PERCEPTION AND PRODUCTION OF L2 MANDARIN STOP CONSONANTS BY MALAYSIAN MALAY SPEAKERS

MAGDALENE A/P VETHANAYAGAM JAMES

FACULTY OF LANGUAGES AND LINGUISTICS
UNIVERSITI MALAYA
KUALA LUMPUR

2023

PERCEPTION AND PRODUCTION OF L2 MANDARIN STOP CONSONANTS BY MALAYSIAN MALAY SPEAKERS

MAGDALENE A/P VETHANAYAGAM JAMES

DISSERTATION SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF ARTS (LINGUISTICS)

FACULTY OF LANGUAGES AND LINGUISTICS
UNIVERSITI MALAYA
KUALA LUMPUR

2023

UNIVERSITI MALAYA

ORIGINAL LITERARY WORK DECLARATION

Name of Candidate: Magdalene a/p Vethanayagam James

Matric No: TOA160046 (17050188)

Name of Degree: Master of Arts (Linguistics)

Title of Project Paper/Research Report/Dissertation/Thesis ("this Work"): Perception and Production of L2 Mandarin Stop Consonants by Malaysian Malay Speakers)

Field of Study: Language Acquisition

I do solemnly and sincerely declare that:

(1) I am the sole author/writer of this Work;
(2) This Work is original;
(3) Any use of any work in which copyright exists was done by way of fair dealing and for permitted purposes and any excerpt or extract from, or reference to or reproduction of any copyright work has been disclosed expressly and sufficiently and the title of the Work and its authorship have been acknowledged in this Work;
(4) I do not have any actual knowledge nor do I ought reasonably to know that the making of this work constitutes an infringement of any copyright work;
(5) I hereby assign all and every rights in the copyright to this Work to the University of Malaya ("UM"), who henceforth shall be owner of the copyright in this Work and that any reproduction or use in any form or by any means whatsoever is prohibited without the written consent of UM having been first had and obtained;
(6) I am fully aware that if in the course of making this Work I have infringed any copyright whether intentionally or otherwise, I may be subject to legal action or any other action as may be determined by UM.

Candidate's Signature Date: 2 January 2023

Subscribed and solemnly declared before,

Witness's Signature Date: 2 January 2023

Name:

Designation:

PERCEPTION AND PRODUCTION OF L2 MANDARIN STOP CONSONANTS BY MALAYSIAN MALAY SPEAKERS

ABSTRACT

Limited studies have been done on second language acquisition in the Mandarin language by Malaysian Malay learners. Therefore, the current study addressed this literature gap by systematically describing Mandarin stop consonant acquisition among Malaysian Malay learners. To ascertain whether Malaysian Malay learners have difficulty perceiving and producing Mandarin voiceless unaspirated and aspirated stops, this study examined the perception and production of L2 Mandarin stop consonants as well as the L2 acquisition of Mandarin unaspirated and aspirated consonants. Ten Mandarin teachers whose first language is Malay participated in this research. The two main data sources of this research were a perception task and a production test. The purpose of the perception task was to test whether Malaysian Malay learners perceptually contrast Mandarin aspirated stops from unaspirated stops by making the participants listen to 24 words and write down *pinyin.* The results showed that out of 240 tokens (24 x 10), only four errors occurred. This indicates that Malaysian Malay speakers can accurately perceive voiceless aspirated and unaspirated stop consonants. On the other hand, the purpose of the production task was to enable Malaysian Malay learners to differentiate aspirated Mandarin stop consonants from their unaspirated counterparts during the production of Mandarin stop consonants in terms of voice onset time (VOT) values. For this task, the participants read 40 Mandarin words. Based on the analysis of VOT, the most difficult consonants for the participants to produce were /p/, while the easiest consonants for the participants to produce were /k^h/ and /t^h/. The overall correct perception rate of Mandarin consonants was 98%, which suggests that Malay Mandarin learners are highly sensitive and intuitive towards the perceptual aspect of Mandarin consonants. The overall correct production rate of Mandarin consonants was 71%, showing that the Mandarin consonants produced by Malay Mandarin learners are at a moderate level. The findings of this study point out that Malaysian Malay speakers have better perception ability than production ability with regard to Mandarin stop consonants. This research adds value to the L2 speech learning literature by examining the perception and production of Mandarin stop consonants by L2 Malaysian Malay learners. Practically, this study provides information on the awareness of Mandarin language perception and production among Mandarin Malay teachers.

Keywords: Mandarin as a second language, Malaysian Malay speakers, stop consonants, perception, production.

PERSEPSI DAN PENGHASILAN KONSONAN HENTIAN MANDARIN OLEH PENUTUR BAHASA MELAYU

ABSTRAK

Kajian terhad telah dilakukan terhadap pemerolehan bahasa kedua dalam bahasa Mandarin oleh pelajar Melayu Malaysia. Oleh itu, kajian semasa menangani jurang literatur ini dengan menerangkan secara sistematik pemerolehan konsonan hentian Mandarin dalam kalangan pelajar Melayu Malaysia. Untuk memastikan sama ada pelajar Melayu Malaysia menghadapi kesukaran dalam mempersepsi dan menghasilkan hentian tidak bersuara Mandarin beraspirasi dan tidak beraspirasi, kajian ini mengkaji persepsi dan penghasilan konsonan hentian L2 Mandarin serta pemerolehan L2 konsonan Mandarin tidak beraspirasi dan beraspirasi. Sepuluh orang guru Mandarin yang bahasa pertamanya adalah Bahasa Melayu telah menyertai penyelidikan ini. Dua sumber data utama penyelidikan ini ialah tugas persepsi dan ujian penghasilan. Tujuan tugasan persepsi adalah untuk menguji sama ada pelajar Melayu Malaysia secara persepsi membezakan hentian aspirasi Mandarin daripada hentian tanpa aspirasi dengan membuat peserta mendengar 24 perkataan dan menulis *pinyin*. Keputusan menunjukkan bahawa daripada 240 token (24 x 10), hanya empat ralat berlaku. Ini menunjukkan bahawa penutur bahasa Melayu Malaysia dapat memahami dengan tepat konsonan hentian tidak bersuara beraspirasi dan tidak beraspirasi. Sebaliknya, tujuan tugasan penghasilan adalah untuk membolehkan pelajar Melayu Malaysia membezakan konsonan hentian Mandarin beraspirasi daripada konsonan hentian tidak beraspirasi semasa penghasilan konsonan hentian Mandarin dari segi nilai VOT. Untuk tugasan ini, para peserta membaca 40 perkataan Mandarin. Berdasarkan analisis masa timbul suara (VOT), konsonan yang paling sukar untuk dihasilkan oleh peserta ialah /p/, manakala konsonan yang paling mudah untuk dihasilkan oleh peserta ialah /k^h/ dan /t^h/. Kadar persepsi betul keseluruhan konsonan Mandarin ialah 98%, yang menunjukkan bahawa pelajar Melayu Mandarin sangat sensitif dan intuitif terhadap aspek persepsi konsonan Mandarin. Kadar penghasilan konsonan Mandarin yang betul secara keseluruhan ialah 71%, menunjukkan bahawa konsonan Mandarin yang dihasilkan oleh pelajar Melayu Mandarin berada pada tahap sederhana. Dapatan kajian ini menunjukkan bahawa penutur Bahasa Melayu Malaysia mempunyai kebolehan persepsi yang lebih baik berbanding kebolehan penghasilan berkaitan konsonan hentian Mandarin. Penyelidikan ini menambah nilai kepada literatur pembelajaran pertuturan L2 dengan mengkaji persepsi dan penghasilan konsonan hentian Mandarin oleh pelajar Melayu Malaysia L2. Secara praktikalnya, kajian ini memberi maklumat tentang kesedaran persepsi dan penghasilan bahasa Mandarin dalam kalangan guru Bahasa Melayu Mandarin.

Kata kunci: Mandarin sebagai bahasa kedua, penutur Bahasa Melayu, konsonan hentian beraspirasi dan tidak beraspirasi, persepsi, penghasilan konsonan hentian.

ACKNOWLEDGEMENTS

First, I extend my tremendous gratitude to my supervisor Dr. Chiew Poh Shin, for her mentorship, support, patience, and encouragement in enabling me to complete this Master's dissertation. I am grateful for her professional feedback and for taking the time to meet with me to discuss my progress. From the selection of the topic to the determination of the research method and data analysis to the multiple revisions of this thesis, she has provided many useful suggestions and substantial guidance. With the supervision and help of my conscientious supervisor, I was able to complete my thesis.

I express my highest appreciation to my beloved husband, Paulvin, and my family for always believing in me and supporting me during this arduous study period. I felt their prayers of support even when they were not all beside me in person.

I would like to thank Miss Sharmiza for her invaluable help in recording the data for the perception test. She handled this lengthy and challenging task with accuracy and perseverance. Finally, I am indebted to all the participants of this study and to those who helped me in ways big and small, throughout this journey.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS AND ABBREVIATIONS

VOT	:	Voice onset time
L1	:	First language
L2	:	Second language
L3	:	Third language
SLM	:	Speech Learning Model
MOE	:	Ministry of Education

LIST OF APPENDICES

CHAPTER 1: INTRODUCTION

1.1 Background of Research

Over the past 20 years, the teaching and learning of the Mandarin language as a key second language or foreign language have proliferated not only in China but across the world, partly due to investments by the Chinese government (Zhao & Huang, 2010). This has prompted a rising number of academics to look into how foreigners learn the Mandarin language. Notably, the majority of previous research on this topic has concentrated on subjects from the USA, the UK, Japan, Korea, and Thailand. Until now, Mandarin learning in Malaysia has received scant empirical attention.

In Malaysia, all vernacular schools and national schools are required to teach the Malay language as a compulsory subject. In contrast, the English language is taught in all government schools as a second language. This means that regardless of ethnicity, Malaysian preschoolers are taught to speak and learn these two languages. In recent years, the Malaysian Ministry of Education (MOE) has also emphasized additional languages in national schools, particularly Mandarin, following the National Education Blueprint 2013-2025 (Malaysian Ministry of Education, 2013). In accordance with this blueprint, the ministry encourages every child from primary school to secondary school to learn an additional language apart from Malay and English by the year 2025.

In addition, according to the National Blueprint 2015-2025 for Higher Education (Malaysian Ministry of Education, 2015), graduates are encouraged to learn one additional global language to communicate better, increase their marketability, and learn new knowledge to advance their careers and personal lives. To this end, universities such as Universiti Teknologi Mara (UiTM) and Universiti Malaya (UM) offer Mandarin as an elective subject, while Universiti Sains Malaysia (USM) requires some of its undergraduates to register for Mandarin courses as a third language.

Indeed, the Mandarin language has become one of the most preferred additional languages among Malaysian university students, owing to the rapid growth of the Mandarin-speaking population around the world (Hoe & Mah, 2009). Cheong et al. (2019) mentioned that the two most important motivational factors for undergraduates at Universiti Teknologi Malaysia Melaka (UTeM) to learn Mandarin are to prepare for their future careers and to be able to communicate with Chinese friends. In another study done by Hiang (2015), both UiTM students (73.9%) and non-UiTM students (76.2%) responded that improving their work chances is an essential consideration for them to learn Mandarin. According to Chua and Muhammad Afiq (2019), the relevance of Mandarin in terms of employability, self-interest, and grade improvement prompts non-native Mandarin learners to keep learning Mandarin. Likewise, Tan, Oii, and Hairul (2012) found that students are highly motivated to enroll in a Mandarin course because they believe learning the language will benefit them in their future careers. In line with this, Lee and Khalid's (2016) study revealed that proficiency in Mandarin increases both Malay and Chinese job applicants' interview prospects. Their study also found that many Chinese-owned companies and about 10% of Malay-owned companies stipulate Chinese language proficiency as a requirement in job postings. Overall, studies have identified a high demand for studying Mandarin as a second language with the aim of furthering one's career in Malaysia.

According to the findings of a study conducted by Lau, Ng, and Lee (2011), Malay students are the most likely to enroll in Mandarin language courses, followed by Indian students and other Bumiputera students. Corresponding to this, Chin et al. (2021) reported that in the year 2020, over 100,000 non-Chinese students were studying in Chinese vernacular schools all over Malaysia, most of whom were Malays. Data from the MOE corroborates that Malay students' intake in Chinese vernacular schools grew by 6.18% in 2020 from 2010, compared to a growth of just 1.08% for Indian students and 0.65% for other races' students (The Straits Times, 2020). Currently, Malay students comprise about 15% of the country's Chinese school students. This indicates that Malays, who are the majority by ethnicity in Malaysia and comprise 69.8% of the Malaysian population, are seeking to be proficient in Mandarin.

The Sun Daily (2016) reported that at the time, there were 814 Malaysian students pursuing Mandarin learning at the Beijing Language and Cultural University (BLCU) and Beijing Foreign Studies University (BFSU), of which the MOE had sponsored 403 students. BFSU has trained nearly 400 Malaysian undergraduate students from 2007 to 2022 in its Chinese Language Teacher Training program (Wang, 2022). Following this program, undergraduates sponsored by the MOE continue a one-and-a-half-year teachers' training program in Malaysia before being placed in national primary schools as Mandarin teachers. However, according to Zhen (2014), there has always been a shortage of Mandarin language teachers, not only in Chinese primary schools but also in national primary schools.

In summary, given the demand for Mandarin learning among Malay students and the role of Mandarin in their careers, numerous studies on the motivation for learning Mandarin as a foreign language have been conducted by Malaysian researchers (Cheong et al., 2019; Hong, 2020; Chua & Azlan, 2019). However, despite the shortage of teachers and the challenges of learning a foreign language, only a few scholars have examined how Mandarin consonants are perceived and produced by Malay learners. Such research is imperative to ensure proper mastery of the language for the students' future needs. Therefore, the lack of research into the perception and production of Mandarin stop consonants in Malaysia is addressed in this study through two tasks, namely perception, and production tasks.

1.2 Statement of Research Problem

Swan (2008) mentioned that it is difficult, complicated, and time-consuming to learn an additional language. Scovel (1969) further asserted in his paper that when adults speak a foreign or second language, it is impossible for them to lose their native accents. Consequently, many adult learners of second languages struggle to understand and communicate with native speakers of the language. Obstruent sounds that do not exist in their native language are particularly challenging for second language learners. Moreover, when learning to perceive and produce sounds in a foreign language, adult learners tend to rely on the rules and categories of their native language rather than those of the second language, which is one of the most significant factors contributing to their difficulty.

In addition, Collins and Vandenbergen (2000) explained that during the learning process of a second language, speakers are known to transfer and apply features from their native language (LI) to the second language, not only on a grammatical level but also on a phonetic level. Due to the fact that these features may not be used in the same way as they are in the first language, transfer frequently leads to errors or misunderstandings in the second language. Similarly, non-native language learners often reproduce their native language's pronunciation traits in their second language.

In this regard, it has been found that the majority of Mandarin L2 learners, especially beginners, face difficulties in verbal communication that may demotivate them from learning and mastering the Mandarin language. Specifically, those learning Mandarin as an L2 language have poor pronunciation of consonants, vowels, and tones (Khor, Mah, & Chow, 2012) and find Mandarin stop aspiration a significant challenge. Soon (2018) found that aspirated voiceless Mandarin stop consonants are especially difficult for Malay students to produce. Goh (2007) also categorized unaspirated and aspirated consonants as difficult pronunciations for Malay learners of Mandarin. This is likely because of

Malay learners' experience with both the Malay and English languages, which do not have such consonants.

Failing to add aspiration in Mandarin may lead to misunderstandings during a conversation. For example, in the sentences 肚子饱了(dù zǐ bǎo le) and 兔子跑了(tù zǐ pǎo le), the first sentence means "stomach is full". Meanwhile, the second sentence means "the rabbit has run away." This shows that the aspirated voiceless stop is a crucial phonological element for the listener of Mandarin, such that communication becomes ineffective without the appropriate aspiration. According to Morley (1991), the most crucial element of communication competence is intelligible pronunciation. Therefore, it is imperative that Mandarin teachers do not neglect the teaching of pronunciation to ensure students' communicative competence.

In this regard, Tchoshanov (2011) emphasized the significance of teachers having a solid understanding of their subject matter for students' academic growth because teachers are students' primary sources of information. Therefore, without adequate preparation, it will be difficult for teachers to pass along knowledge to students. Additionally, teachers' teaching practices reflect their beliefs and knowledge. As such, teachers who currently teach Mandarin to elementary students must be able to properly convey correct sound production as a fundamental phonological skill. Without proper pronunciation training, students' verbal intelligibility and communication proficiency may continue to be impaired.

Despite these challenges, limited research has been done in Malaysia on the production and perception of Mandarin stop consonants. Therefore, in an attempt to bridge this gap in the literature on the phonetic features of Mandarin, this study aimed to explore the production and perception of aspirated and unaspirated voiceless stop consonants in Mandarin by Malaysian Malay learners. Specifically, this preliminary research sought to investigate Mandarin stop consonant realization by Malaysian Malay teachers who are currently teaching Mandarin in schools. Mandarin teachers from the Malay ethnic group were selected to participate in this study to evaluate their awareness of Mandarin stop consonant production and perception as well as their understanding of the interference of different sound systems on Mandarin pronunciation for application in their classroom.

1.3 Research Objectives

This research aimed to investigate how Malaysian Malay L2 learners perceive and produce Mandarin stop consonants through the following objectives:

1. To examine the perception of Mandarin stop consonants by Malaysian Malay learners.
2. To examine the production of Mandarin stop consonants by Malaysian Malay learners.

1.4 Research Questions

In accordance with the research objectives, the research questions were as follows:

1. Do Malaysian Malay learners differentiate Mandarin voiceless unaspirated stop consonants from aspirated stop consonants in auditory perception?
2. What are the acoustic characteristics of Malaysian Malay learners' Mandarin stop consonants?

1.5 Significance of Study

The primary goal of this study was to examine Malaysian Malay learners' ability to recognize and generate Mandarin stop consonants. The importance of the study stems from the fact that there is a shortage of second language acquisition research on Malay learners, specifically on their perception and production of Mandarin consonants. Therefore, the findings in this study will add to the extant understanding of how Malay speakers in Malaysia perceive and produce Mandarin stop consonants, thereby facilitating their successful learning of Mandarin as a second language. This study would further guide academics in conducting further research to expand the body of knowledge in this area.

The role of teachers is seen as essential in the effective learning of their students. However, Mandarin teachers have limited resources to refer to when teaching pronunciation to Malay learners. Ideally, this study will help Mandarin teachers be more aware of how Malay learners' language background may influence their production and perception of Mandarin stop consonants. Through the present findings, Mandarin teachers can gain knowledge of the phonological differences between Malay, English, and Mandarin as well as the significant problem areas for learners in consonant pronunciation. As a result, they can teach Mandarin pronunciation effectively and assist non-native speakers in mastering Mandarin. This study also helps teachers stay abreast of the latest literature on L2 phonology acquisition so that they can develop teaching materials that are up-to-date with current findings. Undoubtedly, their increased awareness and knowledge would improve their second language teaching skills. Finally, the findings of this study highlight the need for Mandarin teachers to receive training, theoretical instruction, and experience in pronunciation teaching in Mandarin for them to be able to effectively teach Mandarin sound production.

1.6 Limitations

Every research is constrained by limitations that may affect its findings. The first limitation of this study is its narrow population and sample size, which both limit the generalizability of the findings. More specifically, this study focused only on Malay learners of Mandarin, of whom only a limited number participated in the study. This means that although the results of this study demonstrate key trends in the perception and production of Mandarin stop consonants by Malaysian Malay learners, they cannot be generalized to the total population of Malay L2 learners or, alternatively, to those learning Mandarin as a second language. Moreover, learners from different regions and backgrounds may exhibit different issues in sound production and perception. As a result, more learners from various races and locations should be recruited for future research on Mandarin stop consonant pronunciation.

The second limitation is that the productions in this study were elicited using a list of only 18 words. Expanding the breadth of this study beyond that list was outside the scope of a Master's thesis. Such an effort would most probably have needed the inclusion of extra reading material and the analysis of a larger number of speech samples. A broader study of this nature can be undertaken in the future to obtain more accurate results.

The third limitation is this study is that it only examined six Mandarin stop consonants due to the limited scope of the Master's research. There are other interesting consonant and vowel sounds that should be studied in the field of Mandarin pronunciation problems.

1.7 Organization of Study

There are five chapters in this research. The purpose of the study is discussed in detail in the first chapter. In the second chapter, the related literature on the topic under study is reviewed. The third chapter details the methodology that was used to obtain and analyze the data. In the fourth chapter, the study's findings are presented and discussed. In consideration of Malaysian Malay speakers who are learning or teaching Mandarin as a second language, the fifth chapter summarizes the key findings and examines their implications.

CHAPTER 2: LITERATURE REVIEW

The concepts of first language, second language, third language, and foreign language are explored and compared in detail in the first part of this chapter. The second section discusses language profiles in Malaysia. Subsequently, the third section discusses foreign accents and production, while the fourth section discusses L2 speech production and perception. The fifth section then delineates consonants in the Malay language, English language, and Mandarin language. Next, the chapter offers explanations of the concepts of Voice Onset Time (VOT), stop consonant, and aspiration. Finally, a detailed review of studies on L2 Mandarin consonants is presented.

2.1 Language Acquisition

2.1.1 First Language (L1)

A person's first language is also known in many other ways, namely as his or her mother tongue, native language, or primary language. A bilingual person's native language is called L1 and, in most cases, is the person's dominant speech medium (Taylor, 1990). In this study, L1 indicates the Malay language of Malaysia, encompassing its various dialects (e.g., Sarawak dialect, Kelantan dialect, Terengganu dialect, standard Malay, etc.). The crucial characteristics that all varieties of L1s have in common are, first, they are thought to be languages that are absorbed in early childhood (i.e., typically before the age of three), and second, they are acquired as a result of growing up around speakers of the language.

2.1.2 Second Language (L2)

A person's second language, or L2, is the language apart from his or her native language that he or she has acquired or learned from the environment, for example, one's nation or locale, where the L2 is predominantly used. L2 can refer to both the study of those who learn a language after their L1 as well as the process of acquiring the language. Ellis (1994) mentioned that L2 plays an institutional and social role in the community. That is to say, the second language may not be widely used in educational or government circumstances, but it is frequently used as a communication tool in people's everyday life.

2.1.3 Third Language (L3)

Third language (L3) refers to the languages learned after the first (Ll) and second (L2) languages are mastered. Regardless of whether a language is the third, fifth, or even tenth to be acquired, it is still referred to as L3 (Jorda, 2005). According to Hammarberg (2010), when considering a multilingual person's linguistic state, L3 is defined as a non-native language utilized or learned when the individual has already acquired more than one L2 and L1. Thus, L3 acquisition indicates learners who are learning one or more languages in addition to their mother tongue and second language (Fouser, 1995; Jessner, 1999). In other words, apart from the learner's native tongue and acquired (possibly incomplete) second language, any language (one or more) he or she is learning is the so-called third language. Jorda (2005) concluded that a third language is a universal concept without a specific number of languages.

2.1.4 Foreign Language (FL)

A foreign language refers to a learned language that is not native to its learners and is not used to communicate in the learners' place of origin (Freudenstein, 1979). It is also known as a language that is learned beyond the boundaries of the learner's culture (Tse, 2000). A foreign language is often taught in the school classroom or in adult courses exclusively for learners to gain competency in the language, which enables them to employ the language in various contexts, such as reading literary and technical works, listening to auditory media, understanding film dialogues, and speaking with others

(Marckwardt, 1965). Despite being taught in school, foreign languages do not hold important roles at the national or societal levels, such that it is not necessary for a typical citizen to know a foreign language for his/her daily life, social activities, or career advancement (Broughton, 1978).

2.1.5 Foreign Language vs Second Language (L2) vs Third Language (L3)

Despite some shared traits, L3 acquisition differs from L2 acquisition in that it is more complex because of the acquisition context, differences in the sequence in which the languages are learned, the perceived gap between languages, and the sociocultural standing of the languages being learned (Cenoz, 2000). Interestingly, Herdina and Jessner (2000) compared L3 learning to L2 learning, reporting that only L3 learners build new capabilities in the form of language learning, language management, and language maintenance.

Meanwhile, Derakshan and Karimi (2015) found that L2 and foreign languages have certain similarities. Both languages are acquired by adult learners at a later stage apart from their L1 or mother tongue. In some regions and literature, a foreign language is not clearly distinguished from a second language, such that both terms are considered to be the same. A foreign language that is called a "second language" indicates a bilingual context wherein speakers actively utilize their L2 as well as a foreign language in their daily communication (Freudenstein, 1979). This generally occurs when the L2 is the instructional medium in the classroom and a lingua franca among speakers of highly varied languages (Marckwardt, 1965) in places of trade, administration, and education (Broughton, 1978). A second language can also be employed for geographical or social purposes (Marckwardt, 1965). Notably, a "second language" can represent any language learned after one's mother tongue (Tse, 2000), even if it is the second, third, fourth, or subsequent language (Ellis, 1997; Liu, 2005).

