-test performance and did not differ from initial post-test. However, the greater improvement size generated by the adaptive staircase procedure did not remain over time as the SED design did not have an effect on the long-term training effect. The identification improvement on the /ε/-/æ/ phonemic contrast in six months post-test was only marginally greater when participants were trained with Adaptive Staircase SED as compared to Constant SED.

Note that there had been a 6 months gap between pre- and follow up six months post-test. It is understandable that advantage of adaptive staircase procedure could fade away especially when participants were not exposed to any further phonetic training. The advantage stemmed from the adaptive staircase procedure relied mainly on the training effect generalized across the untrained phonemic contrast /t/-/d/ final. In relation to that, the reduction of Adaptive Staircase SED's advantage in six months post-test could also be partly attributed to the degeneration of training effect on the untrained /t/-/d/ final phonemic contrast. While the identification improvement of /ε/-/æ/ contrast remained, the /t/-/d/ contrast's identification accuracy showed a tendency to deteriorate back to pre-test performance level six months after training when data from the Adaptive Staircase SED was analyzed together with Constant SED. In order to generate long term training benefit as observed from the /ε/-/æ/ phonemic words, an hour of discrimination training seems to be essential.

Nevertheless, when comparison was made between participants' pre-test and 6 months post-test performance, the data analysis suggested a tendency for the Adaptive Staircase SED to be better than Constant SED in the long term. Although the impact of training design on training effect was just approaching significance, further analysis indicated that participants of Constant SED showed no difference between pre-test and six months post-test identification performance on the untrained materials (novel words and untrained /t/-/d/ final contrast); while participants from Adaptive Staircase SED training continue to perform better in six months post-test when compared to the pre-test.

3.5 General Discussion

3.5.1 Training Effect of SED

In Experiment 2, the Constant SED which can also be known as the High Variability Phonetic Training (HVPT) with constant level of background babble was found to generate a training effect that generalized to novel stimuli of the trained phonemic category, and also to novel speakers whose productions were not included as the training materials. This finding is consistent with previous research on HVPT (Bradlow et al., 1997; Flege, 1989; Handley et al., 2009). In Experiment 3, we examined the effectiveness of Adaptive Staircase SED, in which the background babble level was pre-set for every participant before each training session, using a three-down one-up algorithm adaptive staircase procedure. As with the Constant SED, the training effect of Adaptive Staircase SED was observed and it also generalized to novel stimuli and novel speakers. On top of that, its training effect was found to even generalize across different phonemic contrast /t/-/d/ final which was not directly trained. This training benefit was not seen from Constant SED. Most importantly, when other features of the SED were held constant, Adaptive Staircase SED with the addition of adaptive staircase procedure was more effective as it yielded greater perceptual improvement (about 10%; pre-test 71% to post-test 81%) compared to Constant SED (about 3%; pre-test 73% to post-test 76%). In Experiment 4, the training effect of both SED paradigms was still found in participants six months after the last training session, with training effect still generalizing to novel stimuli. The training effect of Adaptive Staircase SED on the untrained phonemic contrast (/t/-/d/ at word final position) remained after the six months delay.

The results suggest that SED training successfully promote two types of learning generalization. First of all, perceptual learning from the training sessions did generalize to a different variety of English that was not used during the training. Although participants were trained with speech stimuli that were Southern Irish English, Irish English and American English accented across the three training sessions, their identification accuracy on the British English speech stimuli still improved in post-test. Instead of forming new phonemic categories for the training language variety, SED training improved participants' auditory sensitivity towards the acoustical difference between the auditory stimuli.

The identification improvement in post-test also showed the generalization of discrimination learning to participants' identification accuracy. Jamieson and Morosan (1986; 1989) proposed identification training as the better training task because it introduced new phonetic categories while discrimination task promoted intra-phonemic sensitivity that would not help in building phonemic category representations. In contrast with the present findings, their study showed that identification training improved discrimination accuracy (see also Giannakopoulou et al., 2013). The focus of the present study, however, was different from the above studies. The SED training aimed to improve listeners' bottom-up perception sensitivity to discriminate unfamiliar yet similar sounding phonetic sounds. It did not aim to assist in new phonemic category formation. With the discrimination training, participants were expected to identify the phonetic sounds better when they can solidly distinguish them, which were further supported by the present findings. Two related studies by Strange and Dittman (1984) and Handley, Sharples and Moore (2009) also showed this type of generalization. In addition, Handley, Sharples and Moore (2009) had provided preliminary support to the use of discrimination training as it was found to be as effective as an identification task. Because phonetic training studies had shown the effectiveness of both training tasks, it seems that the effectiveness of phonetic training depends mainly on how well it can direct listeners' attention to the relevant acoustic cues that discriminate the sounds phonemically, rather than on the type of training task.

3.5.2 The Advantage of Adaptive Staircase SED

We believe that training participants at their optimal performance in adverse conditions would encourage them to cross over their ability boundary and perform better than before training. Even though individual differences exist even among the same L1 speakers, the adaptive staircase procedure enables SED to personalize the training level (by manipulating the difficulty level/background babble level) to every individual based on their performance. Consistent with our prediction, the results from the present study supported our prediction about the individualized training. Individualized training level improved the effectiveness of HVPT.

In order to examine how far this training program could be extended, more research needs to be conducted on its efficacy in improving speech production and its efficacy across different phonemic materials on different language populations. Previous studies using phonetic training had found generalization of perceptual training effects to speech production accuracy (e.g. Bradlow et al., 1999; Nishi & Kewley-Port, 2007). The upcoming chapters of this thesis will focus on the production intelligibility of Malaysian speakers and the effect of perceptual SED training on their production intelligibility.

Chapter 4: Phonetics Perception and Production Complication of Malaysian Speakers

4.1 Introduction

Phonetics perception and production of Malaysians can be very complicated due to the diverse linguistic environment in the country and the diverse linguistic backgrounds of the speakers.

4.1.1 Phonology of the Three Main Languages in Malaysia

4.1.1.1 Bahasa Malaysia or Malay

Malay is a language spoken by nearly 250 million people living in Indonesia, Malaysia, Brunei and Singapore (Tadmor, 2009; Yap et al., 2010). The Malay language family consisted of the Baku Malay and the Standard Malay varieties. Both varieties significantly differ with each other in terms of their population demographics and their functionality. The Baku Malay is usually spoken to facilitate inter-nation Malay communication. It is more standardized and it is officially taught in Singapore and Malaysia; the language can also be understood by most Indonesians (Yap et al., 2010). On the other hand, the Standard Malay is commonly spoken by Malaysians and Malay speaking Singaporeans in daily communication. It is adapted mainly from the Riau-Johor dialect of Malay and it is used in non-academic and formal contexts (e.g. market place, media, casual conversation with friends and family etc.; Karim, Onn, Musa, & Mahmood, 2015). Baku Malay in general emphasizes more on phonetical accuracy compared to Standard Malay; for instance, some habits or characteristics of everyday speech such as final vowel lowering and the final consonant deletion are common in Standard Malay but are discouraged in Baku Malay.

Malay is usually written in Rumi, a Latin alphabet system. The standard Malay has a total of 25 alphabetical letters (the <x> is not used) and 34 phonemes, with 6 of the phonemes being composed of vowels (/a/, /e/, /ə/, /i/, /o/, /u/) that can be matched up to 7 phonetic sounds (with the inclusion of the /ɛ/ which is an allophone of /e/ for some borrowed words originated from

English; e.g. the word *pek* - */pɛk/* from the English word ‘pack’) and remaining 28 being consonants (see Figure 4.1; Yap et al., 2010; Karim et al., 2015). There are also 3 diphthongs (written as the following digraphs <ai>, <au>, <oi>) in Malay. The two adjacent vowels are sometimes pronounced as two syllables with a gap in between; for instance, the word *la-in* (“different”). Malay also has very few consonant clusters (a combination of two or more consonants within a syllable; e.g. /ks/ - in the word <*teks*>) and most of them are from English load words.

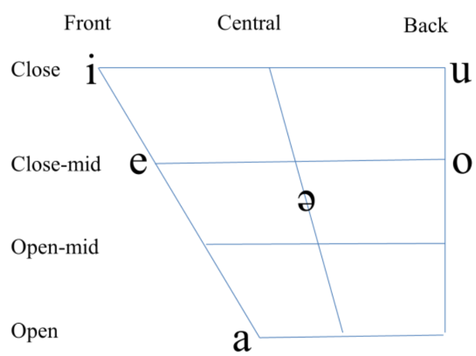


Figure 4.1 The International Phonetic Alphabet (IPA) vowel charts for BM (based on Li, Aljunied, & Teoh, 2005; Yap et al., 2010).

*Note: The horizontal labels – front, central and back refer to the position of the tongue when producing the sounds; while the vertical labels – close, close-mid, open-mid and open refer to the shape of the lips. For instance, when producing the /u/ vowel (as in the English word - *who*), the tongue body is raised as close as possible at the back of the mouth and the lips are simultaneously rounded and protruded (lips close to each other); in contrary, when producing the /a/ vowel (as in the Malay word – *abang*, means “older brother” in English), the highest point of the tongue is at the front of the mouth and the lips are as open as possible.

4.1.1.2 Mandarin

Mandarin is a language written in a logographic script in which each character acts as a grapheme[7] that corresponds to a morpheme[8]. The characters usually represent one syllable of spoken Mandarin and they can either stand as a word on its own or be a part of a polysyllabic word.

Hanyu Pinyin system is a phonetic transcription of Mandarin. It is written in Latin alphabet, along with a few diacritical marks to represent the standard pronunciation of the language. According to Howie (1973), the rules of Hanyu Pinyin only allow around 400 different segmental syllables and less than 1200 tonally differentiated syllables as unit morphemes. A Mandarin syllable can be composed by three parts: the initial, the final and the tone.

All consonants except /n/ and / ŋ/ only occur as initials of a syllable. In Beijing Mandarin however, the common use of retroflex suffix 儿 “er” (/ɚ/) at the end of nouns can also create a consonant-like end of syllable, forming a “retroflex final” phenomenon (Lin, 1989). Due to the fast speaking rate, the originally independent syllable can merge with the preceding syllable and become a segment-affiliated feature in the language (Huang, 2010). For instance, the diminutive form of the Mandarin words 词儿 ‘lyrics’ is often pronounced as merged /tsʰər/ instead of the root sounds /tsʰɨ/ /ər/.

Huang (1992) considers Mandarin’s palatals (/t͡ɕ/, /ɕ^{2}/, /t͡ɕʰ/) and velars (/kʰ/, /k/, /x/) as allophones, treating /t͡s/ as consonant clusters of /t/ and /s/. He proposed a total of 19 consonant and 5 vowel phonemes (/i/, /y/, /a/, /ɤ/, /u/) in which each vowel enjoys a wide range of predictable articulations (see Figure 4.2). On the other hand, Rogers and Dalby (2005) proposed a

---

7 The smallest units of written language / or all written symbols or sequences of written symbols that represent a single phoneme, e.g. *高 ”*high or tall*”* and *山 “*mountain*”* of the Mandarin logographic script system /or *f* and *ph* of the alphabetical system.

8 the shortest speech element that has a meaning or grammatical function that cannot be subdivided into further such elements, e.g. -*ing* and *non*- of English /or the words 冰 “ice” 箱 “box” of Mandarin which means “refrigerator” when combined.

total of 23 consonant and 6 vowel phonemes that form the Mandarin language system. They included /w/ as one of the consonant and the diphthong phoneme /ɑʊ/ as one of the vowels. Most linguists (e.g. Chao, 1968) agree that Mandarin adopts a four tones system: high-level, high-rising, low-dipping and high-falling, with an additional “neutral tone”[9].

Different languages place different emphases on how phonemes are realized phonetically. For instance, spoken Mandarin strongly emphasizes the distinction between aspirated and unaspirated productions of the initial consonant in the syllable, which is usually followed by the voicing of the vowels (Huang, 1992). An aspiration phonetic sound is made when the compressed air is withheld from a consonantal constriction and is released and a “breathy” quality sound is produced (e.g. the voiceless aspirated plosives /p^h/, /t^h/, /k^h/ which are represented by the alphabets <p>, <t>, and <k> in the romanized Chinese orthographic system, as opposed to their unaspirated opponents /p/, /t/, /k/ which are represented by the alphabets <b>, <d> and <g>). On the other hand, the Indo-European languages (including English) also emphasize on voicing, which is one of the distinctive features of the languages (Huang, 1992). The voicing articulation involves the vibration of the vocal folds and the productions of vowels (e.g. /ə/, /i/, /u/), nasals (e.g. /m/, /n/, / ŋ/) and approximants (e.g. /w/, /y/) usually require voicing. The articulation focus thus marks one of the distinctions between English and Mandarin pronunciation.

[9] High level tone/ first tone – a steady high sound; marked by a macron (¯)
High Rising tone/ second tone – a sound rises from mid-level tone to high; marked by an acute accent (´)
Low dipping tone/ third tone – a sound falls from mid-low to low descent tone; marked by a caron/háček (ˇ). High Falling tone/ forth tone – features a sharp downward accent (“dipping”) from high to low; by a grave accent (`). Neutral tone/ fifth tone – light or lack of tone

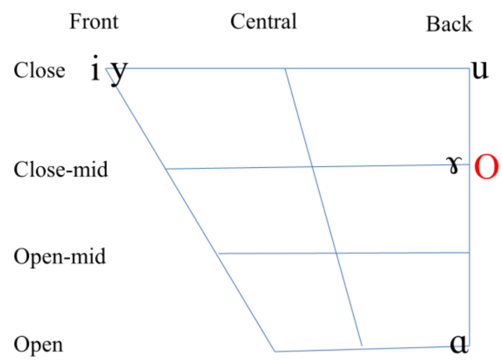


Figure 4.2 The International Phonetic Alphabets (IPA) vowel charts for Mandarin (based on Duanmu & Keith, 2006; Huang, 1992; Zhao & Li, 2009)

* Note: Some researchers include /o/ as one of the Mandarin core vowels (Zhao & Li, 2009)

4.1.1.3 English

Due to the diffusion and localization of English throughout the world, a number of English varieties were formed; with a few varieties (i.e. English from the Inner Circle[10]: British English and America English; see e.g. Kachru, 1985, 1990), better recognized and acknowledged than others (e.g. English from the Outer Circle[11]: Malaysian English and Indian English, or from the Expanding Circle[12]: Japanese English, China English; see Kachru, 1990). The phonetic inventories of the English varieties can vary, and the variation lies mostly on vowels rather than consonants (McMahon, 2002).

English is written in Rumi, a Latin alphabet system. According to Davenport and Hannahs (2011), the "typical" English consists of 31 consonant phonemes. Different varieties of English however,

[10] According to Kachru's Three Concentric Circle Model, the inner circle is dominated by the mother-tongue varieties in which English is spoken as the first language; the countries in the inner circle include United States of America, United Kingdom, Canada, Australia and New Zealand (Kachru, 1990).
[11] The outer circle consists of post-colonial nations (most of the nations were colonies of the United States of America or United Kingdom) that speak English as second language; these countries include Malaysia, Singapore, India, and Kenya (Kachru, 1990; Jenkins 2012).
[12] The expanding circle refers to the countries that use English as a 'foreign' language, but have no history of colonization by the countries from the Inner Circle; including countries such as China, Japan, Saudi Arabia and Israel (Kachru, 1990).

may have different numbers of vowels in their inventories. The vowel inventory of Received Pronunciation English is being referred in this section because this pronunciation model is commonly used to teach and learn the English language in Malaysian schools (Pillai & Jayapalan, 2010). The Received Pronunciation English has a conservative estimation of 19 to 21 distinct vowel phonemes (Davenport & Hannahs, 2011), which include 12 monophthongs (see Figure 4.3) and 9 diphthongs (i.e. /eɪ/, /aɪ/, /aʊ/, /ɔɪ/, /əʊ/, /ɪə/, /eə/, /ɔə/, /ɔə/, /ʊə/).

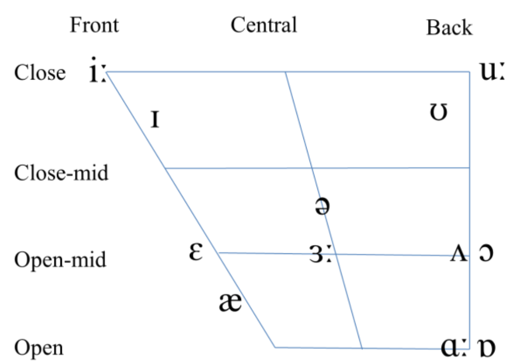


Figure 4.3 The International Phonetic Alphabets (IPA) vowel charts for Received Pronunciation English (based on Roach, 2009; Davenport & Hannahs, 2011)

*Note: Received Pronunciation English refers to the Standard English mostly accepted in United Kingdom (Roach, 2009).

4.1.1.4 Comparisons of Phonetic Inventory between Mandarin, Malay and Standard English

According to Yap et al. (2010), English words demand a heavier cognitive processing load than most other alphabetic orthographies do, due to its deep orthography which can be quantified by its letter to phoneme ratio. Although Mandarin is not an alphabetic language, its orthography depth was approximately measured in this section based on the Latin alphabets used in Hanyu

Pinyin system. Since the adoption of Hanyu Pinyin in Malaysia education system from the year 1982, Malaysian students learn the pronunciation of Mandarin characters with the help of phonetic transcription system (Xiaomei, 2014). Therefore, the transparency of Hanyu Pinyin phonetic script would also reflect the degree of learning difficulty for Mandarin pronunciation.

According to Yap et al. (2010), a language with shallow orthography has an isomorphic relationship between word spelling and sound, where the mappings between orthography and phonology are transparent and predictable. In the English phonetic system which has a relatively deep orthography, a single grapheme in English could represent different sounds in different phonetic contexts (e.g. the alphabet <a> from British English is read as /æ/ in the word <ash>, /ɔː/ in the word <all> and /ɑ/ in the word <father>).

When compared to English, the phonetic script of Mandarin, Hanyu Pinyin has a relatively shallow alphabetic orthography and the language does not have a morphology or affixation system similar to the alphabetic languages (Packard, 2000). Malay also has a relatively shallow alphabetic orthography, simple and short syllable structures and transparent morphology or affixation (Yap et al., 2010).

Looking at the orthographic depth of the three languages: Malay has 25 letters and 34 phonemes thus accompanied with a 0.74 letters to phonemes ratio (LP ratio) (Yap et al., 2010); Hanyu Pinyin (Mandarin) has 25 letters and 29 phonemes, with a 0.86 LP ratio (Chao, 1968; Rogers & Dalby, 2005); English has 26 letters and 44 phonemes, with a 0.59 LP ratio (Yap et al., 2010; Rogers & Dalby, 2005).

The smaller LP ratio of English may yield perception and production difficulties in L1 Malay and Mandarin speakers who are used to their L1 shallow orthography language system. The difficulty arises when the phonemic distinction in English does not contrast distinctively in the Malay and Mandarin language systems and can be treated as allophones to one phonemic category. It is important to keep in mind that the LP ratio calculation only focused on alphabetic orthography and did not take into account the Mandarin tones. The Mandarin tones are also phonemic in Mandarin and they are known to pose perceptual difficulty to non-native Mandarin learners whose native language does not emphasize on tonal or pitch distinction (Huang & Johnson, 2011).

When compared to Malay and Mandarin, English also has a higher syllable complexity and larger grain size. These two measures are used to predict the amount of processing demands a reader would need. The syllable complexity can be quantified by the number of consonant clusters at the initial or final position of a word. Both Malay and Pinyin system of Mandarin have very limited consonant clusters. For instance, Mandarin words only end with vowels or the consonants /n/ and /ŋ/ or also the retroflex /ɚ/ in Beijing Mandarin (Yap et al., 2010; Duanmu, 2006). On the other hand, English syllables often begin and end with a few consonant clusters. Due to the huge amount of possible consonant and vowel clusters of English, a much larger inventory of possible syllables could be formed in English compared to the other two languages (Rogers & Dalby, 2005). The syllable complexity of English further complicates the word recognition process.

Grain size refers to the size of sub-lexical representations or constituent parts of a word is used to optimize reading and language processing (Wydell & Butterworth, 1999). When a language has a deep orthography (e.g. English), readers need to depend on larger units of sub-lexical representations (larger grain size) for more consistent mappings to phonology (see Treiman, Mullennix, Bijeljac-Babic, & Richmond-Welty, 1995 for further discussion), thus resulting in greater processing demands during reading. The deep orthography, high syllable complexity and larger grain size (higher processing demands) of English might be part of the reasons that cause difficulties in the L1 Malay and L1 Mandarin speakers to achieve a native-like fluency for the language.

4.1.2 Perception and Production Complications of Spoken English

4.1.2.1 General Difficulties Experienced by General English Learners

Many L2 English speakers have difficulty differentiating between English vowels as the 6 vowel letters can be matched up to 20 vowel phonemes. Accurate pronunciation of the English vowels depends greatly on the syllable context, or more specifically, the adjacent consonants. Some of the most confusable vowel contrasts include the English /e/-/ɛ/-/æ/ (Richie & Kewley-Port, 2008; Simon & D'Hulster, 2012), /iː/-/ɪ/ (Giannakopoulou et al., 2013; Krebs-lazendic & Best, 2002), /uː/-/ʊ/ (Wang & Munro, 2004; Yang & Fox, 2014), and /ɒ/-/ʌ/ (Zhang, Feng, Dang, Zheng, & Xu, 2015).

Part of the vowel confusion (i.e. the English contrast /iː/-/ɪ/ and /uː/-/ʊ/) can be attributed to the lack of tense-lax distinction in many other languages' vowel systems (e.g. Spanish, Catalan, Polish, and Mandarin). The tense vowels (e.g. /iː/ and /uː/ as in the English words *peel* and *pool*) are pronounced with a relatively longer duration, while the tongue is positioned relatively higher and less centralized than their counterparts, the lax vowels (i.e. /ɪ/ and /ʊ/ as in the English words *pill* and *pull*). Participants from the L1 background that do not distinguish the tense and lax vowels often over rely more on the durational cue rather than the spectral properties (Kondaurova & Francis, 2008; Wang & Munro, 2004); which in turn might not be sufficient to discriminate the similar sounding phonemes. Other most commonly studied difficult contrasts include the consonant pairs /t/-/d/ (e.g. Flege, 1989), /θ/-/ ð/ (e.g. Jamieson & Morosan, 1989) and /r/-/l/ (e.g. Iverson et al., 2005). The discrimination difficulty likely arises due to the absence of the contrast in the speakers' L1.

4.1.2.2 Some Common Difficulties Experienced by Malaysian English Learners

Vowel phonemes for the grapheme <a> and <e> in English could be the most difficult as the number of phonemes they can be matched to out-number those in Malay and Mandarin pinyin system. For instance, the English grapheme <a> is pronounced differently for the respective English words: as /æ/ in the word *cat*, /ɑ:/ in the word *part* and /ə/ in the word *America*; the English grapheme <e> is pronounced as /ɛ/ in the word *bed* and /ə/ in the word *teacher*. Moreover, both Malay and Mandarin languages do not have as many diphthong vowel phonemes as English. Therefore, some of the diphthong vowel pairings such as the /aʊ/-/oʊ/ (as in the English words *couch*-*coach*) and /aʊ/-/a/ (as in the English words *down*-*darn*) may have ambiguous phonemic representation unless the listeners find them phonetically salient. The ambiguity in the phonemic representations could cause inaccurate phonemic perception and production, and also greater cognitive load considering the speakers may need to take into account of the speech and phonetic context in order to identify the words heard.

An example of Malaysians' confusion with the English deep orthographic phonemes was found by Yap, Wong and Aziz (2010). The authors who examined the representation of English vowels in Malay-English bilinguals found that the bilinguals tended to assimilate the two similar sounding phonemes into the low-front vowel category /æ/, which is phonetically closer to the low-front vowel /a/ in their L1 or Malay phonemic system (see Figure 4.4).

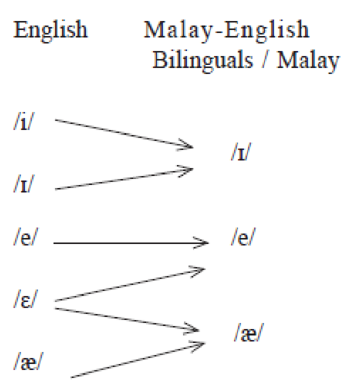


Figure 4.4 Single Category Assimilation of some English phonemes into fewer Malay phonemes (Retrieved from Yap, Wong & Aziz, 2010).

Other than the assimilation of the two front vowels, the results showed that participants had only one phonetic representation for the English tense and lax high front vowels /iː/-/ɪ/ (see Figure 4.4; Yap et al., 2010). The assimilation was probably due to the absence of tense-lax vowels distinction in the Malay language. When teaching Malaysian students, Hart (1969) also realized the same problem in which learners of L1 Malay, Mandarin or any southern Chinese languages (e.g. Cantonese, Hakka) often produced homophones out of the English tense and lax vowels such as the /iː/-/ɪ/, /ɔː/-/ɔ/, and /uː/-/ʊ/ contrasts. The absence of the tense and lax vowel contrast in the Malay and Mandarin language (two of the most widely spoken languages in Malaysia) might be the reason as to why this discrimination difficulty seems to be universal to the general Malaysian population.

Besides vowel perception difficulty, the perception of the /t/-/d/ consonant phonemic contrast (as in *beat* and *bead*) in L1 Chinese languages speakers is another well-studied subject (e.g. Flege, 1989). The English voiceless-voiced /t/-/d/ phonemic contrast can have different distinctive acoustic features depending on the phonetic context. In addition, the voicing feature used to differentiate the phonemic pair can be inapplicable in certain circumstances because the voiced stop /d/ is often partially or completely devoiced in spoken English (i.e. produced without closure voicing through some part or the entire closure interval; Yap, 2013).

Yap (2013) also argued that the more accurate acoustic cue used to differentiate the pair at word initial position should be the presence of aspiration rather than voicing. The L1 Mandarin speakers from Yap's study and this thesis showed near ceiling identification accuracy for the word initial phonemic contrast even though their L1 Mandarin does not encourage voicing discrimination. The ceiling performance may have reflected the presence of another acoustic cue that could assist the discrimination: the aspiration distinction between the aspirated <t> and unaspirated <d> alphabets. The L1 Mandarin speakers are expected to possess a native-like perceptual sensitivity to discriminate the aspirated – unaspirated acoustic cues, because the

Mandarin language system promotes aspiration discrimination (see section 4.1.1.2). Due to the ambiguous presence of the voicing cue in the voiced stop /d/ phoneme and the presence of aspiration distinction, the more appropriate phonemic contrast to represent the English alphabets <t> and <d> at word initial position should be the /t^h/-/t/ pair respectively (Yap, 2013).

The L1 Mandarin speakers however, have difficulty discriminating the /t/-/d/ contrast when they are located at word final position. The difficulty was predicted because Chinese languages (e.g. Mandarin, Cantonese, Hokkien, Hakka) in general do not have consonants as the word-final obstruent (with <ng> or the phoneme /ŋ/ as the exception). Similar as the word initial /t/-/d/ contrast, the voicing distinction between the word final voiceless-voiced pair is often neutralized in spoken English as well (Flege, 1989; Yap, 2013). Flege (1989) proposed that the native English speakers discriminate the word final phonemic contrast using the acoustic information from the preceding vowel (e.g. vowel duration). The non-native English speakers however, do not show the practice.

In Malay language, it is common to find voiceless plosive consonants (e.g. /p/, /t/) at the beginning, medial and final location of the words. The Malay words however, do not end with the opposing voiced plosives (e.g. /b/, /d/), except for the borrowed words such as *Ahad* “Sunday” and *jawab* “answer” which originated from Arabic (Hassan, 2005). In these occasions, the pronunciation of the voiced plosives is often devoiced. Although the voiceless and voiced stops /t/-/d/ phonemes contrasting at word final position is common in English, the contrast does not exist in Mandarin words and rarely exist in Malay words. Thus, the L1 speakers of Malay or Mandarin may have difficulty discriminating the word final phonemic pair .

4.1.3 Research Questions

In this chapter, we aimed to examine the phonetic perception pattern of Malaysian speakers (i.e. L1 Mandarin and Malaysian English speakers) in comparison with British speakers. Findings from this chapter also provided a preliminary overview on the phonetic perception performance of raters recruited for experiments that will be reported in Chapter 5. Production intelligibility of L1 Mandarin speakers to the L1 ME speakers and native British English speakers was also examined in this chapter.

There are three main issues which this chapter addresses: the first being whether L1 Mandarin Malaysian speakers have difficulty discriminating the British English speech production of /ɛ/-/æ/ and /t/-/d/ phonemic contrasts. Due to the lack of contrast in vowel quality between the /ɛ/-/æ/ pair in Malaysia languages, the L1 Mandarin speakers are hypothesized to have one phonetic representation for the two phonemes and will tend to assimilate them to the /æ/ phonemic category (similar to Yap et al.'s (2010) findings). If this is the case, then the identification accuracy of the L1 Mandarin speakers on the /æ/ phoneme would be higher than the /ɛ/ because of their higher tendency to choose /æ/ to represent both phonemes heard. Also, the identification accuracy on the voiceless stop /t/ at word final position would be higher than its opponent voiced stop /d/ because firstly, the /t/-/d/ phonemic contrast at English and Malay words final position is often devoiced in natural speech (see Hassan, 2005; Flege, 1989) and secondly, the lack of use of the preceding vowel acoustic information (e.g. spectral quality, duration and F1 offset frequency; see section 2.4 for further discussion) by the L1 Mandarin speakers. Due to these two reasons, the devoiced /d/ phoneme would probably be misperceived as the voiceless plosive /t/ by the inexperienced L1 Mandarin speakers.

The second issue which this chapter examined was whether the perception pattern differs between L1 Malaysian English (ME) speakers and British English speakers. Like other neighboring varieties of English, Malaysian English has smaller vowel space compared to British English (see Pillai et al., 2010) and its phonetic system is influenced by other local languages spoken in Malaysia (see section 1.1.4 for further discussion on Malaysian English). The two groups of speakers who speak different L1 English varieties are expected to perceive the speech productions differently, with them perceiving their own L1 English variety more accurately.

The third matter of interest relates to the production intelligibility of the L1 Mandarin Malaysian speakers. They were expected to produce the /ɛ/-/æ/ phonemic words with an ambiguous phonemic distinction as perceived by the L1 ME speakers and the native British English speakers. As for the /t/-/d/ phonemic contrast, the L1 Mandarin speakers should have no difficulty in producing highly intelligible phonemic words when the pair locates at the word initial position. On the other hand, some productions of the voiced stop /d/ by the L1 Mandarin speakers would be partially or completely devoiced.

4.2 Experiment 5: Method

4.2.1 Design and Materials

The stimuli consisted of sixty-four spoken English words that were used in pre- and post-test of SED (see Appendix 1 for full list). Two types of stimuli were involved: the British English speech stimuli and the Malaysian English speech stimuli. The sixty-four British English speech stimuli were taken from the pre- and post-test of SED and were spoken by one native British English speaker (female; see section 2.2.1 for recording details). The Malaysian English speech stimuli were taken from recordings made in Chapter 5 and consisted of 192 auditory stimuli which were spoken by fourteen L1 Mandarin Malaysian speakers (9 females). The 192 auditory stimuli were recorded at three different time points: after pre-test, before post-test and before 1 month post-test (see section 5.2.2.1 and 5.2.2.3 for more recording details).

