RICE UNIVERSITY

# The Effect of Chinese Characters on the Speech Perception and Production of Retroflex Sibilants in Taiwan Mandarin

by

**Ru-Ping Ruby Tso**

A THESIS SUBMITTED
IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE

**Doctor of Philosophy**

APPROVED, THESIS COMMITTEE

Nancy Niedzielski, Associate Professor,
Chair, Linguistics

Michel Achard, Professor
Linguistics

Nanxiu Qian, Professor
Chinese Literature, Chao Center for Asian Studies

HOUSTON, TEXAS
April 2017

# ABSTRACT

The Effect of Chinese Characters on the Speech Perception and Production of Retroflex Sibilants in Taiwan Mandarin

by

Ru-Ping Ruby Tso

Evidence has shown that subtle implicit information of a speaker's characteristics or social identity inferred by the listener can influence how language varieties are perceived, and can cause significant effects on the result of speech perception (e.g., Williams 1976; Beebe 1981; Thakerar and Giles 1981; Niedzielski 1999; Hay et al. 2006a; Hay and Drager 2010; Koops 2011).

This dissertation aimed at studying the effects of Chinese orthography on the speech production and perception of retroflex sibilants in Mandarin Chinese. The two variants of written Chinese characters, traditional and simplified, served as subtle implicit information to index speaker's identity of a Taiwan Mandarin speaker or Beijing Mandarin speaker respectively.

The experiment designs were based on the hypotheses that Taiwan Mandarin speakers are aware of the differences between the Taiwan Mandarin dialect and the Beijing Mandarin dialect at both segmental and suprasegmental level. Furthermore, Taiwan Mandarin speakers can activate dialectal features of Beijing Mandarin with the presence of simplified Chinese characters.

In the word-identification tasks of the perception study, a statistically significant relationship between the identification of retroflex phonemes and the variety of written Chinese characters was found for all participants with a Person's chi-square test of association. With a 95% confidence interval, the odds ratio estimated that with the presence of simplified Chinese characters, participants were at least 1.83 times more likely to identify a retroflex audio stimulus with the actual retroflex phoneme instead of its corresponding alveolar sound than with the presence of traditional Chinese characters.

The effect of character variation on speech production was not as straightforward as that in perception. From the data collected in this study, minimal effect was found; however, when taking the speaker's attitude towards different varieties of characters into consideration, personal preferences toward the varieties of characters may lead to a stylistic and intentional variation in speech production of retroflex sibilants.

It was found through the interview with participants of this study that Taiwan Mandarin speakers were fully aware of the variation in the production of retroflex sibilants. They were also aware of the association between simplified characters and the Beijing Mandarin dialect and this association was activated during the speech perception and production experiments of this dissertation.

This study adds to the finding of research in sociophonetic variations that an asymmetry in speech production and speech perception may be a deliberate choice of the speaker instead of a result of unconscious perception and production of speech. In addition, this dissertation also shows that the abundant cultural and ideological values associated with the usage of Chinese written characters and spoken dialects are potential topics of future research.

Keywords: sociophonetic variation, speech perception, speech production, retroflex sibilants, Mandarin Chinese, Taiwan Mandarin

# ACKNOWLEDGMENTS

I owe a great debt of gratitude to many people in the course of completing my Ph.D. dissertation. First and foremost, I would like to express my deepest gratitude to my beloved parents Tai-Yih Tso and Yu-Chang Yang who both gave me unconditional love and unwavering support during the most difficult times when obstacles were confronted. I am in debt to my sister Karen Tso who not only provided me with ideas for experiment design, but most important of all, looked after by baby Aldrich during the final and most frustrating stage of completing my dissertation. I am also thankful to my husband Jamie Chun-Kai Wang who fully understood and supported my passion for research in Linguistics.

My advisor Dr. Nancy Niedzielski is a role model of my career as a researcher. Words alone cannot fully express how lucky I am to have her as my dissertation advisor. She is knowledgeable, talented, and well-recognized in the field of sociolinguistics. She introduced me to this field and led me to discover my passion for studying the interactions between social factors and language variation. During our many discussions, she was always patient with my progress and had confidence in my ability of completing the dissertation. All these characteristics along with her humorous way of comforting me when difficulties were met made her the best advisor anyone could ever wish for when pursuing a doctoral degree.

My two other committee members Dr. Michel Achard and Dr. Nanxiu Qian both inspired me greatly from the beginning of the candidacy. I am thankful to Dr. Achard who always read my manuscripts in detail and provided me with suggestions that helped me formulate the foundations of this dissertation. Dr. Qian, a scholar of Classical Chinese Literature helped me discover the core values of my dissertation and the richness of future research it may lead to.

I am thankful to all current and former faculty members in the Linguistics Department of Rice University, especially Dr. Suzanne Kemmer who taught me how to sort, organize, and make meaningful inferences from empirical data, and Dr. Kevin McGowan who was on my committee of the qualifying paper that led to this dissertation

and provided me with useful technical aids to create audio stimuli. I would also like to thank Ms. Rita Riley who has always been helpful in complying with administrative issues, and kindly replied to my oftentimes last-minute requests.

Finally, I would like to express my gratitude to all my friends who have mentally supported me through the times of frustration before completing my dissertation. I would like to mention in particular Yu-Ying Sarah Chuang, my best friend since college and a talented fellow linguist, who not only spent a great deal of time discussing my dissertation topic with me but also assisted me in the technical aspects of my experiment designs. I would not have been to make all the current achievements without the love and support from all the people mentioned above, and I am extremely grateful to have had them by my side during the completion of this dissertation.

# TABLE OF CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

# Chapter 1

## 1. Introduction

### 1.1 Background and purpose

Variations in speech perception and production are natural developments in a language that give it life and sustainability. While speakers are generally conscious of certain ongoing changes in language, they often do not realize that they themselves are the forces of variations and that there are much more powerful dynamics in the courses of linguistic variations that are beyond their awareness. One of the goals of a sociolinguist is to trace these dynamics, find the cause, and make predictions about future changes. Another goal is to investigate how speakers who inevitably take part in the changes evaluate the variations and make personal or stylistic judgments of them.

The aim of this dissertation is the latter of the previously described goals. As a speaker of the targeted language for investigation as well as a linguist curious in understanding how fellow speakers perceive the variations that I have detected, I would like to devote this dissertation to examining the current situations of one apparent linguistic variation ongoing in Taiwan Mandarin, the merging of the retroflex consonants with their alveolar counterparts. This is not a novel topic of study since there are many available observations of this phenomenon, by which conclusions have been drawn that females generally are more likely to retroflex than males (Lin 1983; Rau and Li 1994; Tse 1998), all speakers use significantly less retroflection in casual speech styles than in formal ones (Lin 1983; Rau and Li 1994), and speakers whose home language is Mandarin are more likely to retroflex than those whose home language is Min (Lin 1983).

Since the basics have already been explored, this dissertation focuses on discovering the inexplicit concepts related to this merging phenomenon of these two sets of consonants. Evidence has shown that subtle implicit information of a speaker's characteristics or social identity inferred by the listener can influence how language

varieties are perceived, and can cause significant effects on the result of speech perception (e.g., Williams 1976; Beebe 1981; Thakerar and Giles 1981; Niedzielski 1999; Hay et al. 2006a; Hay and Drager 2010; Koops 2011). One of the pioneering studies on this subject is Niedzielski's (1997, 1999) Detroit versus Canadian study, where listeners would show different identification patterns for the stimuli depending on who the listeners believed the speaker was. A more complete set of literature review on related topics is provided in the next chapter.

In the same line with these perception studies, this dissertation examines the speech perception and production of retroflex sibilants in Mandarin Chinese under the effect of Chinese characters. There are two variations of Chinese written characters, traditional characters and simplified characters. Traditional Chinese characters have a long history of development and are currently continued to be used in Taiwan, Hong Kong, and Singapore, while simplified characters have only been developed from traditional characters and promoted by the People's Republic of China (PRC) along with the Beijing Mandarin dialect since 1954. The association between Beijing Mandarin and simplified characters has thus begun to establish. One of the goals of this dissertation is to investigate whether phonological features of the Beijing Mandarin dialect can be activated by the presence of simplified written Chinese characters.

## 1.2 Sociophonetic knowledge

In addition to examining how variation in Chinese written characters would affect the speech perception and production of retroflex sibilants of Taiwan Mandarin speakers, this dissertation also aims at investigating the sociophonetic knowledge of native speakers of Taiwan Mandarin. The term *sociophonetic variation*, in the field of variationist sociolinguistics, refers to the correlation between aspects of speakers' social identity and variants of a speech sound (Labov 1966, 1994, 2001). For instance, in a Taiwan Mandarin speech community where male speakers tend to merge initial retroflex consonants with initial alveolar consonants while females tend to keep them distinguished, the social identity of gender together with the alternative realizations of the same phoneme establish a case of sociophonetic variation (Lin 1983; Rau and Li 1994; Tse 1998).

The term *sociophonetic knowledge* focuses on "the knowledge that language users have as listeners of the variation that exists in the speech of other speakers in their speech community" (Koops 2011:2). This particular knowledge of how people's speech is different from one another has been shown to facilitate the phonetic perception of speech (Strand and Johnson 1996; Niedzielski 1999; Drager 2005, 2011; Hay, Warren & Drager 2006; Staum 2008; McGowan 2011 *inter alia*). In other words, the same objectively given acoustic signal may be subjectively interpreted as different when listeners project different social identifications onto the speaker. Therefore, when language users demonstrate their sociophonetic knowledge through speech perception, they are making judgments primarily based on *linguistic decisions* rather than *social decisions*, which is the basis for making judgments of how well a linguistic cue is matched to explicitly given social categories using their knowledge of sociophonetic variation (Koops 2011:18).

Observations of Taiwan Mandarin speakers' sociophonetics knowledge with regard to retroflex sibilants from previous studies include Baran's (2007: 205) conclusion that the "lack of retroflection is the first feature cited by local speakers when they are asked to describe the "biggest" differences between Standard Mandarin and Taiwan Mandarin." Lin (2007) also stated that the first major characteristic of Taiwan Mandarin is the lack of distinction between the retroflex and alveolar consonants. In addition, Chung (2006) claimed that retroflex speech is labeled as being "standard" and viewed as a prestigious form. It was also stated that retroflection is standard for media broadcasting and the presence of retroflection in one's speech is an index of a higher education level.

## 1.3 Hypotheses and overview of this dissertation

The design of this dissertation study is based on two main hypotheses. First, Taiwan Mandarin speakers are aware of the phonological differences between the Taiwan Mandarin dialect and the Beijing Mandarin dialect at both segmental and suprasegmental level. Second, Taiwan Mandarin speakers can activate dialectal features of Beijing Mandarin with the presence of simplified Chinese characters.

The research questions formulated under these two assumptions include:

(1) How is the speech perception of retroflex sibilants of Taiwan Mandarin speakers

affected by the variation in Chinese characters?

(2) How is the speech production of retroflex sibilants of Taiwan Mandarin speakers affected by the variation in Chinese characters?

(3) What is the correlation between speech perception and production of retroflex sibilants by Taiwan Mandarin speakers?

(4) What is considered as the so-called standard production of retroflex sibilants in Mandarin in the mind of Taiwan Mandarin speakers?

The remaining chapters will gradually develop answers to these proposed questions by first reviewing the previous studies of speech perception and production of linguistic variations in chapter 2, followed by a review of studies on the retroflex and alveolar sibilants in Mandarin Chinese in chapter 3. Chapter 4 provides a pilot study to this dissertation as well as explains the methodology of experiment design and general procedure of the experiments. Chapter 5 presents the results of the perception study, and the results from the production study are provided in chapter 6. Finally, chapter 7 includes an integrated discussion of the findings from this dissertation and concludes this dissertation study.

# Chapter 2

## 2. Organization of literature on speech perception

The field of sociolinguistics in the west was pioneered by Labovian variationist studies (Labov 1966, 1994, 2001), addressing issues such as speech production across dialects, speech styles, and speech communities. Later research on speech perception has brought the field of sociolinguistics into a new chapter. Traditionally, the focus of speech perception studies was on the systematic exploration of sound categorizations and identification, which is also the basis of language attitude research. With techniques such as the matched guise (Lambert, Hodgson, Gardner & Fillenbaum 1960) and verbal guise (Cooper 1975), researchers are able to elicit responses that participants often find difficult to describe consciously or voluntarily. During the tasks, listeners consciously evaluate the stimuli and simultaneously, maybe often unconsciously, convey their attitudes toward the social group represented by the speaker. Research on speech perception also provides input to the relation between speech production and speech perception. While many found direct associations between production and speech perception (Labov, Yaeger & Steiner 1972; Willis 1972; Labov & Ash 1997; Flanigan & Norris 2000; Pierrehumbert & Tamati 2008; Fridland & Okamoto 2009 inter alia), further studies pointed out apparent asymmetry between speech production and perception (Janson & Schulman 1983; Labov, Karen & Miller 1991; Thomas 2000).