Although L2 and foreign language have been used interchangeably in many contexts, there are some distinctions between the two (Ellis, 1994; Stern, 1983). While Derakshan and Karimi (2015) pointed out the status and importance of L2 in a social context, Saville-Troike (2006) discussed that the majority of people learn a foreign language in a formal classroom setting and it is not commonly used in the social context. Therefore, one way to differentiate between an L2 and an FL is to consider whether it has a good language environment (Moeller & Catalano, 2015). That is, the foreign language is usually not spoken in the surrounding community, while the second language is usually used in the surrounding community. Accordingly, Stern (1983) pointed out that foreign language does not have official status, whereas L2 is accepted and recognized. Thus, the term "second language" is distinct from the term "foreign language" (Littlewood, 1984; Liu, 2005).

According to Hoque (2017), the process of acquiring languages in addition to one's mother tongue is known as second language acquisition, or SLA. It is the empirical investigation of how individuals pick up languages other than their native tongue. In the late 20th century, most linguists supported the view that L2 and L3 acquisition were the same, such that any language acquired after L1 is considered L2 (Myles, Hooper, & Mitchell, 1998; Singh & Carroll, 1979). However, scholars like Hufeisen and Marx (2004) and De Angelis (2007) contended that learning L3 should not be compared to learning L2 because there are both quantitative and qualitative differences. Additionally, treating them as equals could result in the exclusion of crucial elements of third language acquisition (TLA). As per TLA scholar Cenoz (2013), TLA differs from SLA in the following manner:

"TLA shares many of the characteristics of SLA, but there are also important differences because third language learners already have at least two languages in their linguistic repertoire. Third language learners can use this broader linguistic repertoire when learning a third language. For example, they can relate new structures, new vocabulary or new ways of expressing communicative functions to the two languages they already know, not just to one of them, as in the case of monolinguals" (Cenoz, 2013, p. 4).

With regard to the various language types and acquisitions, Mandarin is considered the native language or L1 of the Chinese ethnic group in Malaysia. However, with the impressive growth of China's economy, non-Chinese Malaysians are becoming more and more interested in learning Mandarin. Currently, Mandarin is the instruction medium in national primary Chinese schools, which are part of the central education system. The Mandarin language is also offered as a second language in national primary schools. This program started in 1996 and expanded in 2003, with Malays and Indians as the target students. In addition, according to Hoe and Mah (2011), the Malaysian government has identified Mandarin as a second language that is required to be taught to Malaysian graduates. Therefore, Mandarin is considered a second language or L2 for Malay and Indian ethnic groups because it is learned after their mother tongues.

2.2 Language Profile in Malaysia

2.2.1 Malay Language

The Malay language is part of the Austronesian family of languages. It developed as a medium of communication for education, religion, law, diplomacy, and business during the 16th century. For hundreds of years, Malay has been used as a regional language in Indonesia, Malaysia, Sumatra, Borneo, and the Malay Peninsula, allowing free trade and cultural exchange. Over 270 million people now use it in Singapore, Indonesia, Brunei, and Malaysia, with each country having its own unique version.

After Malaysia achieved its independence in 1957, Malay became the nation's official language. According to Article 152 of the Malaysian Constitution, the Malay language, as the national language, is indisputable in its purpose and role. However, the native languages of other races in Malaysia, such as Mandarin for the Chinese and Tamil for the Indians, are free to be spoken. According to Watson (1980), the status of Malay as the national language is among the foundations of the endorsed social contract between Malays and other races in Malaysia, which was codified in the National Language Act 1963/1967 mandating the use of Bahasa Malaysia for official functions (Article 152). The Malay language is thus recognized as a building block of the country, particularly in its position as a channel of unity and development, as well as the country's formal language of administration and medium of instruction in the national educational system.

2.2.2 English Language in Malaysia

As far back as the late 18th century, Malaysia was under British administration. Both East and West Malaysia, the latter of which includes the states of Sabah and Sarawak, were colonized. The British anticipated that non-Europeans would have to be English-educated as business, trade, and industry expanded. The British empire then allowed private and mission schools to be launched, where English would be employed as the instruction medium. As the number of these schools expanded, teachers were selected from the local population and taught in local English. Consequently, English was utilized on a daily basis by local citizens (Malays, Chinese, and Indians) engaged in the British Empire's administration in Malaya.

After independence, Malaysia's official national language became Bahasa Melayu (Malay Language), while English was relegated from an alternate official language to a second language under the Language Act of 1967. Indeed, in American and British ex-colony nations (e.g., Malaysia, India, Philippines, and Nigeria), English is considered a second language (L2) (Thirusanku & Melor, 2012). Asmah Hj Omar (1983) stated that English is ranked second by importance among Malaysian languages because of its relative significance to international relations and education. Despite the fact that it is no longer an official language, English holds a unique status in Malaysia. That is, it is a mandatory subject in the country's education system, taught from Standard One to Form Five. Therefore, English language learning takes place in all classrooms in Malaysia.

2.2.3 Mandarin Language as First Language

The standard Chinese language, sometimes referred to as Mandarin, adopts the sound production system of the Beijing dialect as its model of phonology. Mandarin is one of the ten major dialects in China and is widely spoken in northern China, especially in Beijing. Mandarin formally became the national language of China after Dr Sun Yat Sen deposed the Qing Dynasty in 1911.

Currently, Mandarin is the official language in Mainland China and Taiwan and is broadly employed in mass media, schools, and all official functions. There are roughly 1.31 billion speakers of Mandarin (McCarthy, 2020). In Mainland China, Mandarin is called ''Pu tong hua" which means "'Common Speech", while in Taiwan, Mandarin is referred to as "Guo yu", which means "National Language." In Malaysia and Singapore, Mandarin is known as "Hua yu" or "Chinese Language." It has become one of the core symbols of identity not only for Chinese communities in China but also for Singaporean and Malaysian Chinese.

Chinese schools in Malaysia use Mandarin as their primary language, yet the majority of Chinese are actually proficient in Hakka, Cantonese, Hokkien, Teochew, Hainanese, and Hokchiu rather than Mandarin. As far as dialects go, Hokkien, Cantonese, and Hakka can be found in Penang, whereas Mandarin and Hakka are found in the southern parts of West Malaysia and Sabah and Sarawak. Nonetheless, since Chinese schools teach in Mandarin, parents typically speak Mandarin to their children from early childhood, even before they reach schooling age, to ensure they can manage their lessons well. In summary, Mandarin is still considered the first language of the Chinese ethnic group in Malaysia.

2.2.4 Mandarin Language as Second Language

The past decade has witnessed the increasing popularity of learning Mandarin as L2 or FL around the world. In Malaysia, learning Mandarin has also become much more popular nowadays, with the number of students learning Chinese at tertiary educational institutions increasing year by year. This has resulted in the establishment of Chinese classes at Malaysian institutions of higher learning. According to Cheun (2006), non-native speakers in Malaysia learn Mandarin in one of the following four places: 1) Chinese vernacular primary schools; 2) national primary schools; 3) tertiary educational institutions; and 4) private language centers.

In Malaysia, early research on the Mandarin language mainly discussed the motivation of Malay students to learn Mandarin and the development of teaching methods for non-native speakers of Chinese (Teow, 2016; Hoe, 2011). In his study, Huang (2008) compared Mandarin language courses offered by four different universities. The research discussed the number of students enrolled in Mandarin classes, teaching and learning practices, curriculum, teachers, and teaching achievements. This was the first article to compare Mandarin courses across local universities. Some of the main issues that Huang (2008) found were non-uniformity in the Mandarin syllabus, inappropriate textbooks, shortage of teachers, teachers' unprofessionalism, and insufficient teaching hours. He also mentioned that there was no department in the universities that could coordinate Mandarin teaching for non-native learners.

The Malaysian government acknowledges the significance of Mandarin, as evidenced by the language's inclusion in national policy in the National Higher Education Action Plan 2007-2010. As per the Plan, it is mandatory for students in higher education institutions to take up a complementary language other than the national Malay language and L2 English. This initiative was developed under the Malaysian Education Blueprint 2015-2025 (PPPM) via the incorporated Cumulative Grade Point Average (iCGPA) Rubric Learning Outcomes Assessment Guide (Ministry of Education Malaysia, 2015). The PPPM underscores language proficiency in Malay and English as one of its six primary attributes while also encouraging the adoption of an additional second language.

The Malaysian tertiary education system's syllabus for L2 Mandarin instruction differs from the syllabus in China, where language teaching is broken down into the four language aptitudes, i.e., listening, speaking, writing, and reading. In contrast, Malaysia adopts a language teaching methodology that combines all language skills without an emphasis on a specific one (Zhou, 2010, as cited in Hoe & Lim, 2015). In doing so, the Malaysian approach to teaching Mandarin awards more importance to learners' communication and interaction. The disparity between both countries' approaches and focus areas can be explained by the Mandarin language's role in the respective nations; in China, it is the native language, and in Malaysia, it is an L2.

According to Hoe and Mah (2009; 2011), Malaysian undergraduate students share four identical characteristics that best describe their common language profile: 1. Their first language or mother tongue is Malay; 2. Malay is utilized for teaching and learning in primary and secondary education; 3. English is learned as a second language; and 4. They have no or minimal exposure to the Mandarin language.

2.3 Foreign Accent and Production

It has been 30 years since the seminal article was written by Asher and Garcia (1969) about foreign accents in L2 learners' speech during learning. Since then, an expanding body of comprehensive research has emerged on the perception and production of foreign accents in the speech of L2 learners. In general, speakers of a language are categorized into two groups, which are L1 and L2. L1 speakers usually acquire the native language since birth and use it throughout their lives. Meanwhile, L2 speakers acquire the new language at a later stage, usually in a formal classroom or in a foreign country. Speech produced by the non-native L2 speaker will show signs of "a foreign accent." A foreign accent denotes a method of speaking a language that is distinct and different from the way most native speakers of the language do. In this context, foreign language is learned largely in a classroom and not spoken in society. Many linguists have suggested that a learner's foreign accent is influenced by his or her first language. According to Flege (1984), differences in segmental articulation between native and non-native speakers contribute to the foreign accent.

One of the main objectives of a language student is to achieve pronunciation like native speakers in learning a second or foreign language. However, adults who acquire a second language or foreign language in adulthood often struggle to acquire new sound categories. Towell and Hawkins (1994) posited that the majority of L2 learners would never reach the level of capacity of native speakers and often retain a foreign accent even after many years of experience in the L2 language. Correspondingly, Lane (1962) stated that adult learners are unable to be as fluent in a second language's sound patterns as native speakers despite having rigorous distinction training, a vastly articulate vocabulary, and a high level of control over their language learning regime. In other words, although many foreign language students have already mastered the grammar, vocabulary, writing, and reading aspects of the second language, they are still unable to produce native-like pronunciation. Rather, foreign accents can be detected when the learners produce L2 speech sounds that differ substantially from the phonetic system of the L2.

The driver of foreign accents among L2 learners is a topic that has been comprehensively examined. According to Piske, MacKay, and Flege (2001), the extent of a foreign accent is affected by both the age of L2 learning and the level of sustained

L1 usage. Age is indeed frequently cited as an important factor in understanding L2 achievement. Piske et al. (2002) and Mackay et al. (2001) concluded that students who begin learning L2 speech before the critical point of 12 years old perform better in L2 vowel and consonant perception and production than those who learn the language later in life. The former is also known to have a weaker foreign accent (Flege et al., 1995). It has further been found that how well a learner can hear and write in a new language is affected by how well their L1 and L2 phonetic systems work together. This is called the phonetic category of assimilation and disassimilation.

Previous studies about Chinese pronunciation started with research about the problems of teaching Chinese as Foreign Language (CFL) learners pronunciation. Zheng, Song, Fung, and William (2002) proposed that several features make Mandarin different from other languages, such as its retroflex, tones, and syllables' short initial and final structure. As a consequence of these particular features of Mandarin pronunciation, very few Mandarin language learners can adopt native-like or near native-like pronunciation even after a long time of Mandarin language study. Notably, the Mandarin language phonetic system is different from that of the Malay and English languages, which brings many difficulties to Malaysian Malay students in learning Mandarin. Specifically, they often find it difficult to pronounce certain Mandarin stop consonants, especially the aspirated ones.

2.4 L2 Speech Production and Perception

Language study focusing on the spoken (rather than written) word has traditionally been divided into two independent fields: speech perception and speech production. Language learners must be able to comprehend and generate sounds in their second language to be successful in learning it. Compared to the situation in L1 production, L2 learners' understanding of the target language is often incomplete, as these learners typically do not possess the essential linguistic competence to convey their desired message. The difficulties associated with perceiving and creating L2 sounds have been reported to persist throughout time and among advanced learners. One possible explanation for this issue is that adult learners are highly dependent on patterns and categories from their L1 when learning to recognize and generate L2 sounds. According to Fledge (1995), adult L2 speakers' difficulties in perceiving and producing non-native speech signals are significantly affected by their native language (i.e., L1).

2.4.1 Speech Production

Beginning with an idea in the speaker's head, speech production is an iterative process that culminates in the transmission of that idea via speech. Coordinating multiple muscles and advanced cognitive functions are necessary for speech production, as it is one of the most complex human activities. A common description of production is essentially the same process as retrieval, starting with accessing a semantic representation, moving on to a lexical representation, and finally, a sound structure before a word is produced using articulators and syllables. A linguistically (i.e., phonologically) encoded message is transmitted orally through the oral-articulatory system during the process of "production" (or, more simply, speaking). An early description of how one's L1 affects his or her ability to hear and produce foreign sounds comes from Polivanov (1931), according to Bohn (1998). Polivanov postulated that learners' deviant productions are the result of the "subjective nature" of sound perception, which is based on the complexity of language habits each person learns while acquiring their mother tongue.

2.4.2 Speech Perception

Accurate speech perception is compulsory for the successful advancement of communication abilities. As a result, the subject of speech perception development among L2 learners has gained a central focus in the L2 literature. In the context of perception, sounds are categorized according to their phonological structure as they are processed. To correctly recognize a phonetic feature, a learner's underlying phonological system must have a distinct category for it. Numerous studies have demonstrated that one's native language's phonological system has an effect on how phonetic segments, particularly consonants, are perceived in a second language. Early on, Flege and Port (1981) contrasted the phonetic application of stop voicing in English between Saudi Arabians and Americans using the Arabic language as a case study. English stops produced by Saudi speakers were found to have similar temporal acoustic stop voicing as Arabic stops, including VOT, stop closure duration (SCD), and vowel duration (VD). The Americans, in turn, had a hard time identifying the stops made by Saudis, except for /p/, due to the phonetic interference between Arabic and English languages. At stop closure interval times in English, glottal pulsing is frequently used to produce this phoneme, which is not found in Arabic. Though they may have grasped the phonological quality of /p/ (i.e., that the difference between /p-b/ is akin to that between /t-d/ and /k-g/), they could not manage the articulatory dimensions through which this sound was made. According to Aoyama et al. (2004), the most difficult English minimal pairs for Japanese speakers are "rocket" and "locket." The root reason for these issues is that these specific sounds do not contrast in the L1 phoneme repertoires of these students.

According to Strange (1995), numerous problems and variables must be accounted for in predicting the perceptual difficulties encountered by L2 learners. L2 learners face perceptual difficulty when trying to differentiate speech sounds that contrast with their L1 or mother tongue. Nonetheless, many empirical research works have established that L2 students perform better in perception than in production when acquiring the L2 (e.g., Trofimovich & Baker, 2006; Flege, 1993; Flege, Bohn & Jang, 1997; Flege, MacKay & Meador, 1999; Cardoso, 2011).

2.4.3 Relationship between the Perception and Production of Speech

While speech perception and production are frequently examined independently, a large number of studies have demonstrated that these two areas are interrelated. Consequently, how L2 sound categories are seen in relation to L1 sound categories raises questions on how L2 speech perception affects its production. The Perceptual Assimilation Model (PAM) and Speech Learning Model (SLM) both emphasize this connection.

Researchers have reached a variety of conclusions about L2 speech perception and production based on prior studies. Among these conclusions, scholars agree that to master the phonetic expressions of an L2, it is necessary for a learner to master both the perception and production of the L2's sounds. According to Zhang and Yin (2009), a phoneme that is unfamiliar to a learner is frequently perceived as "alien," but once a learner is able to hear a sound, the correct production of the phoneme is possible. In contrast, if a student is unable to perceive a sound, he or she will be unable to produce it. As a result, the perception issue must be addressed first prior to production. However, according to Goto (1971) and Sheldon and Strange (1982), research has also shown that the successful learning of L2 production may precede L2 perception, particularly in perceiving and producing the English /r/ and /l/ by Japanese speakers. Arguments for both these cases, i.e., the precedence of perception over production and vice versa, are presented below.

2.4.3.1 Perception Precedes Production

Most research on L2 phonological learning has found that perception of L2 sounds occurs before L2 production, presumably because the necessary sensorimotor skills for L2 production can be acquired only after the successful perception of L2 sounds (e.g., Flege, Schirru, & MacKay, 2003; Piske, MacKay, Flege, 2001). Substantial cross-sectional research has observed the significant effect of perception on vowel and consonant production, as well as the inclination of L2 learners to perform better in perceiving than in producing (e.g., Baker & Trofimovich, 2006; Cardoso, 2011; Flege, 1993; Flege, Bohn, & Jang, 1997; Flege, MacKay, & Meador, 1999). In particular, the work done in support of Flege's (1995) Speech Learning Model has been highly beneficial in laying the groundwork for this connection. According to the model, if an L2 sound differs from an existing L1 category, new phonetic categories can be easily formed by learners. In their study of Japanese learners, Aoyama et al. (2004) tested the difference in the learning of the English /r/ and /l/. Compared to the Japanese /r/, the English /r/ is more distant than /l/ (which is close to Japanese /r/), which leads to the prediction that it should be possible for Japanese learners of English to more easily develop a new category for the English /r/ than for the /l/. Aoyama et al. (2004) found this to be the case, with evidence of significant improvements in perception over time and, more importantly, improvements in production, offering evidence of the perception–production link. Similarly, Evans and Alshangiti's (2018) recent work on native Arabic speakers' British English vowel and consonant acquisition revealed a link between perception and production, as those who had stronger perception ability with regard to English vowels also had stronger production ability.

2.4.3.2 Production Precedes Perception

Conversely, a number of studies also support the idea that good pronunciation can be achieved without having prior perceptual mastery. In other words, L2 learners are capable of producing sounds that they cannot perceive adequately. Studies done by Borrell (1990), Neufeld (1988), and Brière (1966) have found that the accurate perception of a sound does not guarantee its correct production when learning an L2. Indeed, numerous scholars

(Sheldon & Strange, 1982; Tsukada et al., 2005; Yamada et al., 1994; Linebaugh & Roche, 2015; Flege & MacKay, 2004; Kluge et al., 2007) have found that the production of L2 sounds can precede and also aid perception.

For example, Sheldon and Strange (1982) examined the English /r/ and /l/ production–perception linkage among Japanese adults learning English in the United States. Their study concluded that native Japanese speakers who are able to produce /r/ and /I/ accurately may still erroneously perceive the contrast. The authors mentioned that perceptual proficiency in a foreign contrast does not always influence adult learners' production of appropriate symbols of the contrasting phonemes; rather, perceptual mastery can, at times, be less developed than production mastery. Along the same lines, Tsukada et al. (2005) examined the production and categorial distinction of the English / ε / and / æ / by native Korean adults and children. The author found that the ability of Korean children to produce English vowels surpassed their ability to perceptually discriminate the same vowels. The author concluded that L2 learners are capable of accurately producing L2 vowels prior to honing their perceptual ability of L2 vowels, akin to natives. According to another study done by Yamada et al. (1994), some Japanese learners are successful in producing distinct /r/ and /l/ categories despite the fact that they cannot reliably identify native tokens. The author concluded that their production abilities exceed their perception abilities. In another study done by Linebaugh and Roche (2015), the results show that explicit training in producing difficult sounds in a second language can improve the perception of those sounds.

In light of these findings, it is reasonable to conclude that producing abilities can outweigh perception abilities and hence, that production can come before perception.

2.4.3.3 Summary

Even though the SLM proposes that there is a substantial relationship between speech perception and production, the empirical evidence for such a strong relationship is inconclusive. When it comes to L2 acquisition, in particular, no definitive evidence has been discovered as to whether perception precedes or follows production. However, it is well-acknowledged that the production–perception link is complex and contingent on several factors, such as L2 experience (Llisterri, 1995).

2.5 Theoretical Model of Second Language Speech Learning

Most research on L2 phonology has, in the past, been concerned with establishing the patterns of L1 influence on the L2 sound system. Numerous theoretical models have been presented for this purpose, all of which indicate that prior L1 language expertise influences adult listeners' capacity to detect non-native speech sounds via a perceptual framework, either positively or negatively.

Second language (L2) speech learning models are developed in a manner that considers the phonological acquisition of the L2. In unearthing how these cross-linguistic effects occur in L2 phonetics, various theory-based frameworks have been constructed for L2 speech learning. Eckman (2004) stated that such models generally fall under two key categories: (a) learners' first language (L1) or (b) the general traits of natural language phonologies. According to Gass (1996), the native language's impact on the acquisition of the second language (SLA) is undisputable, and multiple L2 phonologists have conceptualized in-depth theories to explain this effect.

The two foremost frameworks on the impact of L1 on L2 speech learning are the Speech Learning Model (SLM; Flege, 1995, 2003) and the Perceptual Assimilation Model of Second Language Speech Learning (PAM-L2; Best & Tyler, 2007). In this study, the focus was on the Speech Learning Model.

2.5.1 Speech Learning Model (SLM)

Flege (1995) and Flege (2003) aimed for the SLM to serve as an SLA theory that functions beyond the critical time points of acquisition. Their studies frequently examined the way factors like acquisition age or length of stay influence the learners' capacity to achieve sound production and perception in their second language. Furthermore, Flege's body of work covers both production and perception, in contrast to the majority of the other studies covered in this chapter, which are primarily concerned with perception. The SLM focuses, in particular, on foreign speech learning with the main aim of delineating variations in segmental acquisition (production and perception) throughout learners' lifespans. The SLM makes two broad assumptions: 1) bilingual speakers' L1 and L2 phonetic sub-systems are in some way interdependent via a shared "phonological space"; and 2) the abilities that precede successful L1 speech acquisition remain the same throughout learners' lifespans. These abilities include the accurate recognition of a spoken language's featural patterns, the categorical organization of myriad phonetic segments with common features, and the association of speech production with recognized speech input features.

The SLM further classifies L2 sounds into three distinct categories based on how they are perceived:

1) Similar sounds: L2 sounds are considered to be realizations of L1 sounds (diaphones).
2) Identical sounds: Sound similarities are perceived as an equivalence between L1 and L2 (transfer).
3) New sounds: Phonetics that do not match the native language's phonetics and thus constitute a new phonetic category.

Going further, the Fledge Speech Learning Model lays out seven hypotheses that are derived from four postulations (Flege, 1995). According to the first hypothesis, L1 and L2 sounds are associated on a phonological level, and interlingual identification occurs on an allophonic level. The second hypothesis proposes that a bilingual can only create a new categorization for an L2 phonetic when they can distinguish L1 from L2 sounds by observing differences in how the mouth is shaped and positioned in both languages. This means that a student will not generate a new sound category for the foreign language if the English phoneme / ʃ / is believed to realize one of the Mandarin allophones of / ɕ / or a new allophone of the L1's / ɕ / phoneme. Next, the third hypothesis states that people find it simpler to learn L2 phonological systems that are different from their L1 because there are more variances in perception, making it easier to notice the differences between sounds. The fourth hypothesis subsequently posits that as L2 learners age, they lose the capacity to distinguish L1 sounds from L2 sounds. As a result, those who begin to acquire a second language when they are older will be less able to differentiate between the two languages. According to the fifth hypothesis, the equivalence categorization mechanism related to a collection of diaphones in L1 and L2 may prevent the formation of L2 category phonetics. However, a bilingual L2 speaker may have different phonetic categories for L2 sounds than a monolingual speaker because of influences from L1. For example, speakers of Spanish who can also communicate in English as their second language may have higher VOT scores for standard L1 stops compared to Spanish speakers who only speak Spanish as their first language. As a result, the sixth hypothesis is that it is no longer predicted that native-like pronunciation is achievable in the production of specific sounds. Finally, a sound's production will be in accordance with its phonetic category representation, according to the seventh hypothesis.