4.2.2 Participants

Three speaker groups provided perception data in this study: 1) the Malaysian multilinguals who spoke Mandarin as their L1 (L1 Mandarin speakers), 2) the Malaysian multilinguals who spoke Malaysian English as L1 (L1 ME speakers) and lastly 3) the British monolinguals who spoke

British English as their native language (native British English speakers). As the involved participants were recruited in experiments reported in Chapter 5, their further details can be found in section 5.2.2.2 and 5.3.2.2. The L1 Mandarin speakers were recruited as control participants in Chapter 5.

4.2.3 Procedure

The L1 Mandarin speakers named the sixty four stimuli words in three times of test – pre-test, post-test and 1 month post-test without receiving any phonetic training. Their productions were recorded through a microphone. Then, the spoken words were identified and rated by L1 ME speakers and native British English speakers, using a 2 alternatives forced-choice ID task. See Experiment 7 from Chapter 5 for the detailed rating procedure.

All three groups of participants completed the pre- or post-test of SED training (on British English speech stimuli).

4.3 Results

Identification performance of the L1 Mandarin speakers, the L1 ME speakers and the native British English speakers on the British English speech stimuli and Malaysian English speech stimuli was analyzed to examine the intelligibility differences of the speech stimuli. For the following data analysis, mean percentage of correct identification was calculated for every phonemic category, and participants' identification accuracy was compared.

4.3.1 Perception Difficulty of L1 Mandarin Malaysian Speakers on British English Speech Production

As seen in Figure 4.5, identification accuracy on the /t/ and /d/ initial phonemic categories was near to ceiling performance (above 95%). On the other hand, the L1 Mandarin speakers appeared to have the greatest difficulty in identifying the /d/ final phoneme (68%), and followed by the /ɛ/ phoneme (71%).

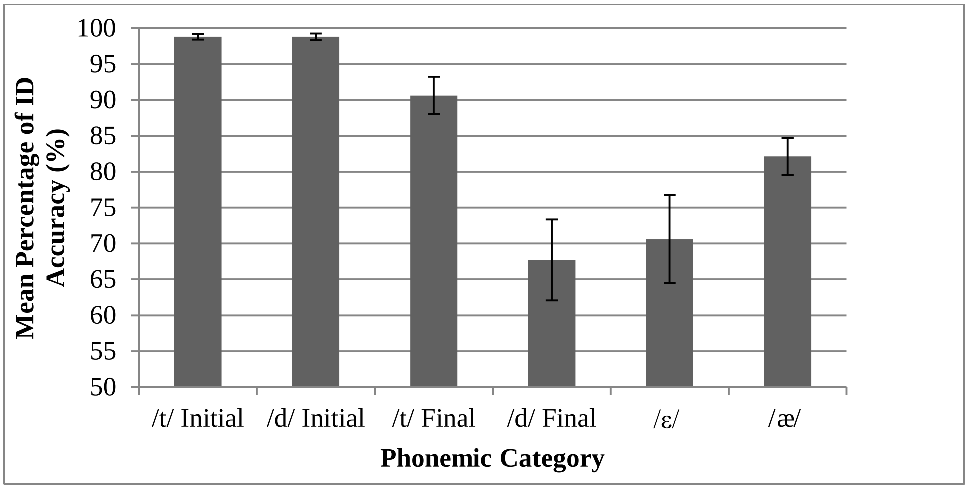


Figure 4.5 Mean percentage of correct identification across pre-test, post-test and 1 month post-test of the L1 Mandarin speakers for every phonemic category (with error bars as the standard error).

The mean percentage of correct identification was subjected to a 3 X 6 repeated-measures ANOVA; with the time of test (pre- vs. post- vs. 1 month post-test) and the phonemic category (/t/ initial vs. /d/ initial vs. /t/ final vs. /d/ final vs. /ɛ/ vs. /æ/) as the within-subject factors. The identification accuracy of the L1 Mandarin speakers did not change across the time of test, $F(2,26) = .84$, $p = .44$ and the consistency of identification accuracy across the time of test was not affected by the phonemic categories, $F(10,130) = .44$, $p = .92$. The identification accuracy however was not uniform across the phonemic categories, $F(5,65) = 18.21$, $p < .001$. Pairwise comparisons (Bonferroni) showed that the identification accuracy on the /t/ initial, /d/ initial and /t/ final was significantly higher than the other three phonemic categories, $p < .01$, but did not differ among each other, $p > .05$. The identification accuracy on the /d/ final, /ɛ/ and /æ/ phonemic categories also did not differ among each other, $p > .05$.

4.3.2 Perception Difference between L1 Malaysian English Speakers and British English Speakers

4.3.2.1 on British English Speech Production

Figures 4.6, 4.7 and 4.8 below illustrate the mean percentage of correct identification for every stimuli word produced by the native British English speaker. From the figures, the British speakers appeared to be more consistent in their identification performance while the Malaysian speakers showed a greater variation in their identification accuracy across the stimuli words. The inconsistency in identification was especially prominent for the /ɛ/ phonemic category, with the word 'men' recognized close to chance level (54%) and four words 'celery', 'merry', 'shell', 'kettle' recognized at the rate of below 80%. In the phonemic category /æ/, only the word 'salary' was correctly identified at about 80% accuracy; other words were relatively better identified than their /ɛ/ counterparts.

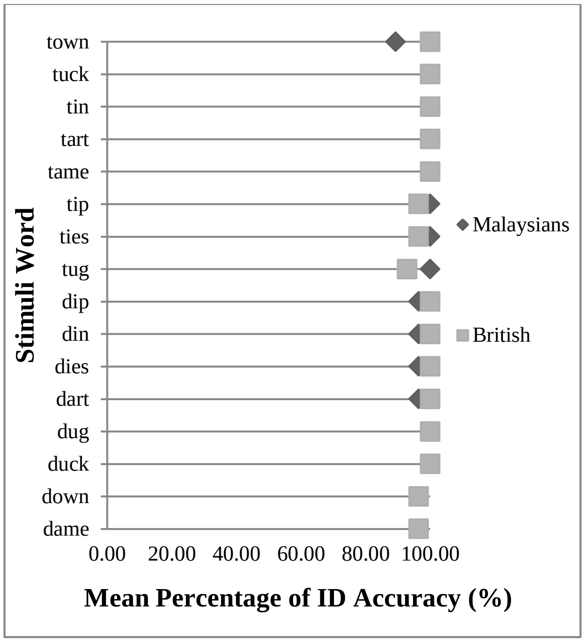


Figure 4.6 Mean percentage of correct identification averaged across the L1 Malaysian English speakers and the native British English speakers for each British English stimuli word that comes from the /t/-/d/ initial phonemic contrast.

Note: The stimuli words were arranged according to their phonemic category and sequenced based on the difference magnitude between the speaker groups. The difference magnitude was calculated by subtracting the Malaysian speakers' identification accuracy from the British speakers'. Starting from the top of the list is the word with the greatest positive difference magnitude (Malaysians < British) and the magnitude reduces across the list to the last word of the phonemic category that either has a zero value (Malaysians = British) or a negative value (Malaysians > British).

*top y-axis: words with /t/ initial phoneme; bottom y-axis: words with /d/ initial phoneme.

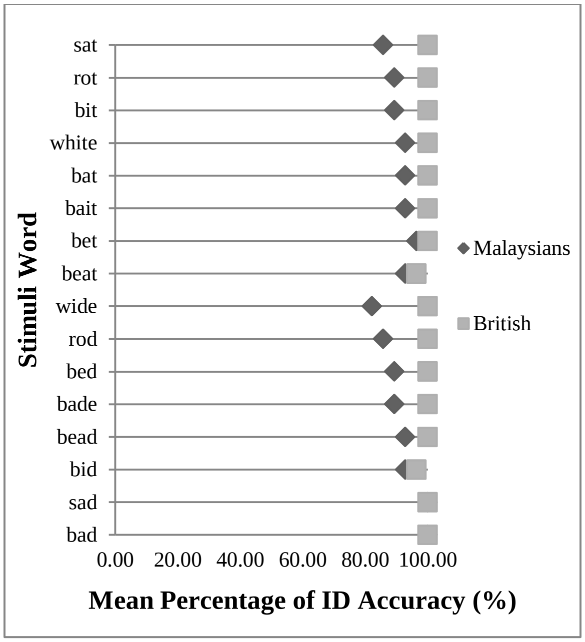


Figure 4.7 Mean percentage of correct identification averaged across the L1 Malaysian English speakers and the native British English speakers for each British English stimuli word that comes from the /t/-/d/ final phonemic contrast.

Note: The stimuli words were arranged according to their phonemic category and sequenced based on the difference magnitude between the speaker groups (see Figure 4.2 for detailed description on the calculation).

*top y-axis: words with /t/ final phoneme; bottom y-axis: words with /d/ final phoneme.

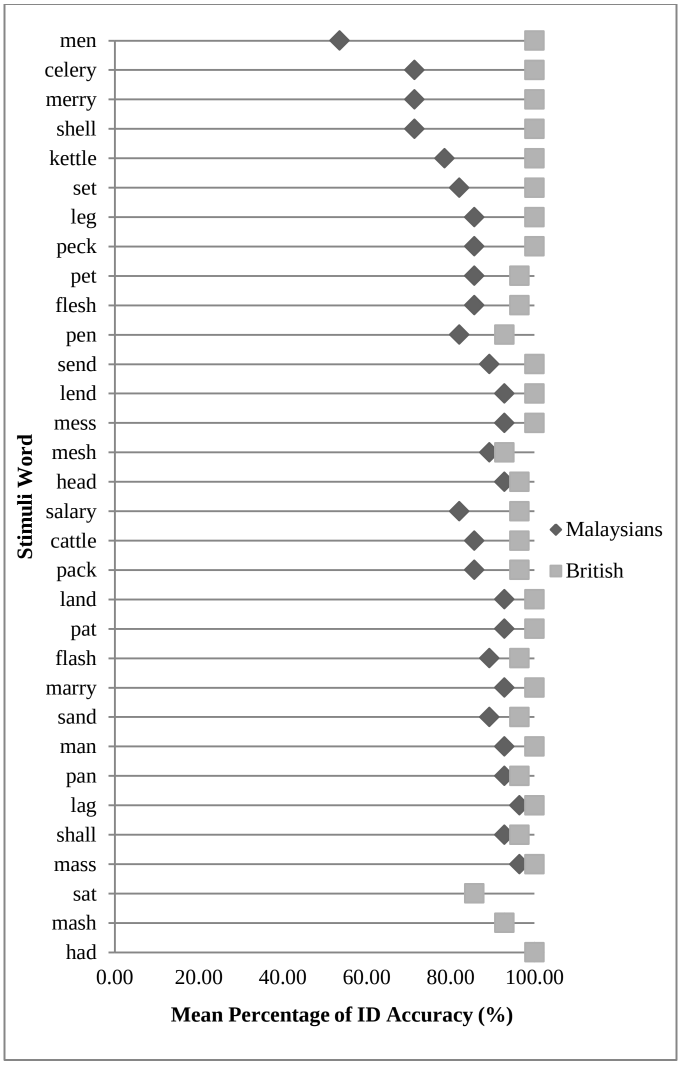


Figure 4.8 Mean percentage of correct identification averaged across the L1 Malaysian English speakers and the British English speakers for each British English stimuli word that comes from the /ɛ/-/æ/ phonemic contrast.

Note: The stimuli words were arranged according to their phonemic category and sequenced based on the difference magnitude between the speaker groups (see Figure 4.2 for detailed description on the calculation).

*top y-axis: words with /ɛ/ phoneme; bottom y-axis: words with /æ/ phoneme.

The phonetic perception of the Malaysian speakers and British speakers was examined using their identification accuracy on the speech stimuli produced by a native British English speaker (see Figure 4.9). The British speakers showed a ceiling performance in the ID task with at least 97% of identification accuracy for every phonemic category. Identification accuracy of the Malaysian speakers fell generally above 90% for all phonemic categories except for the /ɛ/ phoneme which fell at 81%.

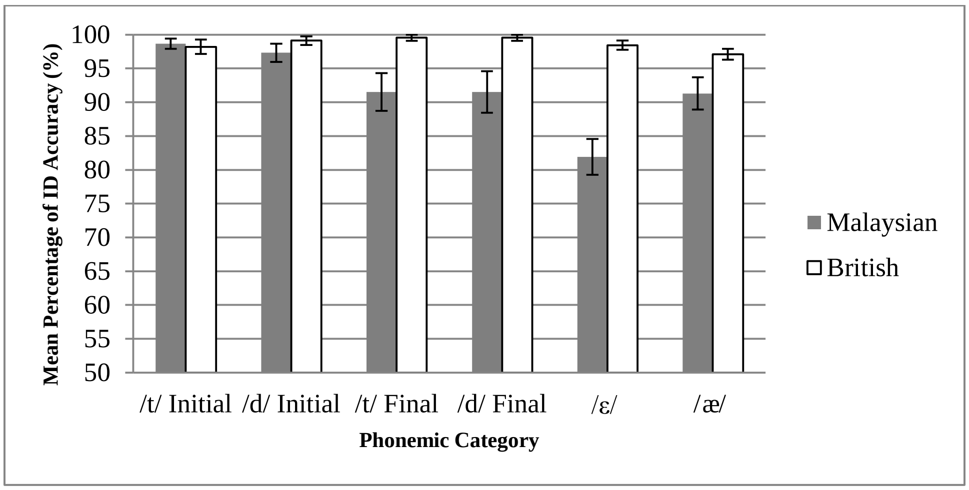


Figure 4.9 Mean percentage of correct identification of Malaysian speakers and British speakers for each phonemic category of the British English speech stimuli (with error bars as the standard error).

The mean percentage of correct identification was subjected to a 2 X 6 mixed-design ANOVA; with speakers (ME vs. British English) as between-subject factor and phonemic category (/t/ initial vs. /d/ initial vs. /t/ final vs. /d/ final vs. /ɛ/ vs. /æ/) as within-subject factor. The analysis revealed a higher identification accuracy of British (99%) compared to the Malaysians (92%), $F(1,54) = 20.05$, $p < .001$. The results also revealed a main effect of phonemic category, $F(5,270) = 8.17$, $p < .001$. Pairwise comparison showed that identification accuracy of the /ɛ/ phoneme was significantly lower than all other phonemic categories ($p < .05$) while identification accuracy of the /æ/ phoneme was significantly lower than the /t/ initial phoneme. An interaction was also found between the speaker group and phonemic category, $F(5,270) = 7.83$, $p < .001$. Post-hoc independent sample t-tests were conducted to compare identification accuracy between the two speaker groups for every phonemic category. The results showed that British speakers identified the British English /t/ final, /d/ final, /ɛ/ and /æ/ phonemes better than Malaysian speakers, ts > 2.30, ps < .05, but they performed similarly on the /t/ and /d/ phonemes at word initial position, ts < 1.21, ps > .23. Post-hoc one-way repeated-measures ANOVA that compared across the phonemic categories within each speaker group showed a main effect of phonemic category for the Malaysian speakers, $F(5,135) = 8.72$, $p < .001$ but not the British speakers, $F(5,135) = 1.90$, $p = .10$. Pairwise comparisons revealed significantly lower identification accuracy for the /ɛ/ phoneme compared to the rest of the phonemes for Malaysian speakers, $p < .05$. No difference in identification accuracy was found between other phonemic categories, $p > .05$.

4.3.2.2 on Malaysian English Speech Production Produced by L1 Mandarin Speakers

Analyses were first conducted to examine whether the production accuracy of the L1 Mandarin speakers changes across the time of test. The mean percentage of correct identification on their speech productions was subjected to two 3 X 6 repeated-measure ANOVAs; with the time of test (pre- vs. post- vs. 1 month post-test) and phonemic category (/t/ initial vs. /d/ initial vs. /t/ final vs. /d/ final vs. /ɛ/ vs. /æ/) included as the within-subject factors. The analyses were conducted separately for the L1 ME speakers and the British English speakers. Importantly, analyses for both speaker groups revealed neither effect of time of test nor any interaction Fs < 1.35, ps > .27. Only effect of phonemic category was found from both analyses. The /t/-/d/ initial phonemes were identified at near ceiling level (> 94%) by both groups and their identification accuracy was significantly higher than all other phonemic categories, $p < .05$. The ME speakers had the greatest difficulty in identifying the phoneme /ɛ/ (66%) among all the phonemic categories; with its identification accuracy being significantly lower than the /t/ initial phoneme ($p < .05$). The difference in identification accuracy between the phoneme /ɛ/ and the phoneme /d/ final also showed a considerable trend toward significance, $p = .06$. Interestingly, the British speakers showed a different perception pattern, in which the phoneme /æ/ turned out to be the most difficult phonemic category (66%), with its identification accuracy being significantly lower than the phoneme /t/ final, $p < .05$.

Due to the results of time of test being very consistent, we collapsed speakers' identification accuracy on the Malaysian English speech stimuli across the three times of test in the upcoming data analysis (see Figure 4.10 & Figure 4.11).

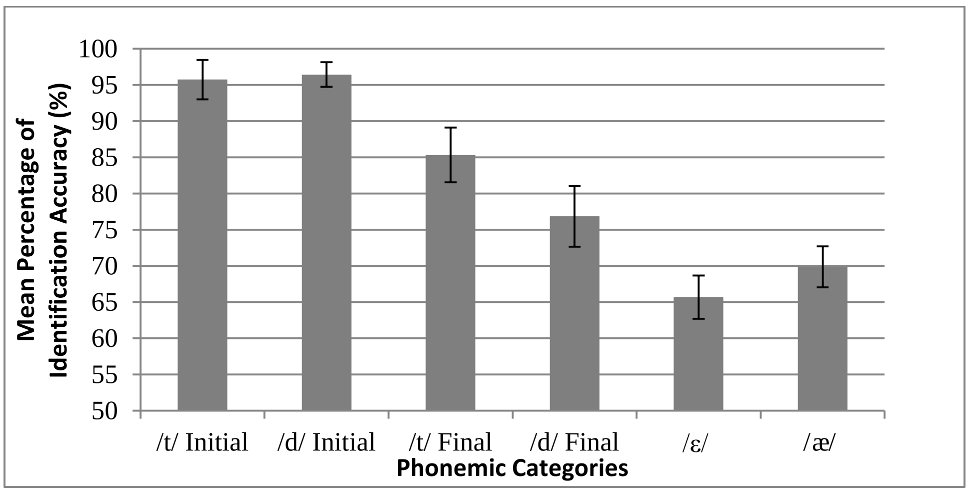


Figure 4.10 Mean percentage of correct identification of L1 Malaysian English speakers across time of test made for each phonemic category, of the Malaysian English speech stimuli (with error bars as the standard error).

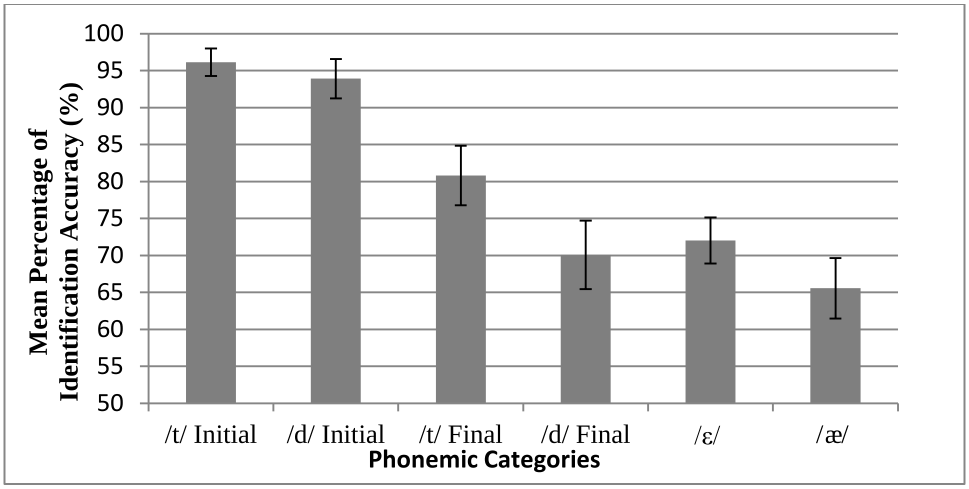


Figure 4.11 Mean percentage of correct identification of native British English speakers across time of test made for each phonemic category of the Malaysian English speech stimuli (with error bars as the standard error).

Figures 4.12, 4.13 and 4.14 below illustrated the mean percentage of correct identification for every Malaysian English stimuli word that was used in the ID task. The data was presented separately for the L1 ME speakers (Malaysians) and native British English speakers (British). Identification accuracy for the /t/-/d/ initial phonemic words appeared to be considerably consistent across the words for both Malaysian and British speakers, with some words starting with the /d/ phoneme being better identified by the Malaysian speakers.

There were more performance variations for the other phonemic categories between the two speaker groups. Similar to the /t/-/d/ initial phonemic words, British speakers showed a higher tendency to identify the /d/ phoneme at word final position as the /t/ phoneme when compared to the Malaysian speaker (see Figure 4.3). The British speakers' tendency to match the /d/ final phoneme to /t/ phoneme, however, did not affect their identification accuracy of the /t/ final phonemes. The Malaysian speakers still identified half of the /t/ final phonemic words (four out of eight stimulus words) better than the British speakers. The identification accuracy for the other four stimulus words was identical between the two groups of speakers.

The British and Malaysian speakers showed the greatest variation in the identification of /ε/-/æ/ phonemic words (see Figure 4.14). The Malaysian speakers showed a higher tendency to identify the /ε/ phonemic words as the /æ/ phonemic words. . At least 5 words from the /ε/ phonemic category – 'pet', 'peck', 'set', 'pen', 'celery' were better identified by the British speakers with at least 10% accuracy (mean 21%). On the other hand, the phonemic category /æ/ was on the whole better identified by the Malaysian participants; with 5 words from the category – 'had', 'sat' , 'man', 'pan' and 'sand', having higher identification accuracy of at least 9% (mean 16%).

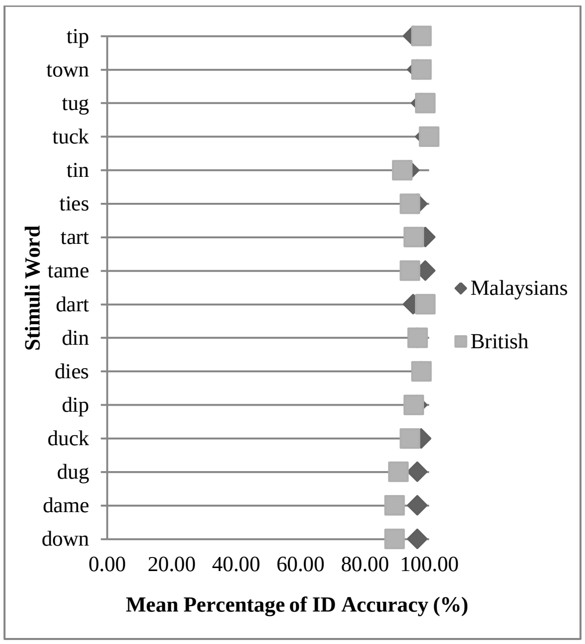


Figure 4.12 Mean percentage of correct identification averaged across Malaysian speakers and British speakers respectively for each Malaysian English stimuli word.

Note: The stimuli words were arranged according to their phonemic category and sequenced based on the difference magnitude between the speaker groups. The difference magnitude was calculated by subtracting the Malaysian speakers' identification accuracy from the British speakers'. Starting from the top of the list is the word with the greatest positive difference magnitude (Malaysians < British) and the magnitude reduces across the list to the last word of the phonemic category that either has a zero difference magnitude (Malaysians = British) or the greatest negative magnitude (Malaysians > British).

*top y-axis: words with /t/ initial phoneme; bottom y-axis: words with /d/ initial phoneme

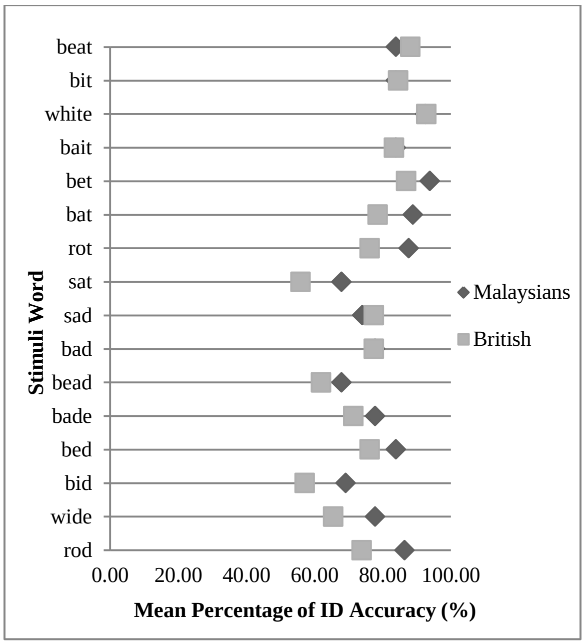


Figure 4.13 Mean percentage of correct identification averaged across Malaysian speakers and British speakers respectively for each Malaysian English stimuli word.

Note: The stimuli words were arranged according to their phonemic category and sequenced based on the difference magnitude between the rater groups (see Figure 4.8 for more details).

*top y-axis: words with /t/ final phoneme; bottom y-axis: words with /d/ final phoneme

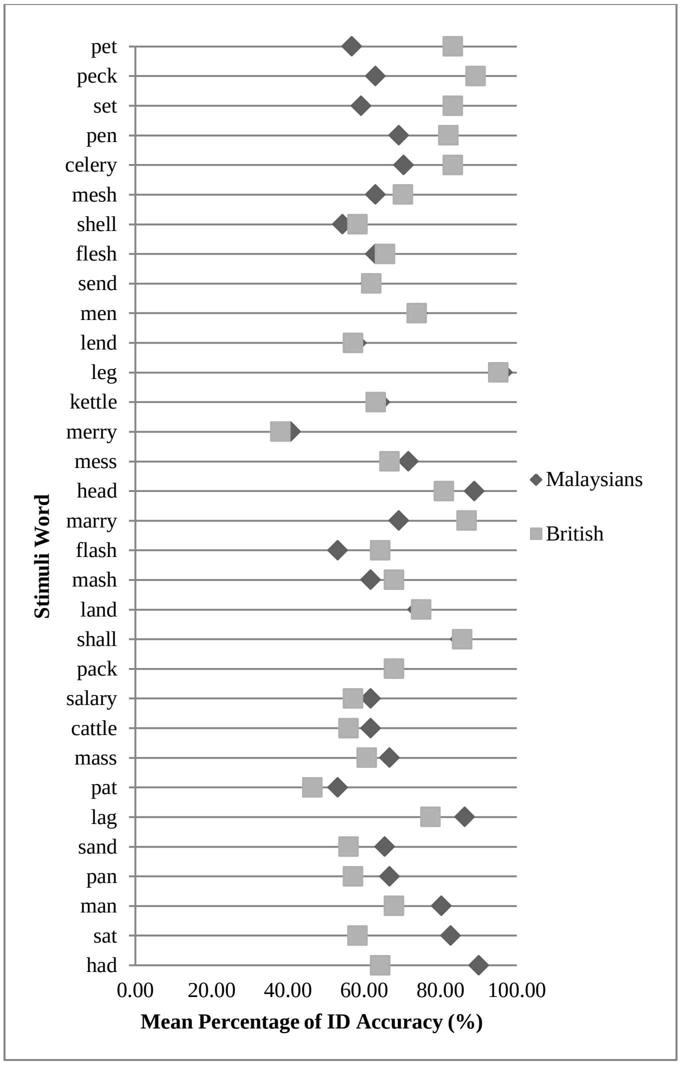


Figure 4.14 Mean percentage of correct identification averaged across Malaysian speakers and British speakers respectively for each Malaysian English stimuli word.

Note: The stimuli words were arranged according to their phonemic category and sequenced based on the difference magnitude between the rater groups (see Figure 4.8 for more details).

*top y-axis: words with /ɛ/ phoneme; bottom y-axis: words with /æ/ phoneme

The phonetic perception of the L1 ME speakers and the native British English speakers was examined by comparing their identification accuracy on the Malaysian English speech stimuli (Figure 4.15). Both speaker groups identified words from the /t/-/d/ initial categories at near ceiling with mean percentage of correct identification above 94%. Identification accuracy of the /t/-/d/ final phonemes (*m* = 78%) was considerably lower than the /t/-/d/ initial phonemes; while the /ɛ/-/æ/ contrasting phonemes (*m* = 68%) appeared to be the most difficult pair to be identified. Referring to Figure 4.15, Malaysians showed the greatest difficulty in identifying the /ɛ/ productions of the L1 Mandarin speakers while the British showed the greatest difficulty in identifying the opponent phoneme /æ/ from the phonemic pair.

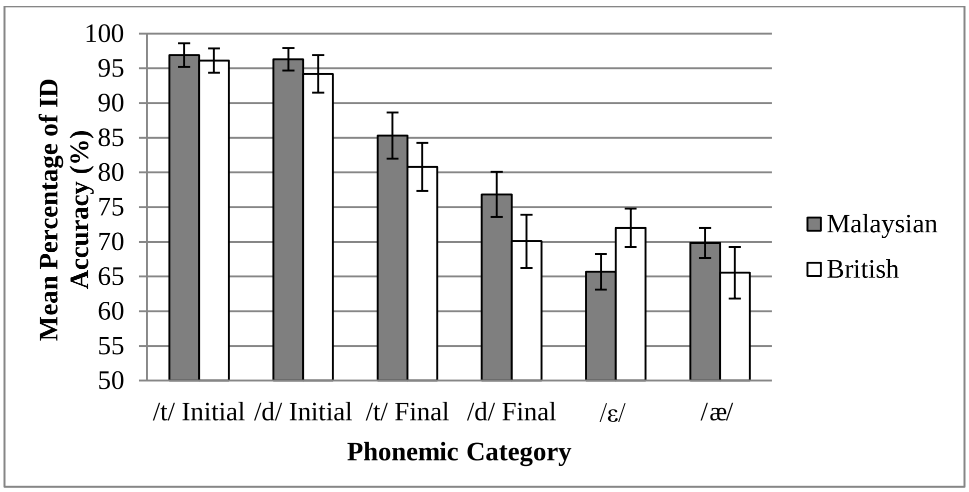


Figure 4.15 Mean percentage of correct identification for each phonemic category of Malaysian English speech stimuli, identified by the Malaysian speakers and British speakers respectively (with error bars as the standard error).

The mean percentage of correct identification on the Malaysian English speech stimuli was subjected to a 2 X 6 mixed-design ANOVA; with speakers (Malaysian vs. British) as the between-subject factor and phonemic category (/t/ initial vs. /d/ initial vs. /t/ final vs. /d/ final vs. /ɛ/ vs. /æ/) as the within-subject factor. A main effect of phonemic category was found, *F*(5,256)

= 44.68, $p < .001$. The speakers identified the /t/ initial and /d/ initial phonemes similarly and at a higher accuracy than any other phonemic categories, $p \leq .001$. There was no difference in identification accuracy between the /t/ final and /d/ final phonemes ($p = .13$) but the /t/ final phoneme was better identified than both /ε/ and /æ/ phonemes (ps < .001). The identification accuracy did not differ among the /d/ final, /ε/ and /æ/ phonemes, ps = 1. Identification accuracy of the Malaysian and British speakers did not differ across the phonemic categories, Fs < 1.42, ps > .22.