Variationist sociolinguistics started to apply a new method to the collection of speech perception data in the 1970s, using synthetic speech. Early examples of applying synthetically-manipulated-speech methodology include Willis' (1972) study of vowel perception by speakers from two adjacent cities across the US-Canada border, and Graff, Labov, and Harris' (1986) classic experiment on ethnic identification in Philadelphia. Contemporarily, with the modern improvement of acoustic-analysis tools, researchers are able to more easily manipulate sounds and also examine the effect of non-linguistic factors on speech perception. With the more sophisticated tools available, there has been

a gradual increase in sociolinguistic studies adopting the method of speech synthesis as shown in later sections (e.g., Strand and Johnson 1996; Niedzielski 1999; Hay et al. 2006a; Walker 2007; Hay and Drager 2010).

Evidence has shown that subtle implicit information of speaker's characteristics or social identity inferred by the listener can influence how language varieties are perceived, and cause significant effects on the result of speech perception (e.g., Williams 1976; Beebe 1981; Thakerar and Giles 1981; Niedzielski 1999; Hay et al. 2006a; Hay and Drager 2010; Koops 2011). One of the pioneering studies on this subject is Niedzielski's (1997, 1999) Detroit versus Canadian study, where listeners would show different identification patterns for the stimuli depending on who the listeners believed the speaker was. Based on Niedzielski (1999), a study in New Zealand English context was conducted by Hay et al. (2006a), which led to further research on the effects of inferred social factors on speech perception. Factors such as nationality (Hay and Drager 2010; Hay et al. 2006a; McGowan 2011), gender (Strand 1999; Levon 2006, 2007), age (Drager 2011) or the perceived social class of the speaker (Hay et al. 2006b) have all been topics for study.

All of these previous works have contributed to the field of sociolinguistic studies significantly in their own way. In order to systematically review these contributions, it is necessary to first understand the process of speech production and perception from a holistic perspective. In daily conversations or speech production and perception situations, it usually involves interlocutors, the conveyed conceptual message, and the exchanged social information of one another whether conscious or not. A typical procedure of such situation can be represented by the following figure of speech production and perception process flowchart.

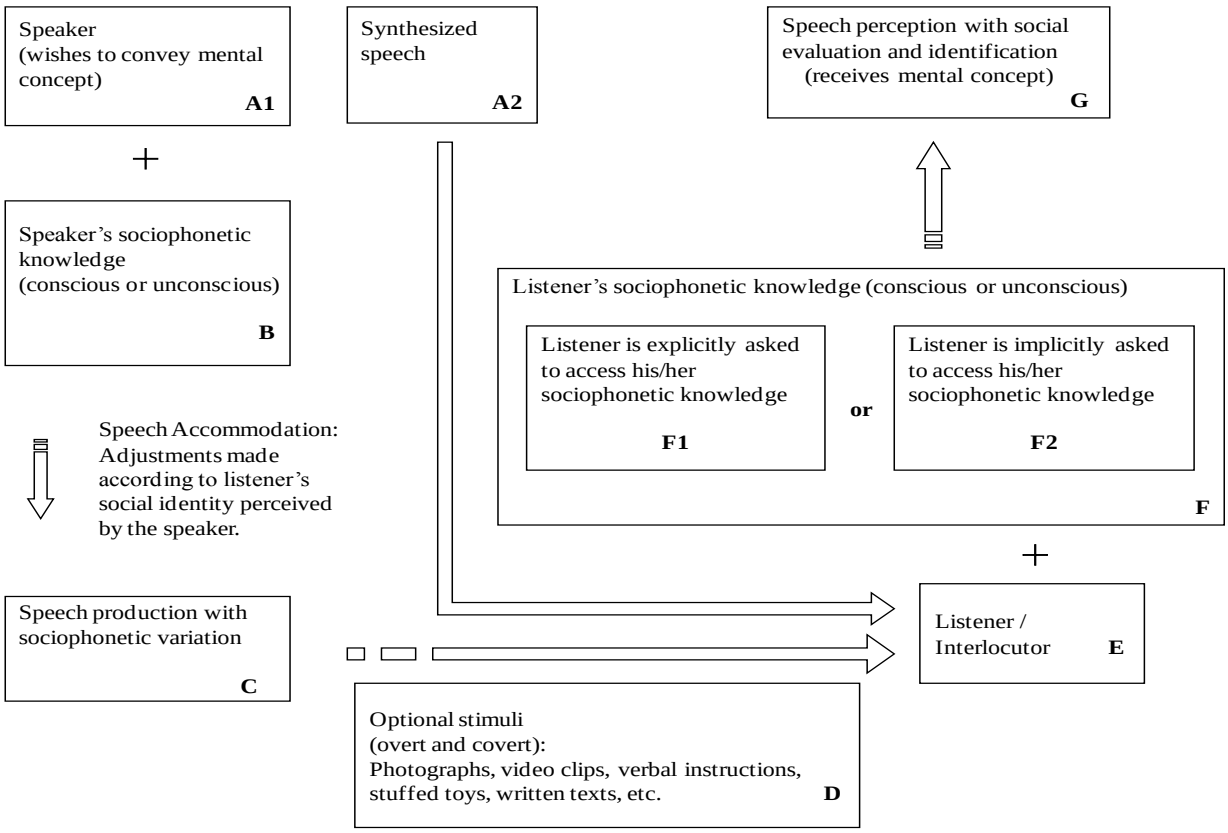



Figure 2.1 Speech production-perception process flowchart

Sociolinguistic studies have examined each of the elements (represented in blocks in the figure) involved in the speech production-perception process, in an attempt to understand what has been conveyed beyond the mental concepts during the process. This flowchart can serve as an outline for the many tasks that were employed in these studies. As shown in the flowchart, the starting point of the speech production-perception process can be either the natural speech of a real speaker or synthesized speech. In natural situations without technical aids, the flow of the speech production-perception process should follow the order of A1-B-C-E-F-G. Studies examining this process can be further divided into two categories according to the nature of the tasks involved in speech perception. If the task explicitly requires listeners to access their sociophonetic knowledge, then the flow of process would be A1-B-C-E-F1-G. Examples of such explicit tasks include asking listeners to identify or label social characteristics such as gender, age, origin, or ethnicity of the speaker they heard according to the speech samples. If the task implicitly asks listeners to access their sociophonetic knowledge, then the flow

of process would be A1-B-C-E-F2-G. For instance, participants may be asked to complete more linguistically driven tasks such as word-identification, sentence-comprehension, or vowel-matching, but the purpose of the tasks remains to examine participants' knowledge of sociophonetic variation.

The speech sample played to the participants can be either naturally recorded speech following the flow A1-B-C or manipulated/synthesized speech starting out from A2 in the speech production-perception process. During the process of speech perception, in order to further test the effects of perceived social characteristics on speech perception, optional stimuli (block D) may be inserted into the process. In either situation, the process proceeds through block F1 or F2 according to the nature of the tasks. The remainder of this section reviews previous studies in accordance to the different flows of speech production-perception process and the tasks involved.

## 2.1 Perception of natural speech

In this section, the research reviewed proceeds through the A1-B-C-(D)-E-F-G sequence in the speech production-perception process flowchart, in which listeners listen to synthesized speech and complete various tasks. The tasks are divided by whether the tasks explicitly or implicitly access participants' sociophonetic knowledge

### 2.1.1 Perception of naturally recorded speech without optional stimuli

Using naturally recorded speech and explicitly asking listeners to identify certain social characteristics of the speaker examines listeners' assumptions of distributions of sociophonetic variations. Williams, Garret & Coupland (1999) asked Welsh teachers and high school students to label the speakers' origins by listening to samples of different dialects of English spoken in Wales. Although the dialects were not robustly identified, it was found that teachers more correctly labeled the speakers' origins than high school students. The suggested reason for this result is that teachers have more exposure and experience with the different dialects, and thus could more correctly identify the speakers'

origins than the students who are less familiar with the different dialects. Another dialect identification task was conducted by Clopper and Pisoni (2004) in which they asked college students of Indiana to categorized six US regional dialects. The error patterns in the results showed that participants were aware of underlying similarities in relevant dialects. It was also found that students who had previously lived in the areas of the tested dialects could more accurately identify them than those who had not have previous exposure. Plichta and Preston (2005), in their study asking listeners to place given speakers on a geographic North-South continuum in the Eastern US based on the different degrees of the monophthongization of the vowel /aɪ/ in the word *guide*, found that listeners are not only aware of the dialectal differences, but show fine-grained knowledge of how sociophonetic variation is distributed in society.

In addition to the speaker dialect identification and categorization tasks conducted through the A1-B-C-E-F1-G procedure in the speech production-perception process, works involving speaker ethnicity-labeling and gender-identification tasks have also been done. Thomas and Reaser (2004) found that North Carolina listeners tend to more accurately label a speaker as African-American when the speaker is an African-American of a younger age rather than older. The suggested reason for such result is that African-American and Anglo speech is fairly similar in the older generations of the relevant speech community, and thus the distinction is less likely to be detected among the older generations. On the other hand, since the two varieties have diverged in recent generations, it is clearer to the listener to label the speaker as an African-American when the speaker is younger.

Foulkes, Docherty, Khattab & Yaeger-Dror (2010) conducted a gender-identification study in the Tyneside region of Northern England. The social distribution of different glottalization patterns for word-medial voiceless stops /p/, /t/, and /k/ is different between male and female. Foulkes et al. hypothesized that a speaker's gender can be identified by local listeners based on his or her glottalization pattern alone if local listeners are sensitive to the correlation between gender and patterns of glottalization. Non-local listeners who are not aware of such correlation should not be able to successfully complete the gender-identification task. In order to test the hypothesis, Foulkes et al. used recordings of preadolescent children since their voice quality is often considered gender-

ambiguous. The results showed that only the local listeners but not listeners from Southern England or from Arizona made gender judgments in accordance to the relevant glottalization patterns

The research reviewed involving the A1-B-C-E-F1-G process show that listeners have extensive sociophonetic knowledge of how variation is distributed in society. In these studies, participants were asked to explicitly access their sociophonetic knowledge. They were given specific social categories and their tasks were to match sample linguistic cues to the provided categories. The hypothesis behind these tasks is that listeners access their sociophonetic knowledge while listening to the stimuli and make assumptions about the speaker's social characteristics. However, in natural discourse settings where interlocutors usually have access to the visual image of one another, social characteristics such as those mentioned above are more often easily identified without accessing their sociophonetic knowledge. In addition, confining participants to a particular set of options to choose from may lead them to exclude all other social factors that would normally be taken into consideration in natural discourse settings. In other words, although participants are able to determine the social characteristics of the speaker during the experiment, it is likely that they do not proceed through the same process in natural discourse settings, or are unable to complete the same tasks outside of the experiment setting.

In order to avoid tapping into listeners' sociophonetic knowledge in such unnatural ways, alternative and more linguistic tasks have been developed to further explore listeners' beliefs of the distribution of sociolinguistic variation. In these tasks, participants were asked to complete linguistic questions based on their sociophonetic knowledge rather than directly identifying social characteristics of the speaker. These kinds of tasks, although still in an experiment setting, are more similar to what actually happens during a speech production-perception exchange. This kind of procedure is illustrated by the A1-B-C-E-F2-G block sequence in the speech production-perception process flowchart. As the following examples show, the tasks applied include asking listeners to identify words or to provide comments about the speakers, which are processes that normally occur in natural discourse settings.

Labov, Yaeger & Steiner (1972) tested the /u/-fronting phenomenon in the North

American chain shift. They recorded Coastal North Carolina speakers producing words containing fronted variants of /u/, and asked Philadelphia listeners to identify the words. The participants showed a tendency to misperceive the stimuli as words containing /i/ because the fronted /u/ variants have an [i]-like quality to their ears. In a more recent study where Campbell-Kibler (2007) asked groups of participants to comment on speakers in the audio clips (e.g., *What can you tell me about Jason?*), it was found that the level of education given to the speaker by the listeners correlated with the variation of the word-final alveolar nasal [n] occurring in gerunds, present participles, and the words *nothing* and *something*. Listeners were more likely to rate the speaker as educated and articulate when hearing the *–ing* guise (e.g., *fishing* in the *–ing* guise) than when hearing the *–in* guise (e.g., *fishin'* in the *–in* guise).