Taking everything into account, this theory is particularly relevant because it focuses on the acquisition of L2 skills specifically. This model claims that L2 speech cannot be accurately produced without an accurate perception. Producing native-like sounds, therefore, has a direct correlation with the learner's ability to perceive in a native-like manner. This theory can answer Research Question One on whether Malaysian Malay speakers' perceptions of L2 Mandarin unaspirated and aspirated consonants correlate with their actual production.

Following this theory, it is imperative to discuss the similarities and differences between Mandarin, Malay, and English. The tenets of the SLM state that any such parallels or discrepancies undoubtedly have an impact on the way L2 Malay Mandarin speakers produce and perceive Mandarin consonants. In terms of the real distinctions across these three groups, the majority of Mandarin stop consonants differ from Malay and English stop consonants in terms of aspirated and unaspirated consonants. Therefore, this study examined all stop consonants produced and perceived by L1 Malay Mandarin learners and attempted to make comparisons of these three languages due to the scarcity of research in this area. Indeed, these categories can answer Research Question Two, which is whether new sounds are easy to produce for Malay speakers and whether similar sounds that occur in the Malay language and English language are difficult for them to produce.

In summary, the purpose of the SLM is to offer a predictive framework for the potential acquisition of new L2 phonetic categories, both for L2 perception and L2 production.

2.6 Consonant System of Malay, English, and Mandarin Languages

Given the multilingual nature of the Malaysian education system, it is likely that students in Malaysia would gain shared sounds among Malay, English, and Mandarin. This means that multilingual Malaysian speakers who simultaneously learn all three languages may have distinct predicted phonological acquisition patterns compared to monolingual learners of any one of those languages.

Therefore, it is imperative to scrutinize the contrasts of Mandarin consonants with English and Malay consonants. First, this study focused on the consonant system of "Standard Malay". Second, in consideration of the various native styles of English around the globe, the present analysis zoomed in on the consonant system of "Standard English" spoken in Malaysia. Lastly, this study reviewed the Standard Chinese spoken in Malaysia. In describing the three languages' sounds, the International Phonetic Alphabet (IPA) was employed.

2.6.1 Consonant System of Malay

The Malay language has 19 native consonants and eight consonants loaned from Arabic and English, which are pronounced in the same manner as English consonants (Wai, Siew, & Roziati, 2007). The sounds of Malay consonants are rather simple as pronouncing them involves only the supraglottal organs (just two glottal sounds), without any pharyngealized sounds (heard in Arabic) or clicks (heard in some African dialects). Nevertheless, the Malay sound system has undergone significant change following language transfer and loans from Arabic and English (see Table 2.1). It now has 26 consonant sounds and six diphthongs.

Table 2.1: Consonant system of Malay language

Manner	Voicing	Place of articulation								
		Bilabial	Labiodental	Alveolar	Post Alveolar	Palatal	Velar	Labial Velar	Uvular	Glottal
Plosive	Voiceless	p		t			k		q	ʔ
	Voiced	b		d			g			
Fricative	Voiceless		f	s	ʃ		x			h
	Voiced		v	z			ɣ			
Affricate	Voiceless					ʧ				
	Voiced					ʤ				
Nasal	Voiced	m		n		ɲ	ŋ			
Trill	Voiced			r						
Approximant	Voiced					j		w		

2.6.1.1 Malay Dialects (Kelantanese and Terengganu Dialect)

(a) ***Kelantanese Dialect***

Abdul Hamid Mahmood (1990), quoting Mario Pei, described a dialect as a variant of a particular language spoken by a cohort of speakers in a single language population. One of the dialects that were mentioned by the participants of this study is the Kelantanese dialect. Kelantan is a state in Malaysia located on Peninsular Malaysia's east coast. It shares a northern border with Thailand. Apart from the Kelantanese, people who live near the borders and in South Thailand's Yala, Sungai Golok, Patani, and Narathiwat districts also speak the Kelantanese dialect (Abdul Hamid, 1993).

The overall number of consonants in these two languages and the placement of consonants within words are the two most obvious characteristics that distinguish the Kelantanese dialect from standard Malay. The standard Malay language uses 25 consonants according to its consonant inventory, whereas the Malay sub-language of the Kelantanese dialect uses 20 consonants (Ajid, l985). The basic consonants used in the Kelantanese dialect are similar to those used in standard Malay (Yunus, 1980), with the exception of /r/, which in the Kelantanese dialect, is velar fricative rather than a roll approximant as it is in standard Malay.

(b) ***Terengganu Dialect***

The Terengganu dialect is spoken predominantly in the Terengganu state, which is situated on Peninsular Malaysia's east coast. Due to migration, some speakers are also found around the borders of the neighboring state of Pahang, as well as further afield in Mersing in the state of Johor.

The Terengganu dialect has 19 consonant phonemes and eight vowel phonemes. According to Ridhuan et al. (2021), there are seven plosive consonants in the Terengganu dialect, which are /p, b, t, d, k, p, and ʔ/. To sum up, the plosive consonants of the Kelantan dialect and the Terengganu dialect are similar to that of the Malay language.

For this study, there are no differences in sound characteristics regarding plosive consonants, which is between the Kelantanese dialect, Terengganu dialect, and Standard Malay. Therefore, for the participants whose first language is Malay dialect, it does not influence the outcome of this research.

2.6.2 Consonant System of English

The English language comprises 23 consonants, excluding glides and semi-vowels. In terms of articulation manner, English possesses two sets of obstruents (i.e., stops, affricates and fricatives). Each set is classified into two categories based on voice features, that is, the presence (voicing) or absence (voiceless) of the vibration of vocal cords. Apart from that, English is made up of nasals, liquids, and glides. In terms of articulation place, English employs the full range of sounds, from bilabials in the front to the glottal at the back. Table 2.2 illustrates the consonant sound system of English.

Table 2.2: Consonant system of English language

Manner	Voicing	Place of Articulation						
		Bilabial	Labiodental	Inter dental	Alveolar	Palatal	Velar	Glottal
stop	Voiceless	p			t		k	ʔ
	Voiced	b			d		g	
Fricative	Voiceless		f	θ	s	ʃ		h
	Voiced		v	ð	z	ʒ		
Affricate	Voiceless					ʧ		
	Voiced					ʤ		
Nasal		m			n		ŋ	
Liquid					l	r		
Glide	w					y		

2.6.3 Consonant System of Mandarin

Most Mandarin consonants exist only in the syllable's initial position. Table 2.3 shows the 21 Mandarin consonants as per the IPA. The consonants present in the inventory of Mandarin Chinese are as stated by Eme and Odinye (2008).

Table 2.3: Consonant system of Mandarin language

Manner		Place of articulation					
		Bilabial	Labiodental	Alveolar	Retroflex	Alveo-Palatal	Velar
Stops	Aspirated	pʰ		tʰ			kʰ
	Unaspirated	p		t			k
Affricate	Aspirated			ts ʰ	tʂʰ	tɕʰ	
	Unaspirated			ts	tʂ	tɕ	
Fricatives	Voiceless		f	s	ʂ	ɕ	x
	Voiced				ɻ		
Nasal	Voiced	m		n			
Latenal	Voiced			l			

2.6.4 Syllables Structure of Mandarin

Pinyin 拼音, which literally means "spell out the sound," is now among the foremost Romanization systems utilized for Mandarin. It is formally called *Hanyu Pinyin*, where *Hanyu* translates as "the Chinese language". This system was designed by a Chinese government board and was first accepted by the national regime in 1958.

Chinese characters are unlike the English alphabet or other languages' letters, wherein one can read a word's sound immediately from its spelling. As a result, the *pinyin* approach was developed to aid in the description of normal Mandarin Chinese pronunciation and has proven to be an extremely valuable instrument for keying-in Chinese-language text into digital systems. The *pinyin* system employs all letters of the English alphabet, with the exception of "v." The relationship between letter and sound, however, differs from that of any other language. When first learning this Romanization system, students must be attentive to the differences between *pinyin* characters and their English letter counterparts. Notably, each Chinese letter consists of a single syllable. A

Mandarin syllable is made up of three parts, which are the beginning (initial consonant), the ending (final vowel), and the tone.

2.6.4.1 Initials

A consonant is an initial (excluding y and w). The initial is generally the lone consonant seen at the start of a syllable and cannot exist on its own. There are a total of 21 Chinese initials. In cases where the initial is empty or a 'zero' 零声母, the initial consonant does not exist, and the syllable commences with a vowel that is still taken as the final component.

2.6.4.2 Finals

The six simple finals, a, o, e, i, u, and ü, are the most basic constituents of Chinese vowels in *Hanyu Pinyin's* single-vowel category. The four hu *四呼* constitute a method of categorizing Standard Chinese syllable finals that have been used for a long time based on the many glides preceding the final center vowel. These *hu*s are:

- *kāikǒu*（开口, "open mouth"), finals sans a medial
  Example: [an] in 兰[lán]word is an open mouth final.

- *qíchǐ* (齐齿, "even teeth"), finals starting with [i]
  Example: [ian] in 天[tiān] word is an even teeth final.

- *hékǒu* (合口, "closed mouth"), finals starting with [u]
  Example: [ui] in 对[duì] word is a close mouth final.

- *cuōkǒu* (撮口, "round mouth"), finals starting with [y]
  Example: [u] in 女[nǚ] word is a round mouth final.

Additionally, simple finals, complex finals, and nasal finals have been established in Mandarin. There is only one vowel in a simple final. Meanwhile, a triphthong (three-vowel sequence) or a diphthong (a two-vowel sequence) makes up a complex final. Lastly, a nasal final consists of one vowel and a nasal consonant added after the vowel.

2.6.4.3 Tones

If the pitch of words in a language modifies their meaning, the language is identified as ‘tonal language’. Mandarin language’s noticeable difference is that it is spoken tonally, rendering pronunciation challenging for learners. Tones exist in Western languages as well; for instance, exclaiming “yes” in a high or low tone portrays distinct implications or feelings, yet the speaker’s tone does not alter or impact the word’s meaning. In contrast, Chinese tones play a vital role in defining meaning. Varied tones for an identical *pinyin* syllable represent different letters in Mandarin. Such is the primary distinction between a tonal language and a non-tonal one in terms of tonal function. Mandarin and English may be the two most extreme examples of tonal versus non-tonal languages.

In essence, the pitch of the Mandarin speaker’s voice distinguishes tones. Two aspects of voice pitch should be addressed. To begin with, tone pitch is a relative rather than an absolute concept. Women often have greater pitch than men, and regardless of gender, one’s pitch can alter depending on the context. To illustrate, people generally speak in a higher pitch when excited and a lower pitch when disappointed. Nonetheless, these variances have no impact on a word’s semantic component. Secondly, changing the pitch at any level should be a seamless and consistent procedure. As a tonal language, Mandarin has four tones: 1. the first tone or the high tone; 2. The second tone or the rising tone; 3. the third tone or the low tone; and 4. the fourth tone or the falling tone. Table 2.4 shows several examples of Mandarin tones.

Table 2.4: Mandarin tones

Tone	Mandarin Character	Note
1	mā (妈）mother	starts high and remains high
2	má (麻） hemp	starts at mid-range and finishes high
3	mǎ (马) horse	starts at mid-range, dips low, then finishes mid-range
4	mà（骂）scold	starts high and finishes low

2.7 Comparison of Stop Consonants in Malay, English, and Mandarin

The Malay language has six oral plosives and one glottal stop. In Malay, /p/, /t/, and /k/ are unaspirated. Consonant /b/ is classified as a voiced bilabial plosive, whereas consonant /p/ is considered a voiceless bilabial plosive. Alveolar plosive consonants can be classified as voiced /d/ and voiceless /t/. Two consonants are categorized as velar plosive, namely consonant /g/ (voiced) and consonant /k/ (voiceless). Next, there are a total of three stops in English, which are bilabial stops, alveolar stops, and velar stops. The phonetic feature that distinguishes between /b/, /d/, /g/ and /p/, /t/, /k/ is voicing. The former is voiced, whereas the latter is voiceless. Finally, Mandarin has six stops consonants which are /p^h/,/t^h/,/k^h/ , and /p/,/t/,/k/. In Mandarin, the phonetic feature that distinguishes between /p^h/,/t^h/,/k^h/ and /p/,/t/,/k/ is an aspiration. Unlike English, all Mandarin stops fall under the voiceless category and can only exist in the word-initial point. The six stops signify an unaspirated–aspirated contrast. A comparison of the three consonant systems is given in Table 2.5.

Table 2.5: A comparison of three consonant systems

Places of articulation / Language	Bilabial	Dental	Labio-Dental	Alveolar-Dental	Palato-Alveolar	Palatal	Velar	Glottal
Malay	p b			t d			k g	
Mandarin	p p^h	t t^h					k k^h	
English	p b			t d			k g	

2.8 Stop Consonants

The phonetic classification of speech sounds based on articulation manner is known as a stop. It typically refers to any sound produced by the full closure of the vocal tract and hence contains the class of plosives. The creation of a stop consonant generally involves three successive phases: 1) the commencement of closure, when one articulator nears the other; 2) the closure, when the articulators are bound together in a manner that fully obstructs airflow and builds up pressure behind the constriction; and 3) the offset of closure, when the articulators move apart once more (Henton, Ladefoged, & Maddieson, 1992).

2.9 Voice Onset Time

A little more than five decades ago, Lisker and Abramson (1964) proposed a direct assessment tool for acoustic differences among voicing categories' stop consonants called Voice Onset Time (VOT). From then to now, hundreds of scholars have adopted this tool. Lisker and Abramson (1964) mentioned that the VOT is a crucial acoustic factor in producing stop consonants. It is essentially a phonetic feature or a unique phoneme expression. While the word "phone" refers to the phonetic element that is quantifiable and trackable, a phoneme provides phonological meaning. Linguists can acquire recordings of phones as speech sounds, as opposed to phonemes which comprise a more theoretical, abstract framework. Phones can therefore be examined for a wide range of features, including the physical qualities of the speaker who generated them and the recorded speech signal. Based on studies of phones both individually and naturally as a component of larger signals, their structure was developed.

Lisker and Abramson (1964) explicated the VOT measurement process by finding, among those regularly distributed, the initial vertical striation that represents glottal pulsing; meanwhile, the release point is determined as the point at which the pattern expresses a sudden change in the general range. Notably, the glottis is the space between the vocal folds situated between the lungs and the mouth. A sound signal is created when air is forced from the lungs through the glottis with sufficient force to cause the vocal folds to vibrate. According to its anatomical, non-acoustical definition, VOT represents the actual amount of time that elapses between the glottis's movements and the mouth's movements through configurations of the tongue and/or lips that embody the speech signal. The phrases used in this acoustical definition indicate particular components of the speech signal that can be seen in the stop consonant's waveform.

After studying 11 languages, Lisker and Abramson (1964) were able to conclude that VOT distinguishes between three categories of stops:

1) plosives with a negative VOT
2) plosives with a slightly positive VOT; and
3) plosives with a clearly positive VOT.

The first category – termed fully voiced – gives rise to a negative VOT and is the result of the production of voicing during the closure. This process is also called pre-voicing. In other words, the vocal folds started vibrating before the release of the initial plosive. Since the release of the plosive counts as the starting point for VOT measurement, the VOT recorded in voiced plosives is negative in the case of pre-voicing. The plosives which are formed where aspiration is limited or non-existent (and thus show only a slightly positive VOT) make up the second category, also labelled as voiceless unaspirated. The third category involves those plosives which lead to the production of a clearly positive VOT as a result of aspiration. This last category is otherwise known as voiceless aspirated. Since the onset of voicing is delayed by the production of aspiration – which is voiceless – the period of voicelessness is longer, i.e., the VOT will be longer than when no aspiration can be detected.

The existence of these three categories of voicing implies that any language could make use of them. Lisker and Abramson (1968) analyzed data on stop consonants across 11 languages and categorized them into three groups (lead, short lag, and long lag), which are each distinguished by VOT values. Specifically, lead refers to when voicing commences about 75-125 ms prior to a consonant's release (negative value VOT). Short lag occurs when voicing commences 0-30 ms following the release, while long lag occurs when voicing commences 60-100 ms following the release (Lisker & Abramson, 1964).

VOT is frequently used to research stops; however, extant studies have utilized slightly different measurement techniques. Several scholars have acquired VOT scores by computing the interval spanning the start of the release burst to the onset of the first vowel formation (Chao et al., 2006; Chen et al., 2007) or the second formation (Cho & Keating 2001) seen in a spectrogram. Other scholars have derived VOT values by evaluating the time interval from the sudden high point that symbolizes the stop's release to the periodic wave's start, representing the start of the vowel (Riney et al., 2007; Whalen, Levitt & Goldstein, 2007). There is also a group of scholars who measure VOT based on waveform data followed by spectrogram validation (Kehoe, Lleó & Rakow, 2004; Macleod & Stoel-Gammon, 2005).

In English voiced stops (/b, d, g/), Lisker and Abramson (1964) outlined two VOT value sets (positive short lag and negative voicing lead). The scholars also implied that every native speaker produces only one type of phonetic representation. Klatt (1975), in turn, calculated the VOT of English stops and reported favorable results for the voiced stops /b, d, and g/ and the voiceless unaspirated stops /p, t, and k/. Similarly, Keating (1984) expressed voiced stops in English to be occasionally spoken with leads but mostly with short and long lag values.

It has been well-established that Mandarin stops are voiceless in terms of phonetics, such that aspiration is the single unique phonetic trait that differentiates its two phonemic classes: voiceless unaspirated /p, t, and k/ and voiceless aspirated /p^h, t^h, and k^h/. In contrast with English, Mandarin stops appear exclusively in the word-initial location and are classified as short lag or long lag patterns.

In comparing the VOT patterns of English and Mandarin, it is observed that voiceless aspirated stops fall under the long lag group in both languages, which contradicts extant findings. According to Chao et al. (2006), the Mandarin /p^h, t^h, and k^h/ belong to the extremely aspirated category, which is not a single long continuum. Average VOT values assessed by Rochet and Fei (1991) further indicate that Mandarin and English dominate distinct areas of the VOT continuum.

Overall, the evidence proves that VOT is a robust measurement tool for the acoustic comprehension of consonantal voicing differences in a majority of languages.

2.10 Aspiration

Aspiration is a crucial distinguishing feature of consonants across numerous languages. It separates stops and affricates into two categories (i.e., aspirated and unaspirated) in relation to their dissimilar speech phonemes. If the aspiration interval is accounted for, aspirated consonants are substantially lengthier than comparable unaspirated ones (Feng, 1985; Wu, 1992).

According to Collins and Mees (2008), the process of aspiration is often referred to as a slight puff of air that is expressed upon releasing the voiceless aspirated stop consonants /p^h/,/ t^h/, and /k^h/. In phonetics, it is symbolized as [h]. In the Malay language, the stops are not normally aspirated. The addition of aspiration does not alter the meaning of the word. To illustrate, take the word /papan/ (‘board’ pronounced with an unaspirated voiceless bilabial stop). Nevertheless, if one pronounces the /p/ sound with aspiration /p^h/, the word’s meaning is not altered. A native Malay speaker still will be able to understand the word, although it sounds uncommon. However, in the English and Mandarin languages, English consonants and Mandarin consonants can be classified as aspirated and unaspirated sounds.

2.11 Previous Studies of L2 Mandarin Consonants

L2 students' difficulties with foreign language sound production have been widely proven in the literature with regard to consonants (Bradlow et al., 1997; Guion et al., 2000; Munro et al., 2015), vowels (Evans & Alshangiti, 2018; Munro & Derwing, 2008; Wang, 1997; Wang & Munro, 2004), and lexical tones (Wang, 2006, 2008, 2013). However, corresponding research on the perception and production of non-native language sounds, especially pertaining to second language learners' issues with Mandarin consonants, remains scarce. The following sections review extant empirical works from the past two decades on Mandarin consonant perception and production.

2.11.1 Studies on L2 Mandarin Consonants

Pengdeng Yin (2021) examined the phonetic errors of northern Myanmar Kachin students from the primary Chinese level at the Mai Ja Yang Institute of Education. The results showed that there are regular, common, and special Chinese pronunciation problems for Kachin students. The author used error analysis to analyze learners' errors and found that n /n/, l /l/, z/ts/, and r/ʐ/ are low difficulty sounds, j /tɕ/,q /tɕ^h/, s /ʂa/, c/ts^h/ and x /ɕ^{j}a/ are medium difficulty sounds, and zh/ʈʂ/, ch /ʈʂ^h/ and sh/ʂ/ are high difficulty sounds. In relation to pronunciation, the degree of difficulty from low to high is blade-alveolar, apical, lingual surface, and blade-palatal. The study shows that the aspirated fricative and aspirated affricate are the most difficult sounds in consonants.

In Taiwan, Lai (2009) examined learners' perceptual challenges in Mandarin's six affricates, namely z /ts/,c/ts^h/, zh /ʈʂ/, ch/ʈʂ^h/and j/tɕ/, and q/tɕ^h/, among L1 speakers of Malay and Burmese. The students and a control cohort (i.e., L1 Mandarin speakers of Taiwan) underwent identical or alternate perception assessments, instantly followed by an assessment to identify the target affricates coupled with various articulation places and manners. The results show that both learner categories could identify unaspirated affricates more accurately than aspirated ones. Both groups also found it more difficult to detect contrasts in the dental-retroflex z /ʈʂ/-zh /ʈʂ/and c / tsʰ /-ch / ʈʂʰ / than in palatal affricates. Lai (2009) surmised that dental and retroflex affricates are combined, wherein the dentalization of retroflex sounds is clarified more effectively by the Markedness theory than by students' native language interpretations. Interestingly, the Mandarin control cohort learners exhibited identical perception integration patterns in their z/ʈʂ/-zh/ʈʂ/ and c/tsʰ/-ch/ʈʂʰ/ identifications due to their similar error rate (approximately 67%) to the two non-native learner groups.

Hao (2012) studied the effect of learners' L2 to L1-sound mapping schemes and L2 experience levels on their Mandarin sound perception. Native English CFL learners' experience with Mandarin learning was divided into three levels: Experienced (5.6 years), Inexperienced (1.5 years), and None (no experience). Based on these groups' perceptual test results, Hao (2012) discovered the significant influence of phonetic context and experience on learners' L1–L2 sound mapping patterns. Specifically, learners with a higher level of experience produced answers with higher consistency in Mandarin-to-English categorizations of sound, being less impacted by phonetic settings than inexperienced learners. Learners with no experience at all integrated the Mandarin /s/ into the English /z/ more than the English /s/, whereas the other two groups observed the Mandarin /s/ as the English /s/. All three learner groups demonstrated the general assimilation of the Mandarin /ʂ/and /ɕ/ into the English /ʃ/; however, learners with no experience divided the categorization of /ɕ/ to /s/ and /ʃ/ equivalently when /ɕ/ preceded the unrounded vowel /i/. Although the Mandarin /ʂ/ and /ɕ/ had been integrated into the English /ʃ/, /ʂ/ is more suitable than /ɕ/ based on its better identification accuracy rate and Category Goodness assimilation score, as per the PAM model. The test of identification further reported that the Mandarin /ʂ-ɕ/ contrast is challenging for learners, especially for the inexperienced group. Nonetheless, all learner groups showed equally satisfactory discrimination performance in differentiating /ʂu-su /and /ʂi-si/ contrasts. Hao (2012) thus confirmed that, in most cases, L2 to L1 patterns of assimilation could not estimate the accurate discrimination of Mandarin contrasts.

More recently, Wang and Chen’s (2019) cross-linguistic perception research on Mandarin consonants found that native English speakers without any Mandarin learning experience could detect 10 consonants in Mandarin syllables (z /tsa/, c /tsʰa/, s /sa/, j /tɕʲa /, q / tɕʰʲa /, x /ɕʲa/, zh /ʈʂa/, ch / ʈʂʰ a /, sh /ʂʰa/, and r / ʐa/) by relating them to the most similar English sounds. The study used a 10-way forced choice activity and then a goodness rating activity ranging from ‘1’ being poor to ‘7’ being good. The fit indexes of sound mapping (identification value multiplied by rating value) were computed for L2 to L1 sounds to measure phonetic intervals between English and Mandarin consonants. The scholars revealed an array of L2 to L1 phonetic intervals, as the fit indexes ranged from 1.0 to 6.3. Categories that received a “poor” match (fit index less than the mean value of 3.7) were x /ɕ/, c/tsʰ/, q /tɕʰ /, zh /ʈʂ /, and j /tɕ/, while categories with a “fair” match (fit index at the mean) were ch /tsʰ/, s /s/, and z /ts/ and categories with a “good” match (fit index one standard deviation above the mean) were r /ʐ/ and sh /ʂ/ (Wang & Chen, 2019). In addition, English CFL learners' perceptions of Mandarin consonants were shown to be highly correlated with their L2 to L1 assimilation patterns in a subsequent study on Mandarin consonant identification across two proficiency levels. The study’s findings reported that beginner learners recorded the lowest identification scores for poor fitting sounds (zh /ʈʂ/, q /tɕʰ /, c /tsʰ/, and x/ɕ /) out of the 10 sounds. The intermediate-level learners, however, exceeded the beginners in the zh /ʈʂ/, q /tɕʰ /,and c / tsʰ / sounds. The results indicate that perceptions of L1 to L2 consonants’ phonetic distances lead to English CFL learners’ difficulties in identifying Mandarin (L2) consonants. Moreover, greater levels of L2 experience were found to enhance perceptual learning.