4.3.3 Production Intelligibility of Native British English Speakers and L1 Mandarin Speakers based on the Degree of Familiarity

In the previous sections, the phonetic perception of British and Malaysian speakers was examined independently on the British English and Malaysian English speech stimuli. In this section, we examined whether speakers could identify speech words better when the speech stimuli was spoken by speakers of the same nationality/language background. The data from the /t/ initial and /d/ initial phonemic categories were excluded from any further analysis due to the ceiling identification performance of the speakers on the British English speech stimuli (> 97%) and also the Malaysian English speech stimuli (> 94%). Identification accuracy was compared using a 2 X 2 X 4 mixed-design ANOVA; with speakers (Malaysian vs. British) as the between-subject factor, English variety (Malaysian English vs. British English) and phonemic category (/t/ final vs. /d/ final vs. /ε/ vs. /æ/) as the within-subject factors. The results are presented in section 4.3.3.1 and section 4.3.3.2.

4.3.3.1 Production Intelligibility based on the Variety of English Spoken

From the ANOVA analysis, the British speakers had a higher identification accuracy when compared to the Malaysian speakers, $F(1,53) = 4.15$, $p < .05$, and the British English speech stimuli were better identified than the Malaysian English speech stimuli, $F(1,53) = 122.14$, $p < .001$ (see Figure 4.16). The analysis revealed an interaction between the speakers and English variety, $F(1,53) = 10.37$, $p < .01$. Post-hoc paired-sample t-tests compared the identification accuracy between the British English and Malaysian English speech stimuli within each speaker group showed that both rater groups identified the British English speech stimuli better than the Malaysian English speech stimuli, *t*s < 4.86, *p*s < .001. The post-hoc independent sample t-tests comparing the identification accuracy between the speaker groups for each English variety showed that the two speaker groups had similar identification accuracy for the Malaysian English speech stimuli, $t(54) = .27$, $p = .79$; however, the British speakers identified the British English speech stimuli better than the Malaysian speakers, $t(54) = 4.37$, $p < .001$.

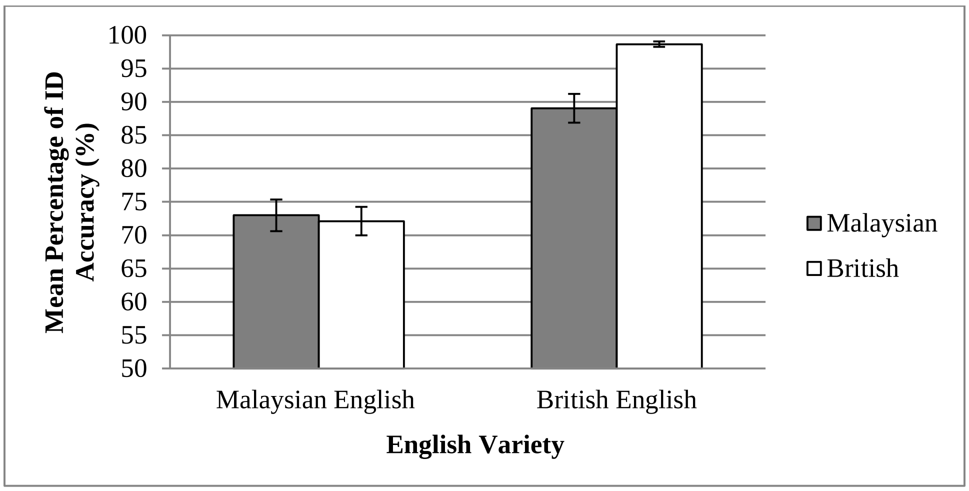


Figure 4.16 Mean percentage of correct identification on the Malaysian English and British English speech stimuli, identified by the L1 Malaysian English and native British English speaking speakers (with error bars as the standard error).

4.3.3.2 Production Intelligibility based on the Phonemic Category

A main effect of phonemic category showed that the identification accuracy of the/æ/ phoneme (81%) was significantly lower than the /t/ final phoneme (89%), while the identification accuracy of the /ɛ/ phoneme (79%) was significantly lower than /t/ and /d/ final (85%) phonemes, $F(3,159) = 13.68$, $p < .001$. There was no difference in identification accuracy between the /t/-/d/ final phonemic pair and between the /ɛ/-/æ/ phonemic pair, *p*s > .17. There was an interaction between the phonemic category and speakers factors, $F(3,159) = 5.36$, $p < .01$, suggesting that the speaker groups had different difficulty levels identifying the phonemic categories. As a result, two one-way repeated-measures ANOVAs were conducted to compare identification accuracy across the phonemic categories for each speaker group respectively. A main effect of phonemic category was found from both the Malaysian speakers and British speakers, $F(3,81) = 8.46$, $p < .001$ and $F(3,81) = 5.07$, $p < .01$ respectively. Results showed that the Malaysian speakers had the greatest difficulty identifying the /ɛ/ phoneme compared to all other categories, *p*s < .05. There was a suggestive trend that the identification accuracy on the /æ/ phoneme was lower than the /t/ final phoneme, $p = .07$. For the British speakers, the /t/ final phoneme was better identified than the /ɛ/ ($p < .05$) and /æ/ ($p < .01$) phonemes. No other significant difference was found between other phonemic categories. Four post-hoc independent samples t-tests were conducted to compare identification accuracy on each phonemic category between the Malaysian and British speakers. The results only revealed a significantly higher identification accuracy of the British speakers compared to the Malaysian speakers for the /ɛ/ phoneme, $t(54) = 5.76$, $p < .001$. No other comparisons were significant, *t*s < 1, *p*s > .05.

There was also an interaction between the English variety and the phonemic category, $F(3,159) = 7.15$, $p < .001$ (see Figure 4.9 and 4.15). Production accuracy across the phonemic categories was compared separately for the Malaysian English and British English speech stimuli in the following analyses. Post-hoc one-way repeated-measures ANOVA indicated that the /t/ final phoneme produced by the L1 Mandarin speakers had the highest identification accuracy compared to other phonemic categories, $F(3,162) = 10.94$, $p <.001$. No other differences were found between the other phonemic categories from the Malaysian English speech stimuli, $p > .05$. On the other hand, the /ɛ/ phoneme produced by the native British English speaker had the lowest identification accuracy compared to other phonemic categories, $F(3,165) = 5.74$, $p = .001$. This finding could be partly attributed to the weakness of Malaysian speakers in identifying the /ɛ/

phoneme (81%); while the identification on the other phonemic categories from the British English speech stimuli achieved near ceiling performance with all means falling above 90%. Post-hoc paired-sample t-tests conducted for the interaction indicated that the British English speech stimuli were better identified by the speakers for all the phonemic categories, ts < 4.19, ps < .001.

There was no significant three-way interaction revealed by the ANOVA analysis, $F(3,159) = .13$, $p = .95$.

4.4 Discussion

4.4.1 Phonological Representation of L1 Mandarin Speakers

The L1 Mandarin speakers seem to have little difficulty in identifying the British English speaker's productions of the voiceless plosive /t/ (mean accuracy 98%) - voiced plosive /d/ contrast at word initial position (mean accuracy 98%) and also the voiceless plosive /t/ at word final position (mean accuracy 91%). Their lowest identification accuracy for the voiced plosive /d/ at word final position (68%) suggests speakers tend to identify the end of the phonemic words heard as the voiceless plosive /t/ instead of the voiced plosive /d/. This tendency might be due to the following three reasons. Firstly, the voicing component of the /d/ final phonemic words which the L1 Mandarin speakers used to identify the /d/ phoneme might not be available in some of the British English speech stimuli. It is especially common for British English speakers to devoice the /d/ phoneme at word final position in natural speech (Flege, 1989). Secondly, when the voicing component is no longer available to assist the identification decision, the native English speakers learnt to use the vowel acoustic properties to discriminate the voiceless-voiced plosives. However, the non-native English speakers were less skillful adopting this strategy (Flege, 1989). The last reason could be attributed to the L1 Mandarin speakers' over-reliance on release burst cues when they identify the voiceless plosive /t/. As a result of the over-reliance, the probability becomes higher for the speakers to choose the voiceless plosive /t/ whenever the release burst was heard.

Another perceptual pattern observed from the L1 Mandarin speakers was their assimilation of the British English /ɛ/ phoneme to the /æ/ phonemic category. The two similar phonemic sounds /ɛ/ and /æ/ were matched to the grapheme <a> more frequently (mean identification accuracy of /ɛ/ - 71%; mean identification accuracy of /æ/ - 82%). Thus, the more operative vowel representation in the L1 Mandarin speakers is more likely to be the low front vowel /æ/. The current finding is consistent with Yap et al.'s (2010) study, in which the L1 Malay speakers were found to perceptually assimilate the /ɛ/ and /æ/ phonemes into the /æ/ category. Although participants from Yap et al. (2010) and this study speak different L1s, the phonology of their L1s (i.e. Malay and Mandarin) have relatively shallow orthography for vowels when compared to the English language; the participants also shared similar linguistic environments. Moreover, the Malay language learnt by the L1 Mandarin speakers as another L2 may also play a role in shaping their phonetic perception. Therefore, it seems plausible for the two language groups to share similar perceptual patterns. The L1 Mandarin speakers showed a certain degree of difficulty in discriminating and identifying the /t/-/d/ final and /ɛ/-/æ/ phonemic words; where the phonological representation of one member of the pair being more prominent than the other (i.e. /t/ final and /æ/ over their opponents).

4.4.2 Phonetic Perception Difference between Native British English Speakers and L1 Malaysian English Speakers

Although the L1 ME speakers and the native British English speakers speak English as their L1, the different English varieties they are exposed to may have shaped their phonetic perception system differently. When the speakers were asked to identify the British English speech stimuli, the native British English speakers had no difficulty and showed ceiling perceptual performance for every phonemic category (m > 98%). On the other hand, the L1 ME speakers showed greater identification variation across the phonemic words, especially for the /ɛ/-/æ/ phonemic contrast (accuracy ranged from 54% to 96%; *m* 87%; SD 9.34). They had the greatest difficulty in identifying the /ɛ/ phoneme. The /ɛ/ phoneme was matched to the grapheme <a> more frequently (*m* 18%) than the /æ/ phoneme was matched to the grapheme <e> (*m* 9%); a similar perceptual pattern observed in the L1 Malay Malaysian speakers (Yap et al., 2010) and also in the L1 Mandarin speakers from the present study. The native British English speakers and the L1 ME speakers identified the British English /t/-/d/ initial phonemic contrast at a similar level but the native British English speakers identified the rest of the phonemic contrasts better than the L1 ME speakers (i.e. /t/-/d/ final & /ɛ/-/æ/).

4.4.3 Production Ambiguity of L1 Mandarin Speakers

The L1 ME speakers did not show any advantage in identifying the Malaysian English speech stimuli, even though they were more familiar with the pronunciation. The fact that L1 ME speakers had difficulty in perceiving their own L1 English variety more accurately might be attributed to the general production ambiguity of the L1 Mandarin speakers. In accordance with one of our earlier predictions, the judges had no difficulty identifying the speakers' production of /t/-/d/ initial phonemic words (near to ceiling identification performance, $m > 94\%$). Generally, the L1 Mandarin speakers failed to produce phonemes accurately when they were ambiguously represented in their English phonetic system. The ambiguous phonemic categories showed by the present study included the /d/ final, /ɛ/ and /æ/ phonemes.

The L1 Mandarin speakers had the greatest difficulty in producing the distinct /ɛ/-/æ/ phonemic contrast. The production of the phonemic pair did not resemble one particular phoneme more as the identification accuracy for the two phonemic categories was similar (*m* 69% for /ɛ/; *m* 68% for /æ/). The ambiguous phonetic representation of the speakers resulted in ambiguous production of the phonemes. The findings on the /t/-/d/ final phonemic contrast production supported the last hypothesis/prediction. The L1 ME speakers and native British English speakers were more inclined to recognize the L1 Mandarin speakers' production of the /d/ final phoneme as its counterpart – the voiceless plosive /t/. It is still unknown whether the assimilation was due to the missing closure voicing (through devoicing) or the lack of vowel distinction (e.g. spectral quality, vowel duration); the vowel acoustical properties were essential cues used by the native English speakers to assist the voiceless-voiced plosive contrast discrimination (see Flege, 1989). Acoustical analysis in the future work would clarify the production difference between the L1 Mandarin speakers and the native British English speakers for the /t/-/d/ final phonemic contrast.

Besides poorer intelligibility of the speech production, inconsistency in production intelligibility was also observed for these difficult phonemes. The production inconsistency was particularly obvious for the /ɛ/-/æ/ phonemic words, in which some words turned out to be more intelligible than the other words of the same phonemic category. Due to the ambiguous English vowels representation, speakers' pronunciation of the vowels might be more influenced by their personal experience and exposure with the words and also by their L1 phonetic representation of the vowel phonemes. This in turn may introduce greater variation in non-native speakers' pronunciation.

4.4.4 Implications of the Present Finding

Generally, the native British English speakers were better in the phonemic words identification, probably due to their more accurate perception of the British English speech stimuli. The British English speech stimuli were also more intelligible than the Malaysian English speech stimuli, suggesting some production flaws of the L1 Mandarin English speakers. The production of the mid-front and low-front vowels - /ɛ/ and /æ/ had the lowest intelligibility among the other phonemic categories; while the production of the voiced plosive /d/ was always perceived as the voiceless plosive /t/. In the following chapter, we examined whether SED phonetic perceptual training would improve the production intelligibility of the L1 Mandarin English speakers.

The present chapter provided some basic background information about the phonetic perception patterns of the raters (L1 ME speakers and native British English speakers). The results from this chapter suggest that both the native British English speakers and L1 ME speakers have the necessary skill set/perceptual skills to evaluate the speech production of the L1 Mandarin speakers in Chapter 5 as they had minimal problem in identifying the British English speech production. Any improvements in intelligibility in the speakers' production should be able to be detected by the raters. Nevertheless, it is important to note that the identification accuracy of the L1 ME speakers on the /t/-/d/ final and the /ɛ/-/æ/ phonemic contrasts did not match up with the ceiling performance of the native British English speakers. The mid-front vowel /ɛ/ was exceptionally difficult for the L1 ME speakers. Hence, caution might be needed when interpreting their evaluation results in Chapter 5.

Moreover, the native British English speakers did not have exceptional difficulty in perceiving the Malaysian English speech production when compared to the L1 ME speakers. It is thus, reasonable to also accept native British English speakers as the raters.

Chapter 5: Effect of SED Training on Speech Production Intelligibility

5.1 Introduction

5.1.1 Effect of Phonetic Training on Speech Production

One of the goals of perceptual training with L2 speakers is to provide acoustic targets for speech production that might serve to reduce foreign accent when speaking in their L2. Evidence of changes in speech production accuracy as a result of perceptual training would be a powerful indicator of the efficacy of a perceptual training program. Numerous studies have successfully shown that perceptual phonetic training can improve speakers' production accuracy (Bradlow et al., 1999; Bradlow et al., 1997; Lambacher, Martens, Kakehi, Marasinghe, & Molholt, 2005) and reduce their foreign accent (Hanulikova, Dediu, Fang, Bašnaková, & Huettig, 2012; Herd, 2011), as evaluated by native speakers. In the study by Bradlow and colleagues (1997), participants' post-test production of the two phonemes were rated "better" and were identified more accurately by the native American English as a result of the perceptual training. As the participants from Bradlow et al. (1997) received no explicit training in the speech production domain, the production improvement observed was attributed to the transfer of knowledge learnt from the perceptual training.

In another related study by Herd (2011) who compared the effectiveness between perception and production training and their cross-modal transferability (i.e. perceptual learning transferred to production performance or production learning transferred to perception performance) trained native American English speakers to distinguish the Spanish /ɾ/-/ r/-/d/ phonemic contrasts (as in the Spanish words *coro-corro-codo* respectively) . The perception training involved a high variability forced-choice identification task whereas the production training involved participants inspecting own production's waveform and spectrogram on Praat, with native speakers' production spectrogram as reference.

When participants had perception training, perceptual learning transferred to both perception and production modalities whereby production performance for all contrasts either improved significantly or exhibited no decline. However, when participants had production training, their perception accuracy of the Spanish /ɾ/-/ d/ contrast (as in *coro-codo*) declined significantly. The Spanish /ɾ/-/ d/ phonetic sounds are two allophonic variants of the English /d/ phoneme and the English speakers were believed to assimilate them into a single category – the English phoneme /d/ (Best's PAM; see section 1.3.2.2). According to Best's PAM (2001), the single category assimilated phonetic sounds are the most difficult phonetic pair to be discriminated. Therefore, the /ɾ/-/ d/ contrast could show the strongest resistance for any learning changes. In conclusion, Herd concluded that perception training was more beneficial than the production training.

5.1.2 Relationships between Perception and Production

Most linguistic theorists have no serious disagreements about the fundamental duality of linguistic processes that suggest the close relationship between speech perception and production. For example, articulatory strategies adopted by speakers were often reflected in their perceptual strategies (Bell-Berti, Raphael, Pisoni, & Sawusch, 1979); the perception phonological performances (measured in discrimination and rhyme judgments) were also found to correlate with the production phonological performance (measured in picture naming; see Martin & Saffran, 2002).

Speech perception and production are clearly linked, in terms of behavioral performance and also functional anatomy. A wealth of neuroimaging data has shown the connection. For instance, studies have shown the involvement of left supratemporal plane in speech perception, as well as in task related to the production mechanism (e.g. lip reading and subvocal productions of words); the involvement of inferior frontal lobe and the Broca's area is also observed in phonological encoding during speech perception and in generation of motor/ articulator output (see Levelt, 2001 for review). The overlap of anatomical activation observed suggests that the speech perception and production processes do not work as one but share modality.

5.1.2.1 The General Auditory Approach

The auditory approach of speech perception assumes speakers perceive speech sounds from acoustic signals, similar to the mechanisms of general audition and perceptual learning that deal with other environmental sounds (Diehl, Lotto, & Holt, 2004). This approach proposes that the perception and production performance are interrelated, and that both performances are mediated by each other. According to the approach, qualitative contrast in phonetic sounds has to be acquired auditorily before qualitatively different speech sounds can be produced (production based on perception). To promote the perceptual phonemical distinctiveness, the perceptual distance between phonemes can be seen maximized near the phoneme boundaries to improve the intelligibility of utterance (for more details see section 1.3.2.3 or Kuhl, 1991). Inversely, the second claim describes how the acoustic correlates or regularities from speakers' own overt speech production assists the perception mechanism in forming phonemic boundaries, and also in judging phonemic content of speech signals (perception based on production).

The theory of production by Levelt (1989) suggested a dual-feedback loop in the perceptual processing system that serves to provide information for self-monitoring speech production (see Figure 6.1). In the external loop, the speakers' own overt speech is used as the object of self-monitoring to guide or repair the articulatory performance; while in the internal loop, the self-generated phonological representations act as the reference object. In this model, the perception ability and sensitivity in both feedback loops plays an important part in determining the monitoring quality on one's speech articulation and production.

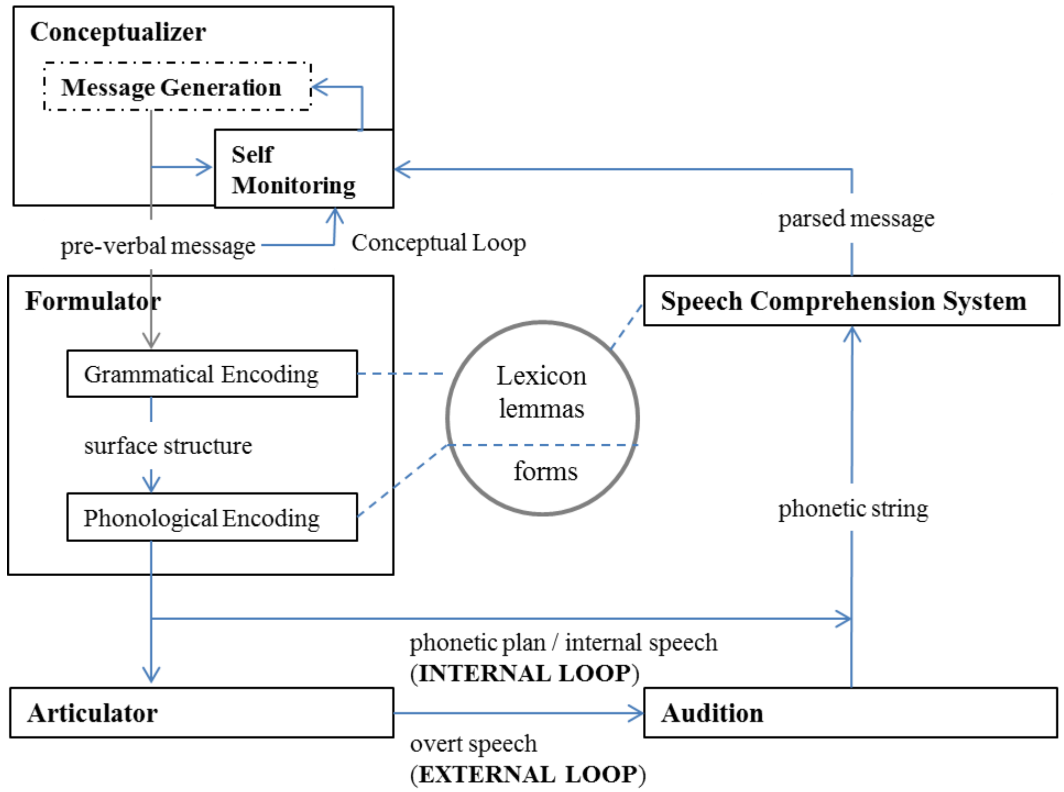


Figure 5.1 Levelt's Model of Speech Production (adapted from Levelt, 1989; Levelt, Roelofs & Meyer 1999)

5.1.2.2 The Motor Theory of Speech Perception

The motor theory assumes that the "objects" listeners perceived from speech are articulatory events rather than the acoustic or auditory properties of the speech sounds (Liberman & Mattingly, 1985). According to Liberman, the acoustic patterns of speech are too varied to be the objects of speech perception; while the fixed neuro-motor commands (e.g. tongue backing, lip rounding etc.) are relatively more invariant. Hence, it is more likely that the speakers are perceiving speech in terms of articulatory gestures. According to this theory, the ability to detect the intended gestures of speakers is an innately built, speech-specific system that is unique to human beings.

As it is the movement of the articulators that the listeners are perceiving (including one's own articulation movement), the regularities of the speech production should be highly correlated with listeners' perceptual judgment. The motor theory by Liberman hereby posits an integrated perception and production systems and claims that the link is innately specified (Liberman & Mattingly, 1985).

To summarize, the speech production models proposed a close relationship between the speech perception and production modalities. The performance is often interdependent and studies had shown behavioral changes in one modality also led to changes in another (e.g. Bradlow et al., 1997; Nobre-Oliveira, 2008).

5.1.3 General Methods for Word Production Evaluation

In order to examine the effects of phonetic training, previous studies had examined three aspects of speech production: the acoustical properties (e.g. Nobre-Oliveira, 2008; Aliaga-Garcia & Mora, 2009), the production quality and the production accuracy/ intelligibility (e.g. Bradlow et al., 1999). The term “production intelligibility” is preferred over the term “production accuracy” because British English and Malaysian English involved in this study are acknowledged as two English varieties with distinct phonological systems; British English is not viewed as the higher “quality” form of English. Therefore, the production improvement was judged based on the “intelligibility” of the productions to a target speaker group (e.g. native British English speakers) in the present study.

The acoustical analysis on the speech production can be considered the most informative evaluation method as it reveals specific acoustical changes after training. However, the analysis is substantially time consuming and it does not take into account whether the productions before and after training are perceptually different to the listeners.

Generally, three varieties of evaluation task are used within the literature. Production quality is usually judged based on the strength of native accent perceived (Hanulikova et al., 2012; Herd, 2011) or on the degree of “goodness” (Bradlow et al., 1999, 1997) of the pronunciation. To evaluate production quality, a rating scale is usually implemented. For instance, raters in the study by Hazan et al. (2005) rated the degree of goodness of the talkers' pronunciation on a 7-point scale; in which 1 was chosen to indicate bad phonetic sound realization in the word (i.e. /l/ and /r/), and 7 was chosen to indicate excellent realization.

The production intelligibility can be evaluated using either a minimal pair identification task (e.g. Lambacher et al., 2005) or an open-set transcription task (e.g. Bradlow et al., 1999). Using these tasks, the probability that the production can be recognized as the intended words is measured in multiple sessions to examine improvement in production. The forced-choice identification task is similar to the task used in the SED pre-test and post-test. Raters identify the auditory word by choosing the orthography of the intended word from a minimal pair presented on a monitor. This method is preferred when the intention of the evaluation focuses on the segment specific improvement (e.g. the vowels of the minimal pairs in Lambacher et al., 2005).

Another task used to evaluate production intelligibility is the open-set transcription task. Unlike the identification task, the open-set transcription task that stresses on the overall word intelligibility allows inspection on a wider scale (Bradlow et al., 1999). It is an objective measure because raters have to write down the intended words based only on the acoustical signals perceived, without any visual cues or hints. Moreover, the word recognition process involved in the transcription task is what people usually experience in real life discourse context. Nevertheless, the task is the least popular method used among the literature as the trial by trial recognition procedure requires an enormous amount of time and effort.

5.1.4 Rationale of Present Chapter

The present chapter aimed to address two main questions raised from Chapter 3. The first question centers on whether the perceptual improvement observed from previous SED training studies is due to the exposure effect rather than the training effect of the SED paradigm. Secondly and most importantly, would the SED training improve speakers' speech production as demonstrated by the previous literature?

In Experiment 6, a new group of L1 Mandarin speakers participated in the SED training and their production of the stimuli words was recorded. In order to rule out the suspected exposure effect, a control group was also included to measure their perceptual performance across time of tests without being given any SED training. We hypothesized that the perception and production performance of the control group would remain consistent across pre- and post-tests even though they were also exposed to the British English speech materials.

Experiment 7 was conducted to examine the effect of SED training on speech production recordings acquired from Experiment 6. Similar to previous studies such as Bradlow et al. (1997; 1999) and Hazan et al. (2005), the production improvement of the speakers will be measured in two ways – the production quality and intelligibility. Due to the large amount of speech stimuli acquired from Experiment 6 (a total of 7872 speech stimuli), the open-set transcription task was deemed unsuitable for the present study. Moreover, the present chapter was more interested in the segment specific improvement of the productions (i.e. the phonemic pronunciation: /t/-/d/ and /ɛ/-/æ/ of the minimal pairs) and not the pronunciation of the whole word. Therefore, the quality rating and the 2 alternative forced-choice identification tasks were implemented in Experiment 7 to evaluate speaker's speech production. Similar to the study by Hanulikova et al. (2012), productions from each speaker were evaluated by four raters.

5.2 Experiment 6 (SED Training and Speech Production Recording)

5.2.2 Training and Recording Methods

5.2.2.1 Design and Materials

The word stimuli involved in the word naming task consisted of the same sixty-four English words that were used in the SED training as outlined in chapter 2 (see Appendix 1).

Participants' speech production was recorded in a quiet chamber using a microphone (M7) and a 24-bit Audiobox USB interface (Presonus AudioBox) connected to a HP SR1913WM computer. Speech was recorded at 44.1.hKz (sampling depth: 32 bit) using Audacity(R) (version 2.0.5).

5.2.2.2 Participants

43 participants who speak Mandarin as their first language were recruited from the University of Nottingham, Malaysia campus and were paid for their participation. Participants were recruited into three groups: the adaptive staircase group (n 15; age range: 18-23, mean age 20.80; 9 females), the constant group (n 14; age range: 18-23, mean age 20.29; 9 females) and the control group (n 14; age range: 18-20; mean age 18.14; 9 females). All participants reported to have normal eyesight or corrected to normal vision, and had no history of any hearing, speech or reading disorders.

Participants completed a language questionnaire, providing information about their language background and formal education (see Appendix 4 for an overview of their language backgrounds across the different groups). They also completed the Lextale test, a standardized test of English vocabulary knowledge to provide a general measure of their English proficiency (Lemhöfer & Broersma, 2012). All participants reported to speak at least three languages; with one language being English (other languages included Malay, Cantonese and Hokkien). Their self-rated English and Mandarin proficiency and age of English acquisition (AoA) were compared across the three groups, together with the three groups of participants from Chapter 3 (Constant Low Babble Group and Constant High Babble Group from Experiment 2 and Adaptive Staircase Group from Experiment 3) to ensure that the participants came from a similar language background and that the training results were comparable. The AoA, self-rated English proficiency and self-rated Mandarin proficiency did not differ across the six groups of participants, *F*s < 1.32, *p*s > .05. The Lextale score also did not differ across the three groups of participants from the present experiment, $F(2,40) = .69$, $p = .51$.

5.2.2.3 Procedure

5.2.2.3.1 SED Training

The training groups completed 5 training sessions across 5 days. The adaptive staircase group received SED training with adaptive staircase procedure (see Experiment 2 in Chapter 3) and the constant group received SED training with constant background babble (see Experiment 1 in Chapter 3 for the detailed procedure). The pre-test, post-test and 1 month post-test of SED training involved a 2 Alternative Forced-Choice (2 AFC) identification task while the training sessions involved a Same-Different discrimination task.

The control group was tested using the 2 AFC identification task at the same time intervals as the training groups, without going through any phonetic training. The perceptual tests took place at three different time points: pre-test, post-test (3 days after the initial testing session) and 1 month post-test. Participants from all groups took part in the 1 month follow-up test which was conducted at least 30 days after each participant's post-test.

5.2.2.3.2 Word Naming Task (Production Recording)

The stimuli words were presented randomly at the center of the computer screen using PsychoPy (version 1.80.01; Peirce, 2007). Each stimulus word lasted 5 seconds on the computer screen before the next stimuli word was presented. Participants read out the 64 English words and their productions were recorded by a microphone that was placed approximately 15 cm before their lips. The entire word naming task lasted about 5 minutes.

Participants were instructed to read out the words loud and clear at their normal speaking pace. Before the word naming task, there were five practice trials whereby participants had to read out five different English words in the presence of the experimenter. During the practice trials, participants' production volume and sitting position in relation to the location of the microphone was advised by the experimenter. The recordings from the practice trials were played to some participants if there was a problem with their recording (e.g. excessive body movement, inconsistencies in the production volume). After the experimenter left the room, participants initiated the word naming task by pressing the space bar.

All participants did the word naming task and were recorded three times in total: after pre-test, before post-test and before 1 month post-test of SED training. In order to measure participants' initial state of perceptual sensitivity before the phonological representations of the speech sounds were activated (through explicit naming of the words), participants did the word naming task after the ID task in pre-test. On the other hand, the participants completed the word naming task prior to the ID task in post-test and 1 month post-test so that any possible changes in production could be recorded before being affected by the British English speech in the perceptual tests.

5.2.3 Training Behavioral Results

The mean percentage of correct identification was calculated for each of the three phonemic contrasts in the pre- and post-test of the two groups of participants (see Figure 5.2 and Figure 5.3). Data from the /t/-/d/ phonemic contrast at word initial position was excluded from further analyses because of ceiling performance in each group (mean above 95%).