To sum up, the studies involving the A1-B-C-E-F-G sequence in the speech production-perception process showed that listeners are aware and have extensive knowledge of the distribution of sociolinguistic variations. They are able to hypothesize certain social categories of the speakers based on their sociophonetic knowledge, but they also make biased linguistic decisions when they are primed by given social information of the speaker. However, in order to test more precisely how fine grained their knowledge is about the sociolinguistic variations, more fine tuned variants need to be provided to the listeners. As a result, manipulating the linguistic cues by synthesizing recorded natural speech becomes a method to achieve this goal as demonstrated by the studies reviewed in section 2.2

### 2.1.2 Perception of naturally recorded speech with optional stimuli

In this section and also in 2.2.2, research reviewed includes stimuli of visual aids or verbal instructions during the experiment. The optional stimuli may or may not correspond with the audio stimuli provided to the participants. Listeners' expectations and assumptions about the speaker are examined in this type of experiment setting. In order not to bias participants' evaluation of speakers' social characteristic during the process, the optional stimuli are most often accompanied by linguistic tasks that do not directly ask for listeners' social judgment of the speaker. Specifically covert questions

about the speakers' social identity are sometimes asked during interviews or debriefings after the experiment.

In research involving the A1-B-C-D-E-F2-G procedure in the speech production-perception process, listeners listen to naturally recorded speech and complete tasks implicitly accessing their sociophonetic knowledge while given optional stimuli. Works with such experiment design include Hay, Warren & Drager's (2006b) study using the word-identification task, Staum's (2008) work utilizing the sentence comprehension task, and Koops, Gentry, and Pantos' (2008) eye-tracker research monitoring the process of word-identification. Hay et al. (2006b) investigated the perception of diphthongs /iə/ and /eə/ in New Zealand English. Participants listened to recorded readings of minimal pairs containing the diphthongs, such as *here* and *hair*. The speakers chosen to record the stimuli showed varying degrees of vowel merger, yet each stimulus could be correctly identified with linguistic information alone. However, listeners seem to have more difficulties distinguishing the diphthongs when they are presented along with photographs of males and females of different ages and in different surroundings. The researchers measured the participants' correct identification rate, which was found to be affected by the photos. The result showed that participants were less likely to distinguish /iə/ and /eə/ when the photo represented a person belonging to the social group that is more likely to merge the diphthongs in reality.

Staum (2008) investigated the "t/d-deletion", the production or omission of /t/ and /d/ in word-final consonant clusters, in two dialects of American English: African-American and Anglo dialects. The assumed pattern is that speakers of African-American English are more likely to produce words such as *mast* as [mæs] than speakers of Anglo dialects. In order to make the variable less identifiable to the participants, Staum used a sentence comprehension task to study how the perception of such words is affected by additional social cues. For instance, participants heard sentences beginning in words such as "The [mæs] lasted …" where [mæs] is ambiguous between *mass* (deletion variant) and *mast* (non-deletion variant), and ending strings such as "…through the storm" or "until noon on Sunday" where the meaning is disambiguated. Along with the audio input, participants also saw photos of Anglo or African-American. The response time that participants took to determine whether the sentence was semantically meaningful was

measured. The results show that the processing time was longer when participants saw a photograph of an Anglo paired with the 'deletion' variant and an African-American photo paired with the 'non-deletion' variant. The longer response time explains that a conflict occurs between what is expected and what is provided by the audio and visual input. When participants see a photo of an Anglo speaker and hearing [mæs], *mast* was activated rather than *mass*, because they expect Anglo speakers to preserve /t/ at word-final positions. On the other hand, participants activated *mass* instead of *mast* when given the same audio input but paired with a photo of an African-American speaker. The implication of the results is that participants' expectations according to their sociophonetic knowledge, though not explicitly accessed, have an effect on speech perception.

While Staum (2008) showed that perceived social identity has an effect on speech perception by measuring the response time, Koops, Gentry, and Pantos (2008) conducted an eye-tracker experiment which measures the time fixation of a word on the monitor to more directly measure the online processing of word identification. Koops et al. hypothesized that the pre-nasal merger of /ɛ/ and /ɪ/ (as in *dentist* and *dinner*) is correlated with age in the speech of native Anglos of Houston, Texas. Older speakers of this dialect tend to merge the two vowels while younger speakers made a distinction. During the experiment, participants listened to recordings of female speakers (young or old) producing the target words and simultaneously looked at photos of women of three different ages, relatively younger, mid-age, or older. The eye-tracker recorded the fixation-time on the target (actual) word and the competitor word. As predicted, the fixation-time on the competitor word was relatively longer when hearing an 'older voice' paired with an 'older photo'.

It can be concluded from these three studies that subtle implicit information of the speaker's social identity causes an effect on the perception of speech. With the visual input of a photograph, simulating real life conversations where interlocutors often see the images of one another, listeners' activate their sociophonetic knowledge of the distribution of variations and make assumptions about the speaker. In other words, listeners hold expectations towards the speaker and these expectations would affect the perception of their speech. Even though the tasks did not directly ask participants to

evaluate the social characteristics of the speaker or make a social decision, participants still activated the link between social identity and variation of speech production in their minds, which was proven through the linguistic oriented tasks.

## 2.2 Perception of synthesized speech

In this section, the research reviewed proceeds through the A2-(D)-E-F-G sequence in the speech production-perception process flowchart. Listeners in these studies listen to synthesized speech and complete various tasks. The tasks are once again divided by whether the tasks explicitly or implicitly access participants' sociophonetic knowledge

### 2.2.1 Perception of synthesized speech without optional stimuli

Research reviewed in this section includes experiments conducted using synthesized speech to examine how fine grained the sociophonetic knowledge is in the minds of the participants. Graff, Labov & Harris (1986) asked participants to identify speakers' ethnicity and Walker's (2007) study asked participants to label speakers' age and rate their social class. In Graff, Labov, and Harris' (1986) experiment, Philadelphia listeners were asked to identify a speaker as either Anglo or African-American by listening to synthetically manipulated variants of /aʊ/ and /oʊ/. The results show that listeners have a higher tendency of matching the more fronted variants to Anglo categories, which corresponds to the patterns of speech production. In Walker's (2007) New Zealand English study, it was found that listeners are more likely to label the speakers as younger and lower in social class when manually cutting the burst of /t/ at the end of intonation units, creating unreleased allophones. On the other hand, listeners rated the speakers as older and higher in social class when hearing the conservative released version of phrase-final /t/.

Research using synthesized speech can also be applied to tasks that implicitly access listeners' sociophonetic knowledge (A2-E-F2-G), such as vowel categorization tasks and again word identification tasks. Willis (1972) recruited high school students from two adjacent cities across the US-Canada border, Fort Erie, Ontario, and Buffalo, New York

and asked them to categorize vowel tokens as /æ/ or /ɛ/. It was found that the listeners from different speech communities categorized the synthesized vowel tokens of /æ/ and /ɛ/ into different phonetic categories that correlated to the production norms of their own speech community. For instance, Buffalo listeners were more likely to categorize tokens intermediate between [æ] and [ɛ] as /æ/ than Fort Erie listeners since the Buffalo speech community had a more raised /æ/.

With a word identification task, Janson and Schulman (1983) found that the link between speech production and perception may not be as symmetrical as expected. In their study of two merging vowels short /e/ and short /ɛ/ in dialects of Swedish, they used a speech synthesizer to create a continuum of vowel qualities in a /sVt/ frame. Listeners from two different dialect communities, Stockholm and Northern Sweden, were asked to distinguish tokens of these two vowels. Both groups of participants failed to differentiate short /e/ and short /ɛ/ even though the Northern Sweden dialect maintained the distinction in speech production, while the Stockholm dialect did not. Janson and Schulman concluded that the distinction although maintained in speech production, was not utilized in speech perception.

Janson and Schulman's (1983) study was replicated by Labov, Karen & Miller (1991), but used a less direct word identification task. They studied the merger of /ɛ/ and /ʌ/ before /ɹ/ as in words such as *merry* and *Murray* in the Philadelphia area, which has come to be known as the pilot of "near-merger" studies. In their devised 'coach test', participants listened to a story about a coach in which the ending of the story based on whether a particular phrase included *Merion* or *Murray in*. They then answered questions to which the answers were based on which phrase they had heard. Each participant repeated the process for the second time, except they heard the opposite phrase in the repeated process which they did not know during the time of experiment. The results were similar to Janson and Schulman's (1983) study. It was found that Philadelphians could not recognize the difference even if they maintained a distinction between the two vowels in speech production.

In summary, with aid of acoustic-analysis tools, researchers are able to more easily manipulate sounds and conduct in depth experiments to examine listeners' sociophonetic knowledge. In some studies it was found that there is a direct correlation between speech

production and perception, while in others an asymmetry was found between production and perception.

### 2.2.2 Perception of synthesized speech with optional stimuli

Previous research that utilized both synthesized speech and visual aids, procedure A2-D-E-F1-G in the speech production-perception process, is reviewed in this section. Strand and Johnson (1996) and Drager (2005, 2011) applied word-identification tasks to test listeners' gender-specific and age-specific assumptions respectively, while Niedzielski (1997, 1999), Hay et al. (2006a), and Hay and Drager (2010) conducted vowel-matching tasks to test dialect-specific assumptions. These studies, similar to the studies in section 2.1.2, also provided participants with additional visual stimuli to investigate how implicit information of social identity can affect listeners' perception of speech.

Strand and Johnson (1996) examined listeners' gender assumptions toward the production of the fricatives /s/ and /ʃ/ in Ohio English. Instead of having participants listen to recordings of real speakers producing words with these fricatives, they provided audio inputs of synthesized words with these initial consonants. They first recorded real speakers (whose gender were less easily identifiable by other speakers) producing the words *sod* and *shod*. By systematically manipulating the spectral quality of the initial consonants /s/ and /ʃ/, they created a synthesized [s] to [ʃ] continuum, from which they spliced tokens onto the original -/ɑd/ sequences. The participants' task was to identify whether the word they heard in the recording was *sod* or *shod*. The twist to the experiment was that while listening to the recordings, participants also looked at video clips of male or female speakers producing the words. The results showed that the same fricative could be interpreted as either /s/ or /ʃ/ depending on the perceived gender of the speaker. When viewing clips of a male speaker, participants biased toward hearing more tokens as *sod*. On the other hand, when given clips of a female speaker, they were more likely to identify the tokens as *shod*. It was concluded that Ohio English speakers had gender-specific expectations toward the production of /s/ or /ʃ/, and used this sociophonetic knowledge to categorize words when the gender of recorded voice was not easily identifiable.

Drager (2005, 2011) studied listeners' perception of /ɛ/ and /æ/ in New Zealand English. It was first found that recent decades of chain shift in the speech of New Zealanders of European decent has moved both /ɛ/ and /æ/ upwards in F1/F2 space. The degree of the raising is, however, different across speakers' age: higher in younger speakers but not as high in older speakers. In order to test whether listeners are aware of such age difference in the production of the two vowels, Drager paired synthesized tokens with photographs of a younger or older speaker and asked participants to categorize the tokens. The synthesized tokens were created by manipulating the synthetic vowel continua from [æ] to [ɛ] in the words *bad* and *bed* as well as *had* and *head* for both male and female speakers. It was found that listeners were more likely to identify tokens as /æ/ when seeing photos of a younger speaker than an older speaker, especially when the tokens were taken from the intermediate range of the vowel continuum.

In Niedzielski's (1999) study of the perception of the Northern Cities Chain Shift in Detroit, subjects were asked to match the vowel sound in a target word to one of the vowels in a series of synthetic vowel stimulus tokens. Participants were separated into two groups with one group being told that the speaker was from Detroit and the other the speaker was from Canada even though the speaker was the same female speaker with Northern Cities shifted vowels. The effect of manipulating the speaker's identity resulted in different choices of vowel selection. The subjects in the group that believed they listened to a Detroit speaker tended to select canonical unshifted vowels, while subjects in the group they believed listened to a Canadian speaker tended to select the most acoustically-similar vowels.