Yang and Yu (2019) conducted a parallel study with more Mandarin affricate contrasts to examine how beginner and intermediate native English CFL learners perceive and produce six Mandarin affricates, z /ts/, c/tsʰ/, zh /ʈʂ/, ch /tsʰ/, j /tɕ/, and q /tɕʰ/. Both groups equaled native Mandarin speakers' perception accuracy resulting in discrimination but could not do so in the identification of target sounds. Articulation place and aspiration showed significant but varied effects in the model. For instance, the unaspirated palatal j /tɕ/ had higher identification compared to the aspirated palatal q /tɕʰ/; however, the aspirated retroflex ch /tsʰ/ showed stronger identification than the unaspirated zh /ʈʂ/. The researchers surmised that distinct affricates exert unique acquisition challenges for English L2 learners. The production assessment further showed that intermediate learners performed better than beginners in mimicking native speakers when producing several (but not all) of the acoustical features being studied, suggesting that the learners failed to fully acquire the affricates. The study's data also reported that differentiating palatal affricates from retroflex ones is more problematic for learners because, according to the PAM, both classes are assimilated into identical English post-alveolar affricates (i.e., two-tone SC assimilation).

According to Lin (2001), the Mandarin consonants that are uniquely difficult for native English speakers include ts, tsʰ, tɕ, tɕʰ, ɕ, ʈʂ, ʈʂʰ, ʐ, x, and ÿ as per the IPA. She mentioned that the challenges generally stem from the incongruence between both languages' consonant inventory, as well as between their feature system, consonant distribution, and within-syllable features (i.e., phonotactic limitations). She concluded that the degree and type of challenges English speakers face in learning Mandarin consonants might shift with the increase in their proficiency in Mandarin as q Second Language (MSL). Early-stage problems, typically from L1 and L2 disparities, may be eliminated altogether with time. However, aspects that seem to be easier in the beginning, generally related to similarities between both languages, may take a far greater duration to grasp.

Liu and Jongman’s (2012) production research looked into the time- and space-related qualities of Mandarin dental affricates (z /ts/ and c /tsʰ/) that L1 English speakers with varying degrees of mastery produce. They revealed that beginner and experienced learners alike exhibited durational discrepancies in the /ts/ and /tsʰ/ contrast. Advanced students were the only ones to master the spectral (center of gravity) contrast across targeted acoustic pairs. The significance of each temporal and spectral cue for target contrasts’ perceptual accuracy is not clear, as a perception test was not performed to assess learners’ production accuracy. Moreover, their research involved only one Mandarin affricate contrast pair at the dental articulation place.

Khor, Mah, and Chow’s (2017) survey of undergraduate students from the Engineering Campus of USM showed that a majority of students made substantial errors when pronouncing retroflex consonants because of their mother tongue’s interference. For example, retroflex consonants r /ʐ/ can be mistakenly referred to as "ri" in the Malay language, such that 生日 “shēngrì” can be mistakenly referred to as “shēngrì” (risau). Additionally, not only do students’ first and second languages lack certain vowels that exist in Mandarin, but students also find it hard to pronounce tones accurately; these difficulties further hamper students’ mastery of the Mandarin language.

Khor, Arriaga, and Mah (2013) examined sound production issues affecting non-native-speaking students, mainly regarding their consonants, vowels, and tones. Following the *Hanyu Pinyin* system, the results reported the highest errors in consonants (68.29%), while tone (13.66%) and vowels (18.5%) had far fewer errors. Of the consonants, most students erroneously pronounce zh /ʈʂ/ (16.43%), followed by z /ts/ (13.57%). Most errors for vowel pronunciation occur for the last phoneme s /s/ (82.14%). In terms of tone, the second (56.76%) and third (24.32%) tones suffer from the most inaccuracies. Therefore, the conclusion was reached that undergraduate students are greatly challenged in Mandarin pronunciation mastery, mostly as a result of the Malay language being their mother tongue.

Hou (2019) combined pre-study and interviews (i.e., recording and listening experiments and speech-acoustic analysis methods) with SPSS statistical analysis of VOT, occlusion duration, aspirated duration, voiced interval, and vowel duration to acknowledge the acquisition of Chinese Mandarin plosives and affricates in three syllables, single syllable, two syllables, and sentences produced by Vietnamese students. The following conclusions were drawn. First, the plosives of Chinese Mandarin are a major difficulty for Vietnamese students to learn. In the survey, the highest rate of bias was for zh /tʂ/, ch /ts^{h}/, c /ts^{h}/,q /tɕ^{h}/, and k /k/. In terms of pronunciation errors, Vietnamese students often mix the same parts with aspiration sounds and unaspiration sounds, especially turning aspirated sounds into unaspirated ones. In the pronunciation part, the Mandarin Chinese supra-dialecticals and the retroflexes, the apical consonants, and the lingua-palatal phonemes are mixed, and the plosives and affricates are made into fricatives. The perception and output bias rate of the plosives and affricates did not show a complete one-to-one correspondence. Indeed, the VOT and duration of plosives and affricates of Vietnamese students are quite different from those of Chinese native speakers except for some individual tones, especially affricates. The differences are as follows: the average VOT and duration of aspirated sounds are smaller than those of native Chinese speakers, while the unaspirated sound is larger than those of native Chinese speakers. This indicates that Vietnamese students have a low perceptual sensitivity to the categorization of "aspirated and unaspirated", which is most obvious in the monosyllabic affricate, followed by disyllabic affricate, and then in the sentence context.

Next, Vietnamese students were found to be quite different from Chinese native students in the pattern of plosives and affricates. This difference is mainly manifested in the following aspects: the unaspirated sound ratio of Vietnamese students is larger than that of Chinese native speakers, while the aspirated sound ratio is smaller than that of Chinese native speakers. The unreasonableness of rhyme proportion is one of the reasons that lead to the pronunciation of the "Vietnamese accent" in Chinese Mandarin. In addition, from the perspective of the occlusion time, in the three contexts, the difference between the plosives and the affricates is greater. The consistent performance of the Vietnamese appeared to be larger than that of the Chinese. From the perspective of voiced intervals, the study found that in the three contexts, the difference between the plosives and the affricates is greater. The main difference is as follows: for the aspirate sound, the Vietnamese are larger than the Chinese. For the two-syllables aspiration sound, the Vietnamese are smaller than the Chinese, and for the sentence, the Vietnamese are larger than the Chinese.

Another finding of the study is that the reasons for the errors of Vietnamese students are mainly due to the features of the antagonistic differences between the Chinese-Vietnamese phonology. Examples include the low perceptual sensitivity of the aspirated and unaspirated sounds, the difference between the Chinese *Pinyin* Scheme and the corresponding actual phonemes in the Vietnamese Latin alphabet, and the fact that there is neither affrication nor retroflex in Vietnamese.

Song (2018) studied the effect of Malay learners' perception of Mandarin affricate consonants using the perceptual experiment methodology. The research mainly employed the two popular SLA theories, the PAM and SLM, to design two experiments: a perceptual assimilation experiment and a perceptual discrimination experiment. Participants were asked to make judgments on their perception of Chinese consonants in the native speech category in the two experiments. The results of the first experiment on the perceptual assimilation of Mandarin consonants as a stimulus can predict the second experiment's result. The perceptual assimilation outcomes showed that Malay students assimilated Chinese aspirated and non-aspirated stops into two categories: Malay voiced and voiceless stop consonants. Significant differences exist between the students at the elementary level, intermediate level, and advanced level in perceiving Mandarin consonants. Language proficiency continues to improve in the intermediate stage of learning (within two or three years). On the other hand, the perceptual discrimination experiment showed that Malay students could distinguish between Chinese plosives and aspirated and non-aspirated affricates. The study also concluded that the language environment has no significant impact on Malay students' distinction between Chinese plosives and affricates.

Huang (2006) examined the difficulties of Indonesian speakers in learning and pronouncing Mandarin because of the difference in sound systems of the two languages. The pronunciation study discussed the sound system of two languages, i.e., Indonesian and Mandarin, using both phonetic and phonological approaches. It is natural that when learning Mandarin pronunciation, Indonesian speakers tend to use Indonesian speech sounds that are similar to Mandarin speech sounds. Some Mandarin speech sounds that are not found in the Indonesian sound system often bring about the inaccurate pronunciation of Indonesian speakers who are learning Mandarin pronunciation. The most difficult Mandarin sounds to pronounce are the fricatives and the affricates, especially the retroflex ones. This is due to the absence of retroflex sounds in the Indonesian speech sound system. Indonesian speakers tend to replace the retroflex sounds in Mandarin with non-retroflex sounds in Indonesian that sound similar. The presence of tones in Mandarin is another hurdle in learning Mandarin pronunciation. Indonesian speakers are often confused by the tones in Mandarin. The most obvious and common difficulty is evident in how the Mandarin tones during spontaneous speech are replaced by Indonesian intonation.

Liu Li (2016) examined how Chinese EFL learners perceive and produce English and Mandarin consonants. The study summarized that the overall correct perceptual rate of Mandarin consonants is 96.57%, which suggests that Chinese EFL learners are highly sensitive and intuitive about the perceptual aspect of Mandarin consonants. Among the five categories of Mandarin consonants, the most well-perceived Mandarin consonantal category by Chinese EFL learners is approximant (98.61%), while the worst is nasal (95.37%). The overall correct productive rate of Mandarin consonants is 97.94%, which shows that the Mandarin consonants produced by Chinese EFL learners are generally acceptable and that the Chinese EFL learners' productive competence of Mandarin consonants is higher than their perceptive competence.

Lim Hui Woan (2010) examined the concurrent phonological acquisition of English, Mandarin, and Malay among Chinese ethnic children. The author concluded the existence of a significant age impact in all three languages. The author also compared both monosyllable structures and disyllable structures between the three languages, revealing significant relationships. However, while contrasting trisyllable structures between English and Mandarin, the author concluded that there is no significant relationship.

2.11.2 Studies on L2 Stop Consonants

Guoqing Shen (2020) studied Chinese phonetic mistakes among learners from the primary level at the Confucius Institute of the Universiti Malaysia Pahang. The paper used the method of comparative analysis and error analysis to analyze the phonetic data. The results showed that the participants produced voiceless unaspirated stop consonants /p/ /t/ /k/ into voiced unaspirated stop consonants /b/ /d/ /g/ and voiceless aspirated stop consonants /p^h/ /k^h/ /t^h/. Some of the participants also produced aspirated voiceless stop consonants /p^h/ /k^h/ /t^h/ into unaspirated voiceless stop consonants.

Yang (2018) conducted comparisons of stop consonants' temporal measurements among 12 adults and 29 children (ranging from three to six years old), all of whom were Mandarin speakers. Every respondent was able to produce 18 Mandarin disyllabic words comprising six stop consonants /p, p^h, t, t^h, k, k^h/, and subsequently, three vowels /a, i, u/ in the first syllable's word-initial point. VOT temporal measurements, burst number, total burst duration, mean time per burst, and VOT-lag time were recorded. Despite all the children achieving short-lag VOTs like adults, long-lag VOTs were more prevalent among younger children and showed a gradual evolution into a focused distribution among older children. Deeper analyses of the burst and VOT-lag indicated that children are more inclined than adults to generate shorter burst durations and lengthier VOT-lags. The findings imply that, unlike adults, children in this age group have yet to develop laryngeal–oral temporal patterns and the airflow control required to produce stops.

Tanaka, Chen, and Hsu (2019) researched the features of stop precision and replacement trends among Mandarin–Japanese bilingual toddlers. Specifically, they examined Japanese and Mandarin word-initial stops produced by 36 bilingual children aged from three to six who had acquired both languages concurrently from birth. Aiming to understand bilingual speakers' phonological development, the study's results showed that: 1) by the age of three, a majority of the bilingual toddlers could produce Mandarin and Japanese, whereby target stop accuracy developed with time; (2) the age at which target consonants develop varies slightly across both languages; and (3) each language's replacement pattern portrays a combination of child-specific trends, language-specific systems, language impacts, and individual discrepancies. The findings suggest that bilingual children own a distinct system of phonological development that constitutes a monolingual pattern with cross-linguistic interactions.

Jiang (2014) investigated the pronunciation errors made by Malaysian students in the primary stage of their acquisition of Chinese. Based on the theory of error analysis, the paper analyzed Malaysian students' phonetic errors from multiple perspectives. Errors that are most frequent among Malaysian learners are stop consonants. Specifically, the learners pronounced aspirated consonants into unaspirated consonants and unaspirated consonants into aspirated consonants. The second error is that the learners pronounce voiceless aspirated consonants as voiced consonants.

Shahidi et al. (2012) studied how Malay speakers produced and perceived English word final stops. According to the study's findings, similar phonemes in L1 and L2 are realized in the same way as L1 sounds. This study specifically implies that the speaker's native language has a major impact on the phonetic qualities produced in English. This process is most likely made easier by the similarity between the phonemic systems (especially the consonants) of the two languages. Therefore, it is determined that L1 is the only source of L2 sound production for current speakers (that is, the speakers are merely matching the L2 phonology with their L1 and generating the relevant L1 phonetic realizations in response).

Li (2013) studied the influence of gender on Mandarin speakers' VOT in stop production. Word-initial stops were drawn from 20 Mandarin speakers (10 from each gender) via a word-repetition activity. The findings demonstrated varied VOT patterns across genders based on raw VOT scores for all four lingual stops. Upon controlling for speech rate differences, gender-specific variations were only present for voiced stops, such that males had longer VOTs than females. To conclude, Li (2013) offered new proof of the effect of speakers' gender on Mandarin VOT, which is different across languages. This implies that language and/or cultural factors (e.g., sociolinguistics, style) are the cause of gender-based stop production variations, as opposed to biological/anatomical factors.

Chao and Chen (2008) examined the VOT patterns of Mandarin and English. They concluded that there are significant differences between Mandarin and English, especially for voiceless aspirated stops. The study further revealed that three-way universal VOT categorization is not suitable for distinguishing Mandarin voiceless stops. Chen, Chao, and Peng (2007) also examined the VOT production of Mandarin and English voiceless stops among L1 Chinese speakers. Their findings showed that voiceless stops generated by native Chinese belong to the same long-lag category in Mandarin and English. However, the average VOTs of Chinese speakers' English [p] and [t] production are relatively subtle. The study concluded that individuals' first language could affect their L2 productions.

Lai (2013) compared VOT values in Mandarin stops and affricates by CSL learners of different first language backgrounds with native Chinese speakers. The results showed that native Chinese speakers could differentiate Mandarin stops and affricates with or without aspiration in their VOT values. However, this study also reported that the CSL Korean learners failed to distinguish Mandarin stops with or without aspiration. The Vietnamese Mandarin learners, however, could not differentiate between aspirated and unaspirated affricates.

2.12 Summary

Based on the discussion above, it can be summarized that the effect of one's first language on L2 production is well-known (Chao & Chen, 2008). When learning to speak any foreign language, the initial step is always learning to pronounce the language's words. To do so, learners often turn to their native language's sound system (Flege, 1995; Eckman & Inverson, 1993). Studies have demonstrated that the transfer of first language phonology is the main factor affecting second language learners' L2 pronunciation (Tarone, 2005). This transfer can be beneficial or detrimental, such that L2 words with similar syllables or phonemes to one's native language initiate a positive transfer, while L2 words with syllables or phonemes that are missing in one's native language initiate a negative transfer. This research has observed and experienced that Malay students do exhibit transfer from their native language in their Mandarin pronunciation. For instance, Malay students are able to easily and accurately enunciate Mandarin syllables such as ma [mᴀ], mu [mu], fu [fu], he [xɤ], ni [ni], ne [ne], and li [li], on the basis that these sounds exist in Malay. In most events, however, they face barriers in the pronunciation of syllables that contain aspirated consonants (p, t, q, ch, c, k), palatal consonants (j, q, x), apical vowels (-i [ɿ], -i [ʅ]), and rounded high front vowels (ü), as despite being anchored in similar roman alphabets, these syllables are absent in the Malay language. These findings concur with that of Goh (2007), who classified b, d, g m, n, l, f, s, h, a, e, i, o, and u as simple to pronounce for Malay speakers and p, t, k, z, c, zh, ch, j, q, x, sh, r, ü, -i [ɿ], and -i [ʅ] as arduous to pronounce. The author also expressed that pronunciation challenges stem from the first language's interference and the confusing nature of Mandarin's alphabetical phonetic, which is the *Hanyu Pinyin* spelling system.

In conclusion, based on the lack of research on Mandarin stop consonants, especially by Malay speakers, this study attempted to bridge this gap and investigate the acoustic properties of stop consonants produced by Malay speakers along with their perceptions of Mandarin stop consonants.

CHAPTER 3: RESEARCH METHODOLOGY

There are a total of six major sections that make up this chapter. Section one introduces the background of the participants, while section two describes the research materials employed to collect data. The third, fourth, and fifth sections discuss the procedures used for data collection and analysis. The relevancy of the theoretical framework is addressed in the sixth section.

3.1 Participants

This study looked at the perception and production of Mandarin stop consonants. Therefore, the participants did not only have to have received standard language training but also to have experience communicating with L1 Mandarin speakers in their native setting and environment. As such, only participants who met certain criteria were allowed to participate in this research. That is, the participants were selected based on their fulfilment of the following requirements: (1) they had been taught Malay as their first language and English as their second; (2) they had received standard language training in Mandarin in the Beijing Foreign Studies University for five years and completed their Bachelor's Degree in China; (3) they had received their Diploma in Teaching Chinese at a teacher's training university in Malaysia after two years of teacher's training; (4) they had not studied Mandarin before going to China; and (5) they had undergone a Chinese Proficiency Test (HSK) in both speaking and writing, achieving at least Band 5.

HSK is a globally standardized test taken by non-native Chinese learners to test their Mandarin language proficiency. Achieving the HSK-5 credential means that the participants are able to read Mandarin newspapers and magazines, watch Mandarin movies and television programs, and give more complete speeches in Mandarin than those who do not have the credential. Additionally, it means they have mastered 2,500 frequently used words and associated grammar patterns. As can be seen, all five criteria used to select the participants were relevant to the current research's questions and objectives.

The research setting was around the Klang Valley. From this region, 10 Mandarin teachers whose first language is Malay participated in this research. They ranged from 22 to 30 years old and were recruited using the snowball sampling approach. This approach is often used in qualitative research, wherein the researcher reaches out to people who then connect them to other suitable respondents rather than directly recruiting respondents (Marcus et al., 2017; Parker, Scott, & Geddes, 2019). The snowball sampling technique was chosen for this research due to the low number of potential participants. After observing the initial subject, the researcher sought assistance from the participants to identify others who fulfil the criteria given. The participants were reached via a mobile phone application (app), namely the WhatsApp Messenger application. Once the participants agreed to take the survey, a Google Form link was sent for them to fill up the consent form and their language background details. Finally, they were contacted through WhatsApp to do the remaining production test and perception test.

Table 3.1 lists the Malaysian Malay speakers' age, gender, HSK level, first language, and years of studying Mandarin. Following that, Table 3.2 shows the information on the participants' daily usage of language, whereby most appear to prefer using the Malay language and Malay dialect, followed by English and then Mandarin. The table further suggests that most of the participants use the Mandarin language in school during Chinese lessons with students. Only Participant 1 uses the Mandarin language at home because she is staying with friends who speak Mandarin. The table also clearly indicates that all 10 participants speak either Malay or a Malay dialect as their first language. The Malay dialects mentioned in this study were the Kelantan and Terengganu dialects (for further information, please refer to the sub-section on these dialects in Chapter Two). The usage of dialects does not influence this study because the aforementioned dialects have the same stop structure as Mandarin. Lastly, a majority of the participants use the Mandarin language with friends and office colleagues.

Table 3.1: Speakers' demographic background

Participants	Age	Gender	Duration of learning Mandarin (years)	HSK Band	Experience as a Mandarin teacher
1	24	F	5	5	1
2	28	F	5	5	5
3	26	F	5	5	2
4	26	F	5	5	2
5	26	F	5	5	2
6	26	F	5	5	2
7	26	F	5	5	2
8	27	F	5	5	3
9	28	F	5	5	4
10	29	F	5	5	6

Table 3.2: Summary of usage of Mandarin language

Participants	Daily usage of languages	Places of Mandarin language usage	Using the Mandarin language during communication
1	1. Malay 2. Mandarin 3. Malay Dialect 4. English	At home At school	friends
2	1. Malay 2. Mandarin 3. English	At school	friends
3	1. Malay Dialect 2. Malay 3. English 4. Mandarin	At school	friends
4	1. Malay Dialect 2. Malay 3. English 4. Mandarin	At school	office colleagues
5	1. Malay dialect 2. Malay 3. English 4. Mandarin	At school	friends
6	1. Malay dialect 2. Malay 3. English 4. Mandarin	At school	friends office colleagues

Table 3.2 continued

7	1. Malay 2. English 3. Mandarin	At school	friends
8	1. Malay dialect 2. English 3. Malay 4. Mandarin	At school	office colleagues
9	1. Malay 2. Malay Dialect 3. Mandarin 4. English	At school	friends
10	1. Malay 2. English 3. Mandarin	At school	friends

3.2 Materials

This research studied English and Mandarin consonants from both perceptual and productive perspectives. Two tasks were accordingly designed to answer the two research questions: a perception task and a production task.

The purpose of the perception task was to compare Malaysian Malay speakers' perception of Mandarin aspirated stops to their perception of Mandarin non-aspirated stops and to explore how well the Malay Mandarin learners perceive Mandarin consonants. The perception task data was used to address the first research question.

Next, the purpose of the production task was to test Malaysian Malay speakers' ability to differentiate aspirated from unaspirated Mandarin stop consonants during the production of Mandarin stop consonants in terms of VOT values. This task was designed to address the following issues. First, it explored which stop consonants are difficult to be produced by Malaysian Malay speakers, along with the acoustic features of Mandarin consonants pronounced by Malay Mandarin learners. The second issue it examined was the influences of L1 in the production of L2 stop consonants. Lastly, this task investigated the phenomena of Malaysian Malay speakers' differences in L2 speech learning. The production task data was to answer the second research question. Prior to the perception and production tests, each participant filled out a consent form and questionnaire regarding their language usage and background.

3.2.1 Questionnaire

In order to contribute to this research, the participants who were selected had to give their consent willingly before any further procedures. Therefore, a consent form was distributed to these participants through Whatsapp for them to approve (refer to Appendix 1 for the Informed Consent Form). This step assured them of the anonymity of their personal information in this study. Next, the participants completed a language background questionnaire in order to attain their demographic data (refer to Appendix 2). Malaysian Malay learners' background in language learning, the quantity of time spent in an L2 environment, and their skill level in the language were the primary goals for the questionnaire. Additionally, participants also provided personal data, including gender, age, and level of education, as well as information about where and when they learned each language and their HSK band.

3.2.2 Perception Test

The audio file for the perception test was recorded by a Mandarin language teacher. The teacher is a female aged 31 who was born and raised in China and was living in Malaysia at the time of recording. She is a native speaker of the Mandarin language from China who works at the Universiti Malaya's Kong Zi Institute as a language teacher. This institute is a member of the international CI network developed by Hanban (Beijing) and the Beijing Foreign Studies University (BFSU) for people who wish to learn the Chinese language. The teacher has experience teaching Mandarin as a second language for eight years and possesses a Master of Chinese as a Second Language degree. She was thus deemed a suitable candidate to pronounce the words in the material for the perception test, especially because she had taken the Putonghua Proficiency Test (PSC).

The PSC is a recognised examination of Standard Chinese (Mandarin) speaking proficiency designed for native Chinese language speakers. All government employees and teachers in China must have PSC certification of a particular minimum level, as per a number of relevant implementation procedures by the government. The test is entirely composed of spoken Mandarin. It necessitates rigorous adherence to the Standard Chinese phonology, which includes elements like retroflex initials, weak syllables, and *erhua*. The PSC consists of five sections. The first section requires test-takers to read 100 monosyllabic words to test their pronunciation; the second section demands them to read 100 polysyllabic words for the same purpose; the third section calls for them to read out the right answer from multiple choices to assess their vocabulary and syntax; the fourth section involves reading a 400-word essay to measure their fluency; and the last section is to speak for three minutes about a subject selected by the examiners.