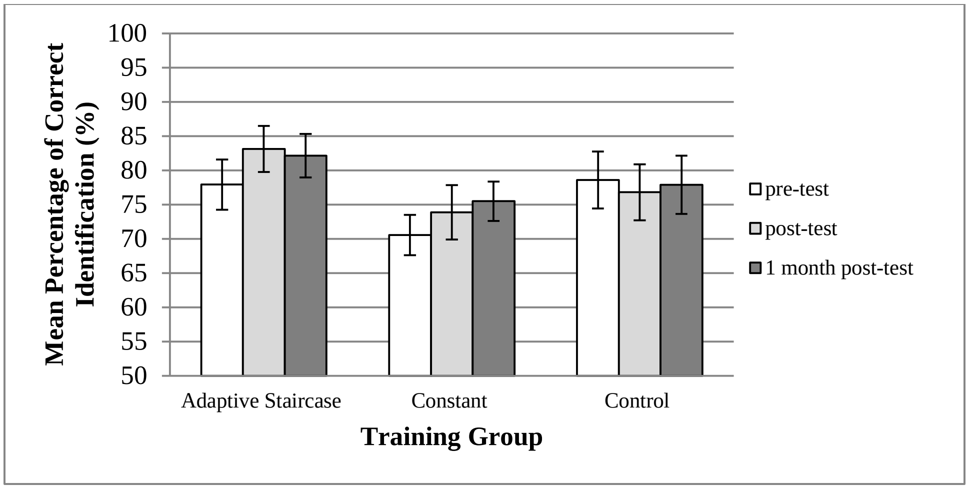


Figure 5.2 Mean percentage of correct identification on the /t/-/d/ final phonemic contrast made by each training group in the respective pre-test, post-test and 1 month post-test (standard errors as the error bars).

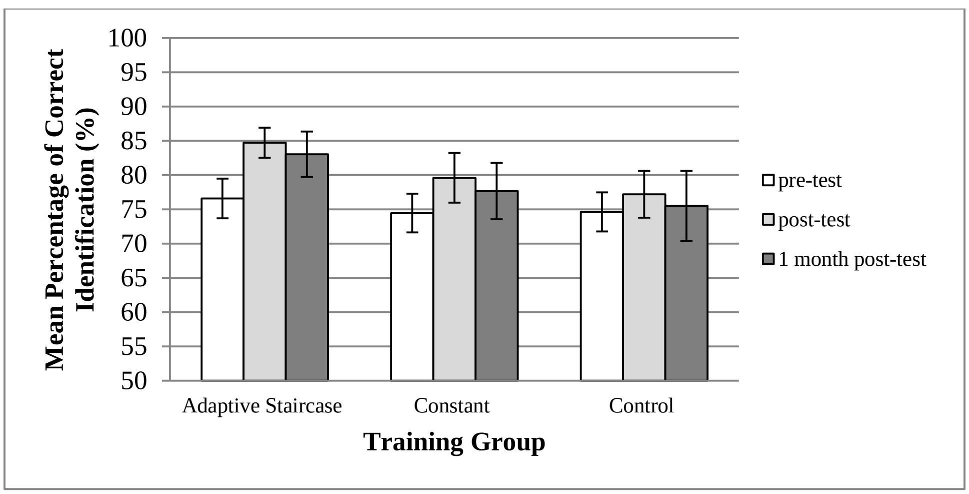


Figure 5.3 Mean percentage of correct identification on the /ɛ/-/æ/ phonemic contrast made by each training group in the respective pre-test, post-test and 1 month post-test (standard errors as the error bars).

5.2.3.1 Training Effect of SED

The training Effect of SED was examined in a 3 X 2 X 2 mixed-design ANOVA with the training group (adaptive staircase vs. constant vs. control) as the between-subject factor, the phonemic contrast (/t/-/d/ final vs. /ε/-/æ/) and the time of test (pre- vs. post-test) as the within-subject factors. Overall, participants identified the phonemic words better in post-test compared to pre-test, $F(1,40) = 14.89$, $p < .001$. However, their performance across time of test differed according to their training group, $F(2,40) = 3.51$, $p = .04$. Post-hoc paired-sample t-tests conducted showed that the adaptive staircase group (pre-test 78%, post-test 84%) and the constant group (pre-test 72%, post-test 77%) scored significantly higher in post-test identification accuracy compared to pre-test, $t(14) = 3.71$, $p < .01$ and $t(13) = 2.46$, $p = .03$ respectively. In contrast, the control group identified the phonemic words similarly in both pre-test and post-test (pre-test 77%, post-test 77%), $t(13) = .87$, $p = .40$. In addition, two post-hoc one-way ANOVAs were conducted to compare the identification accuracy across the training groups in pre-test and post-test separately. The analyses did not show any significant effect, *F*s < 2.21, *p*s > .05. No other effect or interaction was found, *F*s < 1.18, *p*s > .05.

5.2.3.1.1 Adaptive Staircase SED vs. Constant SED

The training effect of Adaptive Staircase SED was compared to the Constant SED to see whether the present trainings still show the advantage of adaptive staircase procedure as outlined in Chapter 3. The training effects were compared in a 2 X 2 X 2 mixed-design ANOVA with the training group (adaptive staircase vs. constant) as the between-subject factor, the phonemic contrast (/t/-/d/ final vs. /ε/-/æ/) and the time of test (pre- vs. post-test) as the within-subject factors. There was a main effect of time of test, $F(1,27) = 19.06$, $p < .001$ whereby the post-test identification accuracy was significantly higher than the pre-test. No other effect or interaction was found, *F*s < 2.72, *p*s > .05.

5.2.3.1.2 Retention of Training Effect

To examine whether training effect of SED was retained after a one month delay, and whether the control group would still perform similarly, a 2 X 2 X 2 mixed-design ANOVA was conducted with the training group (adaptive staircase vs. constant vs. control) as the between-subject factor, the phonemic contrast (/t/-/d/ final vs. /ε/-/æ/) and the time of test (post-test vs. 1 month post-test) as the within-subject factors. No significant effect or interaction was revealed, Fs < 1.71, ps > .05.

5.2.3.2 Generalization to Novel Words, Novel Speakers and Untrained Phonemic Contrast /t/-/d/ Final

In order to examine whether training effect generalized equally to the novel or untrained /ε/-/æ/ phonemic words, a 2 x 2 within-subject ANOVA was conducted for the adaptive staircase group and the constant group separately; with the stimuli set (trained vs. untrained /ε/-/æ/) and the time of test (pre- vs. post-test) as the within-subject factors. The identification accuracy of the adaptive staircase group in post-test was higher than the pre-test, $F(1,14) = 12.65$, $p < .01$. Surprisingly, main effect of time of test was not found for the constant group, $F(1,13) = 2.97$, $p = .11$, although the post-test identification accuracy was higher than the pre-test (see Figure 5.2). Importantly, there was no interaction between the two factors for both adaptive staircase and constant groups, Fs < 1, suggesting that the training effect generalized to the novel stimuli was similar to the trained stimuli. No other effect or interaction was found from both analyses, Fs < 2.97, ps > .05.

Two more 2 X 2 within-subject ANOVAs were conducted for the adaptive staircase and constant groups separately to examine how well training on /ε/-/æ/ phonemic words affected the identification performance on the untrained phonemic contrast, /t/-/d/ final, with untrained stimuli (untrained /ε/-/æ/ vs. /t/-/d/ final) and time of test (pre- vs. post-test) as the within-subject factors. The purpose of comparing the identification accuracy on the untrained /ε/-/æ/ and /t/-/d/ final phonemic words across time of test was to examine whether both types of untrained stimuli improved similarly in post-test. The adaptive staircase and constant groups identified the untrained materials better in post-test when compared to pre-test, $F(1,14) = 7.00$, $p = .02$ and $F(1,13) = 5.52$, $p = .04$ respectively. Nevertheless, no other effect or interaction was found from both analyses, Fs < 1.85, ps > .05.

5.2.4 Discussion

The SED training improved L1 Mandarin speakers' discrimination sensitivity and identification accuracy in post-test; the training effect observed retained after a month delay. Importantly the control group did not show any changes in their perceptual performance when they were tested at a similar time intervals as the training groups.

Initially, the effectiveness of adaptive staircase SED and constant SED seems to be comparable when the their training effects were examined on overall testing materials (i.e. including /t/-/d/ final and /ɛ/-/æ/ phonemic words) . However, when examining the training effect solely on the /ɛ/-/æ/ phonemic contrast, perceptual improvement induced by the constant SED did not remain while the improvement induced by the adaptive staircase SED still remained significant. This finding marks the substantial advantage and effectiveness of the adaptive staircase procedure used to implement the background babble in SED. Nevertheless, the training effect of both SED designs generalized equally well to novel words of the same /ɛ/-/æ/ phonemic contrast and also to the untrained phonemic contrast /t/-/d/ final.

5.2.4.1 Comparison between Present Study and Previous SED Training

Before proceeding to analyze the speech production, it was important to ensure findings of the present study were comparable to the SED training studies reported in Chapter 3. The comparison across SED training studies would illustrate the consistency of the training effect and strengthen the rationale to examine the perceptual training effect on speakers' speech production.

Linguistic information and pre-test performance of the L1 Mandarin speakers from both chapters were compared using between-subjects one way ANOVAs. Results showed no significant difference in age of English acquisition, self-rated English proficiency and self-rated Mandarin proficiency across the speakers, *F*s < 1.67, *p*s > .05. Importantly, the speakers also performed similarly in identifying the /t/-/d/ initial, /t/-/d/ final and /ɛ/-/æ/ phonemic words in the pre-test, *F*s < 1. The results had suggested that speakers involved in both studies shared similar linguistic background (see Appendix 4 for further details) and had started off with similar perceptual performance before training.

SED training from both studies showed that participants' perceptual performance improved and the improvements generalized to novel words, novel speakers and also to the untrained phonemic contrast /t/-/d/ final. There was however, one inconsistent finding between the studies. In Chapter 3, Adaptive Staircase SED was more effective in terms of generating greater training effect size or perceptual improvement (10%) when compared with Constant SED (3%). This advantage of Adaptive Staircase SED, however, was not found in the present study, with the improvement generated by the Adaptive Staircase SED being 6% in comparison with the Constant SED 5%. One reason to explain the incongruity was the smaller improvement space allowed by the adaptive staircase group from the present study (pre-test identification accuracy 78%) compared to the adaptive staircase group from Chapter 3 (pre-test identification accuracy 71%), although the difference was not significant.

The advantage of Adaptive Staircase SED observed in Chapter 3 stemmed from its generalization effect across the untrained phonemic contrast /t/-/d/ final. However, the present study showed that when performance on the /t/-/d/ final phonemic contrast was removed from the analysis, the training effect of Constant SED on the /ɛ/-/æ/ phonemic words became non-significant. The training effect of Adaptive Staircase SED was more robust as it remained significant for the /ɛ/-/æ/ phonemic contrast.

5.3 Experiment 7 (Evaluation of Speech Recordings)

5.3.2 Evaluation Method

5.3.2.1 Design and Materials

The production accuracy of the speakers was compared across the time of test to examine the effects of SED training on their production intelligibility. Speech productions of 2 speakers (1 from the Adaptive Staircase Group and 1 from the Constant Group) were excluded from evaluation because the participants did not return for the 1 month follow-up test.

A total of 7872 Malaysian English speech stimuli (41 participants X 64 English words X 3 times of test) generated from Experiment 6 above were evaluated in the present study. The speech stimuli consisted of 16 words from the /t/-/d/ initial phonemic contrast, 16 words from the /t/-/d/ final phonemic contrast and 32 words from the /ɛ/-/æ/ phonemic contrast (see Appendix 1 for full list) that were recorded during the pre-test, post-test and 1 month post-test. Using Praat, the volume of all Malaysian English speech stimuli were normalized based on the average intensity level of the British English speech stimuli used in the SED training, which was at 68 db SPL.

5.3.2.2 Participants

28 L1 ME speakers (age range: 18-33; mean age 22.21; 15 females) and 28 native British English speakers (age range: 19-30; mean age 23.68; 20 females) were recruited from the University of Nottingham from both the Malaysia and UK campuses. They were paid for their participation. The participants completed a language questionnaire, providing information about their language background and formal education (see Appendix 4 for further details). All Malaysian participants reported to speak at least two languages; with one language being English (other languages included Mandarin, Malay etc.). 20 British participants reported to have learnt a foreign language (e.g. French, Spanish and German) before adulthood. A below average proficiency level was

reported for the foreign language learnt (mean proficiency = 2.33). Participants also completed the Lextale test providing a general measure of their English proficiency. Independent-samples t-tests were conducted to compare the self-rated English proficiency and Lextale scores between the Malaysian raters and British raters. Although the self-rated English proficiency of British raters was significantly higher than the Malaysian raters, $t(54) = 4.10$, $p < .001$, their Lextale score did not differ, $t(54) = 1.33$, $p = .19$

All participants reported to have normal eyesight or corrected to normal vision, and had no history of any hearing, speech or reading disorders.

5.3.2.3 Procedure

The evaluation task involved a 2 Alternative Forced-Choice (2AFC) identification task. There were a total of 576 trials and raters heard, identified and rated every speech stimulus once. In each trial, two words of a minimal pair were presented side by side at the center of computer screen with the speech stimulus played simultaneously. The speech stimuli were presented using SONY MDR-NC8B noise cancelling headphones at a comfortable listening level set by each participant. Participants were asked to indicate which word on the computer screen matched the word they had heard by pressing the keyboard key “A” or the key “L”. After each key pressed, the chosen word remained while the unchosen word disappeared from the screen. The same speech stimulus was played once again and this time participants were asked to rate on a scale from 1 to 7 (1 - Bad, 4 - Average, 7 -Excellent) on how well the auditory word heard was an example of the word they just chose. Participants pressed the number keys 1 to 7 from the keyboard to indicate the rating scores for each speech stimulus. They were asked to rate based on the /t/-/d/ or the /ɛ/-/æ/ realizations in the word rather than on the pronunciation accuracy of the word itself. The next trial began after raters were satisfied with the rating score and had pressed the “Enter” key from the keyboard. No feedback was given at any stage. Before the start of the evaluation task, participants completed six practice trials first in order to familiarize themselves with the task. Speech productions of a novel speaker whose productions were not evaluated by any rater were used for the practice trials. The evaluation task lasted about 50 minutes.

All raters evaluated speech productions of three speakers (one speaker from each of the groups: the adaptive staircase group, the constant group and the control group), which were recorded from the 3 times of test (i.e. pre-test, post-test and 1 month post-test). Presentation of the stimuli was first blocked by the speakers and then by phonemic categories. Hence, raters heard all speech stimuli of one speaker before continuing with speech stimuli of another speaker. The presentation order was counterbalanced at multiple levels. Half of the raters evaluated the productions of the control group speaker first before the training groups' speakers while another half evaluated the productions of the training groups' speakers before the control group speaker. The presentation order of the training groups' speakers was also counterbalanced. Half of the raters evaluated the productions of the adaptive staircase group speaker before the constant group speaker; likewise, another half of the raters evaluated the constant group speaker before the adaptive staircase group speaker. In addition, half of the raters always evaluated the word stimuli from the /t/-/d/ phonemic contrast of a speaker first before the /ɛ/-/æ/ phonemic contrast; and another half did vice versa. The /t/-/d/ phonemic words contrasting at the word initial position were always evaluated before the words contrasting at the word final position.

All raters did the pre-test of SED training after the evaluation task to measure their perceptual performance on the British English speech stimuli.

5.3.3 Evaluation Behavioral Results

Each speech item produced by each speaker was evaluated by 4 raters (2 Malaysian raters and 2 British raters). In order to make up the insufficient number of speakers from the constant group (13 speakers instead of 14 as the other two groups), productions of one speaker from this group were evaluated twice but the second batch of evaluation data was excluded from any further data analysis. The evaluation data was presented in two forms: the identification accuracy of the raters and the rating scores on the speech production quality.

The evaluation data was analyzed separately for the L1 ME raters and L1 British English raters due to their phonetic perception difference as found in the preliminary analysis of Chapter 4 (e.g. L1 ME speakers' identification difficulty of the /ε/ phonemic words). In order to prevent further complications, we do not intend to compare the judgment difference between the rater groups. The following analyses aimed to examine how word production of the participants may have changed in post-test, which might appeal differently to the English speakers of different linguistic backgrounds.

5.3.3.1 Identification Accuracy

The mean percentage of correct identification was calculated for each of the four phonetic categories - /æ/, /ε/, /t/ final and /d/ final phonemes in the pre-, post- and 1 month post-test of the three groups of participants (see Appendix 5). Data from the /t/ and /d/ phonetic categories located at word initial position was excluded from further analyses because of ceiling performance in each rater group (mean above 95%).

5.3.3.1.1 Malaysian Raters

The mean percentage of correct identification on each /æ/, /ε/, /t/ final and /d/ final phoneme was subjected to two separate repeated-measure ANOVAs with time of test (pre- vs. post-test vs. 1 month post-test) and training group (adaptive staircase vs. control or constant vs. control) as within-subjects factors.

Adaptive Staircase SED vs. Control

The analysis comparing identification accuracy on the /æ/ phonemic words that were produced by the adaptive staircase and control group revealed a significant effect of time of test, $F(2,54) = 3.93$, $p < .05$ and an interaction between time of test and training group, $F(2,54) = 3.67$, $p < .05$. Pairwise comparisons (Bonferroni) for the main effect found no significant difference overall between the times of test, *p*s > .05. Two post-hoc one way repeated-measure ANOVAs were conducted separately for each training group to examine the interaction observed. The analyses found a main effect of time of test for the adaptive staircase group, $F(2,54)=7.31$, $p < .01$, but not for the control group, $F < 1$ (see Figure 5.4). Post-hoc pairwise comparisons (Bonferroni) showed no significant difference between pre- and post-test /æ/ identification accuracy in adaptive staircase group ($p = .51$), but a significantly higher accuracy in 1 month post-test compared to pre-test ($p < .01$) and post-test ($p < .05$).

No other significant finding was revealed for other phonemes, *F*s < 2.22, *p*s > .05.

Constant SED vs. Control

The analysis comparing identification accuracy between the constant and control groups revealed no effects or interactions, *F*s < 1.15, *p*s > .05.

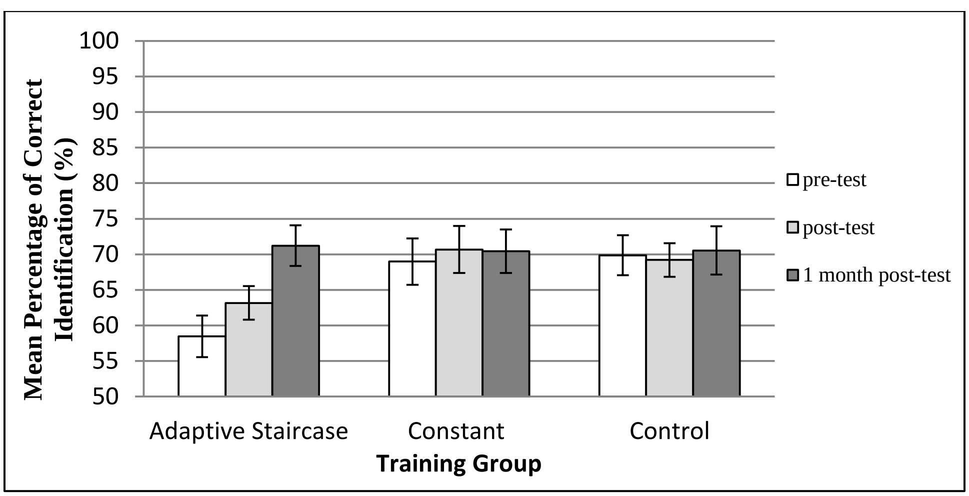


Figure 5.4 Mean percentage of correct identification made by the Malaysian raters on the /æ/ phonemic words.

5.3.3.1.2 British Raters

Adaptive Staircase SED vs. Control

The analysis comparing identification accuracy on the stimuli words produced by the adaptive staircase and control groups revealed no significant effects or interactions, *F*s < 2.35, *p*s > .05.

Constant SED vs. Control

The analysis comparing identification accuracy on the /æ/ phonemic words that were produced by the constant and control group revealed a significant effect of time of test, $F(2,50) = 4.93$, $p = .01$ and a trend towards significant interaction between time of test and training group, $F(2,50) = 2.62$, $p = .08$. Post-hoc pairwise comparisons (Bonferroni) revealed significantly higher post-test identification accuracy (71%) over pre-test (63%), $p < .05$. To examine the approaching significant interaction, two separate one-way repeated-measure ANOVAs were conducted for each training group to compare the identification accuracy across times of test. The analyses revealed a main effect of time of test for the constant group, $F(2,50) = 7.29$, $p < .01$ but not for the control group, $F(2,50) = 1.07$, $p = .35$ (see Figure 5.5). The identification accuracy on the /æ/ phoneme in post-test was significantly higher than pre-test ($p < .01$) and 1 month post-test ($p < .05$).

No other significant finding was revealed for the other phonemes, *F*s < 1.39, *p*s > .05.

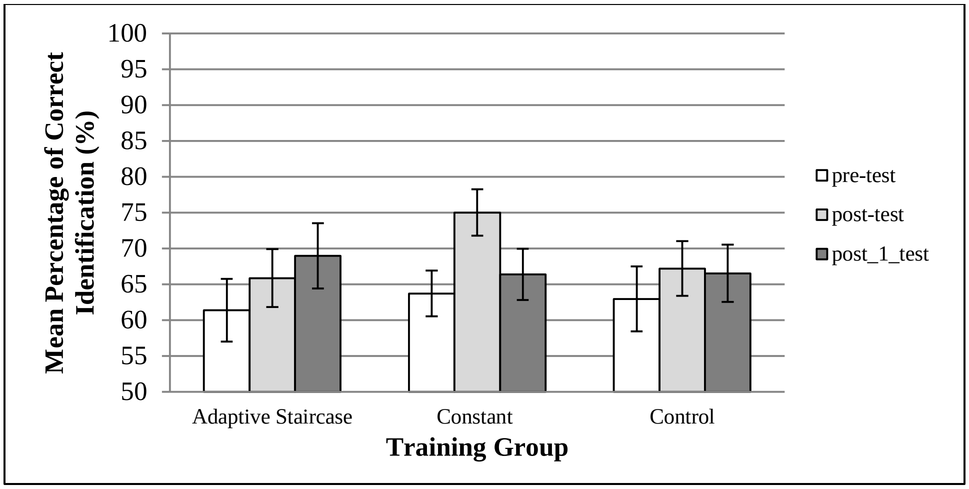


Figure 5.5 Mean percentage of correct identification made by the British raters on the /æ/ phonemic words.

5.3.3.2 Quality Rating

The mean rating score was calculated for each of the six phonetic categories - /æ/, /ɛ/, /t/ final, /d/ final, /t/ initial and /d/ initial phoneme in pre-test, post-test and 1 month post-test (see Appendix 6).

5.3.3.2.1 Malaysian Raters

The mean rating scores on each of the /æ/, /ɛ/, /t/ final, /d/ final, /t/ initial and /d/ initial phoneme provided by the Malaysian raters were subjected to separate repeated-measure ANOVAs with time of test (pre- vs. post-test vs. 1 month post-test) and training group (adaptive staircase vs. control or constant vs. control) as the within-subjects factors.

Adaptive vs Control

The analyses comparing the adaptive staircase and control groups revealed a higher 1 month post-test rating score compared to post-test for the /t/ initial phoneme, $F(2,54) = 3.62$, $p = .03$, and an interaction between time of test and training group for the /t/ final phoneme, $F(2,54) = 5.63$, $p < .01$. Post-hoc one-way repeated-measures ANOVAs were conducted to compare rating scores across time of test separately for the adaptive staircase and control group; the results only revealed time of test effect that was marginally significant for the two groups, $F(2,54) = 2.90$, $p = .06$ and $F(2,54) = 2.59$, $p = .08$ respectively. Three post-hoc paired-samples t-tests were conducted to compare rating score between the two training groups in the respective time of test: pre-test, post-test and 1 month post-test. The rating score on control group was marginally higher than the adaptive staircase group in post-test, $t(27) = 1.99$, $p = .06$. No other effect or interaction was significant for the rest of the phonemes, Fs < 1.41, ps > .05.

Constant vs control

The analyses comparing the constant and control groups showed that the post-test rating score was higher than the pre-test for the /t/ final phoneme, $F(2,50) = 3.81$, $p = .03$. An interaction between the time of test and the training group factors was also found for the /d/ final phoneme, $F(2,50) = 3.47$, $p = .04$. Post-hoc one-way repeated-measures ANOVAs were conducted to compare the rating scores across time of test separately for the constant and control group; the results only revealed a marginally significant effect of time of test for the constant group, $F(2,50) = 2.74$, $p = .07$, in which the 1 month post-test rating score was marginally higher than post-test, $p = .08$. The post-hoc paired-samples t-tests did not reveal any significant difference between the two groups for every time of test, ts < 1.77, ps > .05. There was also a marginally significant interaction between the time of test and the training group factors for the /æ/ and /d/ initial phonemes, $F(2,50) = 3.01$, $p = .06$ and $F(2,50) = 3.03$, $p = .06$ respectively. Nevertheless, their post-hoc tests did not show any significant effect or difference, Fs < 2.60, ts < 1.21, ps > .05. No other effect or interaction was significant for the rest of the phonemic categories, Fs < 1.59, ps > .05.

5.3.3.2.2 British Raters

The mean rating scores on each /æ/, /ɛ/, /t/ final, /d/ final, /t/ initial and /d/ initial phoneme provided by the British raters were subjected to separate repeated-measure ANOVAs with time of test (pre- vs. post-test vs. 1 month post-test) and training group (adaptive staircase vs. control or constant vs. control) as the within-subjects factors.

Adaptive vs control

The analyses comparing the adaptive staircase and control group revealed a main effect of time of test and a main effect of training group for the /d/ initial phoneme, $F(2,54) = 4.43$, $p = .02$ and $F(1,27) = 5.08$, $p = .03$ respectively. Pairwise comparisons (Bonferroni) showed a higher rating score in 1 month post-test compared to post-test, $p < .05$ and a higher rating score for the control group compared to the adaptive staircase group, $p = .03$. The main effect of time of test for the /t/ initial phoneme merely showed a trend towards significance, $F(2,54) = 2.59$, $p = .08$.

Interestingly, there was an interaction between the time of test and the training group for the /ɛ/ phoneme, $F(2,54) = 4.10$, $p = .02$ (see Figure 5.6). Two post-hoc one-way repeated-measures ANOVAs were conducted to compare the rating scores across time of test separately for each training group. The analyses showed an effect of time of test for the adaptive staircase group, $F(2,54) = 4.40$, $p = .02$, but not the control group, $F(2,54) = 1.23$, $p = .30$. The post-test and 1 month post-test rating scores for the adaptive staircase group were significantly lower than the pre-test rating score, *p*s = .04. There was no difference between the post-test and 1 month post-test rating scores, $p = 1$. In addition, post-hoc paired-samples t-tests conducted only showed a marginally higher rating score for the adaptive staircase group compared to the control group in pre-test, $t(27) = 1.98$, $p = .06$. The two groups did not differ in post-test and 1 month post-test, *t*s < 1.

The analysis showed no other effect or interaction for the rest of the phonemic categories, *F*s < 2.44, *p*s > .05.

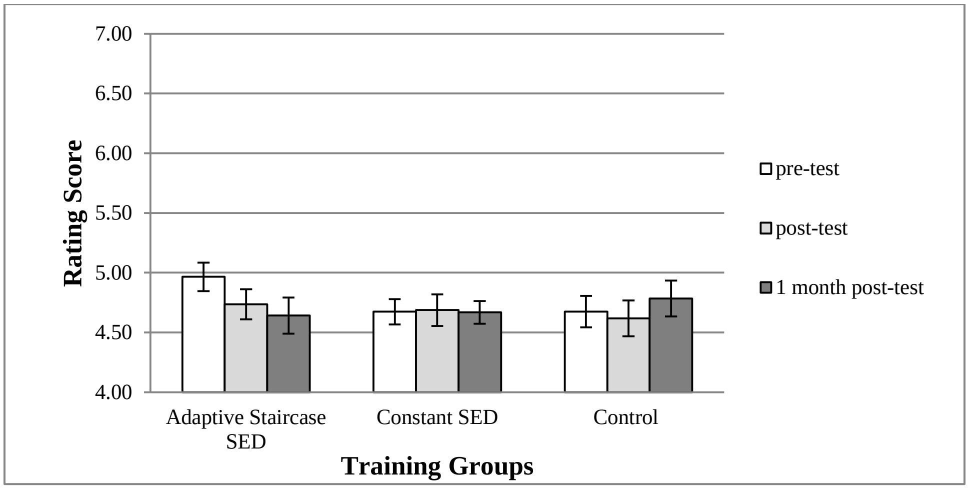


Figure 5.6 Rating scores on the /ɛ/ phonemic words by the British raters (rating scale: 1 – bad example to 7 – excellent example).

Constant vs control

Similar two-ways repeated-measure ANOVA were also conducted to compare the constant and control group for each phonemic category. The analyses showed a main effect of time of test for the /t/ initial phoneme, $F(2,50) = 6.42$, $p < .01$. Pairwise comparisons (Bonferonni) showed that the post-test rating score for the phoneme was the lowest among the other time of test, *p*s < .05; while the pre-test and 1 month post-test rating scores were similar, $p = 1$.

An interaction between the time of test and training group was also found for the /d/ initial and /t/ initial phonemes, $F(2,50) = 6.42$, $p < .01$ and $F(2,50) = 5.38$, $p < .01$ respectively. Post-hoc one-way repeated-measures ANOVAs conducted for the /d/ initial phoneme did not show any significant tome of test effect for both training groups, *F*s < 1.94, *p*s > .05. According to the post-hoc paired-samples t-tests, the rating score for the constant and control group did not differ in pre- and post-test, *t*s < 1; the rating score for the control group, however, was higher than the constant group in 1 month post-test, $p = .02$.

As for the interaction found in the analysis on /t/ initial phoneme, the post-hoc one-way repeated-measures ANOVAs showed a time of test effect for the constant group, $F(2,50) = 8.58$, $p = .001$. Pairwise comparisons (Bonferroni corrected) showed a significantly higher rating score in pre-test compared to post-test, $p < .01$ while the difference between post-test and 1 month post-test showed a marginal significant trend, $p = .07$; the pre-test and 1 month post-test rating scores did not differ, $p = .13$. The time of test effect for the control group was marginally significant, $F(2,54) = 2.88$, $p = .07$. Pairwise comparisons (Bonferroni) only showed a marginal significant difference between the post-test and 1 month post-test rating scores, $p = .06$; the pre-test rating score was similar to both post-test and 1 month post-test, $p > .05$. Besides the post-hoc ANOVAs, paired-samples t-tests were also conducted for the interaction observed. The results showed a marginally significant difference between the two training groups in post-test, $t(25) = 1.88$, $p = .07$. The rating scores in pre-test and 1 month post-test were similar between the two training groups, *t*s < 1.47, *p*s > .05.

The analyses comparing the constant and control group showed no other effect or interaction for the rest of the phonemic categories, *F*s < 2.01, *p*s > .05.

5.3.3.2.3 Inter-rater Reliability across Raters

Inter-rater reliability was assessed using a one-way random model, average measures Intra-class Correlation Coefficient (ICC) to assess the degree that Malaysian and British raters provided consistency in their ratings on the production quality. Four ICC analyses were conducted to examine the rating consistency within each /t/-/d/ final and /ɛ/-/æ/ contrast respectively, and separately for the Malaysian and British raters. The resulting negative ICC estimates from all four analyses, ICC (1, 28) < -.79, indicate systematic disagreement across the raters. The reliability of the production quality rating suggests any conclusions from the current study must be treated with caution.