This study has been replicated for New Zealand versus Australian English by Hay et al. (2006a). The participants were also asked to match a target vowel to tokens from a synthesized vowel continuum. Although all subjects listened to the same New Zealand English speaker, they were divided into two groups as in Niedzielski's (1999) study. Participants in one of the groups were given answer sheets with the word "Australian" on it, while the other group of participants had answer sheet with "New Zealander" written on it. The results showed that despite the fact most of the subjects were aware that they were listening to a New Zealand English speaker, they gave significantly different matches of vowels. For instance, those who were given the "Australian" answer sheet

were more likely to hear a higher fronter /I/ vowel, whereas those working on a "New Zealander" answer sheet more likely to hear a centralized version of the same vowel.

A follow-up experiment also in the New Zealand English setting was conducted by Hay and Drager (2010). The task for participants in this project was still to match natural vowels produced by a New Zealand speaker to tokens from a synthesized vowel continuum. However, instead of giving direct written linguistic cues of nationality, stuffed animals were used to invoke the same concept: stuffed toy kangaroos and koalas implying Australia or stuffed toy kiwis implying New Zealand. The results once again showed that vowel perception shifts accordingly to the set of toys perceived by the subjects.

In summary, research has shown that subtle implicit information of speaker's characteristics or social identity inferred by the listener can influence how language varieties are perceived, and can cause significant effects on the result of speech perception. The implied speaker's identity can be provided through additional visual or oral stimuli ranging from photographs, video clips, verbal instructions, written instructions to even stuffed animals. It seems that even the very implicit cues can activate listeners' sociophonetic knowledge and thereby cause an effect on speech perception. This dissertation aimed to investigate the effects of also a subtle cue, two variants of Chinese orthography, to see whether listener would show a category boundary shift when given audio-visual integrated stimuli in both speech production and perception.

# Chapter 3

## 3. Alveolar and retroflex sibilants in Mandarin Chinese

Mandarin Chinese, consisting of approximately 800 million first-language speakers (70% of the population in China), is the largest spoken variety among the seven dialect groups of the Chinese language (Norman 1988). Among the various local dialects of Mandarin Chinese, the Taiwan Mandarin dialect along with its comparisons with the Beijing Mandarin dialect is the focus of this dissertation. The Taiwan Mandarin dialect was derived from the Beijing dialect in 1949 when the government of the Republic of China (ROC) retreated to Taiwan after the civil war. The later contacts with other Chinese dialects in Taiwan have allowed the Taiwan Mandarin dialect to take in elements from Hakka and Min. This new gradually developed dialect of Mandarin Chinese, Taiwan Mandarin, is officially known as *Guóyǔ*. The Beijing Mandarin dialect, spoken in the People's Republic of China (PRC), is the basis to the so-called standard dialect of Mandarin Chinese in everyday English. It is known as *Pǔtōnghuà*, meaning 'spoken by all people' or 'common speech', and referred to as *Guanhuà* in some literature, meaning the 'official dialect.'

This chapter will begin with a general introduction of the sound inventory in Mandarin Chinese. Subsequently, a brief review of the historical development of Taiwan and the Taiwan Mandarin dialect will be provided together with a comparison between Taiwan Mandarin and Beijing Mandarin. Finally, a collection of previous studies on the production and perception of Mandarin alveolar and retroflex sibilants will be reviewed.

### 3.1 Syllable structure and sound inventory of Mandarin Chinese

The maximum syllable structure of a Mandarin syllable can include an initial consonant, a medial glide, a nucleus, and the coda consonant as illustrated in Figure 3.1. Within this basic syllable structure, only the nucleus vowel is indispensable while all others are optional.

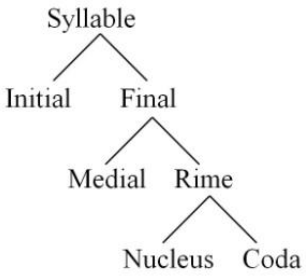



Figure 3.1 Basic Mandarin syllable structure

According to Duanmu (2007:42), a system of twenty-five underlying phonemes, including 19 consonants and 6 vowels, in Standard Mandarin Chinese is "assumed in most studies and needs little comment." However, the exact number of underlying vowels phonemes still remains controversial across studies. It may be as many as 11 (Shih 1995), or as few as six (Cheng 1973), four (Wu 1994), and even none (Pulleyblank 1984). Huang, Li, and Simpson (2014:371) provided a list of previously proposed "vowel systems in order of decreasing frequency of occurrence [which] are 8-vowel (18), 10-vowel (13), 5-vowel (8), 6-vowel (8), 7-vowel (7), 9-vowel (6), 11-vowel (3), 12-vowel (3), 4-vowel (2), 3-vowel (1), 13-vowel (1)." "The main difference is in the apical vowels, and in a large majority of cases, they are allophones of /i/ or /y/" (371). When excluding the retroflex vowel [ɚ] and apical vowels/syllabic consonants[1], Duanmu (2007:35) agrees with many of previous studies (Chao 1948; Cheng 1966; Lin 1989; Wan & Jaeger 2003) that there are five vowel phonemes in Mandarin Chinese.

When assuming a five-vowel system, there are typically three high vowels /i, y, u/, one mid vowel /ə/, and one low vowel /a/ as shown in Table 3.1.

| | |
|---|---|
| High vowels | /i/ → [i], [j] |
| | /y/ → [y], [ɥ] |
| | /u/ →[u], [w] |
| Mid vowel | /ə/ →[ə], [e], [o], [ɤ], [E] |
| Low vowel | /a/ →[a], [ɑ], [ɐ], [æ], [A] |
| Apical vowels/syllabic consonants | |

Table 3.1 Mandarin vowel phonemes and allophones based on Duanmu (2007: 39)

[1] The apical vowels/syllabic consonants are perceived as vowels by Howie (1976:10) and Cheng (1973:13), but considered as syllabic consonants by others (Chao 1968; Pulleyblank 1984; Lee & Zee 2003; Duanmu 2007; Lin 2007). Some proposed that the apical vowels/syllabic consonants could be derived from the underlying vowel /i/ (Cheng 1966; Wu 1984; Wan and Jaeger 2003), while others advises that they are formed by the process of epenthesis of a syllable nucleus and assimilation of consonant features.

The high vowels /i/ and /u/ in Mandarin Chinese are similar to those in English, and [y] is a high front rounded vowel. All three of them can precede another non-high vowel, and thus surface as glides. Therefore, glides are listed here as allophones of the high vowels "when they are in non-nuclear position, for example /uai/ → [wai]" (Huang et al 2014: 403). While the analysis of high vowels is generally consistent across studies, there are alternative perspectives toward the transcription of mid and low vowels.

The transcription and total number of surface vowels for the mid vowel is discordant across literature. According to Xu (1980) and discussed in Duanmu (2007), there are five variants of the mid vowel. The symbol [E] represents a vowel that is higher than [ɛ], but lower than [e]. However, Chao (1968) and Chen (1973) denote both [e] and [E] with [e] without making a distinction. The general agreement is that the surface realizations of the mid vowel are derived from the assimilation process of /ə/ in frontness or rounding with the adjacent consonant phoneme (Duanmu 2000). As Duanmu (2007:38) concluded: "Perhaps it is unrealistic to expect the phonemes of a language to be perfectly clear and evenly distributed."

Similar with the mid vowels, there is no unified treatment of the low vowels in Mandarin Chinese. Xu (1980:33) suggested five surface realizations of the mid vowel, and the variants are represented as the following: [a] is a front low vowel, [ɑ], is a back low vowel, [ɐ] is a central mid low vowel, [æ ] is a vowel between [a] and [ɛ], and [A] is a central vowel. Again, Duanmu (2007:39) concluded: "Overall, the low vowel remains low and unrounded, but could probably change its backness."

There are 26 surface consonants in Mandarin Chinese when including the zero initial since not all syllables require an initial consonant. In the cases where there are no initial consonants, "the tradition is to describe them initials as 'zero'" (Li & Thompson 1981:3). Among the other 25 consonants, there are 20 consonants that only occur at syllable initial position, three prenuclear glides [w, ɥ, j], the alveolar nasal [n] that can occur as both syllable initial and coda consonant, and the velar nasal [ŋ] that only occurs at syllable final position. Aspiration is a distinctive feather in Mandarin Chinese, and thus creating six pairs of contrasting stops and affricates. The following Table 3.2 provides the collection of Mandarin consonants.

| | | Place of Articulation | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Manner of Articulation | | Bilabial | Labiodental | Alveolar | Alveopalatal | Retroflex | Palatal | Velar |
| | Stop | p $p^h$ | | t $t^h$ | | | | k $k^h$ |
| | Fricative | | f | s | ɕ | ʂ | | x |
| | Affricate | | | ʦ $ʦ^h$ | ʨ $ʨ^h$ | ʧ $ʧ^h$ | | |
| | Nasal | | m | n | | | | ŋ |
| | Approximant | w ɥ | | l | | ɹ | j | |

Table 3.2 Mandarin Chinese consonants

As shown in the table above, the alveolar, alveopalatal, and retroflex fricative and affricates form a three-way coronal contrast. The alveopalatal fricative and affricates [ɕ, ʨ, $ʨ^h$] only occur when followed by high front vowels [i, y] and glides [ɥ, j], e.g., [ɕi, ʨi, $ʨ^h$i] and [ɕja, ʨja, $ʨ^h$ja] but not directly by other vowels, such as *[ɕa, ʨa, $ʨ^h$a]. The phonological status of the alveopalatal sibilants are discordant across literature and still remain in debate. Cheng (1973) and Li (1999) perceived them as synchronically independent phonemes. Duanmu (2007: 31-34) treated the alveopalatal sibilants as complex consonants derived from combining the dental sibilants with glides [ɥ] or [j], and "have a special relation with dentals." Lin (2007) viewed them as allophones of the dental sibilants, while others have considered them to be the result of palatalization of the underlying velar obstruents [k, $k^h$, x] (Chao 1934, 1968; Hsueh 1986) because the velar series only occur before non-high-front vowels/glides.

The alveolar and retroflex fricative and affricates, in complementary distribution with the alveopalatal sibilants, are allowed to precede [a, u, ə], apical vowels, and the glide [w]. Depending on the articulatory observations of Mandarin Chinese speakers from various regions, the alveolar sibilants, fricatives and affricates, have been discussed in terms of dentals (Chao 1968; Duanmu 2007; Lin 2007), denti-alveolars (Lee & Zee 2003) and alveolars (Luo & Wang 1981; Pulleyblank 1984; Ladefoged & Maddieson 1996; Li 1999). As for the retroflex sibilants, divergent terms from various studies have been used. They have been referred to as apical post-alveolars [ʃ̺, ʧ̺, ʧ̺$^h$] (Lee & Zee 2003; Lin 2007), laminal post-alveolars [ʂ̻, ʧ̻, ʧ̻$^h$] (Ladefoged & Maddieson 1996), prepalatals (Xu 1980; Wu 1992; Cao 2002), or simply retroflexes (Duanmu 2000). The term 'retroflex' will be

used in this study to describe this series of consonants in view of a broader definition of the term which exceeds indicating the articulatory place but designates the articulatory gesture (Hamann & Fuchs 2010). It also corresponds to the common terminology employed by Mandarin speakers, *juǎnshé yīn* ‘curled-tongue sound’. A more in depth discussion of the articulatory and acoustic properties of Mandarin alveolar and retroflex sibilants will be provided in section 3.3.

Phonologically, the three prenuclear glides [w, ɥ, j] have been analyzed as homorganic glides of the vowels [u, y, i] respectively and some have treated them as part of the syllable initial consonant (Duanmu 1990, 2007; Bao 1996; Lin 2007). There are three nasal consonant phonemes in Mandarin /m/, /n/, and /ŋ/. The bilabial nasal stop only occurs as syllable initial position. The alveolar nasal stop is allowed to occur at both syllable initial and syllable final position. When it occurs as syllable final initial, it forms a minimal with the velar nasal stop /ŋ/, which only occurs at syllable final position.

### 3.2 The Taiwan Mandarin dialect vs. The Beijing Mandarin dialect

Though speaking various dialects, the broad Chinese community originally used the same orthographic system, the traditional characters. However, the People’s Republic of China (PRC) began promoting simplified characters along with the Beijing Mandarin dialect in 1954, while Taiwan, Hong Kong, and Singapore continued to use traditional characters. The connection between simplified characters and the Beijing Mandarin dialect began to formulate since then. The Taiwan Mandarin dialect was gradually developed after three waves of settlement from Mainland China to Taiwan (Huang 1993; Su 2005).