Candidates receive a Certificate of Putonghua Proficiency Level upon passing the exam. There are three levels of scoring for the PSC exam, and each level has two classes with grades A and B. It can be summed up that Level 1-A, which is 97% accuracy, is necessary for nationwide and province-level radio and television presenters. Level 1-B, which is 92% correct, is necessary for teachers of the Chinese language in the north of China, while Chinese language teachers in the south of China require Level 2-A, which is 87% correct. Level 2-B, which is 80% correct, is necessary for teachers of other languages in the country. Next, at 70% and 60% accuracy, respectively, Level 3-A and Level 3-B are demanded to work in the civil service. The language teacher chosen in this study for the perception test had achieved Level 1-B in her PCR test, which shows that she has a good level of mastery of the Mandarin language.

Before beginning the recording, a consent form was distributed to the teacher. The materials for the perception test were recorded using a SONY R33021 recorder with a TAKSTAR SGC-578 external microphone with a 44.1kHz sampling rate in a quiet room at Universiti Malaya. The teacher was asked to read 24 words towards the microphone, whereby the distance between her mouth and the front of the microphone was around 30cm. She read each word individually, clearly, and naturally. The recording was transferred from the recorder to a USB flash drive. After the recording was done, three other native speaker participants acted as production test judges and evaluated the language teacher's utterances through Google Forms. Specifically, the native speakers were asked to listen to the language teacher's utterances and write down the words they heard.

The 24 words in the test contained the six /p/ /t/,/k/,/p^h/, /t^h/, and /k^h/ stop consonants. (refer to Appendix 3). Conventionally, a Chinese syllable comprises two sections: the initial (59heng mǔ 声母) and the final (yùn mǔ 韵母). The former is generally the sole consonant at the start of a syllable, whereas the latter represents whatever follows. The finals used in this test were classified into two groups: the first one is open-mouth finals (kāi kǒu hū 开口呼), which commence with a non-high vowel (i.e., [a] or [o]), and the second one is close-mouth (lip-rounding) finals (hé kǒu hū 合口呼), which commence with a high-back rounded vowel [u]. For example, in the word 土（tǔ）, "t" is the initial (consonant), and "u" is the close-mouth final (monophthongs). Another example is the word 跑 "pǎo", where "p" is the initial (consonant) and "ao" is the open-mouth final (diphthongs). A detailed explanation of the syllables of the Mandarin language has been provided in Chapter Two.

3.2.3 Production Test

In the second task of this study, the participants were required to take a production test, which was recorded in a quiet room. The participants were requested to utter a collection of target terms located in the middle of carrier sentences (refer to Appendix 4). The stimulus material in this research consisted of a wordlist of Mandarin six stops consonants, with three tokens for each stop consonant. The targeted stop consonants always occurred in the second syllable's word-initial position. Each of the target words contained either diphthongs or monophthongs. For example, the word *太 tài* contains diphthongs (two or three vowels together), while the word *顾 gù* contains monophthongs (single vowel). The syllables were either a CV or a CVV in terms of structure.

3.3 Research Design

The goal of a research design is to make sure that the data collected allows the researcher to provide a clear response to the original research questions (de Vaus, 2001). To answer the two research questions of this study as unambiguously as possible, data on Malay Mandarin teachers' backgrounds, as well as their ability to perceive and produce Mandarin stop sounds, was needed. Accordingly, this research comprised three key phases. First, the preliminary phase was the development of the research instruments, which included the generation of words and sentences for the production test and listening words for the perception test. The second phase was the collection of data, during which the researcher carefully selected the participants according to the criteria given. Then, the participants were made to undergo the perception test to examine whether Malaysian Malay speakers can differentiate Mandarin voiceless unaspirated and aspirated stop consonants in auditory perception. Following that, the participants underwent the production test so as to discover the acoustic characteristics of Malaysian Malay speakers' Mandarin stop consonants. Finally, the third phase was the analysis of the data and presentation of the findings. In this phase, the researcher identified error patterns and presented them in tables and figures. Then, the researcher described and explained the patterns shown by the participants using the SLM. The results on the participants' production and perception of stop consonants, as well as the link between these two, were also discussed. Subsequently, the researcher identified possible factors that cause errors in production and perception. Lastly, conclusions were drawn based on the goals and findings of the overall study. The figure below shows the diagram of the research design of this study:

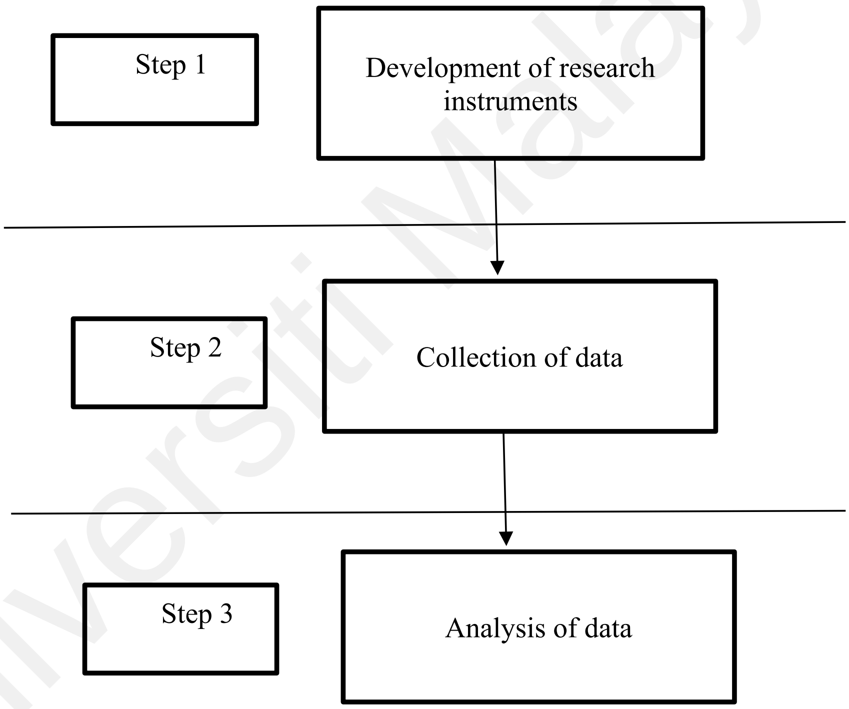

Figure 3.1: Research design

3.4 Procedure for Data Collection

The researcher took one day to complete the data collection process. Every participant received an identification number that was to be written down on their questionnaire, paper sheet (perception test), and recording (production test). As it is necessary for research ethics to treat data with confidentiality and to conceal the relationship between the data and participants by preserving their anonymity, the identification number technique was employed to secure the participants' identities. Therefore, without using the respondents' real identities, the identifying numbers allowed for the detection of the subjects' responses throughout data analysis. Prior to conducting the tests, the researcher gave the participants a short briefing about how to carry out the perception test and production test.

3.4.1 Perception Test

The perception test was conducted by the researcher. The test was done in the school computer lab. First, the participants listened to each word just one-time using headphones. Next, they wrote down the *pinyin* of the words on a piece of paper. The participants took the perception test only once. Each participant was awarded 10 minutes for finishing the task. The perception test consisted of one task, which was the identification task. On response sheets, participants wrote down the *pinyin* for each word they heard on the recording. *Pinyin* is a romanization system based on the pronunciation of Chinese characters. The term literally translates to "spell sound" in Mandarin Chinese.

3.4.2 Production Test

The production test was conducted and recorded by the researcher at the school lab. The production test of each participant was recorded using a SONY R33021 recorder and a TAKSTAR SGC-578 external microphone with a 44.1kHz sampling rate in a quiet room. Each participant had 10 minutes of preparation before recording. They were requested to read out 18 sentences three times toward the microphone placed around 30 cm from their mouth. The participants were asked to read the sentences individually, clearly, and naturally. Participants were also informed before the task began that their responses would be recorded. The recording provided a sample of highly formal or controlled speech. If the participants made mistakes because of nervousness during recording, they were allowed to record once again. This task took about 15 minutes per participant.

3.5 Data Analysis

3.5.1 Perception Test

There were three steps in the data analysis. First, the researcher sorted out the 24 tokens by each participant. Next, the perception data was transferred into an Excel sheet, and the tokens were listed for each participant. Third, the author counted the total number of right and wrong *pinyin* the participants had written.

The correct rate of stop consonants written in *pinyin* for each participant and the overall correct rate in the perception test were calculated in the following way:

a) The correct rate of stop consonants written in *pinyin* for each participant =

$$\frac{\text{the total number of incorrect stop consonants perceived by one participant}}{\text{the total number of the stop consonants contained in one material}} \times 100\%$$

b) The overall correct rate of stop consonants written in *pinyin* for all the participants =

$$\frac{\text{the total number of correct stop consonants perceived by all the participants}}{\text{the total number of the stop consonants contained in one material}} \times 100\%$$

Figure 3.2 is an example of how the perception data were analyzed. The participants had to be able to discriminate the aspirated and unaspirated stop consonants by writing down the *pinyin*. Figure 3.3 shows that the individual participant's correct identification of every target consonant was computed as their percentage of accuracy score. The target sounds that were wrongly identified were calculated as the error rate percentage.

PARTICIPANTS	杯	包	布	爸			叛	泡	床铺	店铺
PARTICIPANT 1	bei	bao	bu	ba			pan	pao	pu	pu
PARTICIPANT 2	bei	bao	bu	ba			pan	pao	pu	pu
PARTICIPANT 3	bei	bao	bu	ba			pan	pao	pu	pu
PARTICIPANT 4	bei	bao	bu	ba			pan	pao	pu	pu
PARTICIPANT 5	bei	bao	bu	ba			pan	pao	pu	pu
PARTICIPANT 6	bei	bao	bu	ba			pan	pao	pu	pu
PARTICIPANT 7	bei	bao	bu	ba			pan	pao	pu	pu
PARTICIPANT8	bei	bao	bu	ba			pan	pao	pu	pu
PARTICIPANT 9	bei	bao	bu	ba			pan	pao	pu	pu
PARTICIPANT10	bei	bao	bu	ba			pan	pao	bu	pu
PARTICIPANTS	独	蛋	队	刀	毒		图	头	特	驼
PARTICIPANT 1	du	dan	dui	dao	du		tu	tou	te	tuo
PARTICIPANT 2	du	dan	dui	dao	du		tu	tou	te	to
PARTICIPANT 3	du	dan	dui	dao	du		tu	tou	te	tuo
PARTICIPANT 4	du	dan	tui	dao	du		tu	tou	de	tuo
PARTICIPANT 5	du	dan	dui	dao	du		tu	tou	te	tuo
PARTICIPANT 6	du	dan	dui	dao	du		tu	tou	te	tuo
PARTICIPANT 7	du	dan	dui	dao	du		tu	tou	te	tuo
PARTICIPANT8	du	dan	dui	dao	du		tu	tou	te	tuo
PARTICIPANT 9	du	dan	dui	tao	tu		tu	tou	te	tuo
PARTICIPANT10	du	dan	tui	dao	du		tu	tou	te	tuo

Figure 3.2: Screenshot of analysis of perception data

PARTICIPANTS	RESULTS	PERCENTANGE
PARTICIPANT 1	10	100%
PARTICIPANT 2	10	100%
PARTICIPANT 3	10	100%
PARTICIPANT 4	9	90%
PARTICIPANT 5	10	100%
PARTICIPANT 6	10	100%
PARTICIPANT 7	10	100%
PARTICIPANT 8	10	100%
PARTICIPANT 9	8	80%
PARTICIPANT 10	9	90%

CORRECT RATE OF PERCEPTION

$$\frac{236 \text{ TOKENS}}{240 \text{ TOKENS}} \times 100\% = 98\%$$

Figure 3.3: Screenshot of percentage rate of stop consonants

3.5.2 Production Test

Target words were extracted from the carrier sentences through Windows Media Player and saved as WAV files. Praat 6.0.49 was used to analyze the WAV files that were uploaded (Boersma & Weenink, 2022). For each of the tokens obtained from the production tasks, three acoustic properties encoding Mandarin stop consonants were measured using PRAAT. This approach was adopted for segmenting and annotation, and the boundaries of the segments were manually drawn based on visual examination of the spectrographic and waveform data. They were thereafter divided into individual utterances and coded for each speaker separately. A text grid was subsequently created for each word. The text grid function was employed to annotate each stimulus on one text grid line, using a second text grid line to measure the acoustics of the stop consonants.

A two-tiered window was utilized to display speech recordings. There were two tiers of data: the top one being waveforms and the bottom one being spectrograms. The start and end of a specific sound measurement were assessed from both the waveform and spectrogram tiers. Waveform analysis of the temporal space from the release of the burst to the beginning of the subsequent vowel can be used to determine the VOT value. VOT was determined to the nearest 0.1 ms in this study by positioning a left cursor at the beginning of a sharp increase in energy, signaling the release of a voiceless stop. The cursor was placed on the right side at the initial upward zero-crossing, at which point the periodicity signaling the vowel on set began. The start of a burst release determined the beginning of VOT, while the start of the waveform periodicity reflecting glottal vibration in the next vowel determined the ending.

The VOT was classified as positive when the wave shape or spectrogram showed no signs of pre-voicing and when there was any amount of delay between the release of the stop and the beginning of the subsequent vowel. On the contrary, when participants voiced before releasing the stop consonants, negative VOT was taken into account. The data measurements were extracted using a Praat script developed by DiCanio (2020) and then imported into Excel after evaluating every stop consonant at the start of the word. The following categories were included in the data measurement: starting of a syllable, ending of a syllable, duration of a syllable, starting of close duration, ending of close duration, duration of close duration, starting of VOT, ending of VOT, and duration of VOT. Figure 3.4 shows the Praat script that was used to automate the creation of speech sounds that vary by VOT. Such scripts automate a series of actions, saving Praat users' time and energy.

```
form 对话框
real the_HanZi_TierIndex_in_TextGrid 1
real the_ShengYun_TierIndex_in_TextGrid 2
real the_Analyse_TierIndex_in_TextGrid 3
endform

hanZitierNum= the_HanZi_TierIndex_in_TextGrid
shengYuntierNum= the_ShengYun_TierIndex_in_TextGrid
analysetierNum= the_Analyse_TierIndex_in_TextGrid

for ifile to fileNum
select Strings list
fileName$ = Get string... ifile
newFileName$ = fileName$ - ".TextGrid"
textGridFileName$= newFileName$ +".TextGrid"
textGridFileName$ = sourceDirPath$ +textGridFileName$
saveFileName$= newFileName$ +"_TextGrid.csv"
saveFileName$ = targetDirPath$ +saveFileName$

filedelete 'saveFileName$'
fileappend 'saveFileName$' 文件名
fileappend 'saveFileName$' ,
fileappend 'saveFileName$' 音节名称
fileappend 'saveFileName$' ,
fileappend 'saveFileName$' 音节起点
fileappend 'saveFileName$' ,
fileappend 'saveFileName$' 音节末点
fileappend 'saveFileName$' ,
fileappend 'saveFileName$' 音节时长
```

Figure 3.4: Screenshot of sample Praat script

Rater reliability for VOT results was evaluated by randomly choosing 10 measured tokens from each of the 10 participants. These responses were randomly assessed by the researcher's supervisor. Figures 3.5, 3.6, and 3.7 depict samples of annotations for voiceless aspirated stop [kʰ]. For each of the tokens, the VOT is marked in red. Left to right indicates the order of time. The large dark region begins to fill the upper half of the image halfway over the spectrogram, moving from left to right. This shows where the consonant k sound starts (refer to Figure 3.5). To the right of this, the thick bar with graph points starts. This denotes the start of voicing or the beginning of the vowel sound (refer to Figure 3.6). As a result, the interval between the start of the darker section and the start of the thick graph line represents the time interval between the start of the voiceless stop (/kh/) and the start of the subsequent vowel, or the VOT.

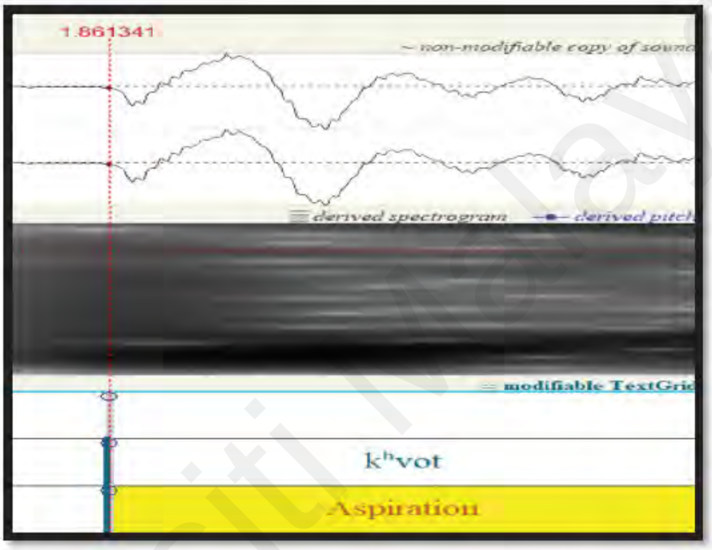

Figure 3.5: Release burst of /k/ followed by aspiration

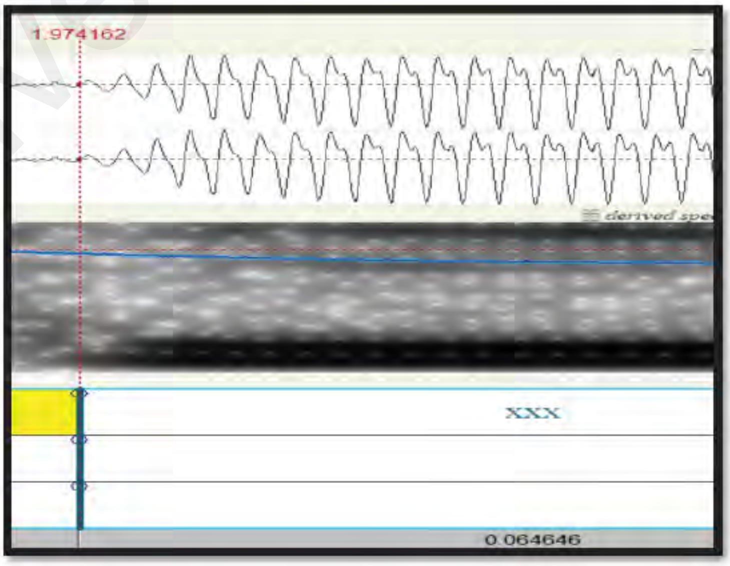

Figure 3.6: Vocal fold vibration, associated with the production of vowel (/u/)

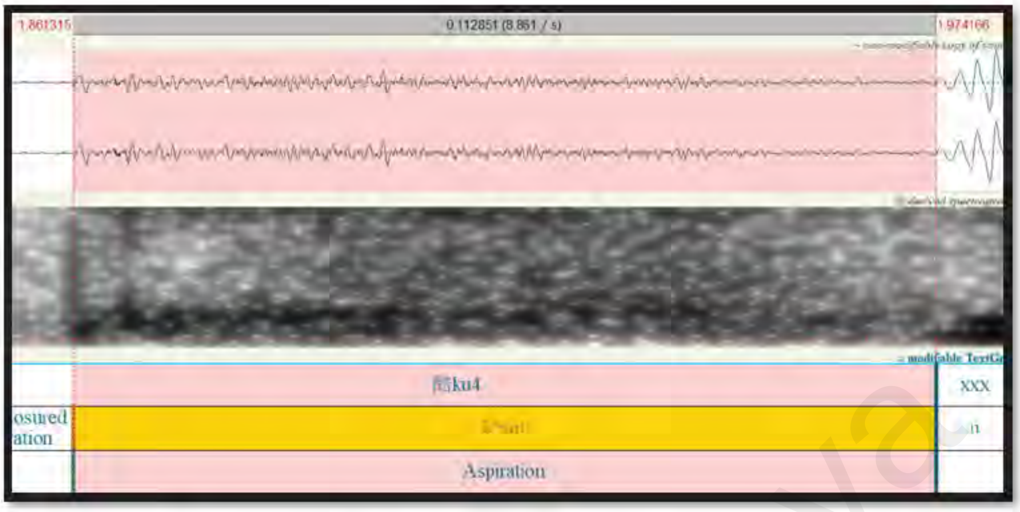

Figure 3.7: An acoustic waveform (top tier) and spectrogram (bottom tier) are shown for the word "ku4" (酷)

The waveform and spectrogram of the Mandarin word "ku4"(酷) produced with aspiration are shown in Figure 3.7. Vocal fold vibration, associated with the production of the following vowel (in this case, /u/), starts later (112 ms after the burst). The highlighted wave represents the area from which the measurement for this sound was taken. A positive VOT is clearly visible in the waveform, which indicates that this particular / k^h / has a noise duration (aspiration) after the stop burst and before the beginning of voicing. Kent and Read (2002) stated that long-lag VOT signifies an early onset of vocalization that is far behind the transient, indicating that the stops' vocalization release time is lengthy (exceeding around 35 ms).

In Figure 3.8, the waveform and spectrogram of the Mandarin word 龟 (gui1) are shown, which is produced without aspiration. As the figure shows, voicing starts almost immediately (23 ms) after the burst. The voiceless unaspirated stop is determined with a voice onset time of zero (or close to it). Plosive release and vocal fold vibration begin at the same time, whereby the short lag VOT is used as the time between plosive release and vocal fold vibration is very short (less than 30 ms) (Kent & Read, 2002).

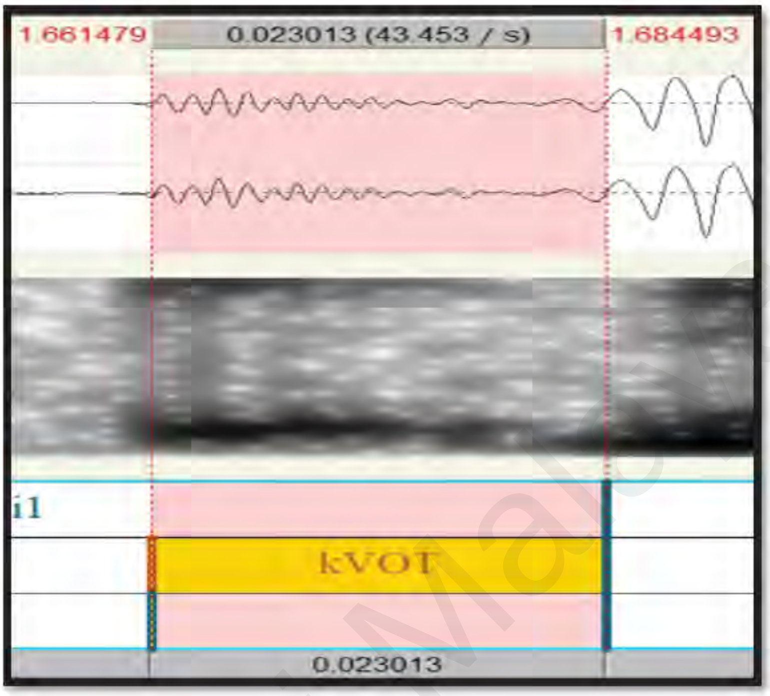

Figure 3.8: An acoustic waveform (top tier) and spectrogram (bottom tier) shown for the word "gui1" (龟)

Lastly, the author made some statistical calculations with the help of Office Excel 2016. The overall correct rate of production of stop consonants was calculated by the following formula:

a) The overall correct production rate of VOT for all the participants =

$$\frac{\text{the total number of correct stop consonants produced by all the participants}}{\text{the total number of the tokens contained in one consonant production}} \times 100\%$$

3.6 Theoretical Framework

As previously mentioned, this study used Flege's SLM taxonomy of errors to identify the causes of mistakes and provide an explanation. Based on how the L2 sounds are interpreted, the SLM defines three potential groups: 1) Sounds that are similar to L1 sounds that are considered to be realizations of L1 sounds (diaphones); 2) L2 sounds that are perceived to be identical to L1 sounds (transfer); 3) New phonetic categories that are generated for phones that are not matched with those in the mother tongue. The model contends that accurate L2 speech output is dependent on accurate perception. In other words, native-like production and native-like perception are mutually exclusive (refer to Chapter Two for more details).

According to the SLM, /p^h, t^h, and k^h/ are new sounds to Malay Mandarin learners. The Malay learners either establish new categories or assign these three sounds to their L1 by the so-called 'equivalence classification' process, depending on how phonetically dissimilar Malay speakers find aspirated and unaspirated stops. Chinese aspirated stops are referred to as new phones which have no L1 counterpart. However, due to the fact that Malay Mandarin learners have English experience, Malay Mandarin speakers with such experience might exhibit a different acquisition of Mandarin stop consonants. This idea is prompted by the theory that learners are influenced by other foreign languages (L2) that are closer to the L3 (Ellis 1994).