5.3.3.3 Relationships between English Proficiency, Perception and Production

Additional statistical analyses were conducted to examine the relationships between speakers' English proficiency, perception and production performance. The perception scores, the production accuracy scores and the production rating scores were first computed by deducting the pre-test scores (either in percentage or in rating score ranging from 1-bad to 7-excellent) from the post-test scores. The scores were then transformed into Z-scores for the following analyses.

5.3.3.3.1 The Relationship between English Proficiency, Perception Gain and Production Gain from SED training

Pearson correlation coefficient was conducted to assess the relationship between speakers' Lextale scores, perception scores, production accuracy scores and production rating scores on the SED speech stimuli (including /t/-/d/ final and /ɛ/-/æ/ phonemic contrasts). Data from /t/-/d/ initial contrast was omitted because the perception scores were at ceiling level. A negative relationship was found between the Lextale and perception scores, $r(82) = -.23$, $p = .04$, suggesting those who have relatively low proficiency in English benefited more from the SED training. Consistent with the findings from Bradlow et al. (1999), there was no relationship between the perception and the production accuracy scores, $r(82) < 0.01$, $p = .98$, or with the production rating scores, $r(82) = -.15$, $p = .18$. There was, however, a positive relationship between the production accuracy scores and the production rating scores, $r(82) = .24$, $p = .03$.

5.3.3.3.2 The Effect of SED on the Quantity of Production Improvement Cases

Speakers were categorized according to their performance changes in post-test perception (improved or deteriorated) and production (improved or deteriorated). Four performance categories of change pattern were generated and the number of occurrence who improved in perception and production, improved in perception but deteriorated in production, deteriorated in perception but improved in production or deteriorated in both perception and production were calculated and presented in Figures 5.7, 5.8, 5.9 & 5.10.

From the figures, a considerable amount of individual variation can be observed across participants in perception and production changes from pre-test and post-tests, regardless of whether they received any SED training. When looking at the perception performance solely (perception axis), more participants from the Adaptive Staircase SED training improved after training (9 from 14) compared to the Constant SED training (5 from 13) and the control group (4 from 14).

Figure 5.7 showed little correlation between the perception and production intelligibility z-scores based on Malaysian raters' ratings. From the graph, the numbers of cases that fell on the improvement and deteriorate sides were relatively uniform. Production intelligibility of six Adaptive Staircase SED participants, six Constant SED participants and seven control group participants improved after training.

Figure 5.8 showed a weak negative correlation between the perception and production rating z-scores for the three training groups, based on Malaysian raters' ratings. There were more deteriorate cases however from the Adaptive Staircase SED (11) and Constant SED training groups (8) compared to the improved cases. Surprisingly, more control group participants (9) were given positive rating z-scores compared to negative scores.

Figure 5.9 was constructed based on British raters' data. Although the figure showed positive correlation between the perception and production intelligibility z-scores for the Constant SED and control groups, the z-scores size for the cases were rather restricted and were mostly clustered at the center. The two positive slopes might be driven by the three extreme influential points (two most negative cases on production axis from control group and one most positive case on production axis from Constant SED). Besides that, the number of cases that improved or deteriorated was rather equal. Production intelligibility of seven Adaptive Staircase SED participants, seven Constant SED participants and seven control group participants improved in post-test.

Based on British raters' ratings, Figure 5.10 showed little correlation between the perception and production rating z-scores for the Adaptive Staircase SED and Constant SED groups, but a positive correlation for the control group. The improved and deteriorated cases from the Adaptive Staircase SED and control groups were rather equal, but the deteriorated cases (8) from Constant SED were more than the improved cases (5).

Four Fisher's exact tests (two on production accuracy and two on production rating) were conducted to examine whether participants' performance pattern differs according to their training groups (adaptive staircase vs. constant vs. control; see figures below). The production data provided by the Malaysian and British raters were analyzed separately. The tests showed no significant association between the training groups and their performance patterns in terms of their production intelligibility, $p = .65$ (Malaysian raters) and $p = .09$ (British raters); nor in the production rating, $p = .13$ (Malaysian raters) and $p = .12$ (British raters).

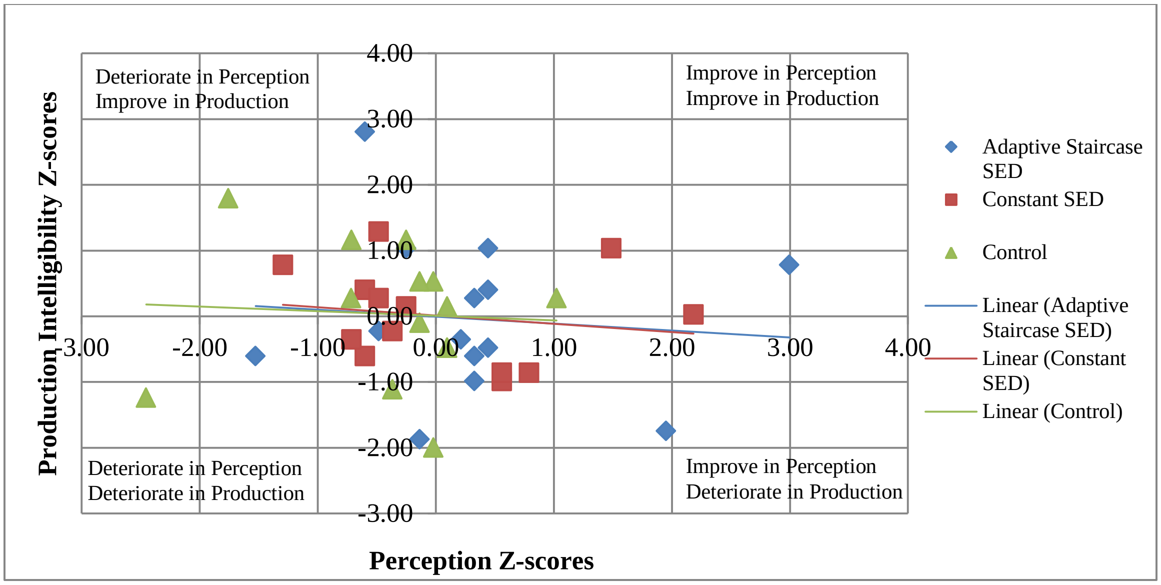


Figure 5.7 Perception z-scores against production intelligibility z-scores based on identification accuracy of the Malaysian raters.

*Note: Perception and production intelligibility scores (in percentage) were first calculated by subtracting pre-test performance from post-test before being converted into z-scores.

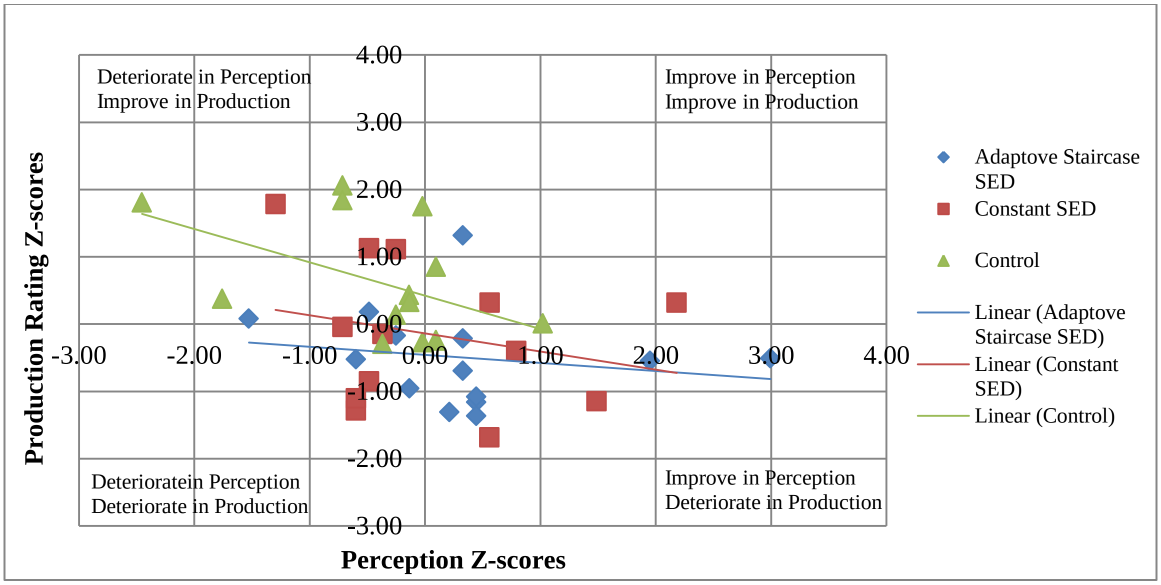


Figure 5.8 Perception z-scores against production rating z-scores as evaluated by Malaysian raters.

*Note: Perception and production rating scores (from a scale of 1 = bad to 7 excellent example) were first calculated by subtracting pre-test performance from post-test before being converted into z-scores.

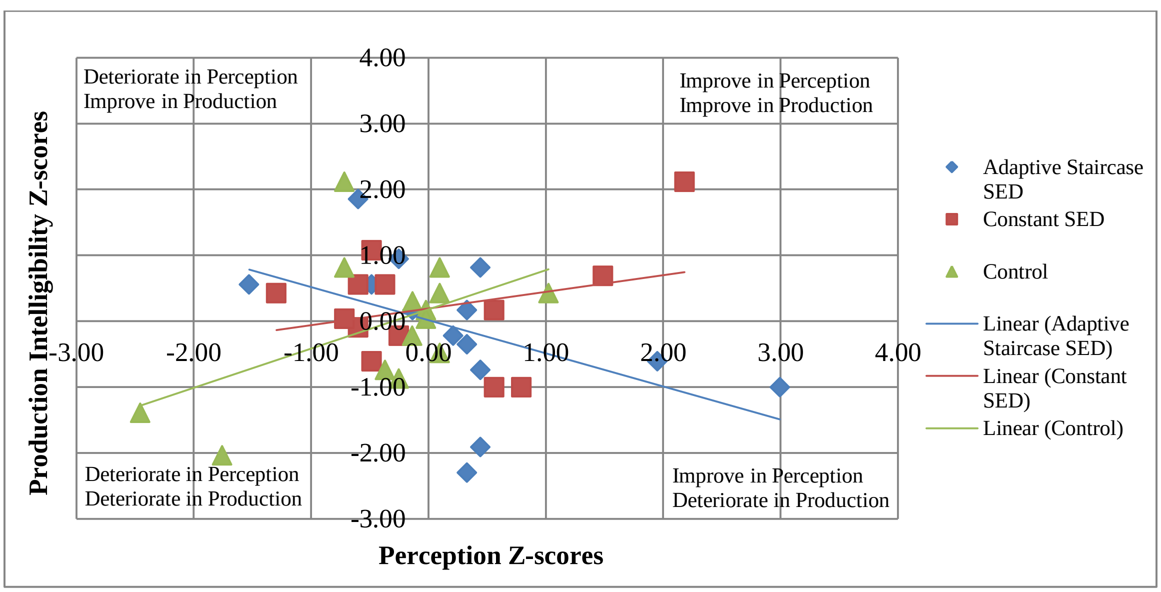


Figure 5.9 Perception z-scores against production intelligibility z-scores based on identification accuracy of the British raters.

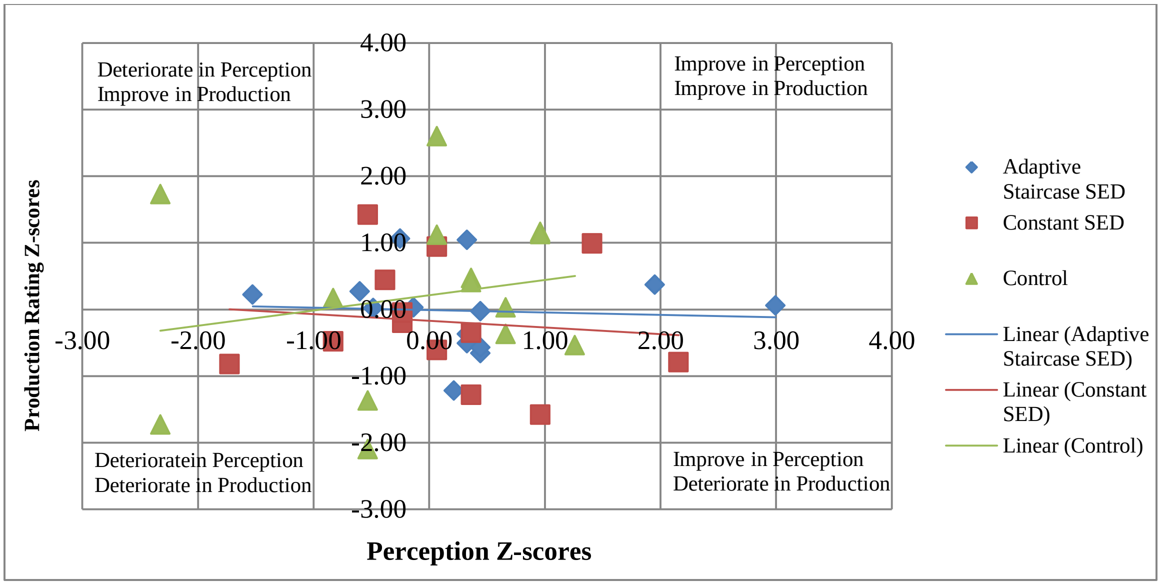


Figure 5.10 Perception z-scores against production rating z-scores as evaluated by British raters.

Four Pearson chi-square tests were conducted to examine whether the training group affects the number of improvements in post-test production. The Pearson chi-square tests were conducted on a 2 (production performance: improve vs. deteriorate), X 3 (training groups: adaptive staircase vs. constant vs. control) cross-tabulation. The production accuracy and production rating data were analyzed separately for each group of raters.

For Malaysian raters, number of improvement cases did not differ by the training group in production accuracy, X ²(2, N=41) = .62, *p* = .73, but the two factors were significantly associated in production rating, X ²(2, N=41) = 7.33, *p* = .03. In terms of production quality (based on rating scores), although the odds of improvement cases were lower if they were trained with adaptive staircase SED (z = -1.3) than if they received no training (z = 1.6), the standardized residuals were not significant. There was also a tendency for the Adaptive Staircase SED to generate more deterioration cases (z = 1.1) compared to the control group (z = -1.4) although the standardized residuals were again not significant.

For the British raters, results from the Pearson chi-square tests showed that the performance in production did not differ by the training group in production accuracy, $X^2(2, N=41) = .37$, $p = .92$ or in production rating, $X^2(2, N=41) = 3.33$, $p = .23$.

5.3.4 Discussion

5.3.4.1 Training Effect of SED on Production Intelligibility

The production intelligibility of the speakers was measured based on the identification accuracy of the raters. The results suggest some improvements in the production intelligibility of the /æ/ phonemic words. It is interesting to note that the L1 British English raters seem to have different perception judgements on the speakers' production improvement compared to L1 ME raters. They found the constant SED design to be more effective in generating an improvement in intelligibility. According to the Malaysian raters, only speakers who were trained with the adaptive staircase SED improved in their /æ/ phonemic words production, in which the 1 month post-test production was better identified than the pre-test. On the other hand, the British raters found production improvement only in the constant SED training group; their production of the /æ/ phonemic words was more intelligible in post-test compared to pre-test. This finding was especially meaningful because the primary aim of SED training in Experiment 6 was to improve the phonetic salience of the /ɛ/-/æ/ phonemic pair in L1 Mandarin speakers. The improvement in /æ/ phoneme production showed the transferability of the perceptual learning to the production modality. This finding also supports the idea that through SED training, the L1 Mandarin speakers learnt the criterial acoustical cues that distinguish the phonemic contrast, and capable to use that knowledge to guide their production. The SED training, however, did not have any effect on the /ɛ/, /t/ final, /d/ final, /t/ initial and /d/ initial phonemic words production.

5.3.4.2 Training Effect of SED on Production Quality

The effect of training on the production quality of the speakers is not as clear cut as the intelligibility results. Production quality was measured based on how well the raters agreed the productions represent the matched word. According to the raters of both nationalities, only the word productions of the /t/ and /d/ phonemic categories showed a tendency to change across time of test. Most of the tendencies were either not significant or not likely to be the result of the SED training. For instance, one month post-test productions were rated higher than the post-test productions for phonemic categories that were not used as training materials (e.g. the /t/ initial phoneme rated by the Malaysian raters and the /d/ initial phoneme rated by the British raters). The improvement from post-test to 1 month post-test did not seem to be induced by the SED training because participants received no training in between these two time points. Rather, the judgment difference might be related to the production quality (e.g. unclear pronunciation) instead of the phonemic quality.

Further acoustical analysis on the productions might reveal the possible changes in productions that affect the rating judgment. If acoustical analysis does not reveal any changes in the acoustical properties, the rating judgment might have been influenced by the environmental elements such as the inevitable extraneous noise that came with the recording procedure.

Nevertheless, it was worth noting the production changes in the /ε/ phonemic words. According to the British raters' ratings, the production quality of the phoneme deteriorates in post-test and 1 month post-test when the speakers were trained with adaptive staircase SED. The intelligibility of the /ε/ phonemic words however did not change across the time of test (see section 5.3.4.1). The reason behind the deterioration in the rating scores remains unknown. It could be attributed to the potential production changes or the relatively higher rating score given to the adaptive staircase group's pre-test production as compared to the other two training groups.

It was also important to note that analyses conducted to examine the inter-rater reliability showed a systematic disagreement across the raters. The finding may have suggested the difficulty of raters in observing or quantifying the construct of interest and they may make the rating judgments based on inconsistent features of the acoustical stimuli. Recruiting trained phoneticians in the future might provide a more reliable evaluation for the speech stimuli.

5.3.4.3 Implications from the Production Evaluation

Due to the inconsistencies between the findings on production intelligibility and quality, the rating on production quality will not be discussed further. The irregularity of the quality rating might have reflected the diversity of recording quality (e.g. extraneous noise during the recording procedure) rather than the acoustical changes in speakers' speech production. Importantly, neither intelligibility nor quality (mostly) of the control group's speech production changed across time of test (see Chapter 4). Therefore, it would be justifiable to attribute any production intelligibility changes observed in the training groups to the effects of the SED perceptual training.

The two designs of SED improved different aspects of L1 Mandarin speakers' production. However, the pattern of improvements was rated differently depending on the nationality of raters. When comparing the pre-test (58%) and post-test (63%) identification performance, the Malaysian raters identified the /æ/ phonemic words better in 1 month post-test (71%) when the speakers were trained with adaptive staircase SED. Although improvements were expected in the post-test, the difference in identification performance between pre-test and post-test was not significant. Nevertheless, there was a trend in improvement in identification accuracy observed across the time of test. Moreover, the intelligibility improvements observed in 1 month post-test (an increment by 13% compared to pre-test) was rather impressive and such improvement had not been observed anywhere else, especially in the control group. This trend was also observed across time of test when the British raters identified the /æ/ production of the adaptive staircase group.

However, the identification accuracy difference between the pre-test (61%), post-test (66%) and 1 month post-test (69%) did not differ statistically.

On the other hand, the British raters found a production intelligibility improvement on the /æ/ phonemic words when speakers were trained with the constant SED. Though the improvement observed in post-test (11%; pre-test 64% - post-test 75%) deteriorated after one month of resting period (66%). Interestingly, the Malaysian raters did not find any changes in the constant group's production. The evaluation discrepancy between the Malaysian and British raters suggests that they identified the /æ/ phonemic words based on different acoustical properties.

The present findings also suggest that the different methods of adjusting the level of background babble in SED training did allocate speakers' attention to different acoustical cues in pronouncing the phonemic words. If the aim of the phonetic training was to improve speakers' production intelligibility among listeners who are familiar with the target language used in the training (e.g. L1 British speakers as the listeners who should be familiar with the Southern Irish English, American English, Irish English speech productions used in SED), constant SED was more effective in achieving this aim compared to the adaptive staircase SED. Implementing background babble using adaptive staircase procedure was only capable of improving the speakers' production intelligibility among listeners who are familiar with the speakers' production.

5.3.4.4 The Link between English Proficiency, Perception and Production

When plotting the perception and production scores of the speakers as evaluated by the Malaysian raters, it appears there is a negative relationship between the two factors (see Figure 5.7 & 5.8). This observation is rather surprising because speakers were expected to improve in their speech production when they perceived the target language more accurately after training. The plotting of perception and production intelligibility scores of the speakers as evaluated by the British raters, however, showed a positive relationship (see Figure 5.9). Nevertheless, the statistical results failed to show any significant relationship between the perception and production changes in speakers, regardless of the nationality of the raters. A similar finding was reported in Bradlow et al. (1999) whereby the perception improvement did not correlate with the production improvement. Speakers benefited more from the SED training perceptually when their English proficiency was lower; probably due to the greater room available for potential improvement.

5.5 General Discussion

Once again, the effectiveness of implementing the adaptive staircase procedure in SED training was shown. The procedure generated more robust perceptual improvements on the trained /ɛ/-/æ/ phonemic contrast when compared to the Constant SED (Experiment 6). Importantly, the performance of the control group remained consistent across the three testing sessions, suggesting mere exposure to the speech materials was insufficient to induce perceptual improvement. It was thus reasonable to attribute the perceptual improvement of the training groups observed to the training effect of SED.

The effect of SED on speakers' production intelligibility and quality were examined in Experiment 7. It is clear that the perceptual improvement generally exceeded the production improvement. This phenomenon seems justifiable as the L1 Mandarin speakers received extensive training in speech perception, but not in speech production. Any improvement in production domain should be a transfer of knowledge gained from the perceptual learning.

The perceptual learning had transferred to the production of /æ/ phonemic words. The initial identification difficulty in L1 Mandarin speakers with the /ɛ/-/æ/ phonemic contrast is due to the ambiguous phonological representation of the two phonetic sounds. They tend to assimilate the two phonetic sounds and matched them more frequently to the <a> English alphabet (see Chapter 4 for more information on phonological representation of L1 Mandarin speakers). After an hour of SED discrimination training, their identification accuracy on the contrast and their production intelligibility of the /æ/ phonemic words significantly improved; probably due to the phonological representation of /æ/ phonemic category (also known as the self-monitoring reference object, see Levelt, 1989) becoming more prominent.

Bradlow and colleagues (1997; 1999) claimed that production improvement observed was the product of modification of the underlying phonological representation because the production changes took place immediately after the perceptually oriented training. Similar to their studies, the speakers in the present study were not given any chance to revise their production acuity through output monitoring (by referring to own overt speech production; see Levelt 1989) before being recorded. Hence, the production changes observed were attributed to the effect of internal-monitoring process in which self-generated phonological representations were used as reference (see chapter 6 for further discussion).

The generalization of phonetic training to production was inconsistent across the training materials. It should be noted that no production improvement was observed for phonemic categories other than the /æ/. The production intelligibility of the L1 Mandarin speakers on the phonemic words involving the /t/-/d/ initial phonemic contrast achieved ceiling performance (above 95% after averaged across all raters) in pre-test; thus no improvement was expected. The production intelligibility of the other phonemic categories (i.e. /t/ final, /d/ final & /ɛ/), however, did not improve along with their perceptual accuracy.

It is still uncertain whether SED training modified the underlying phonological representation. The absence of production improvement might show that the relevant underlying phonological representation and articulatory mechanism for these phonetic sounds were too deep-rooted to change over a short training period, even though speakers' perceptual sensitivity had improved. It is, however, important to note that the six reviewed published papers that had reported the successful production generalization effect had trained their speakers with a relatively longer duration (mean 9 hours, ranging from 4 to 22.5 hours; see Bradlow et al., 1999; 1997; Hazan et al., 2005; Nobre-Oliveira, 2008; Aliaga-Garcia & Mora, 2009; Lengeris & Hazan, 2010; Iverson et al., 2012). If the SED training duration was lengthened to better match other studies, its effect on production might be more robust.

Interestingly, the effectiveness of the SED paradigms in improving production intelligibility differed according to the rater groups (Experiment 7). The design of SED background babble seems to direct speakers' attention to different acoustic cues of the speech sound and alter different properties of the speakers' word production that appealed differently to the raters depending on whether their L1 was British or Malaysian English.

The implementation of background babble using adaptive staircase procedure improved speakers' /æ/ phonemic words production intelligibility to the Malaysian raters, while the implementation of background babble under a constant level improved the production intelligibility to the UK raters. The primary objective of SED training was to improve speakers' perceptual accuracy and production intelligibility when the communication is conducted in an unfamiliar English variety. Although the adaptive staircase procedure was shown to generate greater and more robust perceptual learning, constant SED was more relevant in meeting the objectives with its ability to improve speakers' production intelligibility among native speakers of the target language (i.e. native British English speakers).

To summarize, the advantage of training speakers at individualized level of 80% identification accuracy (Adaptive Staircase SED) lies on its effectiveness in improving speakers' general perceptual sensitivity under adverse conditions. This advantage would benefit a wide range of users as perceiving L2 speech in adverse situations is a common weakness among non-native speakers. On the other hand, the advantage of training speakers at a constant low babble level (also a relatively more quiet condition; Constant SED) lies on its effectiveness in directing speakers' attention to the relevant acoustical cues that were used by the native speakers to distinguish the phonemic contrasts.

The difference in evaluation performance between the raters who speak different L1 English variety also implied that they used different acoustical cues as the /ɛ/-/æ/ phonemic contrast identification criteria (see also Chapter 4 for their perceptual patterns on native British English speech words). It was still unclear which acoustical properties of the speech words were involved in the production changes.

Chapter 6: General Discussion and Conclusion

6.1 Training Effect of SED

6.1.1 Summary of the Training Experiments

SED training, a newly designed high variability phonetic training was investigated in this thesis. We investigated the impact of the level of background babble noise (constant vs adaptive staircase) and speaker variability to see what roles they play in improving phonemic sensitivity. Table 6.1 below summarized the main information and findings of SED trainings that had been conducted as part of the thesis.

6.1.2 The Advantages of Using the Adaptive Staircase Procedure in Improving Perceptual Performance of L2 Speakers under Adverse Conditions

SED training with the addition of constant level background babble successfully improved Malaysian speakers' identification accuracy of the British English /ɛ/-/æ/ phonemic words (Experiment 2). However, with the implementation of an adaptive staircase procedure, the effectiveness of SED training was further improved whereby the improvement magnitude and the generalization benefit towards the untrained phonemic contrast /t/-/d/ final was enlarged (Experiment 3). The training effect of SED was once again shown when the control group did not show any significant changes in their identification performance, even though they were also exposed to the British English speech stimuli during pre- and post-tests (Experiment 6). The training groups were exposed to the speech materials 60 minutes longer than the control group. One may argue that the difference between the training groups and control group in post-test may be attributed to the length of speech exposure rather than the presence of training.

Table 6.1 Summary of SED training experiments reported in this thesis, detailing participants' first language (L1), SED babble design, number of speaker used for the training procedure, phonemic contrast trained, improvement size (in percentage) for each phonemic category involved, generalization of training effects across novel/untrained stimuli and to production, and the general findings.

Experiment	**L1 of Participants**	**Babble Design**	**No. of Speaker**	**Phonemic Contrast**	**Improvement Size (%)**		**Generalization of Training Effect**		**Findings**
					/t/-/d/ Final	**/ɛ/-/æ/**	**Across Novel Stimuli**	**To Production**	
Experiment 1	M. English Mandarin	Constant	1	/t/-/d/ Final	5.35	4.35**	N/A	N/A	L1 ME speakers improved on /t/-/d/ final identification in post-test
Experiment 1	M. English Mandarin	Adaptive Staircase	1	/t/-/d/ Final	0	3**	N/A	N/A	L1 Man speakers showed no improvement on /t/-/d/ final identification
Experiment 1	M. English Mandarin	Constant	3	/t/-/d/ Final	2.68	3.57**	N/A	N/A	
Experiment 1	M. English Mandarin	Adaptive Staircase	3	/t/-/d/ Final	-3.79	4.46**	N/A	N/A	
Experiment 2	Mandarin	Constant (Low Level)	3	/ɛ/-/æ/	1.30	4.10*	Yes	N/A	The babble level (10% vs. 50%) did not affect the training effect.
Experiment 2	Mandarin	Constant (High Level)	3	/ɛ/-/æ/	0.50	6.10*	Yes	N/A	

Experiment 3	M. English Mandarin Malay	Adaptive Staircase	3	/ε/-/æ/	5.93*	7.77*	Yes	N/A	All L1 language groups improved in post-test. L1 Malay and Mandarin speakers improved significantly more than the L1 M. English speakers for the /ε/-/æ/ contrast. Training effect size of the /t/-/d/ final contrast was similar as the /ε/-/æ/ contrast.
Experiment 4	Mandarin	Constant	3	/ε/-/æ/	1.90	4.90*	No	N/A	A tendency for generalization benefit of Adaptive Staircase SED towards the untrained stimuli (/t/-/d/ final and /ε/-/æ/) to retain in six months post-test; but not the generalization benefit of Constant SED.
Experiment 4	Mandarin	Adaptive Staircase	3	/ε/-/æ/	5.40*	10.80**	Yes	N/A	
Experiment 6	Mandarin	Constant	3	/ε/-/æ/	3.35*	5.13	Yes	Yes (only for /æ/ phonemic words)	Perceptual training effect of Constant SED on the /ε/-/æ/ phonemic contrast was not significant when analyzed separately.
Experiment 6	Mandarin	Adaptive Staircase	3	/ε/-/æ/	5.2*	8.13*	Yes	No	Only Constant SED improved participants' production intelligibility of the /æ/ phonemic words among the native British English speakers.
Experiment 6 (Control Group)	Mandarin	N/A	N/A	N/A	-1.78	2.58	N/A	No	

Notes:

1) Improvement Size was calculated by finding the difference between pre-test and post-test identification accuracy. Negative values referred to deterioration in performance at post-test.

2) Improvement size marked with * was statistically significant with p-value < 0.05; ** refers to p-value < 0.001.

A related study by McCandliss et al. (2002) did find that mere exposure to the speech materials was sufficient to induce a small perceptual improvement. In fact, participants from Experiment 1 also showed perceptual learning for the /ɛ/-/æ/ phonemic words even though they were merely exposed to the speech tokens during the pre- and post-test. If mere exposure is sufficient to induce perceptual learning, the necessity of addition training would be questioned. However, other studies (e.g., Bradlow et al., 1997; J. Werker & Tees, 1984) demonstrated that exposure alone (from the testing procedure) was insufficient to induce perceptual changes in the control groups. These studies implemented a similar procedure as ours in which the control group did not receive any extra e training sessions. The results from Experiment 1 (see section 6.1.4 for further discussion) also supported that mere auditory exposure may not be sufficient to generate the perceptual improvements. Participants were trained on the British English /t/-/d/ final phonemic words for the same training period as the other training experiments, but they did not show any training benefits.

Importantly, the effect of feedback given throughout the training was shown to be critical in inducing the robust perceptual training effect. Both experimental and control groups in an early study by Flege (1989) heard the same amount of speech tokens from the experimental procedure. However, the experimental group who received trial-by-trial feedback (training) showed a higher identification rate (an increment of 9% increase) on the speech stimuli in post-test while the control group showed little or almost no improvement (an increment of 1%) when no feedback was given (no training). He pointed out that it was not the amount of exposure but the design of the training that determined the effectiveness of the phonetic training. The majority of the HVPT studies provided feedback during the phonetic training because feedback is believed to guide perceptual improvement. The effectiveness of SED training may be partly attributed to the effects of feedback as well. A more recent study by Perrachione et al. (2011) also showed that multiplying the amount of training or exposure per training session would accelerate the learning but not induced greater improvement size.