Taiwan, before the presence of Chinese people was occupied by tribes of the Austronesian family, who are currently divided into 14 groups of lowland peoples and 9 groups of mountain peoples (Stainton 1999). The first Chinese settlement on the island was in the 17$^{th}$ century. Approximately 200,000 people migrated from the southwestern coast of China to Southern Taiwan and established the first Chinese administration of Taiwan.

The second wave of immigration took place in the late 18$^{th}$ century (the Ching Dynasty). It was recorded that the population reached a climax of 2,500,000 by the end of

the 19th century (Huang 1993). During this wave of settlement, the new settlers migrated mostly from the Fujian Province and secondly from the eastern Guandong Province, which is next to Fujian. The major dialect of Chinese that came to Taiwan along with the immigrants from Fujian Province was Southern Min, which also became the dominant language of Taiwan at that time. The other dialect also introduced to Taiwan was Hakka from the Guandong Province. These two dialects were the major dialects that came into contact with the Beijing Mandarin dialect spoken in Taiwan after the third wave of migration, and later established the modern Taiwan Mandarin dialect.

The third wave of immigration took place during the post-colonial period in 1949 when the Republic of China's Nationalist Government, *Kuo Ming Tang* (the KMT party), lost the civil war to the Chinese communists, who later formed the government now known as the People's Republic of China (PRC). The Nationalist Government retreated to Taiwan, which thus became the last territory of the Republic of China (ROC). Between 1945 and 1949, a record of 600,000-1,000,000 Chinese people from all over Mainland China migrated to Taiwan (Feifel 1994), which is also the end of large-scale Chinese settlements in Taiwan. According to Huang (1993), starting from that point in history, the population of the residents in Taiwan could be divided into four major ethnic groups: Mainlanders (13%) who came the latest, Southern Min (73.3%) and Hakka (12%) who came in the 19th century, and the original Austronesian inhabitants (1.7%).

In order to promote Mandarin Chinese in Taiwan, a "Mandarin only" language policy was effective from 1956 to 1987. The purpose of this policy was to promote Mandarin as the one and only legitimate official language. As a result, all other languages, including the most used Min and Hakka were banned from schools and institutional settings. Many whose first language spoken at home was not Mandarin suffered from this policy, and they struggled during the process of learning it even after entering elementary school. The Mandarin dialect spoken by the teaching authorities at that time is definitely different from the Mandarin dialect spoken in contemporary Taiwan. While the speech of teaching authorities preserved most of the features of the Beijing Mandarin dialect, the contemporary Taiwan Mandarin dialect has acquired many of its own features due to sixty years of contact with other languages on the island.

The Mandarin only language policy, effective between 1945 and 1987, was also the

period when Taiwan was under the Emergency Decree. During this crucial period, people in Taiwan were not allowed to make contact with people in Mainland China. People were not allowed to visit or communicate via letters, and thus there was generally no language contact between Taiwan and the Mainland. News and updates regarding Mainland China were blocked or filtered by the government; therefore, people in Taiwan during that period had little or no information about the events occurring in Mainland China. The one thing that people in Taiwan did have information about was that the government of the People's Republic of China (PRC) officially started promoting simplified Chinese characters since 1954. However, traditional written characters remained to be used in Taiwan, Hong Kong, and Singapore. Hence, the connection between simplified Chinese characters and the social identity of Mainland Chinese people or Beijing Mandarin speakers was formed.

The textbooks for teaching Mandarin in Taiwan were designed to be based on the Beijing dialect; however, the inevitable language contact with the other local dialects of Taiwan still led to the standard Beijing dialect undergoing many changes (Huang 1993; Su 2005). As observed in many southern dialects of Mandarin, Taiwan Mandarin now differs from the Beijing dialect in many phonological features. One of the most notable differences is the merging of the retroflex series of the consonant initials (/tʂ/, / tʂʰ/, /ʂ/, and /ɹ/) with the alveolar series (/ts/, /tsʰ/, /s/, and /l/) (Chen 1999: 41-49). This phenomenon is also known as de-retroflection, which is claimed by many researchers to be the most notable difference between Taiwan Mandarin and Beijing Mandarin (Kubler 1979, 1986; Lin 1983; Rau and Li 1994; Li 1995; Su 2005; Chung 2006). Duanmu (2000: 26) stated that "[t]he retroflex series [tʂ, tʂʰ, ʂ, ʐ] is a major characteristic of SC [Standard Chinese] speakers from Beijing. SC speakers from other places often do not have [tʂ, tʂʰ, ʂ, ʐ] in their native dialects, so they often replace [tʂ, tʂʰ, ʂ, ʐ] with the dentals [ts, $ts^h$, s, z], or they may use [tʂ, tʂʰ, ʂ] for [ts, $ts^h$, s] in hypercorrection". Baran (2007: 205) observed that the "lack of retroflection is the first feature cited by local speakers when they are asked to describe the "biggest" difference between Standard Mandarin and Taiwan Mandarin." Lin (2007) also stated that the first major characteristic of Taiwan Mandarin is the lack of distinction between the retroflex and alveolar consonants. While some simply substituted the retroflex series with the alveolar, others may pronounce both series

somewhere in between the actual retroflex and alveolar phonemes (Lin 2007:268). Therefore, the present study was designed to examine this sociolinguistic phenomenon with a different approach and explore the social factors related to the production and perception of the retroflex series in Taiwan Mandarin.

In addition to the alveolar and retroflex merger in Taiwan Mandarin, there are two segmental differences and one suprasegmental difference between the two dialects of Mandarin. One of them is the commonly known “erhua” or “erhuayin” phenomenon that occurs often in Beijing dialect but rarely in Taiwan dialect. It refers to a phonological process that adds r-coloring or the "ér" [ɚ] sound to syllables in spoken Mandarin Chinese. This final ‘r’ is generally a non-phonemic suffix that is common in Northern varieties of Mandarin, especially in the Beijing dialect. Li and Thompson (1981) pointed out that this r-suffixation is not a regular morpho-phonological rule. The suffix is represented in characters as 兒, and it is suffixed to some but not all words generally as a diminutive suffix for nouns (Wei 1994). For example *hua* [xwa] ‘flower’ may be pronounced as *hua-r*. Despite the fact that the *-r* suffix is still presented in textbooks, seen in lyrics, and taught to foreigners, it is rarely used by Taiwan Mandarin speakers in contemporary everyday speech. The usage of this suffix in Taiwan may actually index a Mainland or Beijing Mandarin speaker identity (Wei 1994).

Still another segmental difference between these two dialects is the final [ŋ]-[n] merger. Although the distinction is taught in elementary school, it is not distinguished by the majority of speakers in Taiwan, including representatives of prescriptive pronunciation such as anchors (Su 2005:150). From the perspective of speech perception, many speakers cannot hear the difference between the two when asked to distinguish them even though they are consciously aware of their prescriptive difference (Baran 2007).

Finally, despite the fact that the same four lexical tones are used in both Taiwan Mandarin and Beijing Mandarin, there are differences in the tonal realizations. Tone realization in Beijing Mandarin, based on Chao’s (1930) tone digits are 55 (tone 1 / level tone), 35 (tone 2 / rising tone), 214 (tone 3/ dipping tone), and 51 (tone 4 / falling tone),

in which 1 represents the lowest pitch and 5 the highest. However, Fon and Chiang (1999) suggested that the tone values in Taiwan Mandarin should be modified as 44, 323, 312, and 42. Hence, the tone production in Taiwan Mandarin is distinguished from Beijing Mandarin in three characteristics: narrower tonal range, lower tonal heights, and flatter tonal contours. These suprasegmental differences between Taiwan Mandarin and Beijing Mandarin may be due to the language contact of Taiwan Mandarin with Southern Min (or Taiwanese). Stated in Huang's (1993) study of the language usage in Taiwan, Min is spoken at various levels by approximately 83% of the population. According to Cheng (1968), the tone values in Min are 44, 24, 31, and 52, which are generally lower than Mandarin. The contact of Taiwan Mandarin and Min could be a plausible factor contributing to the contemporary shape of tonal realizations of Taiwan Mandarin. Huang, Wu, and Fon (2012) further studied the effects of Min on Taiwan Mandarin speakers in terms of tone production and perception. They found that speakers of higher Min proficiency tend to have a lower pitch register for Mandarin level tones as well as a narrower pitch range and flatter contour than speakers of lower Taiwanese proficiency.

The different degrees of retroflection, the "erhuayin" phenomenon, and the final [ŋ]-[n] merger are the most noticeable differences between the Taiwan Mandarin dialect and the Beijing Mandarin dialect. The variation definitely goes beyond the above stated when taking other varieties of Mandarin Chinese into consideration, such as the dialect spoken in Singapore and Cantonese spoken in Hong Kong. Despite the many different dialects, the one feature that once tied all of them together was the orthographical system, the traditional Chinese characters. However, due to political and historical reasons, while the traditional Chinese characters are still used in Taiwan, Singapore, and Hong Kong, simplified Chinese characters have been used in Mainland China since 1954.

Taiwan Mandarin speakers are aware of the differences between the Taiwan Mandarin dialect and the Beijing Mandarin dialect (Baran 2007). The aim of this study is to evoke the respondents' awareness of these differences, their sociophonetic knowledge of these two dialects, via the orthographic system. The fact that the Taiwan Mandarin dialect and the Beijing Mandarin dialect currently use different written characters is the key to this study. Speakers of these two dialects are aware that the writing systems are

different, and thus the written characters play the role of implying different dialects and different nationality as well. Traditional characters are expected to imply the identity of a Taiwan Mandarin speaker, while simplified characters are expected to activate the identity of a Beijing Mandarin speaker.

## 3.3 Previous studies on the production and perception of Mandarin alveolar and retroflex sibilants

The articulatory description of alveolar fricatives typically consists of two dimensions, the constriction location and the portion of the tongue used for the constriction. Although some prefer the term alveolar fricatives, it was found in articulatory studies that the Mandarin /ts, $ts^h$, s/ alveolar series are virtually more dental than alveolar from the perspective of articulation. Lee and Zee (2003) categorized them as either apico-laminal or laminal denti-alveolar sibilants from observing palatographic and linguographic data. Ladefoged and Wu (1984) found from X-ray images (Figure 3.2) of native Beijing Mandarin speakers that /s/ was produced with the tip of the tongue placed at various places of constriction ranging from the teeth (speaker 1), slightly behind the teeth (speaker 2), and further back on the alveolar ridge (speaker 3). Dental affricates /ts/ and /$ts^h$/ were not found to display a different place of constriction from that of the fricative /s/ in Ladefoged and Wu (1984).

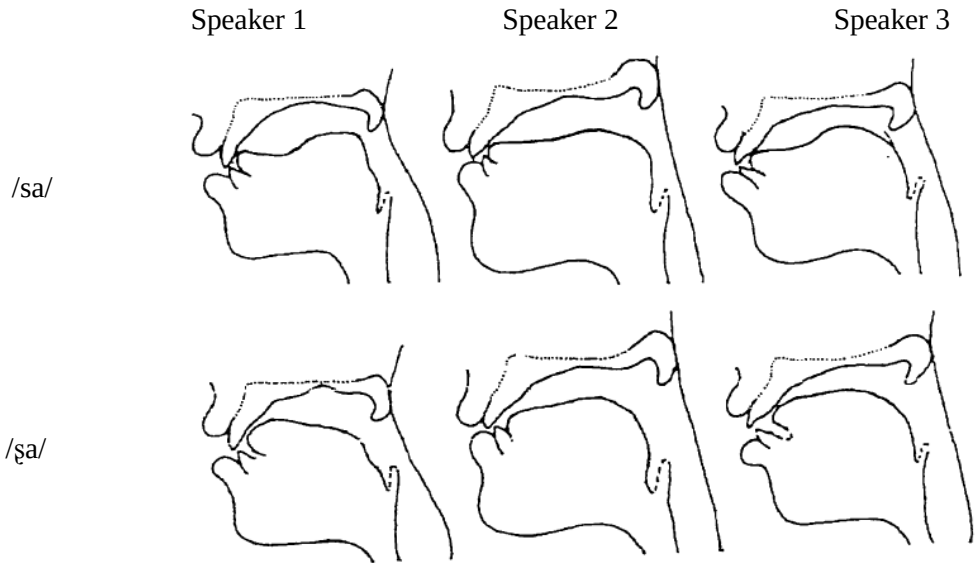


Figure 3.2 X-ray tracings of three male speakers producing /sa/ and /ʂa/ (based on Ladefoged and Wu 1984)

As for retroflex sibilants, Ladefoged and Wu (1984) illustrated with x-ray images that the constriction was made with the tongue blade approaching the area around the center of the alveolar ridge for all three speakers. Ladefoged & Maddieson (1996:152-156) pointed out that there is no contact or approximation between the lower surface of the tongue tip and the palate when articulating the generally known "retroflexes" in standard Chinese. Hence the retroflex series in Mandarin are technically not retroflexes, but "laminal (flat) post-alveolar sibilants." Chung (2006:200) observed that the retroflex series has "a fairly broad range of realization in Taiwan Mandarin." Based on the degree of the tongue retraction, it could vary "from highly retracted, through the palate-alveolar area [tʃ], [tʃʰ], [ʃ] all the way to straight dentals" (Chung 2006:200). It was noticed that speakers of Taiwan Mandarin are able to make the conscious choice of the degree of retroflection and they are aware that different choices index different social identities (Baran 2007). Sometimes, speakers are so conscious about meeting the prescriptive standard of retroflection that they hypercorrect the non-retroflex dental initials into their corresponding retroflex sounds (Chung 2006).