3.7 Statistical Analyses

A paired sample T-test was carried out in Graph Pad Prism (2021) for the production task to see if there were any significant differences in VOT data between the two group means (the unaspirated and aspirated stop consonants). The findings were interpreted as statistically significant only if the probability of an error occurring was less than 0.05%.

3.8 Conclusion

This chapter has delineated the study's methodology, from the selection of participants to the data collection and the analysis methods of Mandarin stops produced by Malay speakers. The findings of the study are deliberated in depth in the next chapter.

CHAPTER 4: FINDINGS

The current chapter presents and discusses the findings from this research, which are based on perception and production tasks. The result of perception errors committed by each participant pertaining to the six stops consonants is elaborated under the perception task. Meanwhile, for the production task, the average VOT values of every stop consonant containing 18 words for each participant are presented and interpreted. Each perception and production result is discussed overall first and then individually.

4.1 Perception Test

4.1.1 Overall Perception Test Results

This section describes each participant's perception results. In summary, most of the participants did not make any perception errors except for three participants. Participant 9 made two perception errors, while Participant 4 and Participant 10 made the smallest number of perception errors, with one each. The overall accuracy rate was 98% across the participants, showing that these Malay speakers of Mandarin were able to perceive the contrasts of interest in this study with a high degree of accuracy. The misperception rate was only 0.2%. Detailed findings are reported in the sections below.

4.1.1.1 Number of perception errors on /p, p^h / and /k, k^h / across participants

All 10 participants were able to identify all 16 words which contained the consonants /p,p^h/ and /k and k^h/. They could constitute 100% of the words correctly.

4.1.1.2 Number of perception errors on /t/ and /t^h/ across participants

Table 4.1 shows the mispronounced stop consonants of /t/ and /t^h/ by the three participants alongside the corrected consonants.

Table 4.1: Mispronounced stop consonants of /t/ and /tʰ/

Word	IPA transcription	Participant	Manner of articulation each misarticulated sound
毒(dú)	/tu/	No 9	/tʰu/
刀(dāo)	/taʊ/	No 9	/tʰao/
队(duì)	/tuei/	No 10	/tʰui/
特(tè)	/tʰe/	No 4	/te/

Eight participants were able to identify all eight words which contained /t/ and /tʰ/ consonants. From Table 4.1, Participant 9 did not identify two out of eight words. Participant 9 wrote the pinyin of the word 毒(dú) as (tú) and 刀(dāo) as (tāo). Participant 10 did not identify one out of four words. Participant 10 wrote the *pinyin* of the word 队（duì) as (tuì). Moreover, nine participants were able to identify all four words which contained the consonant /tʰ/. Participant 4 did not identify one out of four words. Participant 4 wrote the *pinyin* word 特(tè) as (dè).

4.1.1.3 Summary of Perception Test Result

In the perception test, the overall accuracy rates of bilabial plosives, alveolar plosives, and velar plosives were 100%, 95%, and 100%, respectively. Among the three stop consonants, /k/-/kʰ/ and /p/-/pʰ/ acquired the best perception results while /t/-/tʰ/ received the worst, implying that /k/-/kʰ/ and /p/-/pʰ/ are the easiest to discriminate, and on the contrary, /t/-/tʰ/ are most difficult for learners to discriminate.

4.2 Production Test

As explained in Chapter Three, all the participants produced three tokens of six stop consonants (three repetitions x 10 subjects x 18 stop consonants). The mean and standard deviation (SD) values of VOT measurements are presented in milliseconds (ms) for each participant. First, the overall mean VOT values for six stop consonants of all the participants are discussed below, followed by individual production results.

4.2.1 Overall Production Test Results

4.2.1.1 Production of /p/ and $/p^h/$

Figure 4.1 illustrates the overall mean and SD for /p/ and $/p^h/$ of the 10 participants. The overall mean value of VOT for /p/ was 46 ms (SD 31 ms), and the mean value for $/p^h/$ was 57 ms (SD 28 ms).

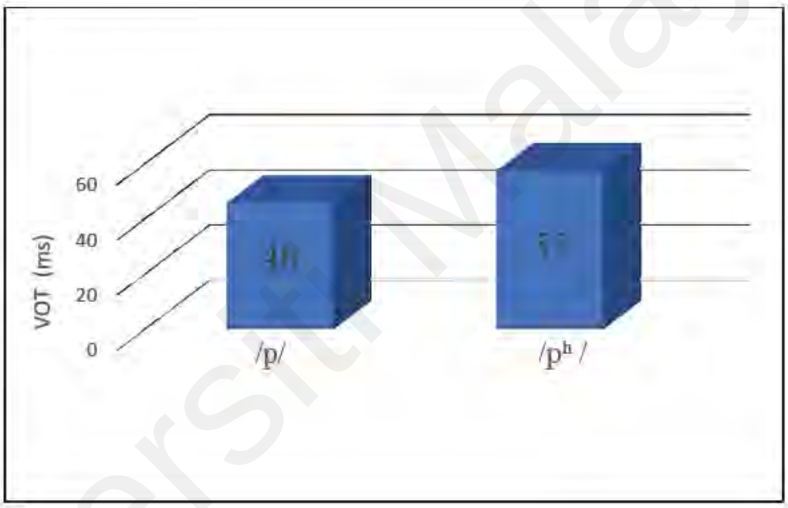

Figure 4.1: Mean VOT of /p/ and $/p^h/$

The statistical significance of the difference between /p/ and $/p^h/$ was inspected through a paired sample T-test. There were no significant production variations between /p/ and $/p^h/$ ($t(9)=1.83, p=.100 < .05$). This indicates that some Malay participants face more difficulty in producing the aspirated voiceless $/p^h/$.

4.2.1.2 Production of /t/ and /t^h/

Figure 4.2 illustrates the overall mean and SD for /t/ and /t^h/ of the 10 participants. The overall mean values of VOT for /t/ was 25 ms (SD 17) and for /t^h/ was 73 ms (SD 37).

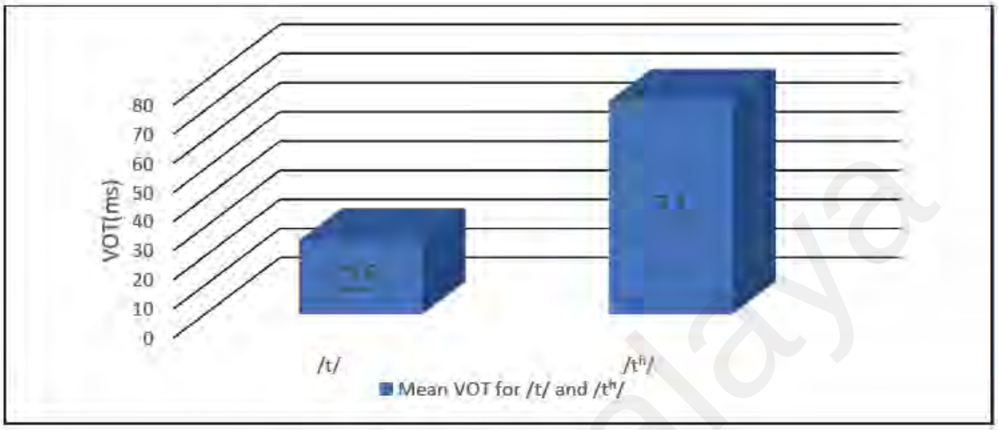

Figure 4.2: Mean VOT for /t/ and /t^h/

The statistical significance of the difference between /t/ and /t^h/ was measured through a paired sample T-test, which showed that there were significant production differences between /t/ and /t^h/ ($t(9)=3.28$, $p=.009 < .05$). This indicates that Malay participants can differentiate aspirated voiceless /t/ from the unaspirated /t^h/ stop consonants.

4.2.1.3 Production of /k/ and /k^h/

Figure 4.3 illustrates the overall mean and SD for /k/ and /k^h/ of the 10 participants. The overall mean value of VOT for /k/ was 37 ms (SD 13 ms), and for /k^h/ was 80 ms (SD 25 ms).

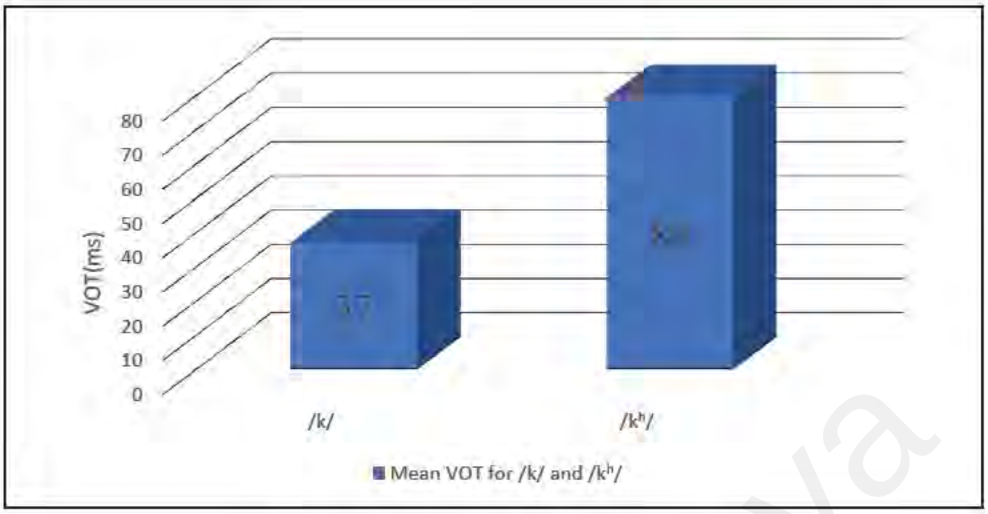

Figure 4.3: Mean VOT for /k/ and /kʰ/

The difference between /k/ and /tʰ/ was assessed for statistical significance through a paired sample T-test, which revealed that there were significant production differences between /k/ and /kʰ/ ($t(9)=3.84$, $p=.004 < .05$). This indicates that Malay participants can differentiate aspirated voiceless /k/ from the unaspirated /kʰ/ stop consonants.

4.2.2 Individual Production Results

This section presents each participant's production accuracy rate. Individual results were discussed because when it comes to learning a second language (L2) speech, Fledge argues that the input quantity and quality are significantly more essential compared to the age of initial exposure. Each participant's experience may therefore differ. L2 input could further be affected by the resident's length of stay (LOR).

4.2.2.1 Production of /p/ and /pʰ/

The results for each of the individual participants for /p/ and /pʰ/ are presented in Tables 4.2 and 4.3. Table 4.2 summarizes the average VOTs of the Mandarin /p/ stop consonant produced by the 10 participants. The mean VOT values ranged from 6ms to 141ms. As seen in the table, the highest VOT value was produced by Participant 9, whose

VOT value was found to be 141ms. Similarly, Participant 2 had a high VOT value of 140ms. A strikingly low VOT of 6ms was produced by Participant 6. In general, seven out of 10 participants pronounced /pɑu/ as an aspirated voiceless stop consonant, while four out of 10 participants pronounced /pai/ as an aspirated voiceless stop consonant. Only one out of 10 participants pronounced /pu/ as an aspirated voiceless stop consonant.

Table 4.2: VOT of /p/ consonant

Participants	Mean VOT(ms)			Mean VOT
	/pɑu/	/pai/	/pu/	
1	46 (49)	7 (0.2)	15 (2)	23 (21)
2	140 (7)	53 (9)	15 (5)	69 (64)
3	13 (3)	10 (4)	14 (5)	13 (2)
4	120(3)	10 (0.5)	15 (2)	49 (62)
5	107 (11)	42 (57)	10 (5)	53 (49)
6	6 (1)	95.5(9)	18 (12)	40 (49)
7	18 (4)	12 (1)	17 (4)	15 (3)
8	100 (21)	7 (0.3)	27(4)	45 (49)
9	141 (13)	13 (5)	74 (53.2)	76 (64)
10	116 (16)	85 (9)	16 (2)	72 (51)
Avg(SD)	81(54)	34 (34)	22(19)	46 (31)

Note* Numbers in parentheses show the standard deviation (SD)

As seen in Table 4.3, the mean VOT values for /pʰ / ranged from 9ms to 103 ms. The highest VOT value was produced by Participant 10, which was 103 ms. By contrast, the lowest VOT value was produced by Participant 1, which was 8ms. In general, nine out of 10 participants pronounced /pʰəi/ as a voiceless unaspirated stop consonant, while two out of 10 participants pronounced /pʰA/ as an unaspirated voiceless stop consonant. Four out of 10 participants pronounced /pʰan/ as an unaspirated voiceless stop consonant.

Table 4.3: VOT of /pʰ/ consonant

Participants	Mean VOT (ms) (SD)			Mean (SD)
	/pʰəi/	/pʰa/	/pʰan/	
1	44 (40)	8 (3)	11 (2)	21 (46)
2	92 (14)	101(4)	102 (1)	98 (6)
3	9 (1)	13 (0.2)	56 (68)	26 (26)
4	55 (25)	6 (0.2)	9(0.4)	23 (27)
5	55 (10)	81 (8)	101 (5)	79 (23)
6	40 (48)	69 (53)	101 (26)	70 (31)
7	68 (11)	17 (4)	30 (13)	39 (27)
8	68 (27)	75 (11)	98 (18)	80 (16)
9	58 (10)	68 (61)	85 (7)	71 (14)
10	77 (57)	103 (82)	20 (5)	66 (42)
Avg (SD)	56 (23)	54 (39)	61 (40)	57 (28)

4.2.2.2 Production of /t/ and /tʰ/

As seen in Table 4.4, the mean VOT values for /t/ ranged from 9ms to 116ms. The highest VOT value was produced by Participant 9, which was 116 ms. By contrast, the lowest VOT value was produced by Participant 3 and Participant 5, which was 9ms. In general, seven out of 10 participants pronounced /tA/ as an unaspirated voiceless stop consonant. One out of 10 participants pronounced /tɑu/ as an aspirated voiceless stop consonant. Likewise, only one out of 10 participants pronounced /tuo/ as an aspirated voiceless stop consonant.

Table 4.4: VOT of /t/ consonant

Participants	Mean VOT (ms) (SD)			Mean (SD)
	/ta/	/tɑu/	/tuo/	
1	10 (2)	10 (2)	37 (27)	19 (16)
2	38 (43)	12 (0.5)	18 (2)	23 (14)
3	9 (0.6)	9 (0.3)	15 (3)	11 (3)
4	10 (2)	11 (4)	25 (7)	15 (8)
5	12 (2)	9 (3)	14 (2)	12 (3)
6	17 (3)	12 (0.8)	21 (5)	17 (5)
7	51 (47)	12 (0.7)	15 (5)	26 (22)
8	22 (3)	10 (3)	19 (6)	17 (6)
9	12 (0.5)	16 (5)	116 (93)	48 (59)
10	66 (52)	108 (7)	18 (4)	64 (45)
Avg(mean)	25 (15)	21 (31)	30 (40)	25 (17)

As seen in Table 4.5, the mean VOT values for /tʰ/ ranged from 11ms to 168ms. The highest VOT value was produced by Participant 8, which was 148 ms. By contrast, the lowest VOT value was produced by Participant 3, which was 11ms. Three out of 10 participants pronounced /tʰi/ as unaspirated voiceless stop consonants. Only one out of 10, namely Participant 10, pronounced /tʰu/ as an unaspirated voiceless stop consonant. Five out of 10 participants pronounced /tʰai/ as an unaspirated voiceless stop consonant.

Table 4.5: VOT of /tʰ/ consonant

Participants	Mean VOT (SD)			Mean (SD)
	/tʰi/	/tʰu/	/tʰai/	
1	17 (5)	97 (64)	16 (10)	43 (46)
2	119 (4)	148 (6)	105 (3)	124 (22)
3	46 (1)	78 (13)	11 (3)	45 (34)
4	18 (6)	97 (64)	16 (10)	44 (46)
5	90 (19)	138 (14)	106 (11)	111 (24)
6	59 (22)	96 (20)	17 (2)	57 (40)
7	45 (15)	89 (8)	53 (26)	62 (23)
8	89 (12)	168 (18)	123 (7)	127 (40)
9	100 (13)	98 (7)	79 (18)	92 (12)
10	17 (2)	22 (6)	24 (9)	21 (4)
Avg(mean)	60 (37)	103 (41)	55 (44)	73 (38)

4.2.2.3 Production of /k/ and /kʰ/

Table 4.6 shows that the average VOT values for /k/ ranged from 18ms to 105.8ms. The highest VOT value was produced by Participant 10, which was 106ms. By contrast, the lowest VOT value was produced by Participant 9, which was 18ms. Only one out of 10 participants pronounced /kuəi/ as an aspirated voiceless stop consonant. Half or five out of 10 participants pronounced /ku/ as a slightly aspirated consonant. Only one participant pronounced /kɑu/ as a slightly aspirated consonant.

Table 4.6: VOT of /k/ consonant

Participants	Mean VOT(SD)			Mean (SD)
	/kuəi/	/ku/	/kɑu/	
1	48 (62)	24 (1)	35 (1)	35 (12)
2	35 (7)	22 (2)	23 (3)	27 (7)
3	32 (8)	52 (18)	19 (2)	34 (16)
4	48 (20)	24 (1)	36 (0.8)	36 (12)
5	23 (2)	25 (4)	19 (2)	23 (3)
6	36 (5)	41 (9)	23 (4)	34 (9)
7	46 (8)	56 (61)	27 (4)	43 (15)
8	38 (4)	32 (29)	26 (2)	32 (6)
9	42 (86)	57 (8)	18 (3)	37 (22)
10	106 (74)	48 (6)	57 (6)	70 (31)
Avg(Mean)	39 (22)	38 (14)	28 (13)	37 (13)

Table 4.7 shows that the mean VOT values for /kʰ/ ranged from 31ms to 144ms. The highest VOT value was produced by Participant 10 at 144ms. By contrast, the lowest VOT value was produced by Participant 3 at 31ms. Two out of 10 participants pronounced / kʰai/ as an unaspirated voiceless stop consonant. Similarly, two out of 10 pronounced / kʰɣ / as an unaspirated voiceless stop consonant, while three out of 10 participants pronounced /kʰu/ as an unaspirated voiceless stop consonant.

Table 4.7: VOT of /kʰ/ consonant

Participants	Mean VOT(SD)			Mean (SD)
	/kʰai/	/kʰɣ/	/kʰu/	
1	74 (12)	83 (41)	37 (41)	64 (24)
2	85 (24)	106 (13)	83 (13)	91 (13)
3	36 (2)	31 (3)	66 (3)	44 (19)
4	72 (13)	85 (46)	37 (46)	65 (25)
5	71 (9)	144 (31)	116 (31)	110 (37)
6	84 (6)	96 (16)	120 (16)	100 (18)
7	67 (13)	115 (21)	64 (21)	82 (29)
8	64 (7)	116 (14)	125 (14)	102 (33)
9	83 (12)	118 (11)	91 (11)	97 (18)
10	36 (6)	43 (6)	38 (6)	39 (4)
Avg(mean)	67 (18)	94 (35)	78 (35)	80 (25)

4.3 Discussion

The discussion is divided into three sub-sections. First is a discussion about the perception test, followed by a discussion about the production test. Finally, the relationship between the perception and the production of this study is examined in further detail.

4.3.1 Perception Test

The perception test results indicate that three out of 10 participants did not identify a few Mandarin alveolar stop consonants. Specifically, Participant 4 did not identify the word 特(tè); Participant 9 did not identify the words 毒(dú) and 刀(dāo); and Participant 10 did not identify the word 队(dui). According to Stevens, Keyser, and Kawasaki (1986), unlike bilabial and velar stops, alveolar stops have less contact with the palatal region of the tongue. This means that they have more burst energy and a faster tongue release than bilabial and velar stops (larger linguopalatal contact). A significant reduction in proper discrimination may be caused as a result of the significant alveolar burst energy interfering with listeners' access to VOT for the alveolar-lenis-aspirated contrast. Thus, this reason may have caused Participant 4 and Participant 9 to fail to identify the alveolar stop consonants.

Overall, there were only a few errors in this perception test. Out of 240 possible items across the 10 participants (24 x 10), only four errors occurred. The overall accuracy rate of 98% shows that these Malay speakers of Mandarin Chinese could perceive the aspirated and unaspirated stop consonants in this study with a high level of precision. The misperception rate was only 0.2%, which suggests that Malay Mandarin learners are relatively sensitive and intuitive toward the perceptual aspect of Mandarin consonants.

Notably, the Malay Mandarin learners in this study could perceive aspirated stops better than unaspirated stops. This study's result is completely different from that of Lai

(2009), which was mentioned in Chapter Two. Lai's (2009) results showed that native Malay speakers could identify unaspirated consonants more accurately than aspirated ones. It can be concluded that although the stop consonant system of the Malay language has only voiced and voiceless unaspirated stops, the participants are also familiar with English, which contains aspirated stops, even though the VOT is slightly different. Prior research has reported that the English /p/, /t/, and /k/ are produced with VOTs that typically are higher than 62 ms (Docherty, 1992; Weismer, 1979). The Mandarin /p^h, t^h, k^h/ are considered aspirated or very aspirated, whereas the English /p, t, k/ fall under the slightly aspirated or aspirated level (Chao & Chen, 2008).

Therefore, it is reasonable to think that the English experience has helped the participants to become familiar with the feature of aspiration. The advantages of knowing the English language have enabled Malaysian Malay speakers to perceive Chinese stop consonants more accurately. The perception result of this study is consistent with the experiment done by Song (2019) on perceptual discrimination, which also showed that Malay students could distinguish between Chinese plosives, which are aspirated and non-aspirated.

Besides that, in SLA research, the term "length of residency" (LOR) has been employed extensively as an indicator of final success in learning L2 phonology (Flege, 2009; Moyer, 2009). The Malay Mandarin speakers in this research had studied the Mandarin language in Beijing for five years, indicating that LOR has a positive correlation with the L2 input quantity an L2 learner has received; the more L2 input gained by an L2 learner, the greater the likelihood that the L2 learner will master the L2.

4.3.2 Production Test

The Malay Mandarin learners in this study have distinct phonetic characteristics in their stop production since they had completed a vital period of language learning necessary to become fluent as native speakers. According to Lenneberg (1967), a foreign accent in an L2 is "evident" if it is acquired after puberty, as brain development and lateralization necessary for language function have been completed at that age.

For the discussion, this study used the reference values of stop VOTs (ms) for /p/ /t/ and /k/ and /p^h/ /t^h/ /k^h/ from Ran and Shi (2007) to determine whether the stops produced by the participants are short-lag or long-lag. VOT values were classified into three major categories by Lisker and Abramson in the 1960s: voicing-lead (-75 ms), short-lag (0 – 25 ms), and long-lag (> 60 ms). Table 4.8 shows the reference values of VOT in the Mandarin language.

Table 4.8: Reference values of stops' VOT (ms) /p/ /t/ /k/ and /p^h/ /t^h/ /k^h/ from Ran & Shi（2007）

Stops	/p/	/t/	/k/	/p^h/	/t^h/	/k^h/
VOT (ms)	12.9	13.4	30.1	105.6	103.6	111.7

4.3.2.1 Overall Accuracy Rate of Production Test by Malay Mandarin Learners

In this section, the overall accuracy rates of the production of Malay Mandarin learners are calculated, and the comparison between the production of aspirated and unaspirated stop consonants is made to explore their differences. As mentioned earlier, the aspirated and unaspirated features can be observed from VOT values.