The advantages of a phonetic training paradigm lies on its effectiveness in improving one's perceptual performance beyond one can achieve from ordinary exposure to the speech materials (e.g. the control group or daily exposure through media play and education material in language classes). The perceptual improvement generated by Adaptive Staircase SED was significantly greater than the Constant SED (Experiment 3 vs. 2) and the control group (Experiment 6). Although the exposure effect could account for some of the perceptual improvement, the effectiveness of the Adaptive Staircase SED should not be overlooked. Furthermore, the adaptive staircase procedure also improved the flexibility of SED training through individualizing the difficulty level of the training based on each participant's performance. With this feature, more participants with varying levels of proficiency could benefit from phonetic training.

6.1.3 Effects of SED Training on Speech Production (Experiment 7)

Improvements in L1 Mandarin speakers' /ɛ/-/æ/ and /t/-/d/ spoken utterances as a consequence of SED perceptual training was examined using identification and rating tasks. Both native British English and L1 ME speakers rated the speech output (Experiment 7). The L1 Mandarin speakers started off with lower production intelligibility (for the /t/-/d/ final and /ɛ/-/æ/ phonemic words) compared to the speech production of a British English speaker (see Experiment 5).

The effectiveness of SED training in improving speech production intelligibility differed according to the background babble design. According to British English speakers, the Constant SED improved L1 Mandarin speakers' production intelligibility of the /æ/ phonemic words. The accomplishment can only be achieved when the speakers learnt the criterial acoustical characteristics that set the individual categories apart from the others. The phonological representation of the /æ/ phoneme might also become more robust compared to its initial ambiguous representation which is often confused with the /ɛ/ phonemic category.

Inconsistent with other HVPT studies (e.g. Bradlow et al., 1999; 1997; Hazan et al. 2005), the adaptive staircase SED training program did not change the speakers' production intelligibility according to the native British English speakers. Assuming perception and production mechanisms are closely related to each other and both mechanisms share the same phonological resources, if the perceptual improvement observed after the training was initiated by the reformation of phonological representation, the production performance should be affected along with the perceptual changes. In a related study by Brosseau-Lapre et al. (2011), the perceptual training effect also failed to generalize to the speech production. Further acoustical analyses revealed attempts to differentiate the vowel productions, although the changes were not intelligible to the native speakers. It is still too soon to conclude that perceptual training failed to modify the deep-rooted phonological representation. Perhaps more articulatory guidance and production monitoring are required to instigate meaningful changes in speech production as the processes involved in speech production (e.g. formation of the articulators and vibration of the vocal cords) are more complex than speech perception.

6.1.4 Reformation of Underlying Phonological Representations or Better Skills in Attention Shifting?

SED training improved L2 speech perception under adverse conditions – a situation that resembles real-world communication settings. It was, however, uncertain whether the perceptual changes were stimulated by the reformation of underlying phonological representations or the refined skills in attention shifting. Bradlow et al. (1999; 1997) claimed that HVPT modifies the underlying representations because the perceptual improvement was followed by immediate production improvement. According to them, the participants were not given any chance to rehearse their production through output monitoring method (i.e. using acoustical information of own overt production as a feedback mechanism to guide production changes). Thus, the production improvement must be a consequence of the modified underlying phonological representations. However, Iverson et al. (2012) argued that it could be the categorization strategy that the training is changing, not necessarily the representations themselves. Iverson et al. (2012) found inconsistent generalization patterns from perception to production across different phonemic categories involved. In another words, the perceptual improvement achieved in their study did not guarantee a corresponding production improvement, which was similar to our findings reported in Experiment 7. To account for the inconsistent generalization pattern, Iverson suggested that the perception and production modalities may operate on different resources (i.e. the underlying phonological representation). This is not consistent with the general duality theory of linguistic processes that proposes an inseparable relationship between the two modalities (e.g. Liberman's Motor Speech Theory and Levelt's Theory of Production).

On the other hand, other researchers have attributed the effectiveness of phonetic training to its ability to change listeners' perceptual strategy to resemble those of the native speakers. For instance, Logan et al. (1991) claimed that HVPT re-allocated participants' attention to the criterial acoustic cues across a wide range of phonetic sounds that belong to one phonemic category; while Giannakopoulou et al. (2013) believed the HVPT improved participants' identification accuracy by altering their perceptual cue weighting. The L1 Mandarin speakers began with native-like ability to discriminate the /t/-/d/ phonemic contrast at word initial position, which might have suggested that their ability to discriminate the aspirated-unaspirated plosive contrast /t/-/d/ at word initial position might not be very much different than that of native speakers of English (Avery & Idsardi, 2001). The speakers' discrimination and identification performance deteriorated when the contrast was located at word final position, but improved after SED training (see Experiment 3 and Experiment 6). However, their production intelligibility of the /t/-/d/ final phonemic words did not improve correspondingly (Experiment 7).

We first assumed that the underlying phonological representation has to change before production changes can take place. Based on this assumption and the limited production improvement observed from training (i.e. production intelligibility on the /t/-/d/ final phonemic words), we assumed the phonological representation did not change as a consequence of the SED training. If the discrimination ability of the L1 Mandarin speakers remained intact for the general voiceless-voiced plosives /t/-/d/ (or better known as the aspirated-unaspirated contrast in the word initial context), the performance difference between them and the British English speakers should be attributed to their different perceptual strategy (e.g. utilizing acoustical information from preceding vowel as secondary cues; see section 6.2.4 for further discussion). Therefore, the perceptual improvement observed after SED training is more likely to be induced by the improved perceptual strategy (i.e. ability to inhibit the interference of background noise and in attending to the relevant acoustical cues that assist the identification process).

To summarize, training speakers who received an individualized level of babble noise (i.e. Adaptive Staircase SED) turned out to be more effective in diverting speakers' attention away from the noise (Experiment 3), but not in attending the criterial acoustical sounds that are attended by the native speakers (Chapter 6; their production intelligibility improved for the Malaysian English speakers but not for the British English speakers). Inversely, training speakers under relatively less adverse conditions (i.e. Constant SED) like most of the previous training studies (e.g. Lively et al., 1993;1994; Bradlow et al., 1999;1997) was less effective in improving speakers' ability to ignore the background noise (smaller and less robust perceptual improvement compared to Adaptive Staircase SED; see Experiment 2 and Experiment 6), but more effective in directing speakers' attention to the criterial acoustical cues that the British English speakers are attending (Chapter 6; production intelligibility improved according to the British English speakers).

6.1.5 Understanding the Fundamental Perceptual Difficulty Experienced by the L2 Speakers

We attributed the lack of a training benefit in Experiment 1 to the use of inappropriate training materials. Flege (1989) pointed out that L1 Mandarin speakers' difficulty in discriminating the /t/-/d/ phonemic contrast at word final position might be accounted by their lack of use of the preceding vowel acoustical cues (a strategy adopted by the native speakers), rather than the loss of sensitivity to discriminate the contrast. Training participants to discriminate the plosives contrast (i.e. /t/-/d/ at word final position) in Experiment 1 was not the most effective as the fundamental weakness of the L1 Mandarin speakers turned out to be the preceding medial vowels perception, not the plosives contrast. On top of that, the final release burst /t/ and closure voiced /d/ is often dropped in daily-life conversational speech. Improving perceptual sensitivity to discriminate the final burst and voiced closure at word final position might not do much to help in daily-life speech perception.

Identification improvement on the /t/-/d/ final contrasts was found after speakers were trained with /ɛ/-/æ/ phonemic words (Experiment 3 and Chapter 6). As the speakers were not trained to perceive the closure phonetic sounds, the improvement observed was suspected to be the product of improved sensitivity (or attention shifted) towards the medial vowels, rather than the word final consonants.

Before considering the training material, the primary reason for the discrimination difficulty needs to be identified. In this case, the L1 Mandarin speakers should learn to allocate attention to the medial vowels and not to the voiced and voiceless plosives, in order to cue accurate voicing judgments. In order to promote native-like performance, researchers should first understand the perception strategy difference between the native and non-native speakers, before identifying the most ideal training material. Acoustical analysis can be first conducted to compare speech production of the native and non-native speakers. Any discrepancy in acoustical signals could mean a different linguistic representation in the non-native speakers and actions can be taken to modify the non-native speakers' perception and production performance to the pattern similar as the native speakers. It is also important for researchers to be aware of the pronunciation habits in spoken language, besides mastering the phonological system of the target language. Taking the /t/-/d/ final phonemic contrast as an example, knowing that the voiced and voiceless plosive contrast is often dropped and unavailable in everyday speech would imply the L1 Mandarin speakers' fundamental perception weakness; which is utilizing acoustical cues from the preceding vowel.

6.2 Comparing SED Training Effect across Other Phonetic Training Literature

6.2.1 Comparisons across Training Studies

With the widely varied training paradigms used in the literature, it would be interesting to identify the most effective training method. Nevertheless, the variability of training from the literature makes comparisons across the studies difficult. The majority of phonetic training studies involved a large number of training sessions across a number of weeks and was relatively more time consuming (see Table 1.1 and Table 1.2 in Chapter 1). In comparison, SED training from the present study is more time efficient, with training spread across 3 sessions over 3 consecutive days that can be completed in an hour.

In order to make the studies comparable, we normalized the training effect size reported by the studies and the normalized scores are shown in Figure 6.1. Note that only a few studies were included in the comparison because information provided by other studies was insufficient for the effect size calculation. We first calculated the training effect size (Cohen's *d*) for each reviewed study using an online effect size calculator (Becker, 1999). The mean and standard deviation values in pre- and post-tests reported in the papers were used to compute the effect size; except for the study by Perrachione et al. (2011) in which the effect size, Cohen's *d*s were already provided in the paper. We then divided the Cohen's *d* values by the number of training hours. Only the greatest effect size reported by each study was included in the comparison; unless the findings were meaningful enough to be reported (e.g., two different training paradigms for differently proficient language speakers in Perrachione et al., 2011).

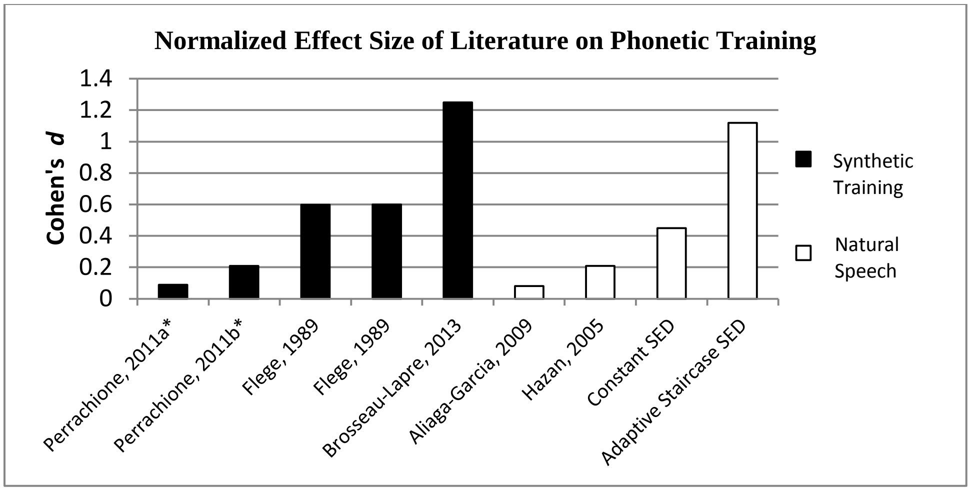


Figure 6.1 Literature arranged according to the type of training materials (black bars on the left – synthetic stimuli, white bars on the right– natural speech stimuli) and their normalized effect size scores.

The effect size of SED is represented by data from Experiment 2 (Constant SED) and 3 (Adaptive Staircase SED) because the effect sizes generated were the largest. The values were calculated using the identification accuracy of L1 Mandarin speakers only on the /ɛ/-/æ/ phonemic contrast. The performance of L1 Malay and L1 ME speakers were not included in the comparison because they performed relatively better in the pre-test (L1 Malay 81%, L1 English 82%) which in turn allowed a smaller space for improvement. The L1 Mandarin speakers who had the greatest difficulty on the /ɛ/-/æ/ phonemic contrast in pre-test (71%) generated the greatest training effect size in post-test. Furthermore, the L1 Malay and L1 ME speakers also did not participate in the Constant SED.

Training using synthetic stimuli generally produced a larger effect size but with limited generalizability. For instance, McCandliss et al. (2002) only reported generalization of training effect to another untrained synthetic continuum of the same /r/-/l/ phonemic contrast (e.g. trained on road-load, generalized to rock-lock contrast); particularly in the experimental group who received feedback during training. In fact, the large normalized effect size of McCandliss et al.'s study might be partly attributed to the adaptive staircase procedure applied to adjust stimuli difficulty level during training. Likewise, although the normalized effect size of Brosseau-Lapre et al.'s (2011) study is also relatively greater than SED, their training effect generalized only to the novel words, but not to novel speakers (see Figure 6.1 and section 1.3.3.6 for more information of the study).

From Figure 6.1, the normalized score of the Constant SED appears to be more comparable to other available studies while the normalized score of Adaptive Staircase SED appears to be relatively greater than other natural speech training studies. Aliaga-Garcia and Mora (2009) trained their bilingual Catalan or Spanish participants on the English oral stops (/p/, /t/,/ l/, /b/, /d/, /g/) and vowels (/æ/-/ʌ/-/ ɑː/, /uː/-/ʊ/, /iː/-/ɪ/), with various perception and production tasks preceded by an introductory theoretical session consisting of description on articulatory movements, exposure to native speaker's model and contrastive analysis on the training materials. Their perception training tasks were more extensive and included identification and discrimination tasks, phonetic transcription, exposure to native speakers' productions and others. Although the improvement size reported was greater than Adaptive Staircase SED's (10% from their training vs. 8% from Experiment 3), their training effect normalized score was reduced by the long training duration (six sessions of two hours training) due to the variety of training tasks imposed. It was however important to note that their training focused on more contrasts compared to the present studies. The duration of time spent to train each contrast was not specified.

A related study by Hazan et al. (2005, see Figure 6.1) also used a variant of HVPT similar to SED. The main features they investigated were, however, different. Hazan and colleagues (2005) investigated the effect of presenting visual cues (speakers' face producing the auditory stimuli) along with the spoken stimuli in phonetic training. Participants who were trained with both auditory and visual stimuli benefited more from the training with an average improvement of 17% (pre-test 60%, post-test 78%) in identification accuracy. The Japanese speakers started off with lower pre-test identification accuracy compared to the present study, thus allowing a greater improvement space. Nevertheless, their training effect size was still reduced by the length of their training duration.

Although the training study by Wang and Munro (2004) was not included in the comparison figure due to insufficient information for effect size calculation, it was particularly relevant to our SED training. Their training paradigm and training population were highly similar to ours. They trained L1 Mandarin or Cantonese speakers who spoke advanced L2 English using a similar concept of self-paced, computer-based perceptual training. The participants were trained on the English /iː/-/ɪ/, /ɛ/-/æ/ and /ʊː/-/u/ vowel pairs, using identification task with a perceptual fading method. Participants began training with very distinct exemplars of training stimuli before gradually proceeding with less acoustically distinct exemplars. The training progressed from low variability training with synthetic stimuli, to high variability training with natural stimuli produced by four speakers. After an intensive and lengthy training that took up to 24 hours over 24 sessions, a great improvement in identification accuracy was observed (about 20%; pre-test 69% to post-test 90%) that was retained for 3 months (an improvement of 16% compared to pre-test). The training effect was also found to generalize to novel words produced by novel speakers. The factors that contributed to their success might be their use of both synthetic and natural speech stimuli and the intensive training participants undertook. Nevertheless, the amount of time and training sessions involved in the training can be a big drawback. The training duration on each contrast (24 hours averaged by 3 phonemic contrasts = 7 hours) still exceeds the 1 hour training duration of SED.

When the period of training was taken into account, Adaptive Staircase SED from the present study was relatively more time effective in generating observable perceptual improvement (see Figure 6.1). The inclusion of background babble during training could have encouraged participants to inhibit irrelevant noises and re-allocate attention to the relevant acoustic cues. However, merely increased the difficulty of training task in Experiment 2 did not improve the training effect. It was the individualized training feature introduced by the adaptive staircase procedure which improved participants' perceptual performance at a greater extends.

Consistent with previous literature, the High Variability Phonetic Training (HVPT) approach adopted by SED might also partly contribute to its effectiveness (e.g., Lively et al., 1993, 1994; Bradlow et al., 1999, 1997; Callan et al., 2003; Hazan et al., 2005; Iverson et al., 2005). SED was designed by maximizing the variability of training features that can be categorized into 3 domains: 1) number of speakers, 2) gender of speakers and lastly 3) variety of English heard in the training. The high variability training materials might also encouraged training generalizability, similar to other HVPT studies.

6.2.2 SED vs. Existing Training Paradigms

What set Adaptive Staircase SED apart from other existing training paradigms is the use of background babble and the adaptive staircase procedure. Daily communication is often conducted in adverse conditions and L2 speakers' disadvantageous in speech perception manifests when the listening conditions are not in absolute quiet (Lecumberri et al., 2010). When compared with L1 speakers, L2 speakers tend to be more affected by adverse environments, especially when conversing in a noisy public area or receiving an unclear speech signal after being transmitted through electronic devices (e.g., mobile phones; Lecumberri et al., 2010). Training L2 speakers in adverse conditions (e.g. background babble used in the present study) that resembles more of the real world discourse context might be more relevant to improve L2 speakers' real life speech perception performance.

Constant SED was more similar to the existing HVPT paradigm but with the addition of background babble. The Adaptive Staircase SED on the other hand, had the background babble implemented with the adaptive staircase procedure using a three-down, one-up algorithm. The procedure measured participants' detection threshold by manipulating the volume of background babble where participants can achieve an average of 79% accuracy (Leek, 2001). None of the natural speech phonetic training studies showed in Table 1.2 implemented an adaptive staircase procedure in their training. It was important to know that the procedure requires systematically spaced stimuli in terms of their acoustical qualities. However, manipulating acoustical properties of natural speech tokens would reduce the naturalness and variability of the training materials. Therefore, the adaptive staircase procedure in SED was used to adjust the volume of background babble rather than the natural speech stimuli.

Unlike most training studies that used an ID task for training (all studies from Table 1), SED training uses a DT that is more commonly used by earlier studies that trained on synthetic stimuli (e.g., Strange & Dittman, 1984; Kraus et al., 1995). Kraus et al. (1995) trained participants to discriminate between two similar-sounding synthetic speech stimuli ($/da/_1$ - $/da/_2$) using DT to examine the effect of DT on the neurophysiology of the central auditory system. In their study, DT resulted in significant improvement in discrimination of the two speech stimuli and notable change in the mismatch negativity (MMN) event-related potential that signified learning or auditory experience. Similar to Kraus et al.'s (1995) objective, we aimed to improve participants' L2 perceptual performance from this low-level processing aspect, which we believe is the fundamental limitation of L2 speakers compared to L1 speakers. Flege's Speech Learning Model hypothesizes that the more dissimilar speech sounds are, the higher chance they would be encoded into two distinctive phonological categories and identified as distinctive phonemic sounds. By improving participants' sensitivity towards meaningful cues in L2 speech signals that distinguish the phonetic sounds, we expect the discrimination improvement to facilitate the formation of the higher-level linguistic representations of L2, which was assessed using a word ID task in pre- and post-tests.

The earlier studies preferred the ID task over the DT (e.g. Jamieson & Morosan, 1986; Logan et al., 1991) because the DT was claimed to increase intra-phonemic sensitivity rather than refining the boundaries of individual phonemic category. A more recent study by Handley et al. (2009) however, supported the use of either task, given that the tasks are able to direct participants' attention to the right acoustical cues. Handley and colleagues compared the effectiveness of ID task and DT used HVPT paradigm, using the /r/-/l/ phonemic contrast as training materials. Although the identification of the /r/-/l/ contrast did not improve equally across all phonetic contexts, they provided preliminary evidence to support the use of DT in perceptual phonetic training, as the task was showed to be as effective as the ID task.

6.3 Theoretical Implications

According to the Best Perceptual Assimilation Model, L2 speakers have difficulty discriminating similar sounding L2 phonetic distinctions especially when the distinction does not exist in the speakers' L1. The single category assimilation was observed whereby the L1 Mandarin speakers assimilated the British English /ɛ/-/æ/ phonemic sounds into a single category, due to the absence of the categorical distinction in the local Malaysian English (Pillai et al., 2010). The low front vowel /æ/ was later identified as the more operative vowel representation as the two phonetic sounds were matched to the /æ/ phonemic words at a higher rate (see section 4.4.1 from Experiment 4 for further discussion; Yap, Wong & Aziz, 2010).

The SED training conducted aimed to improve non-native British English speakers' realizations of the two phonetically distinct sounds. According to Flege's Speech learning Model, the phonetic salience of the vowel pair was essential and speakers' perceptual sensitivity towards the phonetic distinction needs to be improved before distinct phonetic category can be established (Flege, 1995). In order to improve participants' discrimination sensitivity of the similar phonetic sounds, discrimination training was conducted using SED and it led to improvement in participants' post-test identification accuracy. The production intelligibility improvement of the /æ/ phonemic words after Constant SED training (Experiment 7) further suggests that a new phonetic representation may be formed for the British English /æ/ phoneme.

The Native Language Magnet Model by Kuhl (2008) proposed the existence of perceptual magnet effect whereby the phonetic prototypes that located at the center of the L1 phonetic categories tend to assimilate neighboring phonetic sounds like a magnet (see section 1.3.2.3 for further details). When the operative phonetic prototypes during L2 speech perception originated from the speakers' L1, the L2 similar sounding phonetic sounds are likely to be assimilated to the L1 categories and became less discriminable. If a new L2 phonetic category was formed (i.e. /æ/) as a consequence of the SED training, the new categorical prototype would contain acoustical characteristics learnt from the training speech material. The newly formed prototype will then works as the "new magnet" and assist in identifying the future British English /æ/ instances. Consequently, identification performance of speakers becomes more accurate.

The interference of L1 language system can be both beneficial and problematic at the same time (Bohn & Best, 2012; Kuhl et al., 2008). On a general scale, the L1 phonetic system intervenes especially when there is ambiguity during L2 processing (e.g. disrupted signal perceived due to noisy environment; see Lecumberri et al., 2010). SED training successfully trained listeners to perceive in a more native-like manner by directing them to the criterial acoustical cues the native speakers are attending. When trained participants learn how to acquire sufficient acoustical information to cue accurate judgments, the reliance and interference of L1 can be reduced.

6.4 Practical Implications

6.4.1 Application of Technology in Language Learning

With the advancement of technology, the use of computer has been brought into language learning. Availability of computer assisted language learning (CALL) can overcome many shortcomings of traditional classroom language learning. Some known benefits of computer technology in language learning include: 1) higher motivation for self-learning, 2) flexibility of self-learning in terms of timing and frequency, 3) independent and interactive learning, 4) flexibility in choosing study materials, 5) immediate individualized automatic feedback, 6) availability of lesson repetition as necessary, 7) maximized learning outcome with individualized procedure and lastly 8) enhanced learning achievements (Elimat & Abuseileek, 2014; Warschauer & Meskill, 2010).

Numerous phonetic training studies have already shown the effectiveness of CALL in improving perception and production of an individual's L2 segmental speech contrasts (e.g. Wang & Munro, 2004). The flexible and easily accessible training could be extended and used in academic context such as classrooms and schools.

6.4.2 English as the Lingua Franca

As a social phenomenon, English is often regarded as one language, with the British English and American English being widely favored and recognized as the ideal prestige models. However, adopting the prestige model of English pronunciation can be difficult for non-native speakers due to a few reasons. Firstly, exposure to the "accurate English pronunciation" is often limited within the classroom and the exposure to Englishes spoken in the living environment is often too diverse. In addition, English teachers in non-native speaking countries are often non-native British or American English speakers themselves. Even though these prestige models were often used to teach and assess the learning of English language, there are also some biological (e.g. neural commitment to L1 and lateralization of brain process) and environmental factors (e.g. quality of language input) that the non-native English speakers cannot overcome to adopt the "native pronunciation" (see introduction section 1.2 and 1.3 for more discussion).

English as the global lingua franca[13] (ELF) serves as a common means of communication in many international academic and business settings (Mauranen, Hynninen, & Ranta, 2010), such as the academic conferences and business meetings that involved attendants from various countries who may or may not be native English speakers.

Jenkins, an English as Lingua Franca (ELF) researcher has questioned the viability of the "Native Speaker English" ideology especially when the world and the usage of English has become so interconnected (Jenkins, 2009; Jenkins et al., 2011; Jenkins & Gobbo, 2010). She believes the English language should be *fluid, flexible, contingent, hybrid and deeply intercultural* and all varieties of English worldwide that are relatively well established should be acknowledged (Jenkins et al., 2011). Similar to other ELF researchers who believe in the pluralism of English variety, Jenkins referred all English varieties as World Englishes. Some of the well-established World Englishes include those spoken in the post-colonial nations (e.g. Indian English, Nigerian English and Singapore English), which exist as a product of the English education introduced by the European colonizers. One interesting difference between the many Englishes spoken around the world lies in the great diversity in the spoken phonologies (Carr & Honeybone, 2007).

Generally, ELF users are viewed as highly skilled communicators who prioritize successful communication over the "correctness" of the language spoken (Jenkins et al., 2011). One common attempt made by the ELF speakers includes their continuous effort to accommodate to each other despite their own local English variety (e.g. substituting cultural relevant phrases for

[13] The term lingua franca refers to a common choice of language used for communication among speakers who come from different linguacultural backgrounds.

mutual understanding or adapting own pronunciation habit) for the benefit of interlocutors involved. A flexible training program that could assist in the adaption effort within a very short amount of time would be very beneficial for the ELF users.

6.4.3 Globalization of English Education and Applications of SED Training in Education Institution

Due to the globalization of education, universities become globally intertwined at a fast pace with growing population of international students and staff from all around world (Jenkins et al., 2011). English as the main medium is used to teach academic courses and share literature sources (e.g. research publications) in the international universities. Interaction between academic members who speak non-native English or World Englishes has also become inevitable. The diversity of the many Englishes spoken in the academic discourse context is rather unpredictable and it affects the effectiveness of the communication process between the message transmitter (e.g. lecturers) and the receivers (e.g. students). In such a context, sensitivity in listening becomes essentially important to promote mutual understanding between discourse parties. Humphreys et al. (2012) investigated the role of tertiary international students' English proficiency on their academic performance (as measured in Grade Point Average score, GPA). They found that students' English listening and reading ability can affect their first year academic achievements, but their speaking and writing fluency cannot. English listening proficiency affects international students to understand lectures at the beginning of their enrollment. Therefore, the importance of English language perception of tertiary students should not be underestimated.

SED training was designed with the primary aim to improve conversation intelligibility between speakers who speak various Englishes with different levels of proficiency. In occasions where individuals often need to travel and communicate with people of different cultures and language backgrounds who speak different variants of English (usually for education and business purposes), Adaptive Staircase SED may be beneficial to improve one's perceptual sensitivity towards the target language and promote better conversation quality.

Previous research (except Wang & Munro, 2004) has demonstrated the effectiveness of phonetic training using rigid research paradigms in laboratory settings. As a computer-based program, we aim to introduce SED training in a pedagogically oriented approach, in which learners can choose their own training materials, and go through the training at a self-determined pace. These features improve the feasibility of Adaptive Staircase SED and its suitability to be used by the wider public. Other than the greater training effect size, another advantage of Adaptive Staircase SED is its adaptability to speakers with different proficiency levels. The adaptive staircase procedure adjusts the difficulty level of the phonetic training according to every individual's performance standard. Even though participants from Experiment 3 speak different levels of Malaysian English, SED benefited all; even to highly proficient L1 ME participants who already performed remarkably well in the pre-test (84%). The Adaptive Staircase SED would also be suitable for advanced language learners or for anyone who wish to get adapted to an unfamiliar English accent, due to its ability to improve perceptual sensitivity of each individual with the optimal difficulty level.

6.5 Suggestions for Future Studies

Experiments from this thesis posed some questions that can be answered with future acoustical analysis. For instance, the Adaptive Staircase SED and Constant SED changed different aspects of L1 Mandarin speakers' /ɛ/-/æ/ phonemic words production that appealed differently to the native British English raters and the L1 ME raters. Subsequent acoustical analysis would reveal the acoustical properties of initial pronunciations and the properties involved in the intelligibility discrepancy after training.

It would also be interesting to examine the potential neurological changes before and after SED training. Previous studies had shown that adaptive and enriched linguistic training could induce neural plasticity in L2 learning adulthood (Zhang et al., 2009; see section 1.3.3.4 for more discussion) and activate anatomical areas that are involved in processing the native language (Golestani & Zatorre, 2004). To our knowledge, none of the studies had shown neurological changes for learning to perceive an accented spoken language (i.e. unfamiliar British English). Future investigations with functional imaging techniques would reveal the effect of SED training on neurological functions. Referring to Golestani and Zatorre (2004), suppression of the angular gyri after training was at least expected as speakers from the present studies learnt to inhibit the irrelevant acoustical cues and "noise" from the perceived signals.

Before introducing SED training for pedagogical purposes, more testing is still required to examine how far the training could be extended. For instance, further research should be conducted to examine the efficacy of SED training in improving speech perception and production across different training materials and of speakers who come from different language backgrounds.

The English tense-lax vowel contrast is another well-known difficult pair for the L2 English speakers due to the absence of tense-lax distinction in many language systems (see section 4.1.2 for more details). There are quite a number of tense-lax phonemic pairs in the English language, including: /iː/-/ɪ/ (e.g. sh*ee*p-sh*i*p) , /uː/-/ʊ/ (e.g. p*oo*l-p*u*ll,) and /ɔː/- /ɑ/ (e.g. c*a*ller-c*o*llar). See Appendix 7 for the list of minimal pairs contrasting the English tense-lax vowels.

Another primary purpose of SED training is to help L2 speakers in learning other foreign languages. Examining phonetic training effects on other contrasts of less common languages (e.g. Malay) would reveal how wide the training effect can be extended to facilitate general L2 learning. The most ideal training paradigm should benefit speakers from a wide variety of language backgrounds in perceiving all varieties of languages.

It is still questionable whether training gain in discrimination of speech sounds can improve higher-order listening and speaking skills. Future SED training studies can adopt other testing batteries to examine the potential effects of phonetic training on perception and production of sentences or conversation.

6.6 Conclusion

Laboratory perceptual training is often too rigid and unapproachable to the general public of L2 users. The limitation of laboratory training can be resolved by incorporating perceptual training into CALL technology (i.e. technology-based program such as SED). The SED training studies have shown that interactive technology incorporating individualized training regimes can efficiently enhance spoken English discrimination and identification skills in adult multilingual speakers.

In conclusion, language learning can be accompanied by multiple difficulties. SED training, a flexible and easily-accessed computer or device-based software could come in handy for users who wish to improve their perceptual sensitivity towards the target language variety and further improve effectiveness of the conversation within a short training period.