The distinction between alveolar and retroflex sibilants can also be made by comparing their acoustic features. Spectral measurements, one type of acoustic features, are a series of small snapshots of a sound wave that provide information about the frequency and amplitude of the waveform. These measurements indicate at what frequency the energy is the most concentrated or intensified. Sibilants at different places of articulation exhibit different spectral shape of frication noise, and thus distinction between alveolar and retroflex sibilants can be made by comparing their spectral measurements (Hughes & Halle 1956; Heinz & Stevens 1961). In Stevens' (1989) two-tube model, the sound wave (energy) passes through a vocal tract that is imaginarily divided into two cavities (tubes), front and back, by the constriction. The model showed that a smaller front cavity or a more anterior constriction would indicate more energy in higher frequencies. Using the [ʂ, s] pair as an example, the frication noise of the retroflex fricative is centered at about 2500Hz, while energy is clustered between 5000 and 6000Hz for the alveolar fricative (Ladefoged 2005). Another well discussed spectral

property that differentiates retroflex and alveolar fricatives is the lowering of F3. Research on retroflexes in various languages including Polish and Russian (Hamann 2002), Australian languages (Hamilton 1996), and English (Ladefoged & Maddieson 1996) has reached consensus that the lowering of F3 is a characterizing feature of retroflection.

The most robust and frequently used parameter to distinguish between retroflex and alveolar fricatives, according to Chang & Shih (2015), is the center of gravity (COG) of frication, also known as the 1st spectral moment. The center of gravity is a measurement of the average frequency weighted by intensity of the entire spectrum, which was employed by Lee, Zhang, and Li (2014) to distinguish between retroflex and alveolar consonants in Beijing Mandarin and by Jeng (2006) in Taiwan Mandarin. The length of the front cavity is in negative correlation with the value of COG. Therefore, a shorter front cavity, e.g., alveolar series resulted from a more anterior constriction would raise the frequency at which major energy is concentrated in the fricative spectrum, and thus a higher COG value than the retroflex series.

The lack of retroflection among Taiwan Mandarin speakers has been the center of focus in many previous studies. Conclusions have been drawn that females generally are more likely to retroflex than males (Lin 1983; Rau and Li 1994; Tse 1998), all speakers use significantly less retroflection in casual speech styles than in formal ones (Lin 1983; Rau and Li 1994), and speakers whose home language is Mandarin are more likely to retroflex than those whose home language is Min (Lin 1983). It is also observed in Rau and Li (1994) that retroflexing is more common when preceding back round vowels because of the effect of anticipatory tongue retraction, retroflex initials are more often realized as alveolars when immediately following syllables with alveolar initials due to perseverative assimilation. Examining the retroflex sounds individually, Lin (1983) found that the different sounds were de-retroflexed to a differing degree. Specifically, /ʂ/ exhibited the highest rate of retroflection, while /tʂʰ/ was most often de-retroflexed. While there is significant work on the de-retroflection of retroflex consonant initials in Taiwan Mandarin from the perspective of speech production, little or none has been devoted to speech perception.

# Chapter 4

## 4. Methodology

This chapter begins with explaining the motivations of this dissertation by discussing the results from a pilot study on the effects of the variation in Chinese characters on speech perception of retroflex sibilants in Taiwan Mandarin. The demographics and relationship between the participants of the experiments will be introduced and followed by providing information on the experiment design for a series of speech perception and production tasks. Finally, the general procedure of the experiments will be explained.

### 4.1 Pilot study of the effect of orthography on the speech perception of retroflex sibilants

In a pilot study for this dissertation (Tso 2013), 316 participants listened to the same acoustic stimuli, but were separated into two groups: the traditional-character-reading group (TCR) and the simplified-character-reading group (SCR). Traditional characters are currently used in Taiwan, while simplified characters are used Mainland China. The hypotheses were that (a) the social identity of either a Taiwan Mandarin speaker or a Beijing Mandarin speaker can be inferred by the type of character used, (b) speech perception would be affected by the expectations connected to the suggested social identity, and (c) the effects of suggesting social identity via the writing system on speech perception are reflected in the results of the surveys.

Based on the above hypotheses and the notable phenomenon of the retroflex series merging with the alveolar series in Taiwan Mandarin, the prediction was that participants reading a traditional character survey were more likely to match the retroflex sounds they hear to the alveolar series than those reading a simplified character survey. SCR participants on the other hand were predicted more likely to match the retroflex sounds they hear to the actual retroflex series of the consonant initials.

Three retroflex and alveolar consonant initial pairs /tʂ/ vs. /ts/, /tʂʰ/ vs. /tsʰ/, and /ʂ/ vs. /s/ were examined via a total of 63 words, including 36 target words and 27 filler words. In addition to the three pairs of target sounds, another three pairs were used as fillers to filter out invalid surveys in cases where a participant may have guessed through the survey instead of filling it out as instructed. These three additional filler pairs were /p/ vs. /pʰ/, /f/ vs. /x/, and /tɕ/ vs. /tɕʰ/. All of the words chosen for this project were among the top 900 most frequent words in the Academia Sinica Balanced Corpus of Modern Chinese, and over 50% of the chosen words ranked in the top 500 most frequent words. The chosen words were inserted into carrier sentences but in different positions of the carrier sentences. All carrier sentences were semantically meaningful and different from one another.

The participants were asked to listen to audio stimuli while reading the written transcript and pay attention to the underlined target words. The task was to match what they heard to one of the six consonant tokens synthesized with a Praat script[2]. For the purpose of this project, we first recorded each sound in the six pairs of sounds tested in this experiment: [/tʂ/, /ts/], [/tʂʰ/, /tsʰ/], [/ʂ/, /s/], [/p/, /pʰ/], [/f/, /x/], and [/tɕ/, /tɕʰ/]. The consonant pairs were mixed by four proportions of the sounds in the pairs: 20%-80%, 40%-60%, 60%-40%, and 80%-40%. The four synthesized sounds of each pair were then attached to appropriate vowels. Hence along with the two end points, six continuous choices were provided for the participants to choose from. Participants were also asked to complete a language background survey, which gathered their personal information, language background, and their knowledge toward the differences between the Taiwan Mandarin dialect and the Beijing Mandarin dialect.

The participants were asked to put a checkmark in the box of the sound that they find most similar to the sound of the word in the sentence. Since a chi-square test was necessary to determine whether the written characters have an effect on speech perception, each checkmark was assigned a number value in order to convert the results into an ordinal data format, on which the statistic tests could then be preformed.

---

[2] The Praat script is written by Dr. Chris Darwin, an emeritus professor of experimental psychology at the School of Psychology of the University of Sussex, UK. The website from which the script was retrieved (in June 2012) is as follows:
http://www.lifesci.sussex.ac.uk/home/Chris_Darwin/Praatscripts/Add2_variable

If the box corresponding to the sounds of the retroflex series (/tʂ/, /tʂʰ/, /ʂ/) in the target pairs or /p/, /f/, and /tɕ/ in the filler pairs were checked, the question was assigned an answer number value of 6. If the box corresponding to the sounds on the extreme opposite end of the continuum were checked, the question was assigned an answer value of 1. If participants chose one of the four answers in between the two ends, an answer value of 2, 3, 4, or 5 would be assigned according to its position on the continuum. The closer the answer was to the targeted end, the higher the assigned value. If more than one answer was chosen for a question[3] or the question was left unanswered[4], a zero would be assigned to that question.

A series of chi-square tests were applied to compare the results collected from the group of participants given the traditional Chinese character survey and the group that completed the simplified Chinese character survey. With the statistical results, the first and most fundamental question to answer was whether the results of speech perception among the two groups of participants were significantly different when presented with different written characters of the same context while listening to the same speaker reading the context out loud. It was found that while the speech perception of all female Taiwan Mandarin speakers across age groups were affected by the perceived written characters, only the younger but not the elder male speakers were affected by the perceived characters.

The findings suggested that (1) both male and female SCR participants under the age of 19 responded to the actual retroflex sounds while TCR participants were indecisive among 4 of the synthesized tokens; (2) both male and female SCR participants between the age of 20 and 29 responded to the actual retroflex sounds while TCR participants also responded to the actual retroflex sounds but with a slightly lower frequency; (3) female participants across ages assign different standards of Mandarin to both the speech in the question statements and the synthesized answer tokens, but the similar phenomenon was

[3] The questions that were responded with multiple choices were all retroflex-dental-pair questions. The multiple choices, usually two, were adjacent choices. There was no particular pattern found in the questions that were answered with multiple choices.

[4] The questions that were unanswered were again all retroflex-dental-pair questions. Some participants decided to leave blank all twelve questions testing the same pair, but there was not a trend of which pair was more likely to be answered. Other participants randomly left questions blank without a particular pattern.

not found in the male participants (4) female participants over thirty showed a different trend of direction in their responses than those under thirty, but the results from the male participants, instead of showing a different trend of direction, did not seem to be affected by the different types of characters; and (5) the merging of the retroflex and alveolar consonant initials in male speakers occurred twenty years earlier than the female speakers in terms of speech perception.

The data analyses showed that different types of written characters have different effects across gender and age. However, more details regarding the social background of the participants were collected during the experiment, such as their first language, their everyday language usage distribution, geographical differences, and frequency of reading simplified characters and listening to the Beijing Mandarin dialect. These were also potential directions for future research. In addition, even though the difference between retroflex and non-retroflex (dental) is binary, the majority of respondents did not necessarily make a binary choice when given a semi-continuous range of choices. This seemed to indicate that participants allowed different degrees of retroflection, and whether the different degrees were indicators of various social characteristics was also subjected to further investigation. Furthermore, it was found that the three retroflex sounds tested in this experiment seemed to show different degrees of de-retroflection. In summary, this study contributed to our understanding of the perception of sociolinguistic variation in many aspects. This was a large-scale study on the interaction between social identity and speech perception in Mandarin. The priming of social identity was fulfilled by the writing system, which variants are used by different dialects of Mandarin Chinese. It demonstrated that even with such a subtle hint of nationality, by using orthography, listeners' expectation would be effected by variation in written characters.

The conclusions drawn from the pilot study were based on perception alone; however, whether the relation between speech perception and production is symmetrical or asymmetrical and how they affect one another was not examined. Although there is an adequate amount of previous research on the variation of the production of retroflex initials in Taiwan Mandarin, some are outdated (Lin 1983; Rau and Li 1994; Tse 1998), while others' (Su 2005; Baran 2007) focus were not entirely on the retroflex initials. Therefore, both speech production and perception experiments were conducted in this

dissertation to examine the correlation between speech perception and production of retroflex consonants under the priming of written characters.

In addition to collecting categorical data and performing statistical tests to prove the significance of effects, this dissertation also aimed to utilize the language background information as well as contents from the interview with participants to conduct a more qualitative analysis and search for stylistic variations present among Taiwan Mandarin speakers. Another major difference between this pilot study and the dissertation is the audio stimuli employed for experiment. While the stimuli were created from synthesized speech for the pilot study, this dissertation, in order to examine speech perception in natural settings, only used natural spoken recordings as stimuli.

## 4.2 Participants

A total number of 43 participants were recruited for the experiments of this dissertation. A summary of the participants by age group and gender is provided in Table 4.1. For the convenience of discussion in later chapters, the participants are coded according to their age, gender, and fabricated last name which will be illustrated later in Figure 1.

| | 19 and under | 20-29 | 30-39 | 40-49 | 50 and older | Total |
|---|---|---|---|---|---|---|
| Male | 1 | 2 | 6 | 4 | 4 | 17 |
| Female | 2 | 7 | 4 | 6 | 7 | 26 |
| Total | 3 | 9 | 10 | 10 | 11 | 43 |

Table 4.1 Number of participants by gender and age group (n = 43)

The participants are all related either closed or remotely to one particular family, the TUV family. They were referred by the core members of the family, which includes the father (M58-TUV) who is a university professor, a stay at home mother (F57-XYZ), and two daughters who are both pursuing a graduate degree (F23-TUV and F29-TUV). As shown in the following Figure 4.1, the four members of the TUV family are at the

center of the participant relationship map.