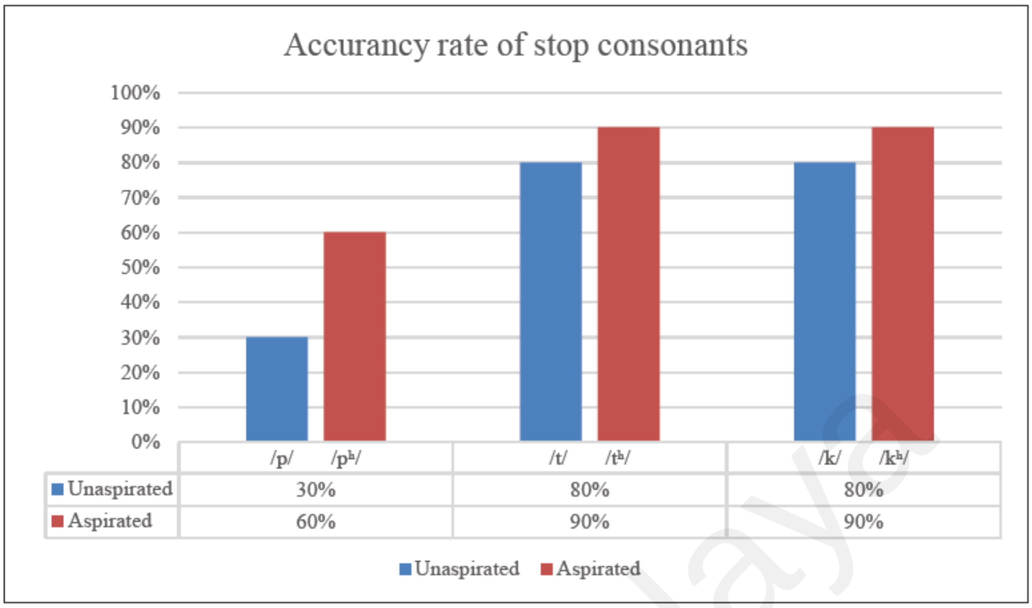

Figure 4.4: Accuracy rate of stop consonants

Figure 4.4 presents the overall accuracy rates of the production test. For the bilabial stops /p/ and /pʰ/, the accuracy rates of production were 30% and 60%, respectively. For alveolar stop consonants, the accuracy rate of /t/ was 80%, and that of /tʰ/ was 90%. Lastly, the accuracy rates for velar stop consonants /k/ and /kʰ/ were 80% and 90%.

Taking the production situation of the articulation place of stop consonants, the velar stop consonants and the alveolar stop consonants were the best produced, while the bilabial plosives were the worst produced. This result is consistent with the study by Shahidi et al. (2012), wherein Malay learners also encountered difficulty in the production of /b/ and /p/ in English.

4.3.2.2 General Discussion of Six Stop Consonants

Table 4.3 shows that Malay Mandarin speakers do have difficulty with /p/. Only 12 out of 30 tokens consisting of /p/consonants were produced as voiceless unaspirated consonants. Seven out of 10 participants produced the word /pau/ as a voiceless aspirated consonant. This basically shows that Malay Mandarin speakers are challenged by this consonant. As seen in the table, there is a wide variation in SD among the subjects. For example, Participant 6 had a VOT value of 6ms, while Participant 9 had, by contrast, a VOT of 141.1ms. The mean value for the /pau/ word was 81ms, which does not conform with Ran and Shi's (2007) categorization of the VOT value for /p/ as 12.9ms. Furthermore, if each participant's VOT value is individually examined, it can be noticed that some participants' VOT values are not in line with that of Lisker and Abramson. For example, Participant 6 has a short VOT (6ms), while Participant 2 and Participant 9 have long VOT values of 141.1ms and 140.2ms, respectively. For the production of the /pau/ word, seven participants produced the word as a voiceless aspirated consonant.

This research found that there is a particular pattern for the words /pau/and /pai/. The consonants with longer VOTs are likely to have longer vowels right after the /p/, and the word /pu/ with the shortest VOTs are likely to have back lax vowels. It can be concluded that when high vowels follow word-initial stops in Mandarin, the VOT is lengthier than when low vowels follow them. In addition, according to SLM theory, due to Mandarin unaspirated sounds being considered similar to those in the Malay language, they may be equivalently classified as an L1 sound, causing seven participants to have difficulty in producing them.

For the production of /p^h/, out of 30 tokens, only nine tokens were produced as unaspirated consonants. Most of the participants could produce /p^h/ as an aspirated bilabial stop. The table shows that the positive VOT is generally aspirated and produced in the long-lag area higher than 30ms, as per Ran and Shi (2007) and Khattab (2002). The VOT value produced by Participant 10, which was 103.5ms, is the nearest to 105.6ms. This VOT value is within the "highly aspirated" range (i.e., higher than 90 ms). This result shows that even though plosive consonants in the Malay language are always unaspirated, Malay speakers still can produce aspirated bilabial consonants very well because they have acquired English since school. This means they have learned to produce their voiceless stops with aspiration from the English language.

Tables 4.4 and 4.5 show that Malay speakers do not have difficulty producing /t/ and /tʰ/. Out of 30 tokens, only six tokens were produced as aspirated alveolar stop consonants. All the participants could exhibit positive VOT results in producing /t/ alveolar stop consonants. On the other hand, out of 30 tokens, only eight tokens of /tʰ/ were produced as unaspirated stop consonants. Compared to the VOT value for /tʰ/, which was 103.6ms **(**Ran & Shi, 2007), the VOT values of most of the participants were not above 100ms. As per Lisker and Abramson's (1964) and Cho and Ladefoge's (1999) classifications, stops in Mandarin and English are situated in close proximity on the VOT continuum; however, they do not fall into the same range on the continuum, particularly for voiceless aspirated occlusives. According to Chao's (2006) research, Mandarin is "highly aspirated" for voiceless aspirated stops, whereas English is "highly unaspirated." In summary, Malay speakers can produce aspirated alveolar stops. However, the VOT value is not equivalent to Mandarin VOT values.

Tables 4.6 and 4.7 show that Malay speakers do not have difficulty producing /k/ and /kʰ/. Out of 30 tokens,13 tokens were produced as long-lag. Numerous scholars have found articulation place to significantly influence the VOT time period for voiced and voiceless stop consonants. Studies have also demonstrated velar stops to record the highest VOT score and alveolar stops to record average VOT scores (Smith, 1978; Fischer & Goberman, 2010; Baum & Ryan, 1993; Kessinger, 1997; Klatt, 1975; Volaitis & Miller, 1992; Robb, Gilbert, & Lerman, 2005; Jancke, 1994). The VOT value produced by Participant 3, which was 31.8ms, is the nearest to 30.1ms. Out of 30 tokens, all the tokens of /kʰ/ were produced as long-lag. The VOT value given by Ran and Shi (2007) is 111.7ms. However, most of the participants' VOT values for aspirated velar stops were below 100ms.

The production results indicate that Malay learners can produce aspirated stops better than unaspirated stops. The Flege Learning Model can provide an explanation for the current findings. Malay speakers perform better on L2 phones that are dissimilar to L1 ones, necessitating the establishment of new phonetic categories, such as aspirated stops. On the other hand, the similar sound /p/ was difficult for Malay speakers to perceive because the similar L2 sound may be equivalently classified as an L1 (Malay language) sound without more detailed refinement.

In summary, a majority of the voiceless stop tokens in the present research were produced as aspirated, whereas several voiceless unaspirated stop tokens were produced as short-lag. It can be debated whether this language phenomenon is related to the absorption of equivalence classification (assimilation) or the creation of new categories (dissimilation). Given the ongoing nature of L2 learning, it is plausible that there is an additional category in the middle. As a result, L2 speech learning frameworks should take this possibility into account when describing "new" and "similar" sound productions.

4.4 The Relationship between Perception and Production in Mandarin Stop Consonants

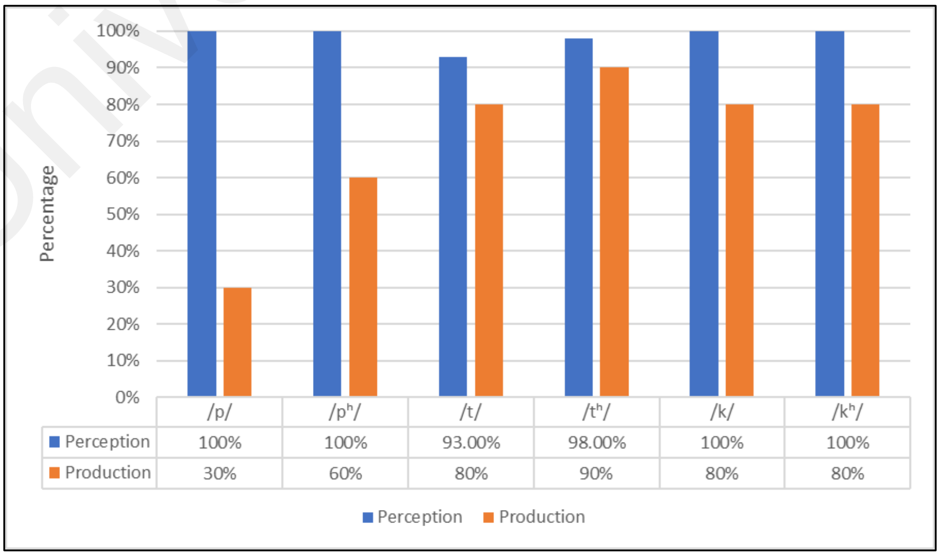

Figure 4.5: Mean percentage of perception and production of 10 participants

Figure 4.5 shows a comparison of the average accurate perception and production percentages for the six consonants, which mainly indicates higher perception scores than production scores. The SLM implies a strong link between perception and production, such that those who have good L2 speech perception are also competent at L2 speech production, as L2 perception can reflect L2 production abilities. This study corroborates extant research evidence on the significant impact of perception on vowel and consonant production, as well as the inclination of L2 speakers to perceive more accurately than produce (e.g., Baker & Trofimovich, 2006; Cardoso, 2011; Flege, 1993; Flege, Bohn, & Jang, 1997; Flege, MacKay, & Meador, 1999). According to Flege (1995), perception should come before production, and if students fail to perceive the target language accurately, they cannot accurately produce sounds in the second language either. This was supported by Chan's (2011) finding of a positive relationship between Cantonese learners' English word-initial consonant perception and production. She found that the perceptual performance of speakers who repeatedly produced the same sounds incorrectly was much lower than that of learners who always produced the sounds accurately. All the studies above support that learners should have adequate perception skills first; then, production skills should follow. Adding to the body of knowledge, this study has shown a small correlation between Mandarin stop perception and production by Malay speakers, which is that speech perception precedes speech production.

CHAPTER 5: CONCLUSION

5.1 Summary

This study attempted to investigate Malay Mandarin learners' perception and production of Mandarin stops. The purpose of this research was to observe their level of perception and production of similar sounds. To emphasize the most important findings, it is necessary to review the research questions outlined in Chapter One and develop a summary.

5.1.1 Do Malaysian Malay Speakers Differentiate Mandarin Voiceless Unaspirated and Aspirated Stop Consonants in Auditory Perception?

The perception results suggest that it is relatively easier for Malaysian Malay speakers to perceive Mandarin voiceless unaspirated and aspirated stop consonants. Indeed, they can master the perception of these phonemes after a time of study. The stop consonant system of the Malay language has only voiceless aspirated stops; however, the participants also know the English language, which contains aspirated stops. The advantages of knowing these two languages enable Malaysian Malay speakers to perceive Chinese stop consonants accurately.

5.1.2 What are the Acoustic Properties of Mandarin Stop Consonants Produced by Malaysian Malay Speakers?

In the production test, the most difficult consonant for the participants to produce was /p/, while the easiest consonants for participants to produce were /k^h/ and /t^h/. In summary, according to SLM theory, most Malay Mandarin speakers can create new phonetic categories for new sounds, which are the aspirated stops. Conversely, for the unaspirated stops like /p/, which are similar to the Malay language and English language, most Malay Mandarin speakers have difficulty in producing the unaspirated bilabial stop.

5.2 Implications

This study aimed to describe how Malay Mandarin learners perceive and produce Mandarin stop consonants based on a production task and a perception test. First, this study's findings add to the limited body of research on how Malay Mandarin learners perceive and produce Mandarin stop consonants. With this knowledge, Mandarin language course designers and teachers of Malay learners would be able to identify and pay closer attention to those sounds that can result in communication problems and the inability to understand each other.

Second, in order to help Mandarin teachers pronounce the stop consonant sounds correctly, teachers should understand the segmental features of consonants and vowels in the Mandarin language. As Roach (2009) stated, teachers should be able to detect several commonly expressed sounds (vowels and consonants). If Mandarin teachers do not understand Mandarin phonology and learn Mandarin by rote, it affects not only their speaking skills but also their listening skills.

Third, the instruction of Mandarin stops is a challenging area, as Malay Mandarin students encounter difficulties in both stop perception and stop production. Language teachers are advised first to comprehend the mapping of phonetic differences between the target language and the student's first languages. Assimilation of phonetic contrasts in the target language with pre-existing categories in the learner's first language is likely to occur during the early stages of language acquisition. In the stop system context, it is important first to determine if the initial and target languages are voiced or aspirating languages, depending on the expected cross-linguistic or mixed cross-linguistic effects.

Next, listening or perception skill is necessary for Mandarin teachers. If they cannot identify and differentiate sounds, it certainly causes problems in Mandarin language teaching and learning. As supported by Pennington (1996), listening and reiterating seem to be a two-way path where practising oral production can enhance auditory perception while concentrated listening can enhance oral production. Therefore, teaching the listening and pronunciation of individual problematic sounds is recommended.

Celce-Murcia (2004) proposed that instructors should determine which contrastive sounds are most difficult for learners for more focused exercises to be designed. In this case, practising minimal pair contrasts is recommended. Regarding pronunciation teaching, as seen from the findings, Mandarin teachers have limited English competencies for production, so teaching phonetics appears to be complicated for them. Therefore, it is a challenge for instructors to develop and design an appropriate syllabus for this group of teachers that addresses not only Mandarin pronunciation but also other skills.

Lastly, based on Nagle's (2020) highlight of the advantages of imitation, teachers could strengthen imitation training to assist students in producing voiced stops more effectively. Imitation could serve as a link between the perception and production of phonetics. It is also important to keep in mind that training results will also be significantly impacted by target linguistic competence. For instance, perception and production may be separate for novice learners, whereas they may be synchronous (or at the very least, there may be a linkage between perceptual precision and production precision) for more experienced learners (Flege, 1995, 2003; Flege et al., 1997) Therefore, when choosing the content and manner of teaching, teachers should consider the competence of students in the target language.

5.3 Recommendations for Further Research

This research has offered several valuable insights and conclusions on the perception and production of Mandarin stop consonants by Malay Mandarin speakers. However, it also has limitations that highlight areas requiring additional research and improvement. The study's sample size was one of its main shortcomings. A larger sample size and a broader diversity of participants are needed for future studies in the same field to generalize better and meet credibility requirements. With regard to the sample as well, the participants in this study were Mandarin Malay teachers in national schools from the Klang Valley. Therefore, future studies can explore the perception and production of Malay Mandarin teachers in different settings, such as teachers from every state in Malaysia or teachers from private schools.

Next, this study did not take into account gender as a potential variable affecting the correct pronunciation of stop consonants. Future studies looking into whether men or women perform better in stop consonants might find it interesting to take gender into consideration. The research on stop consonants in active communication, instead of in a list of words, is also an avenue for improvement. This is to generate findings that reflect the natural expressions of these phones rather than being read from a word list alone.

Finally, this research created only a reading for production text, excluding alternative techniques like interviews and open conversation. As a result, it was harder to get learners to give their pronunciation in a more natural and conversational setting. Future studies in this field are anticipated to use a more organic and communicative approach to assess Malay learners' production of stop consonants.

5.4 Concluding Remarks

This study focused on specific aspects and patterns in Malay Mandarin speakers' perception and production of Mandarin stop consonants. The results of this research may spur further investigation into other areas of difficulty for Malay second language learners in acquiring Mandarin pronunciation.

REFERENCES

Altmann, H., & Kabak, B. (2011). Second language phonology. In B. Botma, N. C. Kula & K. Nasukawa, (Eds.), *The continuum companion to phonology*, (pp. 298-319). Continuum.

Aoyama, K., Flege, J. E., Guion, S. G., Akahane-Yamada, R., & Yamada, T. (2004). Perceived phonetic dissimilarity and L2 speech learning: the case of Japanese /r/ and English /l/ and /r/. *Journal of Phonetics*, *32*(2), 233–250. https://doi.org/10.1016/S0095-4470(03)00036-6

Asher, J. J., & García, R. (1969). The optimal age to learn a foreign language. *The Modern Language Journal*, *53*(5), 334–341. https://doi.org/10.2307/323026

Baker, W., & Trofimovich, P. (2006). Perceptual paths to accurate production of L2 vowels: The role of individual differences. *International Review of Applied Linguistics in Language Teaching*, *44*(3), 231-250. http://dx.doi.org/10.1515/IRAL.2006.010

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In O.-S. Bohn & M. J. Munro (Eds.), *Language experience in second language speech learning: In honor of James Emil Flege* (pp. 31-52). John Benjamins.

Best, C. T., McRoberts, G. W., & Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *Journal of the Acoustical Society of America, 109*(2), 775-794. https://doi.org/10.1121%2F1.1332378

Best, C., Levitt, A. G., & Mcroberts, G. W. (1991). *Examination of language-specific influences on infants' discrimination of prosodic categories* [Paper presentation]. International Congress of Phonetic Sciences, Aic-en-Provence, France. https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS1999/papers/p14_1261.pdf

Boersma, P., & Weenink, D. (2022). *Praat: Doing phonetics by computer* (Version 6.2.07) [Computer Software]. http://www.praat.org/

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*, 109-128. https://doi.org/10.1016/j.wocn.2011.08.002

Bohn, O. -S. (1998). Wahrnehmung fremdsprachlicher laute: Wo ist das problem? In H. Wegener (Ed.), *Eine zweite sprache lernen. empirische untersuchungen zum zweitspracherwerb* (pp. 1-20.). Narr.

Borrell, A. (1990). Perception et (re)production dans l'apprentissage des langues étrangères. quelques réflexions sur les aspects phonético-phonologiques. *Revue De Phonétique Appliquée, 95*, 107-114.

Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., & Tohkura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *Journal of the Acoustical Society of America*, *101*, 2299-2310. doi: 10.1121/1.418276

Brière, E. J. (1966). An investigation of phonological interference. *Language, 42*(4), 768-796. https://doi.org/10.2307/411832

Broughton, D. G., Broughton, G., Brumfit, C., Pincas, A., & Wilde, R. D. (1978). *Teaching English as a foreign language.* (2nd Ed.). Routledge.

Bundgaard-Nielsen, R., Best, C., & Tyler, M. (2011). Vocabulary size is associated with second-language vowel perception performance in adult learners. *Studies in Second Language Acquisition, 33*(3), 433-461. doi:10.1017/S0272263111000040

Bundgaard-Nielsen, R., Best, C., & Tyler, M. (2011). Vocabulary size matters: The assimilation of second-language Australian English vowels to first-language Japanese vowel categories. *Applied Psycholinguistics*, *32*(1), 51-67. doi:10.1017/S0142716410000287

Cardoso, W. (2011). The development of coda position in second language phonology: A variationist perspective. *Second Language Research, 27*(4), 433-465. https://doi.org/10.1177/0267658311413540

Chao, K. -Y, & Chen, L. -M. (2008). A cross-linguistic study of voice onset time in stop consonant productions. *Computational Linguistics and Chinese Language Processing*, *13*(2), 215-232.

Chao, K. -Y., Khattab, G., & Chen, L. -M. (2006). Comparison of VOT patterns in Mandarin Chinese and in English. *Proceedings of the 4th Annual Hawaii International Conference on Arts and Humanities,* 840-859.

Chen, L. -M., Chao, K. -Y., & Peng, J. -F. (2007). VOT productions of word-initial stops in Mandarin and English: A cross-language study. *Proceedings of the 19th Conference on Computational Linguistics and Speech Processing,* 303-317.

Cheong, K. M., Jano, Z., Salam, S., Abdul Bareed, N., Ithnin, F., & Suraidi, N. (2019). Motivational factors among undergraduates learning Mandarin: A qualitative perspective. *International Journal of Academic Research in Progressive Education and Development*, *8*(2), 187–199. http://dx.doi.org/10.6007/IJARPED/v8-i2/5694

Cheun, H. H. (2006). *Problems encountered in teaching Chinese as a second language in Malaysia.* New Era University College Publication.

Chin, H. B., Feng, N. Y., & Yan, C. J. (2021). Investigating the attributes of Chinese primary school adoption among the Malay community in Malaysia. *ESTEEM Journal of Social Sciences and Humanities, 5*(2), 151-164.

Cho, T., & P. A. Keating. (2001). Articulatory and acoustic studies on domain-initial strengthening in Korean. *Journal of Phonetics*, *29*(2), 155-190. https://doi.org/10.1006/jpho.2001.0131

Chua, H. W., & Azlan, M. A. K. (2019). Factors influencing foreign language learners' motivation in continuing to learn Mandarin. *EDUCATUM Journal of Social Sciences*, *5*(1), 1-6. https://doi.org/10.37134/ejoss.vol5.1.1.2019

Collins, B., & Mees, I. M. (2008). *Practical phonetics and phonology: A resource book for students*. (2nd ed.). Routledge.

Collins, B., & Vandenbergen, A. -M. (2000). *Modern English pronunciation: A pratical guide for speakers of Dutch.* (2nd ed.). Academia Press.

Derakhshan, A., & Karimi, E. (2015). The interference of first language and second language acquisition. *Theory and Practice in Language Studies, 5*(10), 2112-2117. http://dx.doi.org/10.17507/tpls.0510.19

DiCanio, C. (2020). *Insert VOT components for stops in Praat*. https://www.acsu.buffalo.edu/~cdicanio/scripts/Insert_components.praat

Docherty, G. J. (1992). *The timing of voicing in British English obstruents*. Foris Publications.

Eckman, F. R. (1977). Markedness and the contrastive analysis hypothesis. *Language Learning, 27*(2), 315-330. https://doi.org/10.1111/j.1467-1770.1977.tb00124.x

Eckman, F., & Iverson, G. (2013). The role of native language phonology in the production of L2 contrasts. *Studies in Second Language Acquisition*, *35*(1), 67-92. http://dx.doi.org/10.1017/S027226311200068X

Eckman, F. (2004). From phonemic differences to constraint rankings: Research on second language phonology. *Studies in Second Language Acquisition*, *26*(4), 513-549. http://dx.doi.org/10.1017/S027226310404001X

Ellis, R. (1994). *The study of second language acquisition.* Oxford University Press.

Ellis, R. (1997). *Second language acquisition.* Oxford University Press.

Eme, C. & Odinye, S. (2008). Phonology of standard Chinese and Igbo: Implications for Igbo students learning Chinese. *Nkuzi Omumu Asusu Journal, 1*(1), 26-37.

Evans, B. G., & Alshangiti, W. (2018). The perception and production of British English vowels and consonants by Arabic learners of English. *Journal of Phonetics*, *68*, 15-31. https://doi.org/10.1016/j.wocn.2018.01.002

Feng, L. (1985). Duration of initials, finals and tones in Beijing dialect. In T. Lin & L. J. Wang (Eds.), *Working Papers in Experimental Phonetics* (pp.131-195). Peking University Press.

Flege J. E. (1984). The detection of French accent by American listeners. *The Journal of the Acoustical Society of America*, *76*(3), 692–707. doi: 10.1121/1.391256

Flege, J. E. (1987). The production of "new" and "similar" phones in a foreign language: evidence for the effect of equivalence classification. *Journal of Phonetics*, *15*(1), 47-65. https://doi.org/10.1016/S0095-4470(19)30537-6

Flege, J. E. (1993). Production and perception of a novel, second-language phonetic contrast. *Journal of the Acoustical Society of America, 93*(3), 1589–1608. http://dx.doi.org/10.1121/1.406818

Flege, J. E. (1995). Second language speech learning: Theory, findings and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 229-273). York Press.

Flege, J. E. (1999). Age of learning and second language speech. In D. Birdsong (Ed.), *Second language acquisition and the critical period hypothesis* (pp. 101–132). Lawrence Erlbaum Associates Publishers.

Flege, J. E. (2003). Assessing constraints on second-language segmental production and perception, In N. Schiller & A. Meyer (Eds.), *Phonetics and phonology in language comprehension and production: Differences and Similarities* (pp. 319-355). De Gruyter Mouton. https://doi.org/10.1515/9783110895094.319

Flege, J. E. (2007). Language contact in bilingualism: Phonetic system interactions. In J. Cole & J. I. Hualde (Eds.), *Laboratory phonology 9* (pp. 353–380). De Gruyter Mouton.

Flege, J. E. (2008). Give input a chance! In T. Piske & M. Young-Scholten (Eds.), *Input matters in SLA* (pp.175-190). Multilingual Matters. https://doi.org/10.21832/9781847691118-012

Flege, J. E. (2018). It's input that matters most, not age. *Bilingualism: Language and Cognition*, *21*(5), 919-920. doi:10.1017/S136672891800010X

Flege, J. E., & Port, R. (1981). Cross-language phonetic interference: Arabic to English. *Language and Speech*, *24*(2), 125–146.