References

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

Alladi, S., Bak, T. H., Duggirala, V., Surampudi, B., Shailaja, M., Shukla, A. K., … Kaul, S. (2013). Bilingualism delays age at onset of dementia, independent of education and immigration status. *Neurology, 81*(22), 1938–1944. http://doi.org/10.1212/01.wnl.0000436609.20587.65

Antoniou, M., Liang, E., Ettlinger, M., & Wong, P. C. M. (2014). The bilingual advantage in phonetic learning. *Bilingualism: Language and Cognition*, *18*(4), 683–695. http://doi.org/10.1017/S1366728914000777

Avery, P., & Idsardi, W. J. (2001). Laryngeal dimensions: Completion and enhancement. In T. A. Hall & U. Kleinhanz (Eds.), *Studies in DIstinctive Feature Theory* (Ed., pp. 41–70). Berlin: Mouton de Gruyter.

Baskaran, L. (1994). The Malaysian English mosaic. *English Today*, *10*, 27–32. http://doi.org/doi:10.1017/ S0266078400000857

Baskaran, L. M. (2005). *A Malaysian English primer: aspects of Malayisan English features*. Kuala Lumpur: University Malaya Press.

Becker, L. A. (1999). Effect Size Calculators. Retrieved from http://www.uccs.edu/lbecker/index.html

Bell-Berti, F., Raphael, L. J., Pisoni, D. B., & Sawusch, J. R. (1979). Some relationships between speech production and perception. *Phonetica*. http://doi.org/542499

Best, C., & McRoberts, G. W. (2003). Infant perception of non-native consonant contrasts that adults assimilate in different ways. *Language and Speech*, *46*(Pt 2-3), 183–216. http://doi.org/10.1177/00238309030460020701

Best, C. T. (1994). The emergence of native-language phonological influences in infants: A perceptual assimilation model. *The Development of Speech Perception: The Transition from Speech Sounds to Spoken Words*, *167*, 167–224.

Best, C. T. (1995). A direct realist perspective on cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (ed., pp. 171–204). Timonium: MD: York Press.

Best, C. T., McRoberts, G. W., & Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *The Journal of the Acoustical Society of America*, *109*(2), 775–794. http://doi.org/10.1121/1.1332378

Best, C. T., McRoberts, G. W., & Sithole, N. M. (1988). Examination of Perceptual Reorganization for Nonnative Speech Contrasts : Zulu Click Discrimination by English-Speaking Adults and Infants. *Journal of Experimental Psychology: Human Perception and Performance*, *14*(3), 345–360.

Bialystok, E. (1999). Cognitive complexity and attentional control in the bilingual mind. *Child Development*, *70*(3), 636–644. http://doi.org/10.1111/1467-8624.00046

Bialystok, E., & Martin, M. M. (2004). Attention and inhibition in bilingual children: Evidence from the dimensional change card sort task. *Developmental Science*, *7*(3), 325–339. http://doi.org/10.1111/j.1467-7687.2004.00351.x

Boersma, P., & Weenink, D. (2007). Praat: doing phonetics by computer (Version 4.5.)[Computer program]. *Retrieved from Http://www.praat.org/*, *5*(9/10), 341–345. Retrieved from papers3://publication/uuid/AF582E4D-2F7A-409E-B4F1-7A10385D9135

Bohn, O.-S., & Best, C. T. (2012). Native-language phonetic and phonological influences on perception of American English approximants by Danish and German listeners. *Journal of Phonetics*, *40*(1), 109–128. http://doi.org/10.1016/j.wocn.2011.08.002

Bosch, L., & Ramon-Casas, M. (2011). Variability in vowel production by bilingual speakers: Can input properties hinder the early stabilization of contrastive categories? *Journal of Phonetics*, *39*(4), 514–526. http://doi.org/10.1016/j.wocn.2011.02.001

Bradlow, A. R., Akahane-Yamada, R., Pisoni, D. B., & Tohkura, Y. (1999). Training Japanese listeners to identify English /r/ and /l/: long-term retention of learning in perception and production. *Perception & Psychophysics*, *61*(5), 977–985. http://doi.org/10.3758/BF03206911

Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., & Tohkura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *The Journal of the Acoustical Society of America*, *101*(4), 2299–310. http://doi.org/10.1121/1.418276

Brosseau-Lapré, F., Rvachew, S., Clayards, M., & Dickson, D. (2011). Stimulus variability and perceptual learning of nonnative vowel categories. *Applied Psycholinguistics*, *34*, 1–23. http://doi.org/10.1017/S0142716411000750

Brysbaert, M., New, B., & Keuleers, E. (2012). Adding part-of-speech information to the SUBTLEX-US word frequencies. *Behavior Research Methods*, (October), 991–997. http://doi.org/10.3758/s13428-012-0190-4

Butler, Y. G., & Hakuta, K. (2006). Bilingualism and Second Language Acquisition. *The Handbook of Bilingualism*, 114–144. http://doi.org/10.1002/9780470756997.ch5

Callan, D. E., Tajima, K., Callan, A. M., Kubo, R., Masaki, S., & Akahane-Yamada, R. (2003). Learning-induced neural plasticity associated with improved identification performance after training of a difficult second-language phonetic contrast. *NeuroImage*, *19*(1), 113–124. http://doi.org/10.1016/S1053-8119(03)00020-X

Carr, P., & Honeybone, P. (2007). English phonology and linguistic theory: an introduction to issues, and to "Issues in English Phonology." *Language Sciences*, *29*(2–3), 117–153. http://doi.org/10.1016/j.langsci.2006.12.018

Catford, J. C. (2002). *A practical introduction to phonetics* (2nd Ed.). Oxford: Oxford University Press.

Central Intelligence Agency. (2015). Malaysia. In The World Factbook. Retrieved April 22, 2015, from https://www.cia.gov/library/publications/the-world-factbook/geos/my.html

Chao, Y. (1968). *A grammar of spoken Chinese*. Berkeley and Los Angeles: University of California Press.

Costa, A., Hernández, M., Costa-Faidella, J., & Sebastián-Gallés, N. (2009). On the bilingual advantage in conflict processing: Now you see it, now you don't. *Cognition*, *113*(2), 135–149. http://doi.org/10.1016/j.cognition.2009.08.001

Crystal, D. (1997). *English as a Global Language* (Cambridge). Cambridge University Press.

Cutler, A., Garcia Lecumberri, M. L., & Cooke, M. (2008). Consonant identification in noise by native and non-native listeners: effects of local context. *The Journal of the Acoustical Society of America*, *124*(2), 1264–1268. http://doi.org/10.1121/1.2946707

Davenport, M., & Hannahs, S. J. (2011). *Introducing phonetics and phonology* (2nd ed). London: Arnold.

de Bruin, A., Treccani, B., & Della Sala, S. (2015). Cognitive advantage in bilingualism: an example of publication bias? *Psychological Science*, *26*(1), 99–107. http://doi.org/10.1177/0956797614557866

Diehl, R. L., Lotto, A. J., & Holt, L. L. (2004). Speech perception. *Annual Review Of Psychology*, *55*(55), 149–179. http://doi.org/10.1146/annurev.psych.55.090902.142028

Duanmu, S., & Keith, B. (2006). Chinese (Mandarin): Phonology. *Encyclopedia of Language & Linguistics*, (February), 351–355. Retrieved from http://www.sciencedirect.com/science/article/B7T84-4M3C3K0-341/2/715160f3e6de93b4ef86d3c8db7c2da7

Eimas, P. D. (1975). Speech perception in early infancy. In *Infant Perception: From Sensation to Cognition* (p. 193 228 ST-Speech perception in early infancy).

Eimas, P. D., Siqueland, E. R., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. *First Language Acquisition: The Essential Readings*, *171*(3968), 279–284. http://doi.org/10.1037/e509922011-001

Elimat, A. K., & Abuseileek, A. F. (2014). Automatic speech recognition technology as an effective means for teaching pronunciation, *10*(1), 21–47.

Flege, J. E. (1989). Chinese subjects' perception of the word-final English/t/-/d/ contrast: Performance before and after training. *Journal of Accoustical Society of America*, *86*(5), 1684–1697.

Flege, J. E. (1991). 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), 395–411. http://doi.org/10.1121/1.400473

Flege, J. E. (1995). Second Language Speech Learning: Theory, Findings, and Problems. *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, 233–277. http://doi.org/10.1111/j.1600-0404.1995.tb01710.x

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

http://doi.org/10.1006/jpho.1996.0040

Flege, J. E., & MacKay, I. R. A. (2011). What accounts for “age” effects on overall degree of foreign accent? *Achievements and Perspectives in the Acquisition of Second Language Speech: New Sounds 2010*, 65–82.

Flege, J. E., Munro, M. J., & Mackay, I. R. A. (1995). Effects of age of second-language learning on the production of English consonants. *Speech Communication*, *16*(1), 1–26. http://doi.org/10.1016/0167-6393(94)00044-B

Flege, J. E., Yeni-Komshian, G. H., & Liu, S. (1999). Age Constraints on Second-Language Acquisition. *Journal of Memory and Language*, *41*(1), 78–104. http://doi.org/10.1006/jmla.1999.2638

Gaudart, H. (1987). English Language Teaching In Malaysia: A Historical Account. *The English Teacher*, *XVI*, 1–17. Retrieved from www.melta.org.my/ET/1987/main2.html

Giannakopoulou, A., Uther, M., & Ylinen, S. (2013). Enhanced plasticity in spoken language acquisition for child learners: Evidence from phonetic training studies in child and adult learners of English. *Child Language Teaching and Therapy*, *29*(2), 201–218. http://doi.org/10.1177/0265659012467473

Golestani, N., Paus, T., & Zatorre, R. J. (2002). Anatomical correlates of learning novel speech sounds. *Neuron*, *35*(5), 997–1010. http://doi.org/10.1016/S0896-6273(02)00862-0

Golestani, N., & Zatorre, R. J. (2004). Learning new sounds of speech: Reallocation of neural substrates. *NeuroImage*, *21*(2), 494–506. http://doi.org/10.1016/j.neuroimage.2003.09.071

Guenther, F. H., & Gjaja, M. N. (1996). The perceptual magnet effect as an emergent property of neural map formation. *The Journal of the Acoustical Society of America*, *100*(2 Pt 1), 1111–1121. http://doi.org/10.1121/1.416296

Handley, Z., Sharples, M., & Moore, D. (2009). Training novel phonemic contrasts: a comparison of identification and oddity discrimination training. In *Speech and Language Technology in Education (SLaTE) 2009*. Wroxall, England.

Hanulikova, A., Dediu, D., Fang, Z., Bašnaková, J., & Huettig, F. (2012). Individual differences in the acquisition of a complex L2 phonology: a training study. *…Learning*, (September), 79–109. Retrieved from http://onlinelibrary.wiley.com/doi/10.1111/j.1467-9922.2012.00707.x/full

Hart, D. C. (1969). Some English pronunciation difficulties in Malaysia. *ELT Journal*, *23*(3), 270–273. http://doi.org/10.1093/elt/XXIII.3.270

Hassan, A. (2005). Vokal dan Konsonan Segmental Bahasa Melayu. In A. Hassan (Ed.), *Linguistik Am* (Eds., pp. 77–88). Malaysia: Professional Training Service, PTS.

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://doi.org/10.1016/j.specom.2005.04.007

Herd, W. (2011). *The perceptual and production training of /d, ɾ, r/ in L2 Spanish: behavioral, psycholinguistic and neurolinguistic evidence*. University of Kansas.

Howie, J. M. (1973). *Acoustical studies of Mandarin vowels and tones*. Cambridge: Cambridge University Press.

Huang, L. M. (1992). Remarks on the phonological structure of Mandarin Chinese. *Bulletin Nat 1 Taiwan Nor. Univ.*, *37*, 363–383.

Huang, T. (2010). Er-suffixation in Chinese monophthongs: phonological analysis and phonetic data. In L. E. Clemes & C. M. L. Liu (Eds.), *22nd North American Conference on Chinese Linguistics (NACCL-22)* (pp. 331–344). MA: Harvard University, Cambridge.

Huang, T., & Johnson, K. (2011). Language specificity in speech perception: Perception of Mandarin tones by native and nonnative listeners. *Phonetica*, *67*, 243–267.

Humphreys, P., Haugh, M., Fenton-Smith, B., Lobo, A., Michael, R., & Walkinshaw, I. (2012). Tracking international students' English proficiency over the first semester of undergraduate study. *IELTS Research Reports Online Series*, *1*, 1–41. Retrieved from www.ielts.org/researchers

IPA. (1999). *Handbook of the International Phonetic Association*. *A guide to the use of the International Phonetic Alphabet*. Cambridge: Cambridge University Press. Retrieved from http://www.worldcat.org/oclc/40305532

Iverson, P., Hazan, V., & Bannister, K. (2005). Phonetic training with acoustic cue manipulations: a comparison of methods for teaching English /r/-/l/ to Japanese adults. *The Journal of the Acoustical Society of America*, *118*(5), 3267–3278. http://doi.org/10.1121/1.2062307

Iverson, P., & Kuhl, P. K. (2000). Perceptual magnet and phoneme boundary effects in speech perception: do they arise from a common mechanism? *Perception & Psychophysics*, *62*(4), 874–886. http://doi.org/10.3758/BF03206929

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

Iverson, P., Pinet, M., & Evans, B. G. (2012). Auditory training for experienced and inexperienced second-language learners: Native French speakers learning English vowels. *Applied Psycholinguistics*, *33*(1), 145–160. http://doi.org/10.1017/S0142716411000300

Jamieson, D. G., & Morosan, D. E. (1986). Training non-native speech contrasts in adults: acquisition of the English /delta/-/theta/ contrast by francophones. *Perception & Psychophysics*, *40*(4), 205–215. http://doi.org/10.3758/BF03211500

Jamieson, D. G., & Morosan, D. E. (1989). Training non-native speech contrasts: A comparison of two techniques. *Canadian Journal of Psychology*, *43*(1), 88–96. http://doi.org/10.1037/h0084209

Jenkins, J. (2009). English as a lingua franca. *World Englishes*, *28*(2), 200–207. http://doi.org/10.1093/elt/cci064

Jenkins, J., Cogo, A., & Dewey, M. (2011). Review of developments in research into English as a lingua franca. *Language Teaching*, *44*(3), 281–315. http://doi.org/10.1017/S0261444811000115

Jenkins, J., & Gobbo, F. (2010). English as a Lingua Franca: Attitude and Identity. *Language*

Problems & Language Planning, *34*(1), 81–83. http://doi.org/10.1075/lplp.34.1.05gob

Jusczyk, P. W., & Luce, P. A. (1994). Infants' Sensitivity to Phonotactic Patterns in the Native Language.pdf. *Journal of Memory and Language*, *33*, 630–645. http://doi.org/doi:10.1006/jmla.1994.1030

Kachru, B. B. (1985). Standards, codification and sociolinguistic realism: The English language in the outer circle. In *English in the world: Teaching and learning the language and literatures* (pp. 11–30).

Kachru, B. B. (1990). World Englishes and applied linguistics. *World Englishes*, *9*(1), 3–20. http://doi.org/10.1111/j.1467-971X.1990.tb00683.x

Karim, N. S., Onn, F. M., Musa, H. H., & Mahmood, A. H. (2015). *Tatabahasa Dewan* (Edisi Keti). Kuala Lumpur: Dewan Bahasa dan Pustaka.

Kondaurova, M. V, & Francis, A. L. (2008). The relationship between native allophonic experience with vowel duration and perception of the English tense/lax vowel contrast by Spanish and Russian listeners. *The Journal of the Acoustical Society of America*, *124*(6), 3959. http://doi.org/10.1121/1.2999341

Kraus, N., McGee, T., Carrell, T. D., King, C., Tremblay, K., & Nicol, T. (1995). Central auditory system plasticity associated with speech discrimination training. *Journal of Cognitive Neuroscience*, *7*(1), 25–32. http://doi.org/10.1162/jocn.1995.7.1.25

Krebs-lazendic, L., & Best, C. T. (2002). Early and Late Bilinguals ' Vowel Perception and Production : English Vowel Contrasts that Give Serbian-English Bilinguals a H (E) AD-ache.

Kuhl, P. K. (1991). Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics*, *50*(2), 93–107. http://doi.org/10.3758/BF03212211

Kuhl, P. K. (2000). Language, mind, and brain: Experience alters perception. In *The new cognitive neurosciences* (pp. 99–115). Retrieved from https://books.google.cl/books?hl=es&lr=&id=0wx17lC075EC&oi=fnd&pg=PA99&dq=experience+and+perception&ots=RnGGxvKXSx&sig=MBzsTZRU6eVjKo5XEV1ULCacu84

Kuhl, P. K. (2004). Early language acquisition: cracking the speech code. *Nature Reviews. Neuroscience*, *5*(11), 831–43. http://doi.org/10.1038/nrn1533

Kuhl, P. K. (2010). Brain Mechanisms in Early Language Acquisition. *Neuron*. http://doi.org/10.1016/j.neuron.2010.08.038

Kuhl, P. K. (2011). Brain Mechanisms Underlying the Critical Period for Language : Linking Theory and Practice. *Human Neuroplasticity and Education*.

Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., & Nelson, T. (2008). Phonetic learning as a pathway to language: new data and native language magnet theory expanded (NLM-e). *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1493), 979–1000. http://doi.org/10.1098/rstb.2007.2154

Kuhl, P. K., Conboy, B. T., Padden, D., Nelson, T., & Pruitt, J. (2005). Early Speech Perception and Later Language Development: Implications for the "Critical Period." *Language*

Learning and Development, *1*(3–4), 237–264. http://doi.org/10.1207/s15473341lld0103&4_2

Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science*, *9*(2). http://doi.org/10.1111/j.1467-7687.2006.00468.x

Kuhl, P. K., Tsao, F.-M., & Liu, H.-M. (2003). Foreign-language experience in infancy: effects of short-term exposure and social interaction on phonetic learning. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(15), 9096–101. http://doi.org/10.1073/pnas.1532872100

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic Experience Alters Phonetic Perception in Infants by 6 Months of Age. *Source: Science, New Series*, *255*(5044), 606–608. http://doi.org/10.1126/science.1736364

Lambacher, S. G., Martens, W. L., Kakehi, K., Marasinghe, C. A., & Molholt, G. (2005). The effects of identification training on the identification and production of American English vowels by native speakers of Japanese. *Applied Psycholinguistics*, *26*, 227–247. http://doi.org/10.1017.S0142716405050150

Lane, H., & Tranel, B. (1971). The Lombard sign and the role of hearing in speech. *Journal of Speech, Language and Hearing Research*, *14*, 677–709. http://doi.org/10.1044/jshr.1404.677

Lecumberri, M. L. G., & Cooke, M. (2006). Effect of masker type on native and non-native consonant perception in noise. *The Journal of the Acoustical Society of America*, *119*(4), 2445–2454. http://doi.org/10.1121/1.2180210

Lecumberri, M. L. G., Cooke, M., & Cutler, A. (2010). Non-native speech perception in adverse conditions: A review. *Speech Communication*, *52*(11–12), 864–886. http://doi.org/10.1016/j.specom.2010.08.014

Leek, M. R. (2001). Adaptive procedures in psychophysical research. *Perception & Psychophysics*, *63*(8), 1279–92. http://doi.org/10.3758/BF03194543

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: a quick and valid Lexical Test for Advanced Learners of English. *Behavior Research Methods*, *44*(2), 325–43. http://doi.org/10.3758/s13428-011-0146-0

Lengeris, A., & Hazan, V. (2010). The effect of native vowel processing ability and frequency discrimination acuity on the phonetic training of English vowels for native speakers of Greek. *The Journal of the Acoustical Society of America*, *128*(6), 3757–3768. http://doi.org/10.1121/1.3506351

Lenneberg, E. H. (1967). The biological foundations of language: A summary. *Hospital Practice*, *2*, 59–67 ST–The biological foundations of language.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

Levelt, W. J. M. (2001). Relations between speech production and speech perception: Some behavioral and neurological observations. *Language, Brain, and Cognitive Development: Essays in Honor of Jacques Mehler*, 241–256.

Li, H., Aljunied, M., & Teoh, B. S. (2005). A Grapheme to Phoneme Converter for Standard Malay. *Research Collection Lee Kong Chian School of Business (Open Access).*, Paper 2781. Retrieved from http://ink.library.smu.edu.sg/lkcsb_research/2781

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perceptions of the speech code.pdf. *Psychological Review*, *74*(6), 431–461. Retrieved from http://www.haskins.yale.edu/Reprints/HL0069.pdf

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, *21*(1), 1–36. http://doi.org/10.1016/0010-0277(85)90021-6

Lin, Y. H. (1989). The retroflex as a complex segment. In *ESCOL* (Vol. 6, pp. 182–193). ESCOL Publications Committee, Department of Linguistics, Ohio State University.

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*(3 Pt 1), 1242–1255. http://doi.org/10.1121/1.408177

Lively, S. E., Pisoni, D. B., Yamada, R. a, Tohkura, Y., & Yamada, T. (1994). Training Japanese listeners to identify English /r/ and /l/. III. Long-term retention of new phonetic categories. *The Journal of the Acoustical Society of America*, *96*(4), 2076–2087. http://doi.org/10.1121/1.410149

Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English /r/ and /l/: a first report. *The Journal of the Acoustical Society of America*, *89*(2), 874–886. http://doi.org/10.1121/1.1894649

Malaysia, M. of E. (2015). Indikator Pengajian Tinggi Malaysia 2013. Retrieved April 22, 2015, from http://www.mohe.gov.my/web_statistik/Indikator_Pengajian_Tinggi_Malaysia_2013.pdf

Martin, N., & Saffran, E. M. (2002). The relationship of input and output phonological processing: An evaluation of models and evidence to support them. *Aphasiology*, *16*(November 2013), 107–150. http://doi.org/10.1080/02687040143000447

Mauranen, A., Hynninen, N., & Ranta, E. (2010). English as an academic lingua franca: The ELFA project. *English for Specific Purposes*, *29*(3), 183–190. http://doi.org/10.1016/j.esp.2009.10.001

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 : JSLHR*, *40*(3), 686–693. http://doi.org/10.1044/jslhr.4003.686

McCandliss, B. D., Fiez, J. a, Protopapas, A., Conway, M., & McClelland, J. L. (2002). Success and failure in teaching the [r]-[l] contrast to Japanese adults: tests of a Hebbian model of plasticity and stabilization in spoken language perception. *Cognitive, Affective & Behavioral Neuroscience*, *2*(2), 89–108. http://doi.org/10.3758/CABN.2.2.89

McMahon, A. (2002). *An introduction to English phonology*. Oxford: Oxford University Press.

Moore, D. R., Rosenberg, J. F., & Coleman, J. S. (2005). Discrimination training of phonemic contrasts enhances phonological processing in mainstream school children. *Brain and*

Language, *94*(1), 72–85. http://doi.org/10.1016/j.bandl.2004.11.009

Morton, J. B., & Harper, S. N. (2007). What did Simon say? Revisiting the bilingual advantage. *Developmental Science*, *10*(6), 719–726. http://doi.org/10.1111/j.1467-7687.2007.00623.x

Muniandy, M. K., Sekharan Nair, G. K., Krishnan, S. K., Ahmad, I., & Mohamed Noor, N. (2010). Sociolinguistic Competence and Malaysian Students ' English Language Proficiency. *Language Teaching*, *3*(3), 145–151.

Narayan, C. R., Werker, J. F., & Beddor, P. S. (2010). The interaction between acoustic salience and language experience in developmental speech perception: Evidence from nasal place discrimination. *Developmental Science*, *13*(3), 407–420. http://doi.org/10.1111/j.1467-7687.2009.00898.x

Nishi, K., & Kewley-Port, D. (2007). Training Japanese listeners to perceive American English vowels: influence of training sets. *Journal of Speech, Language, and Hearing Research : JSLHR*, *50*(6), 1496–1509. http://doi.org/10.1044/1092-4388(2007/103)

Nobre-Oliveira, D. (2008). Effects of perceptual training on the learning of English vowels in non-native settings. In *New Sounds 2007: Proceedings of the Fifth International Symposium on the Acquisition of Second Language Speech* (pp. 382–389). Florianópolis: Brazil: Federal University of Santa Catarina.

Nottingham. (2015). Facts and figures - The University of Nottingham. Retrieved April 22, 2015, from https://www.nottingham.ac.uk/about/facts/factsandfigures.aspx

Omar, A. H. (2016). *Languages in the Malaysian education system: Monolingual strands in multilingual settings*. Routledge.

Paap, K. R., & Greenberg, Z. I. (2013). There is no coherent evidence for a bilingual advantage in executive processing. *Cognitive Psychology*, *66*(2), 232–258. http://doi.org/10.1016/j.cogpsych.2012.12.002

Packard, J. L. (2000). *The morphology of Chinese: a linguistic and cognitive approach*. Cambridge: Cambridge University Press.

Peirce, J. W. (2007). PsychoPy-Psychophysics software in Python. *Journal of Neuroscience Methods*, *162*(1–2), 8–13. http://doi.org/10.1016/j.jneumeth.2006.11.017

Perrachione, T. K., Lee, J., Ha, L. Y. Y., & Wong, P. C. M. (2011). Learning a novel phonological contrast depends on interactions between individual differences and training paradigm design. *The Journal of the Acoustical Society of America*, *130*(1), 461–472. http://doi.org/10.1121/1.3593366

Phoon, H. S., Abdullah, A. C., & Maclagan, M. (2013). The Consonant Realizations of Malay-, Chinese- and Indian-Influenced Malaysian English. *Australian Journal of Linguistics*, *33*(1), 3–30. http://doi.org/http://dx.doi.org/10.1080/07268602.2013.787902

Pillai, S., & Jayapalan, K. (2010). The Teaching of Pronunciation in Malaysia: State of the Art or No State at All? In *Transformations in English Language Education Proceedings of the 19th MELTA International Conference*.

Pillai, S., Mohd. Don, Z., Knowles, G., & Tang, J. (2010). Malaysian English: An instrumental analysis of vowel contrasts. *World Englishes*, *29*(2), 159–172. http://doi.org/10.1111/j.1467-

971X.2010.01636.x

Pisoni, D. B., Aslin, R. N., Perey, A. J., & Hennessy, B. L. (1982). Some effects of laboratory training on identification and discrimination of voicing contrasts in stop consonants. *Journal of Experimental Psychology: Human Perception and Performance*, *8*(2), 297–314. http://doi.org/10.1016/j.biotechadv.2011.08.021.Secreted

Prior, A., & Macwhinney, B. (2010). A bilingual advantage in task switching. *Bilingualism: Language and Cognition*, *13*(2), 253. http://doi.org/10.1017/S1366728909990526

Pulvermüller, F., Huss, M., Kherif, F., Moscoso, F., Hauk, O., & Shtyrov, Y. (2006). Motor cortex maps articulatory features of speech sounds (Vol. 103, pp. 7865–7870). Proceedings of the National Academy of Sciences of the United States of America.

Rajadurai, J. (2006). Pronunciation issues in non-native contexts: A Malaysian case study. *Malaysian Journal of ELT Research*, *2*(March), 42–59. Retrieved from http://www.melta.org.my/modules/tinycontent/Dos/RJoanne.pdf

Richie, C., & Kewley-Port, D. (2008). The Effects of Auditory-Visual Vowel Identification Training on ... *Journal of Speech and Hearing Research*, *51*(6), 1607–1619. http://doi.org/1092-4388_2008_07-0069 [pii]\r10.1044/1092-4388(2008/07-0069)

Roach, P. (2009). *English phonetics and phonology: a practical course* (4th ed.). Cambridge: Cambridge University Press.

Rogers, C. L., & Dalby, J. (2005). Forced-choice analysis of segmental production by Chinese-accented English speakers. *Journal of Speech, Language, and Hearing Research : JSLHR*, *48*(2), 306–322. http://doi.org/10.1044/1092-4388(2005/021)

Rogers, C. L., Lister, J. J., Febo, D. M., Besing, J. M., & Abrams, H. B. (2006). Effects of bilingualism, noise, and reverberation on speech perception by listeners with normal hearing. *Applied Psycholinguistics*, *27*(3), 465–485. http://doi.org/10.1017/S014271640606036X

Simon, E., & D’Hulster, T. (2012). The effect of experience on the acquisition of a non-native vowel contrast. *Language Sciences*, *34*(3), 269–283. http://doi.org/10.1016/j.langsci.2011.10.002

Singleton, D. (2007). The Critical Period Hypothesis: Some problems. *Inter-Linguistisca*, *17*, 48–56.

Skipper, J. I., van Wassenhove, V., Nusbaum, H. C., & 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. http://doi.org/10.1093/cercor/bhl147

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. http://doi.org/10.3758/BF03202673

Tadmor, U. (2009). Malay-Indonesian. In B. Comrie (Ed.), *The world’s major languages* (2nd ed., pp. 791–818). London: Routledge.

Tan, S. I. (2013). *Malaysian English: Language Contact and Change (Duisburg Papers on Research in Language and Culture)* (Volume 98). Peter Lang GmbH.

Thornburgh, D. F., & Ryalls, J. H. (1998). Voice onset time in Spanish-English bilinguals: Early versus late learners of English. *Journal of Communication Disorders*, *31*(3), 215–229. http://doi.org/10.1016/S0021-9924(97)00053-1

Treiman, R., Mullennix, J., Bijeljac-Babic, R., & Richmond-Welty, E. D. (1995). The special role of rimes in the description, use, and acquisition of English orthography. *Journal of Experimental Psychology. General*, *124*(2), 107–136. http://doi.org/10.1037/0096-3445.124.2.107

Wang, X. (2013). Perception of mandarin tones: The effect of L1 background and training. *Modern Language Journal*, *97*(1), 144–160. http://doi.org/10.1111/j.1540-4781.2013.01386.x

Wang, X., & Munro, M. J. (2004). Computer-based training for learning English vowel contrasts. *System*, *32*(4 SPEC.ISS.), 539–552. http://doi.org/10.1016/j.system.2004.09.011

Warren, R. M. (1970). Perceptual restoration of missing speech sounds. *Science (New York, N.Y.)*. http://doi.org/10.1126/science.167.3917.392

Warschauer, M., & Meskill, C. (2010). Technology and Second Language Teaching, (2000), 1–15.

Werker, J. F., & Pegg, J. E. (1992). Infant speech perception and phonological acquisition. In L. M. C. Ferguson & C. Stoel-Gammon (Eds.), *Phonological Development: Models, Research, and Implications* (pp. 285–311). Timonium: MD: York Press.

Werker, J., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, *7*(1), 49–63. http://doi.org/10.1016/S0163-6383(84)80022-3

Wilson, S. M., & Iacoboni, M. (2006). Neural responses to non-native phonemes varying in producibility: evidence for the sensorimotor nature of speech perception. *NeuroImage*, *33*(1), 316–25. http://doi.org/10.1016/j.neuroimage.2006.05.032

Wydell, T. N., & Butterworth, B. (1999). A case study of an English-Japanese bilingual with monolingual dyslexia. *Cognition*, *70*(3), 273–305. http://doi.org/10.1016/S0010-0277(99)00016-5

Xiaomei, W. (2014). Chinese Education in Malaysia. In X. L. Curdt-Christiansen & A. Hancock (Eds.), *Learning Chinese in Diasporic Communities: Many Pathways to being Chinese* (pp. 139–157). John Benjamins Publishing Company.