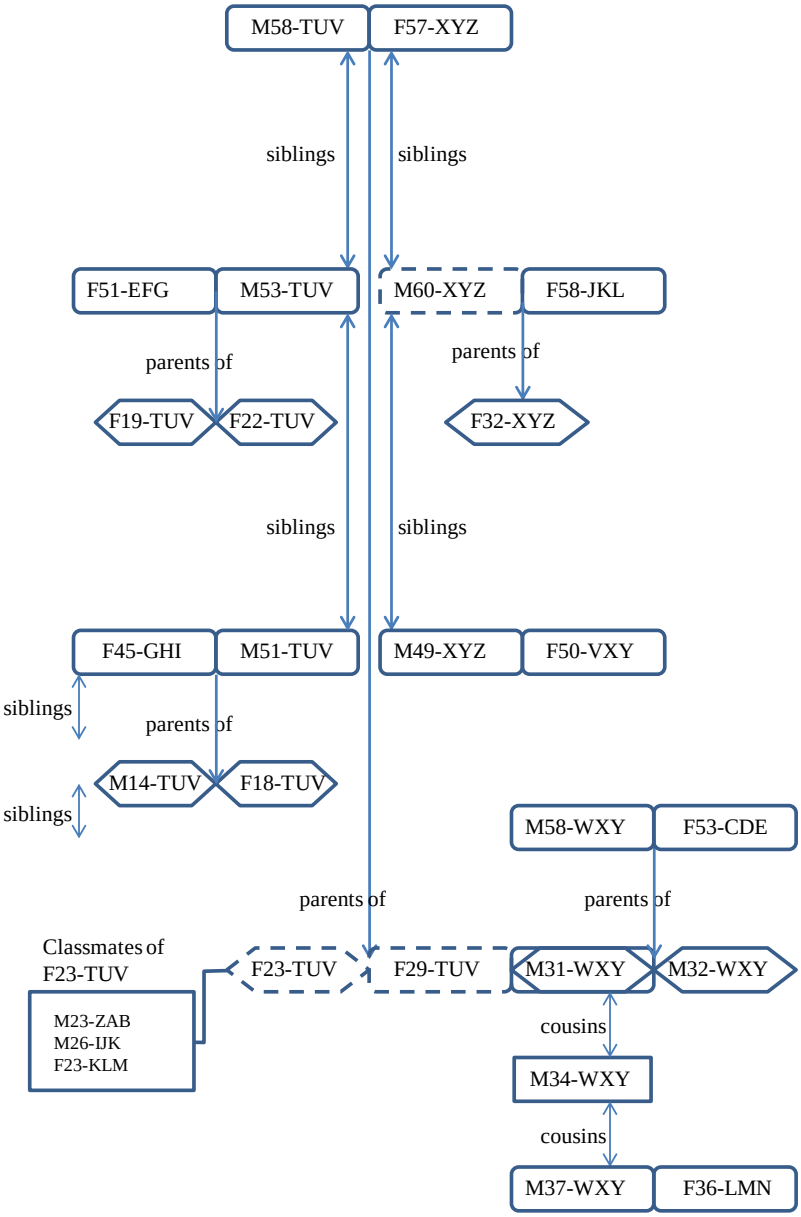



Figure 4.1 Participant relationship map

In the above relationship map each participant is coded by gender, age, and fabricated last name, and represented by a block of a particular shape to illustrate the participant's relationship with one another. Two connected rounded-corner rectangles

indicate married couples, two connected hexagons indicate siblings if not marked by arrow lines, and dotted outlines indicate non-participants who are important nodes to the relationship map and worth mentioning. Take the core TUV family for example, the father (M58-TUV) is a fifty-eight-year-old male speaker married to a fifty-seven-year-old female XYZ. As indicated by a single-head arrow, the couple are the parents of both a twenty-three-year-old female and a twenty-nine-year-old female both of who have TUV as last name, and they are naturally siblings. Neither of the TUV sisters from the core family participated in the official experiments, yet both played an important role in experiment design and held a place in maintaining the integrity of the relationship map.

M58-TUV, a university professor, referred eight of his advisees to participate in the experiments, two males and six females as shown in the block extended from his. M58-TUV also introduced two of his brothers, M53-TUV and M51-TUV, and their family to participate in the experiments. M58-TUV's two sister-in-laws further expanded the network by introducing nine of her (F51-EFG) colleagues, and two of her (F45-GHI) sisters to take part in the experiments. F57-XYZ on the other hand, referred two of her brothers' family to participate in the experiments (even though one brother, M60-XYZ, did not himself participate). F23-TUV asked three of her classmates to participate, and finally F29-TUV invited her husband M31-WXY and his immediate as well as extended family members to contribute to the experiments of this dissertation.

## 4.3 Experiment design for speech perception tasks

The speech perception tasks includes four sections with the first three being word-identification tasks and the fourth a speaker-identity-observation task. The details of the word-identification tasks will be introduced first and followed by an explanation of the speaker-identity-observation task.

### 4.3.1 Word-identification tasks with photographic and orthographic stimuli

The word-identification tasks include three subsections with regard to the application of three visual stimuli: photographic, traditional Chinese characters, and simplified Chinese

characters. All participants completed the three tasks in the same order of procedure by seeing first the traditional Chinese characters stimuli, followed by the photographic stimuli, and finally the simplified Chinese characters stimuli. The lexical items and audio stimuli are identical in all three visual stimuli, yet randomized differently in sequential order.

#### 4.3.1.1 Lexical items

The lexical items employed for the word-identification tasks were selected through the processes of improvisation, word frequency inquiry, and focus group discussion. Three first-language speakers of Taiwan Mandarin including F29-TUV gathered to improvise ten minimal pairs of two-character words for each retroflex/dental sibilant pair and five minimal pairs of two-character words for each of the four pairs of filler sounds that were known to also cause some degree of confusion for certain speakers of Taiwan Mandarin: /f, x/, /n, l/, /i, y/, and /p, $p^h$/. The fifty pairs of words were carefully chosen in a way that the target phonemes were in the same position of the two-character words as well as carrying the same tone. Also, whenever possible, the exact same character among the many homophones was selected as the character with the non-target phoneme for both words in each minimal pair.

These fifty minimal pairs, a total of one hundred two-character words, were then checked for their frequency against the Word List with Accumulated Word Frequency in Sinica Corpus[5]. There is a variety of word lists to choose from based on different Sinica Corpora, for example the word list from the Pre-modern Chinese Corpus, the Old Chinese Corpus, and the 300 Tang Poems. The word list employed for the purpose of this study is the word list based on The Academia Sinica Balanced Corpus of Modern Chinese or the Sinica Corpus, for short. The Sinica Corpus is a 5,000,000 word corpus of modern Chinese that was first developed in 1990, and is currently at its 4.0 version freely available to the public. The data compiled into this corpus is collected between 1990 and 1995. It is a segmented and tagged corpus where all texts were collected from the Taiwan

[5] The Mandarin Chinese version of the website was consulted, and data was retrieved from http://elearning.ling.sinica.edu.tw/CWordfreq.html on May 11, 2013.

Mandarin dialect and are written in traditional Chinese characters.

The data include both written and spoken texts. Written texts account for 90% of the corpus and they were collected from a wide range of genres, including news reports, critiques, advertisements, letters, novels, stories, diaries, guidance handbooks, poetry, play scripts, etc. The topics covered philosophy, natural science, social science, history, art, modern literature, and everyday life. Spoken texts account for 10% of the corpus and they were collected from sources such as spoken monologues and conversations in TV and radio programs, private conversations, and dialogues from meetings. The texts were categorized according to five criteria: genre, style, mode, topic, and source. Hence, the Sinica Corpus is intended to be a representative sample of modern Chinese language in Taiwan.

The goal of consulting the frequency word list of these one hundred two-character words was to narrow the minimal pair candidates for target phonemes from ten to six pairs and filler phonemes from five to three pairs. Since the words were given a numeral ranking in the frequency word list, the selection criteria was to add the value of the two words in the same pair together, which would indicate a general impression of the familiarity of the two words when shown simultaneously. The minimal pairs with the largest four and three sum values of target phonemes and filler phonemes respectively were eliminated from the candidate list of words that were presented to a focus group in the last process.

A focus group of five male and five female, ages between 23 and 26, were gathered and asked for their opinions toward the familiarity of the remaining thirty minimal pairs. They first each filled out a questionnaire asking them to rank the familiarity of the six minimal pairs from each target phoneme and the three pairs from each filler phoneme. They were specifically asked to consider both words in the pairs simultaneously instead of basing their decision solely on the more familiar word of the two. The survey was subsequently followed by a one hour discussion, discussing their impression of each pair with focus on (1) whether all the words were semantically meaningful to them, (2) whether one word of a pair was overtly more frequently used than the other, (3) how likely they were to misidentify one word for the other in the same pair during audio perception, (4) and whether they would suggest alternative words for

the pairs. According to the survey results and feedbacks from the discussion, the number of minimal pairs for the target phonemes was further eliminated from six to three and filler phonemes from three to two. There was one filler phoneme pair /n, l/ that ended up with only one minimal pair because many participants of the focus group mentioned that they rarely confused this pair of phonemes.

The final set of lexical items to be tested throughout all three word-identification tasks in the speech perception experiments is provided in Table 4.2. There were a total of sixteen pairs, nine target phoneme pairs and seven filler phonemes pairs. Among the seven filler phoneme pairs, the two pairs from /p, $p^h$/ were used as example trials for all three word-identification tasks. These two pairs each occurred once at the very beginning of the tasks to demonstrate the procedure of the experiment and test the software and equipment in use. The remaining fourteen pairs each occurred twice in each task but appeared in a slightly different visual layout as explained in the next section 4.3.1.2. Therefore, all participants completed thirty trials, including the two example trials. The occurrence of the pairs of words was also accompanied by different audio stimuli, each word was played once throughout each task. The creation of the audio stimuli is introduced in section 4.3.1.3.

**Target consonants minimal pairs**

| No. | Phoneme IPA / *Pīnyīn* | A | B | C |
|---|---|---|---|---|
| 1 | /tʂ/ *zh* | 知識 *zhīshí*<br>'knowledge' | 治理 *zhìlǐ*<br>'to govern' | 致意 *zhìyì*<br>'to give one's regard to' |
| 2 | /ʦ/ *z* | 姿勢 *zīshì*<br>'posture' | 自理 *zìlǐ*<br>'to take care of oneself' | 恣意 *zìyì*<br>'unbridled' |
| 3 | /tʂ$^h$/ *ch* | 延遲 *yánchí*<br>'to postpone' | 池塘 *chítáng*<br>'pond(s)' | 赤字 *chìzì*<br>'a deficit' |
| 4 | /ʦ$^h$/ *c* | 言詞 *yáncí*<br>'word(s)' | 祠堂 *cítáng*<br>'ancestral shrine(s)' | 刺字 *cìzì*<br>'to tattoo (a) / word(s)' |
| 5 | /ʂ/ *sh* | 詩人 *shīrén*<br>'poet(s)' | 使節 *shǐjiē*<br>'ambassador(s)' | 相識 *xiàngshí*<br>'to be acquainted with' |

| | | | | |
|---|---|---|---|---|
| 6 | /s/ *s* | 私人 *sīrén* ‘private’ | 死結 *sǐjié* ‘fast knot(s)’ | 相似 *xiàngsì* ‘to similar with’ |

**Filler consonants/vowels minimal pairs**

| | | A | B |
|---|---|---|---|
| 7 | /f/ *f* | 發生 *fāshēng* ‘to happen’ | 發展 *fāzhǎn* ‘to develop’ |
| 8 | /x/ *h* | 花生 *huāshēng* ‘peanut(s)’ | 花展 *huāzhǎn* ‘flower exhibition’ |
| 9 | /i/ *yí* | 猶疑 *yóuyí* ‘to hesitate’ | 權益 *quányì* ‘rights and interests’ |
| 10 | /y/ *yú* | 魷魚 *yóuyú* ‘squid(s)’ | 痊癒 *quányù* ‘to fully recover’ |
| 11 | /n/ *n* | 泥巴 *níbā* ‘mud’ | |
| 12 | /l/ *l* | 籬笆 *líbā* ‘a fence’ | |