Flege, J. E., & Port, R. (1981). Cross-language phonetic interference: Arabic To English. *Language and Speech, 24*(2), 125–146. http://jimflege.com/files/Flege_Port_phonetic_interference_L_S_1981.pdf

Flege, J. E., Bohn, O. -S., & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics, 25*(4), 437–470. https://doi.org/10.1006/jpho.1997.0052

Flege, J. E., MacKay, I. R. A., & Meador, D. (1999). Native Italian speakers' perception and production of English vowels. *Journal of the Acoustical Society of America*, *106*(5), 2973–2987. http://dx.doi.org/10.1121/1.428116

Flege, J. E., Schirru, C., & MacKay, I. R. A. (2003). Interaction between the native and second language phonetic subsystems. *Speech Communication*, *40*(4), 467–491. https://doi.org/10.1016/S0167-6393(02)00128-0

Flege, J., & Lui, S. (2001). The effect of experience on adults' acquisition of a second language. *Studies in Second Language Acquisition, 23*(4), 527–552. doi:10.1017/S0272263101004041

Fu, A. H. (2006). *Yin du ni xi ya Hanyu yu yan xue xi zhe han yu fa yin kun nan fen xi* [Analysis of the difficulties of Indonesian speakers in learning Mandarin pronounication]. Hebei Normal University. https://kns.cnki.net/KCMS/detail/detail.aspx?dbname=CMFD0506&filename=2006099585.nh.

Gass, S. (1996). Second language acquisition and linguistic theory: The role of language transfer. In S. W. Ritchie & T. Bathia (Eds.), *Handbook of second language acquisition* (pp. 317-345), Academic Press.

Ghee, T. T., Kee, O. A., & Ismail, H. N. (2012). The orientations for learning Mandarin amongst Malay undergraduate students. *International Journal of Humanities and Social Science*, *2*(12), 104-112.

Goh, S. B. (2007). Analisis kontrastif system fonetik Bahasa Mandarin dan Bahasa Melayu (Unpublished Master's thesis). Universiti Malaya.

Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds "L" and "R". *Neuropsychologia*, *9*(3), 317–323. https://doi.org/10.1016/0028-3932(71)90027-3

Guion, S. G., Flege, J. E., Akahane-Yamada, R., & Pruitt, J. C. (2000). An investigation of current models of second language speech perception: The case of Japanese adults' perception of English consonants. *The Journal of the Acoustical Society of America*, *107*(5), 2711-2724. https://doi.org/10.1121/1.428657

Guo, Q. S. (2020). *Ling chu ji shui ping Hanyu xue xi zhe yu yin pian wu fen xi yu jiao xue dui ce* [The error analysis and teaching strategies in Chinese phonetic for primary stage Chinese learner——Taking Confucius Institute Universiti Malaysia Pahang as examples]. Hebei Normal University. https://overseacnkinet.eu1.proxy.openathens.net/KCMS/detail/detail.aspx?dbcode=CMFD&dbname=CMFD202101&filename=1021510282.nh&uniplatform=OVERSEA&v=Kjh12Q25UT8qvgiRpetMNEM9kofVLAN3VAMdpoe6IjyX3a_irYblNG50-k60akBk

Hao, Y. -C. (2012). *The effect of L2 experience on second language acquisition of Mandarin consonants, vowels, and tones* [Unpublished doctoral dissertation]. Indiana University.

Henton, C., Ladefoged, P., & Maddieson, I. (1992). Stops in the world's languages. *Phonetica, 49(2)*, 65-101. doi: 10.1159/000261905

Hiang, L. L. (2015). Why is Mandarin needed by MFL learners in Malaysia. *Global Journal for Research Analysis*, *4*(12), 269-271. https://www.worldwidejournals.com/global-journal-for-research-analysis-GJRA/fileview/December_2015_1451713109__96.pdf

Hoe, F. T., & Lim, T. H. (2013, May). Chinese/Mandarin as a second language course: A comparison Chinese/Mandarin as a second language course: A comparison between a Malaysian and Chinese University (Universiti Teknologi Mara and Beijing Language and Culture University as an example). *International Conference on Language 2013 (UPAL ICL 2)*.

Hoe, F. T., & Mah, B. Y. (2009). GAT: A teaching module of restructuring basic Mandarin sentences to UITM Malay students. *International Conference on Languages*.

Hoe, F. T., & Mah, B. Y. (2011). Group-arrange-touch-up (GAT): A method of teaching non-native Mandarin speakers in restructuring elementary Mandarin sentence. *Esteem Academic Journal*, *7*(2), 111-124.

Hong, C. C. (2021). Motivation in learning Mandarin as a foreign language among secondary student in international school. *International Journal of Sciences: Basic and Applied Research (IJSBAR)*, *55*(1), 325-337.

Huang, K. (2008). *A study of Malaysian universities Mandarin classes* [Unpublished course work]. Universiti Kebangsaan Malaysia.

Jing, Y. J. (2014). *Chu ji jie duan Ma Lai Xi Ya liu xue sheng Hanyu yu yin pian wu fen xi* [Analysis of Chinese phonetic errors of Malaysian students in primary stage]. Beijing Foreign Studies University. https://kns-cnki-net.eu1.proxy.openathens.net/KCMS/detail/detail.aspx?dbname=CMFD201402&filename=1014255048.nh

Keating, P. A. (1984). Phonetic and phonological representation of stop consonant voicing. *Language*, *60*(2), 286-319. https://doi.org/10.2307/413642

Kehoe, M. M., Lleó, C. & Rakow, M. (2004). Voice onset time in bilingual German-Spanish children. *Bilingualism: Language and Cognition, 7*(1), 71-88. doi:10.1017/S1366728904001282

Kent, R. D., & Charles, R. (2002). *Acoustic analysis of speech.* (2nd ed.). Thomson Learning.

Khor, G. S., Arriaga, L. R., & Mah, B. Y. (2013). Errors analysis in Hanyu Pinyin pronunciation among the undergraduates from Universiti Sains Malaysia (USM), Engineering Campus. *The Asian Conference on Language Learning 2013,* 293-301.

Khor, G. S., Mah, B. Y., & Chow, V. F. T. (2017). Pronunciation errors in Mandarin among the engineering undergraduates: An analytical analysis. *International Academic Research Journal of Social Science, 3*(1), 71-76.

Klatt, D. H. (1975). Voice-onset time, frication, and aspiration in word-initial consonant clusters. *Journal of Speech and Hearing Research*, *18*(4), 686-706. doi: 10.1044/jshr.1804.686

Kluge, D. C., Rauber, A. S., Reis, M. S., Bion, R. A. H. (2007). The relationship between the perception and production of English nasal codas by Brazilian learners of English. *8th Annual Conference of the International Speech Communication Association,*, 2297–2300. doi:10.21437/Interspeech.2007-623

Lado, R. (1957). *Linguistics across cultures: Applied linguistics for language teachers*. University of Michigan Press.

Lai, Y. H. (2013). VOT for Mandarin stops and affricates produced by L1 and L2 speakers of Mandarin Chinese. *Kaohsiung Normal University Journal, 35*, 57-78.

Lai, Y. -S. (2009). Asymmetry in Mandarin affricate perception by learners of Mandarin Chinese. *Language and Cognitive Processes, 24*(7-8), 1265–1285. https://doi.org/10.1080/01690960802113850

Lane, H. (1962). Some differences between first and second language learning. *Language Learning, 12*(1), 1-14. https://doi.org/10.1111/j.1467-1770.1962.tb00759.x

Lau, S. K., Ng, L. L., & Lee, H. L. (2011). *Ma lai xue sheng wu zhong han yu ju xing yu xu pian wu fen xi ji yu xu jiao xue fa*. *Conference on Chinese Language Learning and Teaching in the Digital Age*.

Lee, H. -A., & Khalid, M. A. (2016). Discrimination of high degrees: Race and graduate hiring in Malaysia. *Journal of the Asia Pacific Economy*, *21*(1), 53–76. https://doi.org/10.1080/13547860.2015.1055948

Leigh, L., & Arthur, S. A. (1964). A cross-language study of voicing in initial stops: Acoustical measurement. *Word, 20*(3), 384-422. https://doi.org/10.1080/00437956.1964.11659830

Lenneberg, E. H. (1967). *Biological foundations of language.* John Wiley and Sons.

Li, F. (2013). The effect of speakers' sex on voice onset time in Mandarin stops. *The Journal of the Acoustical Society of America*, *133*(2), 142-147. https://doi.org/10.1121%2F1.4778281

Li, L. (2016). *Zhong guo wai yu xue xi zhe de ying han fu yin gan zhi he chan chu shi yan yan jiu* [Perception and production of English and Mandarin consonants by Chinese EFL learners: An empirial study]. Jiangsu University of Science and Technology. https://kns-cnki-net.eu1.proxy.openathens.net/KCMS/detail/detail.aspx?dbname=CMFD201701&filename=1016736478.nh

Lin, H. (2001). *A grammar of Mandarin Chinese*. Lincom Europa.

Linebaugh, G., & Roche, T. B. (2015). Evidence that L2 production training can enhance perception. *Journal of Academic Language and Learning, 9*(1), A1-A17. https://journal.aall.org.au/index.php/jall/article/view/326

Lisker, L., & Abramson, A. (1967). Some effects of context on voice onset time in English stops. *Language and Speech*, *10,* 1-28. https://doi.org/10.1177/002383096701000101

Lisker, L., & Abramson, A. S. (1964). Stop categorization and voice onset time. *Proceedings of the Fifth International Congress of Phonetic Sciences,* 389-391.

Lisker, L., & Abramson, A. S. (1970). The voicing dimension: Some experiments in comparative phonetics. *Proceedings of the Sixth International Congress of Phonetic Sciences,* 563-567.

Lisker, L., & Abramson, A. S. (1971). Distinctive features and laryngeal control. *Language*, *47*(4), 767–785. https://doi.org/10.2307/412155

Littlewood, W. (1984). *Foreign and second language learning: Language acquisition research and its implications for the classroom.* Cambridge University Press.

Liu, J., & Allard, J. (2013). American Chinese learners' acquisition of L2 Chinese affricates /ts/ and /tsh/. *Proceedings of Meetings on Acoustics Acoustical Society of America*, *18*(1), 060005-060005. http://dx.doi.org/10.1121/1.4798223

Liu, Y. (2005). A case study on using English language movies in teaching English as foreign language experiences. In P. Kommers & G. Richards (Eds.), *Proceedings of ED-MEDIA 2005--World Conference on Educational Multimedia, Hypermedia and Telecommunications,* (pp. 52-58). Association for the Advancement of Computing in Education. https://www.learntechlib.org/primary/p/20054/

Llisterri, J., Mann, R., de la Mota, C., & Rios, A. (1995). Factors affecting F0 peak displacement in Spanish. *Proceedings of the 4th European Conference on Speech Communication and Technology (Eurospeech 1995),* 2061-2064. doi: 10.21437/Eurospeech.1995-493

Loo, W. K., Ow, S. H., & Zainuddin, R. (2007). Building a unit selection speech synthesizer for Malay Language using FESTVOX and hidden markov model toolkit (HTK). *Chiang Mai University Journal of Natural Sciences, 6*(1), 149-158. https://www.thaiscience.info/Journals/Article/CMUJ/10325099.pdf

Mackay, I. R. A., & Flege, J. E. (2004). Effects of the age of second language learning on the duration of first and second language sentences: The role of suppression. *Applied Psycholinguistics*, *25*(3), 373–396. doi:10.1017/S0142716404001171

MacKay, I. R. A., Meador, D., & Flege, J. E. (2001). The identification of English consonants by native speakers of Italian. *Phonetica, 58*(1-2), 103–125. http://dx.doi.org/10.1159/000028490

Macleod, A. A. N., & Stoel-Gammon, C. (2005). Are bilinguals different? What VOT tells us about simultaneous bilinguals. *Journal of Multilingual Communication Disorders, 3*(2), 118-127. https://doi.org/10.1080/14769670500066313

Marckwardt, J. C. G., & Albert, H. M. (1965). *The college teaching of English.* Meredith Publishing Company.

McCarthy, N. (2020). *The world's most spoken languages*. Statista. https://www.statista.com/chart/12868/the-worlds-most-spoken-languages/

Ministry of Education Malaysia. (2013). *Malaysia education blueprint 2013-2025*. Ministry of Education Malaysia.

Ministry of Education (2015). *Malaysia education blueprint 2015-2025.* Ministry of Higher Education Malaysia. https://www.kooperationinternational.de/uploads/media/3.MalaysiaEducationBlueprint2015-2025Higher_Education.pdf

Moeller, A. J., & Catalano, T. (2015). Foreign language teaching and learning. In J.D. Wright (Ed.), *International encyclopedia for social and behavioral sciences* (2nd ed.) (pp. 327-322). Pergamon Press.

Motulsky, H. J. (2021). *Advice: When to plot SD Vs. SEM.* Graphpad Statistics Guide. http://www.graphpad.com/guides/prism/7/statistics/index.htm?statwhentoplotsdvssem.html.

Moyer, A. (2008). Input as a critical means to an end: Quantity and quality of experience in L2 phonological attainment. In T. Piske & M. Young-Scholten (Eds.), *Input matters in SLA* (pp. 159-174). Multilingual Matters.

Munro, M. J. (2008). Foreign accent and speech intelligibility. In G. H. G. Jette & Z. L. Mary (Eds.), *Phonology in second language acquisition* (pp.193–218). John Benjamins.

Munro, M. J., Derwing, T. M., & Thomson, R. I. (2015). Setting segmental priorities for English learners: Evidence from a longitudinal study. *International Review of Applied Linguistics in Language Teaching*, *53*(1), 39-60. https://doi.org/10.1515/iral-2015-0002

Neufeld, G. G. (1988). Phonological asymmetry in second-language learning and performance. *Language Learning*, *38*(4), 531-559. https://doi.org/10.1111/j.1467-1770.1988.tb00166.x

New Straits Times (2020, November 11). More Malaysian Malays studying in Chinese primary schools today versus a decade ago. *New Strait Times*. https://www.straitstimes.com/asia/seasia/more-malaysian-malays-study-in-chinese-primary-schools-today-versus-adecade-ago

Peng, D. Y. (2021). *Chu ji shui ping mian bei ke qin xue sheng Hanyu yu yin pian wu yan jiu* [Phonetic errors research of northern Mymmar Kachin students with primary Chinese level.] Yunnan Normal University. https://kns-cnki-net.eu1.proxy.openathens.net/KCMS/detail/detail.aspx?dbname=CMFD202102&filename=1021621209.nh

Pilus, Z. (2005). Perception of voicing in English word-final consonants: Examining the perceptual assimilation model. *Malaysian Journal of ELT Research, Inaugural*, 1-12. http://www.melta.org.my/Doc/The_Perceptual_Assimilation_Model1.pdf

Piske, T., Flege, J. E., MacKay, I. R. A., & Meador, D., (2002). The production of English vowels by fluent early and late Italian–English bilinguals. *Phonetica, 59*(1), 49–71. http://dx.doi.org/10.1159/000056205

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

Polivanov, E. D. (1931). La perception des sons d'une langue Étrangère. *Traveaux Du Cercle Linguistique De Praque*, *4*, 79-96.

Ran, Q. B., & Shi, F. (2007). The analysis of VOT values of monosyllable in Mandarin. *Nankai Linguistics*, *2*, 21-31.

Reudenstein, R. (1979). *Teaching foreign languages to the very young: Papers from seven countries on work with 4-8 year-olds.* Pergamon Press.

Riduan, M., Ain, A. L., Nur, E. N. B., & Mohd, F. A. (2021). Perbandingan dialek Melayu Kelantan dan dialek Melayu Terengganu. *Jurnal Melayu*, *20*(2), 329-345. http://journalarticle.ukm.my/18378/1/51822-169875-1-SM.pdf

Riney, T. J., Takagi, N., Ota, K., & Uchida, Y. (2007). The intermediate degree of VOT in Japanese initial voiceless stops. *Journal of Phonetics, 35*(3), 439-443. https://doi.org/10.1016/j.wocn.2006.01.002

Rochet, B. L., & Fei, Y. (1991). Effect of consonant and vowel context on Mandarin Chinese VOT: Production and perception. *Canadian Acoustics*, *19*(4), 105–106. https://jcaa.caa-aca.ca/index.php/jcaa/article/view/691

Sakai, M., & Moorman, C. (2018). Can perception training improve the production of second language phonemes? A meta-analytic review of 25 years of perception training research. *Applied Psycholinguistics, 39*(1), 187-224. http://dx.doi.org/10.1017/S0142716417000418

Saville-Troike, M. (2006). *Introducing second language acquisition*. Cambridge University Press.

Scovel, T. (1969). Foreign accents, language acquisition, and cerebral dominance. *Language Learning*, *19*(3–4), 245–253. https://doi.org/10.1111/j.1467-1770.1969.tb00466.x

Shahidi, A. H., Rahim, A., & Ab.Samad, K. (2012). Production and perception of English word final stops by Malay speakers. *GEMA Online Journal of Language Studies*, *12*(4), 1109-1125. http://journalarticle.ukm.my/5775/1/1572-2959-1-SM.pdf

Sheldon, A., & Strange, W. (1982). The acquisition of /r/ and /l/ by Japanese learners of English: Evidence that speech production can precede speech perception. *Applied Psycholinguistics*, *3*(3), 243-261. doi:10.1017/S0142716400001417

Shu, H. H. (2019). *Yuenan liu xue sheng han yu Pu Tong Hua se yin he se ca yin de chan chu shi yan yan jiu* [An experimental study on the production of plosives and affricates by Vietnamese Learners of Chinese Mandarin]. Jinan University. https://kns-cnki-net.eu1.proxy.openathens.net/KCMS/detail/detail.aspx?dbname=CMFD202101&filename=1020705406.nh

Song, W. (2018). *Study on the perception of Mandarin affricated consonant contrast in Malay learners* [Unpublished doctoral dissertation]. Beijing Foreign Studies University.

Stern, H. H. (1983). *Fundamental concepts of language teaching*. Oxford University Press.

Stern, H. H. (1992). *Issues and options in language teaching*. Oxford University Press.

Stevens, K. N., Keyser, S. J., & Kawasaki, H. (1986). Toward a phonetic and phonological theory of redundant features. In J. Perkell & D. Klatt (Eds.), *Invariance and variability in speech processes* (pp. 426-463). Erlbaum.

Strange, W. (1995). *Speech perception and linguistic experience: Issues in cross-language research.* York Press.

Swan, M. (2008). *The influence of the mother tongue on second language vocabulary acquisition and use*. Mikeswan.co.uk. http://www.mikeswan.co.uk/elt-applied-linguistics/influencesecond-language.htm

Tanaka, M., Chen, L. M., & Hsu, C. J. (2019). Correction to: Acquisition of stop consonants in Mandarin-Japanese early bilingual children. *Journal of Psycholinguistic Research*, *48*(2), 331. https://doi.org/10.1007/s10936-018-9608-1

Tarone, E. (2005). Impact of literacy on oral language processing: Implications for second language acquisition research. *Annual Review of Applied Linguistics*, *25*, 77-97. http://dx.doi.org/10.1017/S0267190505000048

The United Chinese School Teachers' Association of Malaysia. (2015). *The evolution and current status of the number of primary school students and schools in Malaysia from 2000 to 2014.* The United Chinese School Teachers' Association of Malaysia. http://web.jiaozong.org.my/doc/2015/jan/2000-2014_sjkc.pdf

Thirusanku, J., & Md. Yunus, M. (2012). The many faces of Malaysian English. *International Scholarly Research Notices, 2012*, 1-14. https://doi.org/10.5402/2012/138928

Towell, R., & Hawkins, R. (1994). *Approaches to second language acquisition*. Multilingual Matters.

Trofimovich, P., & Baker, W. (2006). Learning second-language suprasegmentals: Effect of L2 experience on prosody and fluency characteristics of L2 speech. *Studies in Second Language Acquisition*, *28*(1), 1–30. doi:10.1017/S0272263106060013

Tse, L. (2000). Student perceptions of foreign language study: A qualitative analysis of foreign language autobiographies. *The Modern Language Journal*, *84*(1), 69-84. http://www.jstor.org/stable/330450

Tsukada, K., Birdsong, D., Bialystok, E., Mack, M., Sung, H., & Flege, J. (2005). A developmental study of English vowel production and perception by native Korean adults and children. *Journal of Phonetics, 33*(3), 263–290. https://doi.org/10.1016/j.wocn.2004.10.002

Uffmann, C. (2006). Epenthetic vowel quality in loanwords: Empirical and formal issues. *Lingua, 116*, 1079-1111. https://doi.org/10.1016/j.lingua.2005.06.009

Wang, X. (1997). *The acquisition of English vowels by Mandarin ESL learners: A study of production and perception* [Unpublished Master's thesis]. Simon Fraser University.

Wang, X. (2006). Perception of L2 tones: L1 Lexical tone experience may not help. *Third International Conference on Speech*.

Wang, X. (2008). Training for learning Mandarin tones. In Z. Felicia & B. Barber Hershey (Eds.), *Handbook of research on computer-enhanced language acquisition and learning* (pp. 259-274). IGI Global.

Wang, X. (2013). Perception of Mandarin tones: The effect of L1 background and training. *The Modern Language Journal, 97*(1), 144–160. http://dx.doi.org/10.2307/23361743

Wang, X., & Chen, J. (2019). English speakers' perception of Mandarin consonants: The effect of phonetic distances and L2 experience. In C. Sasha, E. Paola, B. Marija, & W. C. Paul, (Eds.), *Australasian speech science and technology: Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia 2019* (pp. 250–254). Association Inc.

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

Wardhaugh, R. (1970). The contrastive analysis hypothesis. *TESOL Quarterly*, *4*(2), 123-130. https://doi.org/10.2307/3586182

Watson, J. K. P. (1980). Cultural pluralism nation-building and educational policies in Peninsular Malaysia. *Journal of Multilingual and Multicultural Development*, *1*(2), 155-174. https://doi.org/10.1080/01434632.1980.9994008

Weinreich, U. (1953). *Languages in contact*. Mouton.

Wen, Q. K. (2017). *Malai xiya hua wen jiao yu shi zhi fa zhan wenti------jiyu<haiwaihuawenjiaoyudongtai> de yan jiu fa xian hai wai huawen jiaoyu* [The difficulties of teacher development in Malaysia Chinese education-based on overseas Chinese education news]. Xiamen University. https://oversea-cnki-net.eu1.proxy.openathens.net/kns/manage/export?filename=hwhw201703016&dbname=CJFDLAST2017

Whalen, D. H., Levitt, A. G., & Goldstein, L. M. (2007). VOT in the babbling of French and English-learning infants. *Journal of Phonetics, 35*(3), 341-352. https://doi.org/10.1016%2Fj.wocn.2006.10.001

Woan, L. H. (2010). Phonological acquisition in three languages: A cross-sectional study in English, Mandarin and Malay [Unpublished doctoral dissertation]. University of Sheffield.

Wu, Z. (1992). *Essentials of modern Chinese phonetics*. Foreign Language Printing House.

Yamada, R. A., Strange, W., Magnuson, J. S., Pruitt, J. S., & Clarke, W. D. (1994). The intelligibility of Japanese speakers' production of American English /r/, /i/, and /w/, as evaluated by native speakers of American English. *Proceedings of the 3rd International Conference of Spoken Language Processing,* 2023–2026.

Yang, C., & Yu, A. C. L. (2019). The acquisition of Mandarin affricates by American L2 learners. *Taiwan Journal of Linguistics, 17*(2), 91–122. doi: 10.6519/TJL.201907_17(2).0004

Yang, J. (2018). Development of stop consonants in three- to six-year-old Mandarin-speaking children. *Journal of Child Language*, *45*(5), 1091–1115. doi:10.1017/S0305000918000090

Yeni-Komshian, G. H. (1987). Stop consonant voicing in Malay-English bilinguals. *The Journal of the Acoustical Society of America, 82*(S1), S115. http://dx.doi.org/10.1121/1.2024599

Zhang, F., & Yin, P. (2009). A study of pronunciation problems of English learners in China. *Asian Social Science*, *5*(6), 141-146. https://doi.org/10.5539/ass.v5n6p141
.

Zhao, M. J. (2007). Hanyu zuo wei wai yu jiao xue neng li biao zhun shi shuo [A commentary on capacity standards for teaching Chinese as a foreign language]. *Language Teaching and Research*, 2, 1-10.

Zhao, H., & Huang, J. (2010). China's policy of Chinese as a foreign language and the use of overseas Confucius Institutes. *Educational Research for Policy and Practice*, *9*(2), 127–142. https://doi.org/10.1007/s10671-009-9078-1

Zheng, F., Song, Z., Fung, P., & William, B. (2002). Mandarin pronunciation modeling based on CASS corpus. *Journal of Computer Science and Technology*, *17*(3), 249-263. http://dx.doi.org/10.1007/BF02947304

Zhou, X. (2010). *Duiwai Hanyu jiaoxue daolun (Di'er Ban) [Introduction to Chine se as a teaching foreign language]* (2nd ed.n). Shangwu Yinshuguan.