Yahaya, M. F., Mohd Noor, M. A., Mokhtar, A. A., Mohd Rawian, F., Othman, M., & Jusoff, K. (2009). Teaching of Mathematics and Science in English: The teachers' voices. *English Language Teaching*, *2*(2), 141–147.

Yang, J., & Fox, R. A. (2014). Perception of English Vowels by Bilingual Chinese–English and Corresponding Monolingual Listeners. *Language and Speech*, *57*(2), 215–237. http://doi.org/10.1177/0023830913502774

Yap, M. J., Liow, S. J. R., Jalil, S. B., Faizal, S. S. B., & Louis, S. (2010). The Malay Lexicon Project: a database of lexical statistics for 9,592 words. *Behavior Research Methods*, *42*(4), 992–1003. http://doi.org/10.3758/BRM.42.4.992

Yap, N. T. (2013). Perception of “Voicing” Among Malaysian ESL Learners. *Malaysian Journal of Languages and Linguistics, 2,* 94–106.

Yap, N. T., Wong, B. E., & Aziz, A. Y. A. (2010). Representation of english front vowels by Malay-English bilinguals. *Pertanika Journal of Social Science and Humanities*, *18*(2), 379–389.

Zhang, A., Feng, H., Dang, J., Zheng, X., & Xu, Z. (2015). The Perception of English Vowel Contrasts by Chinese EFL Learners and Native English Speakers. *Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS 2015)*.

Zhang, Y., Kuhl, P. K., Imada, T., Iverson, P., Pruitt, J., Stevens, E. B., … Nemoto, I. (2009). Neural signatures of phonetic learning in adulthood: A magnetoencephalography study. *NeuroImage*, *46*(1), 226–240. http://doi.org/10.1016/j.neuroimage.2009.01.028

Zhao, X., & Li, P. (2009). An online database of phonological representations for Mandarin Chinese. *Behavior Research Methods*, *41*(2), 575–583. http://doi.org/10.3758/BRM.41.2.575

Appendices

Appendix 1: English minimal pairs used in Spoken English Discrimination (SED) Training

1.1 Minimal pairs with /t/-/d/ phonemic contrast at word initial position with word frequency based on SUBTLEX-US (Brysbaert et al., 2012).

/t/ Initial	Word Frequency	/d/ Initial	Word Frequency
*t*ame	2.73	*d*ame	13.76
*t*art	2.39	*d*art	1.92
*t*ies	9.84	*d*ies	24.9
*t*in	8.65	*d*in	1.18
*t*ip	27.63	*d*ip	7.96
*t*own	247.92	*d*own	1490.27
*t*uck	7.96	*d*uck	24.76
*t*ug	2.75	*d*ug	8.92

1.2 Minimal pairs with /t/-/d/ phonemic contrast at word final position with word frequency based on SUBTLEX-US (Brysbaert et al., 2012).

/t/ Final	Word Frequency	/d/ Final	Word Frequency
bai*t*	9.73	ba*d*e	0.43
ba*t*	20.63	ba*d*	545.18
bea*t*	131.69	bea*d*	1.12
be*t*	171.82	be*d*	187.12
bi*t*	235.04	bi*d*	12.59
ro*t*	7.73	ro*d*	10.73
sa*t*	28.61	sa*d*	63.37
whi*t*e	171.45	wi*d*e	23.8

1.3 Minimal pairs with /ɛ/-/æ/ phonemic contrast at word medial position with word frequency based on SUBTLEX-US (Brysbaert et al., 2012)

/ɛ/	**Word Frequency**	**/æ/**	**Word Frequency**
h*ea*d	371.51	h*a*d	1675.92
m*e*n	372.76	m*a*n	1845.75
m*e*sh	0.69	m*a*sh	4.39
m*e*ss	78.14	m*a*ss	17.25
p*e*ck	3.53	p*a*ck	43.82
s*e*t	231.47	s*a*t	28.61
sh*e*ll	13.22	sh*a*ll	185.12
s*e*nd	179.78	s*a*nd	20.29
l*e*g	56.51	l*a*g	1.47
m*e*rry	39.1	m*a*rry	104.35
p*e*n	24.73	p*a*n	12.29
fl*e*sh	22.06	fl*a*sh	15.35
p*e*t	20.18	p*a*t	18.29
l*e*nd	11.67	l*a*nd	88.12
k*e*ttle	2.8	c*a*ttle	13.22
cel*e*ry	1.86	sal*a*ry	10.75

Appendix 2: Language History Questionnaire

2.1 Language history questionnaire for participants who took part in the SED training.

Language Background Questionnaire

1. Date of birth (DD/MM/YYYY) : ______________

2. Gender : Male / Female

3. Ethnicity :__________________

4. Country/ City of Origin : __________________

5. Current Residence Country (Family) :___________________

6. Participant ID :___________________

7. What type of program are you officially registered on in the university?

   Pre-sessional/ Pre-U / Foundation
   Undergraduate
   Postgraduate
   Academic Staff

8. Name of course that you are enrolled in (e.g. BSc(Hons) in Psychology)

9. First/Native Language (Language that you speak at home, could be dialects e.g., Hokkien, Hakka, telugu)

10. What language(s) does your Mother speak?
    (List all the languages down)

11. What language(s) does your Father speak?
    (List all the languages down)

12. List ALL the languages you know (even if your listening comprehension/speaking ability is poor, including dialects e.g., Hokkien, Hakka, telugu) in the order of most proficient. Rate your ability on the following aspects in each language. Please rate according to the following scale (write down the number in the table):

1- Very poor
2- Poor
3- Fair
4- Functional
5-

5- Good
6- Very good
7- Native-like

Language	Reading Proficiency	Writing Proficiency	Speaking Fluency	Listening Ability/ Comprehension
First Language				
Second Language				
Third Language				
Forth Language				
Fifth Language				

13. Provide the age at which you were first exposed to each language (including dialects e.g., Hokkien, Hakka, telugu) in terms of speaking, reading and writing, where you have learnt them from and the number of years you have spent on learning each language.

    Enter "0" if language is exposed since birth; enter "never" if language is not learnt for any particular field (e.g. writing).

Language	Age first exposed to the language			Where/ how was the language learnt? (e.g. Home, School, friends)	Number of years learning (Academically/ from formal education)	Number of years learning generally
	Speaking	Reading	Writing			

14. Please provide your education background:

Education Background			
School Level	Name of School	Languages taught	Main medium (Language) used for instructions
Kindergarten			
Primary School			
Secondary School			
College			
University			

15. Estimate how often you use your native language (first language) and other language per day for DAILY ACTIVITIES (eg. Going to classes, writing papers, talking to parents, classmates, or peers).

Language	WHEN/WHERE do you use this language? (can be more than one)	How Often? (Answer with NUMBERS, where 1 = Never, 2 = Rarely, 3 = Sometimes, 4 = Often, 5 = All the time)

16. Please provide information about your English Proficiency qualifications (e.g. GSCE / SPM / IELTS / TOEFL):

Name of Test	Testing Date (mm/yyyy)	Overall Score/ Grade	Listening Score	Speaking Score	Writing Score
e.g. IELTS	09/2009	62	75	58	61

17. In which language (CHOOSE ONE among your best two languages) do you feel you usually do better? Write the name of the language under each condition:

	At Home	At Work/ Study
Reading		
Writing		
Speaking		
Understanding		

18. Do you mix/switch between languages (e.g. within/between sentences) when you speak?

YES / NO

(If NO, proceed to question 20)

19. What languages do you mix and how often do you do that?
Write the main language (which should compose most part of your conversation) in the first column.

Main Language	Minor Language (Different language words/ sentences found in the main language spoken)	How often? (Answer with numbers, 1 = Never, 2 = Rarely, 3 = Sometimes, 4 = Often, 5 = All the time)

20. For the following questions, circle the number that corresponds with the frequency of the language that you use currently , using the scale of 1-5
(1= never, 3 = sometimes, 5 = always).

1. How often do you read texts (e.g. newspaper, books, instructions etc.) written in:

Traditional Chinese	1	2	3	4	5
Simplified Chinese	1	2	3	4	5
English	1	2	3	4	5
Malay	1	2	3	4	5
Tamil	1	2	3	4	5
Hindi	1	2	3	4	5
Arabic	1	2	3	4	5
Others: ________________________	1	2	3	4	5
Others 2: _______________________	1	2	3	4	5

2. How much Malay do you speak now? to

Your housemates/hall mates/ at home	1	2	3	4	5	n/a
Your friends	1	2	3	4	5	n/a
Your course mates	1	2	3	4	5	n/a
Other social context	1	2	3	4	5	n/a

3. How much Malay do the following people speak to you now? to

Your housemates/hall mates/ at home	1	2	3	4	5	n/a
Your friends	1	2	3	4	5	n/a
Your course mates	1	2	3	4	5	n/a
Other social context	1	2	3	4	5	n/a

4. How much English do you speak now?

Your housemates/hall mates/ at home	1	2	3	4	5	n/a
Your friends	1	2	3	4	5	n/a
Your course mates	1	2	3	4	5	n/a
Other social context	1	2	3	4	5	n/a

5. How much English do the following people speak to you now? to

Your housemates/hall mates/ at home	1	2	3	4	5	n/a
Your friends	1	2	3	4	5	n/a
Your course mates	1	2	3	4	5	n/a
Other social context	1	2	3	4	5	n/a

6. How much Cantonese do you speak now? to

Your housemates/hall mates/ at home	1	2	3	4	5	n/a
Your friends	1	2	3	4	5	n/a
Your course mates	1	2	3	4	5	n/a
Other social context	1	2	3	4	5	n/a

7. How much Cantonese do the following people speak to you now? to

Your housemates/hall mates/ at home	1	2	3	4	5	n/a
Your friends	1	2	3	4	5	n/a
Your course mates	1	2	3	4	5	n/a
Other social context	1	2	3	4	5	n/a

8. How much Mandarin do you speak now? to

Your housemates/hall mates/ at home	1	2	3	4	5	n/a
Your friends	1	2	3	4	5	n/a
Your course mates	1	2	3	4	5	n/a
Other social context	1	2	3	4	5	n/a

9. How much Mandarin do the following people speak to you now? to

Your housemates/hall mates/ at home	1	2	3	4	5	n/a
Your friends	1	2	3	4	5	n/a
Your course mates	1	2	3	4	5	n/a
Other social context	1	2	3	4	5	n/a

10. How much ____________ (other known language) do you speak now? To

Your housemates/hall mates/ at home	1	2	3	4	5	n/a
Your friends	1	2	3	4	5	n/a
Your course mates	1	2	3	4	5	n/a
Other social context	1	2	3	4	5	n/a

11. How much ____________ (other known language) do the following people speak to you now? to

Your housemates/hall mates/ at home	1	2	3	4	5	n/a
Your friends	1	2	3	4	5	n/a
Your course mates	1	2	3	4	5	n/a
Other social context	1	2	3	4	5	n/a

21. Do / did you have any reading difficulties in your native language or other languages (e.g. dyslexia - a disorder that involves difficulty in learning to read or interpret words, letters, and other language symbols)? Yes / No

If Yes, Please provide details:

____

____

22. Do you have any listening difficulties in your native or other languages (e.g. unable to accurately perceive/process, understand and respond to sound)? Yes / No

If Yes, Please provide details:

____

____

23. Do you have normal vision? (If you wear spectacles or contact lenses, you are not considered to have normal vision, but are corrected to normal vision.) Yes / No

If No, Please provide details:

____

____

Thank You ☺

2.2 Language History Questionnaire for Raters (Experiment 5 and 7) also include the following questions:

24. Have you stayed in other countries for more than one month?

Yes (Continue to next question)
No

25. If you answered YES in question 24, please state which country and the length of your stay.

	When?	How Long?
Country Name 1:		
Country Name 2:		
Country Name 3:		

Appendix 3: SPSS Results on Training Effect of Adaptive Staircase SED

A main effect of L1 was found, $F(2,37) = 4.48$, $p = .02$. Post-hoc independent-samples t-tests revealed higher identification accuracy of the L1 Malaysian English speakers when compared to the L1 Mandarin Malaysian speakers, $t(23) = 2.93$, $p < .01$. On the other hand, the L1 Malay Malaysian speakers performed similarly as the L1 Mandarin Malaysian speakers, $t(27)= 1.20$, $p = .24$ and the L1 Malaysian English speakers, $t(24) = 1.68$, $p = .11$.

The identification accuracy for the /t/-/d/ final and /ɛ/-/æ/ phonemic contrasts did not differ, $F(1, 37) = 0.54$, $p = .47$. However, there was an interaction between the phonemic contrast and L1, $F(2,37) = 3.67$, $p = .04$. Post-hoc paired-samples t-tests were conducted to compare the perceptual performance of the /t/-/d/ final with the /ɛ/-/æ/ phonemic contrasts, for every language group. The L1 Mandarin and L1 Malaysian English speakers had similar identification accuracy for the phonemic contrasts, *t*s < 1.17, *p*s > .05, but the L1 Malay Malaysian speakers had greater difficulty identifying the /t/-/d/ final compared to the /ɛ/-/æ/ phonemic words , $t(14) = 3.14$, $p < .01$. Post-hoc one-way between-subjects ANOVAs were also conducted to compare identification accuracy across the three language groups, for each phonemic contrast. Main effects of L1 were found for both /t/-/d/ final contrast, $F(2,39) = 4.42$, $p = .02$ and /ɛ/-/æ/ contrast, $F(2,39) = 3.92$, $p = .03$. Multiple comparisons (Bonferroni) showed that the L1 Malaysian English speakers identified the /t/-/d/ final contrast better than the L1 Mandarin and Malay speakers, ps < .05, but the L1 Mandarin and Malay speakers performed similarly, p > .05. L1 Malay speakers identified the /ɛ/-/æ/ phonemic words significantly better than the L1 Mandarin speakers, $p = .04$, the L1 Malaysian English speakers identified the /ɛ/-/æ/ contrast similarly as the other two language groups, *p*s > .05.

Appendix 4: Linguistic details of participants from all training studies (AS refers to Adaptive Staircase).

Exp	Training Group	L1	Number of Participants	Mean Age	Mean Self-rated English Proficiency	Mean Self-rated Mandarin Proficiency	Age of Acquisition	IELTS	Lextale
1	Constant SED; Single Speaker	M.E or Mandarin	14	21.29	6.11	-	2.64	7.21	-
1	AS SED; Single Speaker	M.E or Mandarin	14	21.86	5.45	-	3.21	7.07	-
1	Constant SED; Multiple Speakers	M.E or Mandarin	14	21.07	5.36	-	4.36	6.82	-
1	AS SED; Multiple Speakers	M.E or Mandarin	14	20.64	4.55	-	4	6.82	-
2	Constant SED; Low Babble	Mandarin	14	18.93	4.71	5.66	2.36	7.07	-
2	Constant SED; High Babble	Mandarin	14	18.57	4.75	5.21	2.36	6.96	-
3	AS SED	Mandarin	14	18.64	4.96	5.87	3.86	7.04	-
3	AS SED	Malay	15	18.00	5.18	-	5	6.83	-
3	AS SED	M.E	11	17.81	6.23	3.34	2.18	7.32	-
4	Constant SED	Mandarin	13	18.69	4.56	5.54	2	7.31	-
4	AS SED	Mandarin	14	18.64	4.87	5.87	3.93	7.04	-
6	AS SED	Mandarin	15	20.80	5.13	6	3.87	-	75.47
6	Constant SED	Mandarin	14	20.29	4.89	5.91	2.86	-	76.79
6	Control	Mandarin	14	18.14	5.19	5.55	2.86	-	71.84
7	Malaysian Raters	M.E	28	22.21	6.51	-	0	-	85.89
7	British Raters	British English	28	23.68	6.98	-	0	-	89.22

* Exp – Experiment, IELTS- International English Language Testing Scores, Self-rated Language Proficiency (from scale 1 = very poor to scale 7 = native like)

* Average self-rated English proficiency scores were used to represent participants' overall self-rated language proficiency. The score was calculated by averaging the reading, writing, speaking and listening proficiency, as rated by the participants.

Appendix 5: Identification Accuracy of Raters on Word Productions of Different Training Groups.

5.1 Mean percentage of correct identification made by Malaysian raters for the /t/ and /d/ phonetic categories located at word initial position (with standard errors in brackets).

	/t/ Initial			**/d/ Initial**		
Training Groups	**Pre-test**	**Post-test**	**1 month post-test**	**Pre-test**	**Post-test**	**1 month post-test**
Adaptive Staircase SED	96.88 (2.70)	96.43 (2.01)	96.88 (2.70)	96.88 (1.89)	94.64 (2.27)	97.77 (1.45)
Constant SED	95.67 (2.49)	96.63 (1.63)	95.67 (2.07)	97.60 (0.99)	98.56 (0.80)	97.60 (1.20)
Control	94.20 (4.14)	93.75 (3.72)	92.41 (3.98)	92.86 (3.77)	92.41 (3.98)	93.75 (3.72)

5.2 Mean percentage of correct identification made by Malaysian raters for the /t/ and /d/ phonetic categories located at word final position (with standard errors in brackets).

	/t/ Final			**/d/ Final**		
Training Groups	**Pre-test**	**Post-test**	**1 month post-test**	**Pre-test**	**Post-test**	**1 month post-test**
Adaptive Staircase SED	83.04 (2.67)	82.14 (3.18)	78.57 (3.21)	75 (3.86)	73.66 (4.33)	72.32 (4.03)
Constant SED	86.06 (2.33)	82.69 (2.78)	79.81 (4.16)	73.56 (3.94)	75.48 (3.95)	69.23 (4.56)
Control	79.91 (4.83)	84.38 (4.57)	82.59 (4.83)	73.21 (4.57)	74.11 (4.94)	75 (5.10)

5.3 Mean percentage of correct identification made by Malaysian raters for the /ɛ/ and /æ/ phonetic categories (with standard errors in brackets).

	/ɛ/			/æ/		
Training Groups	**Pre-test**	**Post-test**	**1 month post-test**	**Pre-test**	**Post-test**	**1 month post-test**
Adaptive Staircase SED	67.19 (3.72)	66.52 (2.81)	65.40 (3.30)	58.48 (2.93)	63.17 (2.35)	71.21 (2.86)
Constant SED	66.35 (2.64)	70.67 (3.30)	70.43 (3.07)	68.99 (3.26)	70.67 (3.30)	70.43 (3.07)
Control	66.74 (2.47)	63.39 (3.04)	66.96 (3.43)	69.87 (2.80)	69.20 (2.34)	70.54 (3.40)

5.4 Mean percentage of correct identification made by British raters for the /t/ and /d/ phonetic categories located at word initial position (with standard errors in brackets).

	/t/ Initial			/d/ Initial		
Training Groups	**Pre-test**	**Post-test**	**1 month post-test**	**Pre-test**	**Post-test**	**1 month post-test**
Adaptive Staircase SED	96.88 (1.38)	95.98 (1.58)	98.21 (1.06)	95.54 (1.84)	93.30 (3.04)	95.54 (1.60)
Constant SED	95.67 (1.69)	96.15 (1.51)	98.08 (1.50)	95.19 (1.71)	89.42 (3.58)	89.90 (3.80)
Control	95.09 (2.52)	96.43 (1.42)	96.88 (1.65)	94.20 (2.69)	93.30 (3.25)	94.20 (2.08)

5.5 Mean percentage of correct identification made by British raters for the /t/ and /d/ phonetic categories located at word final position (with standard errors in brackets).

	/t/ Final			/d/ Final		
Training Groups	**Pre-test**	**Post-test**	**1 month post-test**	**Pre-test**	**Post-test**	**1 month post-test**
Adaptive Staircase SED	76.34 (3.36)	80.80 (3.94)	81.25 (3.37)	75.89 (3.52)	74.55 (3.83)	73.66 (4.18)
Constant SED	81.73 (3.27)	82.21 (3.11)	80.29 (2.96)	70.19 (4.86)	72.60 (3.33)	73.56 (4.40)
Control	79.02 (4.68)	80.80 (3.66)	82.59 (3.71)	70.98 (4.02)	66.96 (5.31)	72.32 (4.56)

5.6 Mean percentage of correct identification made by British raters for the /ɛ/ and /æ/ phonetic categories (with standard errors in brackets).

	/ɛ/			/æ/		
Training Groups	**Pre-test**	**Post-test**	**1 month post-test**	**Pre-test**	**Post-test**	**1 month post-test**
Adaptive Staircase SED	71.65 (3.44)	70.54 (3.06)	72.54 (3.31)	61.38 (4.38)	65.85 (4.06)	68.97 (4.56)
Constant SED	68.99 (2.81)	68.51 (2.63)	69.71 (3.19)	63.70 (3.18)	75 (3.22)	66.35 (3.57)
Control	72.54 (3.04)	71.21 (3.00)	72.32 (3.30)	62.95 (4.55)	67.19 (3.80)	66.52 (3.98)

Appendix 6: Quality Rating of Raters on Word Productions of Different Training Groups.

6.1 Mean rating score given by Malaysian raters for the /t/ and /d/ phonetic categories located at word initial position (with standard errors in brackets).

	/t/ Initial			**/d/ Initial**		
Training Groups	**Pre-test**	**Post-test**	**1 month post-test**	**Pre-test**	**Post-test**	**1 month post-test**
Adaptive Staircase SED	5.96 (0.19)	5.77 (0.20)	5.89 (0.18)	5.90 (0.16)	5.79 (0.18)	5.86 (0.19)
Constant SED	5.88 (0.20)	5.81 (0.20)	5.78 (0.21)	5.84 (0.19	5.72 (0.20)	5.59 (0.23)
Control	5.61 (0.29)	5.43 (0.27)	5.68 (0.29)	5.45 (0.28)	5.52 (0.29)	5.58 (0.27)

*GIVE RATING SCALE INFO

6.2 Mean rating score given by Malaysian raters for the /t/ and /d/ phonetic categories located at word final position (with standard errors in brackets).

	/t/ Final			**/d/ Final**		
Training Groups	**Pre-test**	**Post-test**	**1 month post-test**	**Pre-test**	**Post-test**	**1 month post-test**
Adaptive Staircase SED	5.62 (0.18)	5.41 (0.20)	5.52 (0.18)	5.13 (0.21)	4.97 (0.28)	5.18 (0.21)
Constant SED	5.50 (0.20)	5.75 (0.18)	5.57 (0.29)	5.09 (0.26)	4.93 (0.25)	5.28 (0.22)
Control	5.37 (0.25)	5.81 (0.18)	5.66 (0.17)	4.82 (0.28)	4.91 (0.27)	4.78 (0.30)

6.3 Mean rating score given by Malaysian raters for the /ɛ/ and /æ/ phonetic categories (with standard errors in brackets).

	/ɛ/			**/æ/**		
Training Groups	**Pre-test**	**Post-test**	**1 month post-test**	**Pre-test**	**Post-test**	**1 month post-test**
Adaptive Staircase SED	5.11 (0.18)	4.99 (0.18)	5.05 (0.17)	5.11 (0.16)	5.03 (0.18)	5.11 (0.18)
Constant SED	5.09 (0.20)	5.07 (0.19)	5.16 (0.20)	5.30 (0.20)	5.21 (0.19)	5.12 (0.19)
Control	5.18 (0.17)	5.20 (0.17)	5.17 (0.15)	5.15 (0.15)	5.21 (0.17)	5.26 (0.17)

6.4 Mean rating score given by British raters for the /t/ and /d/ phonetic categories located at word initial position (with standard errors in brackets).

	/t/ Initial			/d/ Initial		
Training Groups	**Pre-test**	**Post-test**	**1 month post-test**	**Pre-test**	**Post-test**	**1 month post-test**
Adaptive Staircase SED	5.42 (0.16)	5.26 (0.17)	5.34 (0.14)	5.07 (0.18)	4.98 (0.17)	5.19 (0.15)
Constant SED	5.72 (0.16)	5.19 (0.17)	5.51 (0.18)	5.27 (0.19)	5.17 (0.17)	5.05 (0.19)
Control	5.48 (0.15)	5.44 (0.16)	5.67 (0.15)	5.39 (0.17)	5.26 (0.19)	5.51 (0.17)

6.5 Mean rating score given by British raters for the /t/ and /d/ phonetic categories located at word final position (with standard errors in bracket).

	/t/ Final			/d/ Final		
Training Groups	**Pre-test**	**Post-test**	**1 month post-test**	**Pre-test**	**Post-test**	**1 month post-test**
Adaptive Staircase SED	5 (0.16)	5.10 (0.16)	5.11 (0.16)	4.07 (0.17)	4.12 (0.15)	4 (0.19)
Constant SED	5.23 (0.15)	5.03 (0.17)	5.23 (0.16)	4.06 (0.19)	4.07 (0.17)	3.96 (0.20)
Control	5.09 (0.22)	5.30 (0.17)	5.29 (0.19)	4.18 (0.20)	3.91 (0.23)	4.17 (0.22)

6.6 Mean rating score given by British raters for the /ɛ/ and /æ/ phonetic categories (with standard errors in brackets).

	/ɛ/			/æ/		
Training Groups	**Pre-test**	**Post-test**	**1 month post-test**	**Pre-test**	**Post-test**	**1 month post-test**
Adaptive Staircase SED	4.97 (0.12)	4.74 (0.13)	4.64 (0.15)	4.70 (0.20)	4.63 (0.17)	4.72 (0.21)
Constant SED	4.67 (0.11)	4.69 (0.13)	4.67 (0.10)	5.05 (0.12)	4.97 (0.12)	5 (0.15)
Control	4.67 (0.13)	4.62 (0.15)	4.78 (0.15)	4.85 (0.14)	4.67 (0.19)	4.83 (0.17)

Appendix 7: Tense-lax English Vowel Phonemic Pairs

Appendix 7.1: /i:/-/I/ Minimal Pairs

Tense /i:/	Lax /I/
beach	bitch
bead	bid
beat	bit
cheek	chick
deed	did
deep	dip
each	itch
ease	is
eat	it
feast	fist
feel	fill
feet	fit
green	grin
He’d	hid
heat	hit
heel	hill
he's	his
keen	kin
lead	lid
leak	lick
leased	list
least	list
leave	live
meal	mill
meat	mitt
peach	pitch
peak	pick
peel	pill
reach	rich
seat	sit
seek	sick
seen	sin
sheep	ship
skied	skid
sleep	slip
steal	still
week	wick
we'll	will

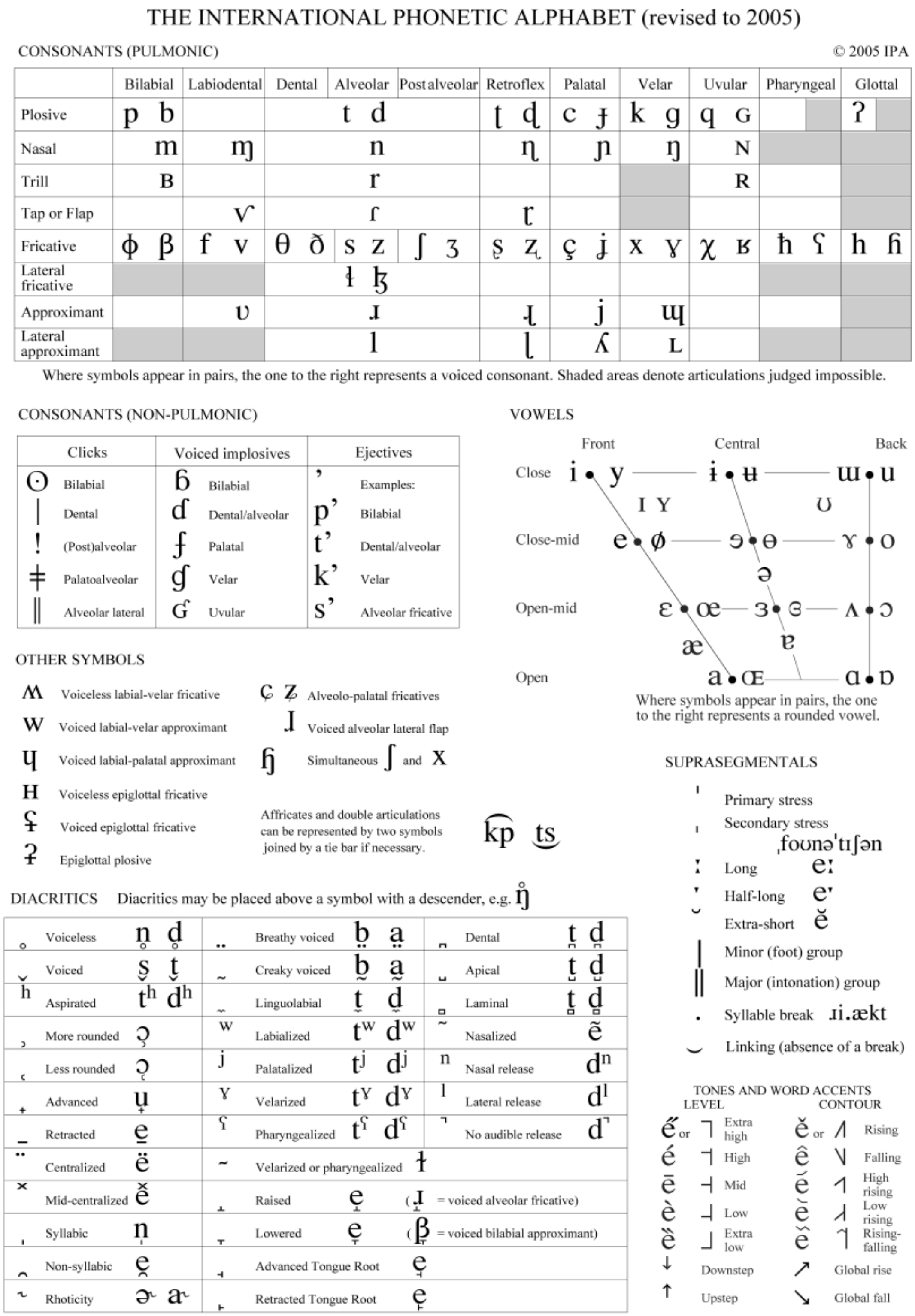
Appendix 7.2: /uː/-/ʊ/ Minimal Pairs

Tense /uː/	**Lax /ʊ/**
boo	book
buhl	bull
cooed	could
food	foot
fool	full
kook	cook
kooky	cookie
luke	look
pool	pull
shoe	shoot
shoed	should
stewed	stood
stoop	stoep
suit	soot
wooed	would/ wood

Appendix 7.3: /ɔ:/- /ɑ/ Minimal Pairs

Tense /ɔ:/	Lax /ɑ/
caller	collar
caught	cot
chalk	chock
cord	cod
cork	cock
corn	con
dawn	don
fond	fawned
force	fosse
gaud	god
hawk	hock
naughty	knotty
nautch	notch
pawned	pond
scorn	scone
short	shot
sort	sot
sport	spot
stalk	stock
sword	sod
sworn	swan
warble	wobble
ward	wad
warp	whop
yon	yawn

Appendix 8: IPA 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

w Voiced labial-velar approximant

ɥ Voiced labial-palatal approximant

ʜ Voiceless epiglottal fricative

ʢ Voiced epiglottal fricative

ʡ Epiglottal plosive

ɕ ʑ Alveolo-palatal fricatives

ɺ Voiced alveolar lateral flap

ɧ Simultaneous ʃ and x

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

*Note: IPA Chart, http://www.internationalphoneticassociation.org/content/ipa-chart, available under a Creative Commons Attribution-Sharealike 3.0 Unported License. Copyright © 2015 International Phonetic Association.