**Example consonants minimal pairs**

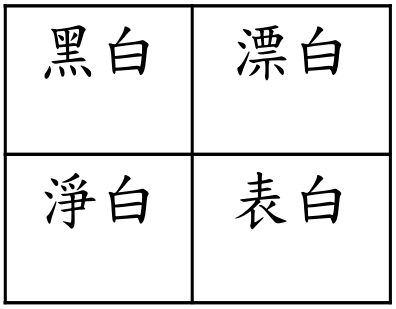
| | | A | B |
|---|---|---|---|
| | /p/ *b* | 表白 *biǎobái* ‘to confess’ | 棒子 *bàngzǐ* ‘stick(s) / club(s)’ |
| | /$p^h$/ *p* | 漂白 *piāobái* ‘to bleach’ | 胖子 *pàngzǐ* ‘an obese person’ |

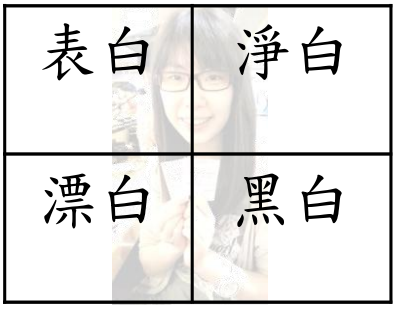
Table 4.2 Lexical items

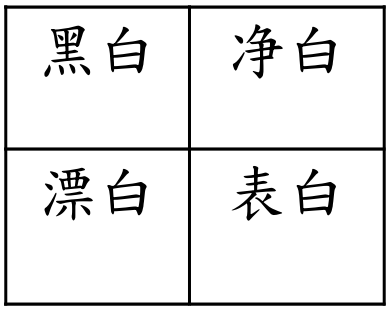
#### 4.3.1.2 Visual stimuli

The word-identification experiments were based on three different stimuli and all participants proceeded in the following order, traditional Chinese character, photographic stimuli with traditional Chinese character, and simplified Chinese character. The three tasks were conducted by utilizing Microsoft Office Powerpoint 2007. On the screen during tasks, participants saw a two-by-two table on a white-background slide for each trial as shown in Figure 4.2. In the cells of each table, there were the two words from the minimal pair to be tested and two additional words filling in the other two cells of the table. The same four-word set appeared twice in each task, once for each word in the same minimal pair, but the allocation of the words randomly changed to different cells in each appearance.

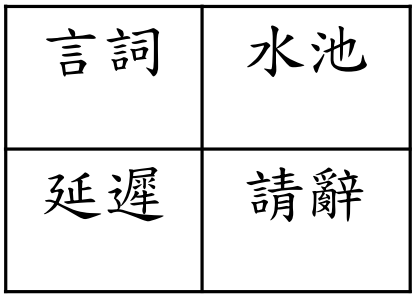
a.




b.






(1) Traditional-character stimuli (2) Photographic stimuli (3) Simplified-character stimuli

Figure 4.2 Examples of visual stimuli

As seen in the two sets of examples illustrated by Figure 4.2, the two additional words were either the exact same character as the non-target character, 白 *bái* 'white' in set (a.), or homophones of the target character, 水池 *shuǐchí* 'water pool' and 請辭 *qǐngcí* 'to resign' as in set (b.). The goal of this design was to prevent the participants from easily detecting the actual purpose of this study since they were told in the instructions that they were providing feedback on the clarity of an artificial machine voice under construction. Before launching each task, participants were asked to read through the same set of instructions. The instructions were written in the corresponding characters of the tested stimuli and traditional Chinese characters were used for giving instructions in the photographic stimuli. The intention was to emphasize the different types of orthography used in the experiments and intentionally activate their sociophonetics

knowledge of dialectal variations associated with the two types of Chinese characters if they were aware of them.

All participants completed the traditional-Chinese-character task first because it is the most natural setting for speakers of Taiwan Mandarin. It was predicted that the results from this particular task should be able to most authentically represent participants' perception of retroflex and dental sibilants in daily situations. The task with traditional Chinese characters also served as the control group of the experiments, while the simplified-Chinese-character task and the task with photographic stimuli were the treatment group of this series of experiments. The pictures of the photographic stimuli were shown as watermarks in the middle and beneath the two-by-two table of words as demonstrated in Figure 4.2. The original photos shown during the task were of three females at the age of early twenties, late twenties, and early fifties as seen in Figure 4.3. It was purposefully designed that the two words from the same minimal pair were stimulated by the same photo. For the three trials of each of the six target phonemes, 3 retroflex sibilants and their corresponding dental sibilants, all three photos occurred once as stimulus. The function of this treatment group was to examine whether the speaker's age had an effect on the choice of phonemes. The participants were asked to estimate the age of each female of which they saw photos after they completed the photographic stimuli task before continuing on to the simplified-Chinese-character task.


Early twenties


Late twenties


Early fifties

Figure 4.3 Photos adopted for the photographic stimuli task

#### 4.3.1.3 Audio stimuli

The audio stimuli for the speech perception experiments were recorded from a 29-year-old female who speaks Taiwan Mandarin as a first language. The sounds were recorded with an Edirol R-09 at a sampling rate of 44.1/48 kHz in WAV format. The speaker read through a word list of each word from all minimal pairs three times at a speed of approximately one word (two syllables) per second. The trial with the best clarity for each word was selected as the audio stimuli for the corresponding word. The stimuli were clipped out from the original recorded sounds and were not processed or synthesized in any way. The stimulus of each word occurred once through each of the three word-identification tasks in randomized order, but structured in a pattern that one filler word appeared after every two target words. The audio stimulus for each slide was designed to automatically play once after a 'beep' sound when the participants proceeded to the next Powerpoint slide

### 4.3.2 Speaker-identity-observation task design with orthographic stimulus

The objective of the speaker-identity-observation task was to examine Taiwan Mandarin speakers' perceptive and ideology toward both the Taiwan Mandarin dialect and the Beijing Mandarin dialect. The participants heard a speaker recite an excerpt from a news passage while reading the passage from a monitor simultaneously. They were then interviewed about their impression on the speaker with regard to the speaker's basic demographic information and other observations. They were also asked questions that were rather irrelevant but served as filler questions in order to disguise the real intentions of this experiment. A list of basic questions that all participants were asked to answer is provided in Appendix 1. In additional to the standard questions, follow-up questions from a great variety of aspects were inquired based on each participant's response to the basic questions.

The passage read and shown to the participants was an excerpt from a piece of

news published in 南方人物周刊 Nánfāng Rénwù Zhōukān ‘Nanfang People Magazine’ on November 22, 2011. The original title of the article was 英国学者看鸦片战争 Yīngguó Xuézhě Kàn Yāpiàn Zhànzhēng ‘An English Scholar’s Perspective on The Opium War’. The first section of the article, approximately 750 words, was extracted and used as the material for the speaker-identity-observation task (Appendix 2). This topic is of an important event in the Chinese history and well-emphasized in all textbooks of History in the Taiwan education system. It is also a topic that does not indicate any specific locations in modern Taiwan or China, and thus would not be inappropriate for a news anchor from either location to report on such an historical event. The article was originally published in simplified Chinese characters and transformed into traditional Chinese characters (Appendix 3) so that both versions were available for the purpose of this study.

The audio recordings were recorded by two female speakers, one native Taiwan Mandarin speaker and one native Beijing Mandarin speaker. Both speakers were in their late twenties at the time of recording. The Taiwan Mandarin speaker was born and raised in Taipei and also speaks nearly conversable proficiency of Southern Min. The Beijing Mandarin speaker was born and raised in Shandong Province until 18, by which she moved to Beijing to attend college. She is able to speak fluently in both the Shandong dialect and the standard Beijing Mandarin dialect. She was specifically asked to recite the passage in the standard Beijing Mandarin dialect. The speakers were asked to read the passage in their most natural voice and tone without intentionally impersonating the intonation of a news anchor. The recording device was also Edirol R-09 as introduced in the previous section.

The center of gravity measurements (COG) of both speakers were obtained from the retroflex sibilants and alveolar sibilants that occurred in the extracted passage. COG values were taken from a total of 130 retroflex phonemes and 29 alveolar phonemes. The mean COG of Beijing Mandarin speaker’s retroflex sibilants was 4132 Hz and alveolar sibilants was 8904 Hz. As for the Taiwan Mandarin speaker, the mean COG of retroflex sibilants was 5276 Hz and alveolar sibilants was 9135 Hz. The Beijing Mandarin speaker made a greater distinction between these two sets of sibilants with a mean difference of 4772 Hz, while the Taiwan Mandarin speaker showed a smaller mean COG difference of

3859Hz.

Since there were two versions of the passage, one traditional and one simplified characters, and two dialects of recitation of the passage, four combinations of the passage and audio clip were possible. Therefore, the participants were divided into four groups according to the combination of passage and audio clip they were presented with: traditional-character passage with Taiwan Mandarin audio clip, traditional-character passage with Beijing Mandarin audio clip, simplified-character passage with Beijing Mandarin audio clip, and simplified-character passage with Taiwan Mandarin audio clip. The two combinations with discordant characters and audio clip were of greater interest in this experiment; therefore, in cases where there were less than four people in a certain age-gender group, the conflicting settings were presented to participants with priority.

## 4.4 Experiment design for speech production tasks

The goal of the speech production study was to compare speakers' production of Mandarin retroflex sibilants under different formality settings and priming with different Chinese characters. Two types of formality settings were tested in the speech production experiments in this dissertation, story-telling and passage-reading, with the assumption that passage-reading is more formal than story-telling. Each type was tested twice to collect the necessary data, and thus participants completed four tasks in total for the speech production study of this dissertation.

### 4.4.1 Story-telling tasks

It is hypothesized that story-telling is one of the settings in which participants are able to speak naturally. Hence, among the series of production experiments for this dissertation, two story-telling tasks were designed to collect natural spoken data from the participants. The tokens collected from these two tasks served as the control group for all collected spoken data since they best represent the speakers' characteristics of speech production in daily situations. All participants were provided with two animation clips that were between two to three minutes long, in which the characters did not conduct any

conversation and thus no verbal stimuli was given in either spoken or written format.

Since the main purpose was to collect tokens of retroflex sibilants spoken in a natural setting, the participants were asked to describe the content of the series of events that took place in the animations they saw after watching each clip on a monitor. The two animation clips were composed of main characters that needed to be described by using the retroflex sibilants, /tʂ/ and /tʂ$^{h}$/ in the first clip and /ʂ/ in the second clip. The main characters in the first clip are a spider and some bugs that were stuck on a web and gradually eaten one by one. The Mandarin noun for spider is 蜘蛛 *zhīzhū*, in which the initials of the two-syllable word are both /tʂ/, and the Mandarin noun for bugs (larvae) is *chóngzǐ*, in which the first character of the word begins with /tʂ$^{h}$/. This animation, titled *Spider*, is one episode of the comedy television series, *Larva*, that were computer-animated by Tuba Entertainment in Seoul, South Korea. *Larva* is mainly concentrated on the adventures and events encountered by two larvae. This particular episode is the 72$^{nd}$ episode from season 1[6] that was released to the public in 2011 and available on Youtube.

The second clip presented to the participants was a cartoon of which the main storyline focused on a competition between a lion and a leopard that were both trying to chase and capture a turtle. The Mandarin noun for lion is 獅子 *shīzǐ* in which the initial consonant of the first syllable is /ʂ/. As the participants tried to tell the story they inevitably needed to use the noun frequently to describe the development of the story. This animation was chosen from a French series of animation produced by Studio Hari for children's television that was released in 2009. The name of the series is *Leon*, also known as *Léon, (t)erreur de la savane*, which is the name of a funny lion living in the African savannah. The series features his interactions with other animal characters living in Africa. The particular episode shown to the participants was the 26$^{th}$ episode titled 4 × 400 m[7], which indicates four 400-meter dashes between the lion and a leopard in the attempt to chase and capture a turtle.

The two clips were each shown once to the participants and they were asked to

[6] This episode is officially available on Youtube at the URL https://www.youtube.com/watch?v=uRQpL65-6FQ&index=72&list=PLxrv_vTPuZ6Qi6Hx868jli1m5CDCEDfxj

[7] This episode is available on Youtube at the URL https://www.youtube.com/watch?v=_qiTBo_i3_0&index=19&list=PL9L-6-P3Ey8mTq-p8BdhYZ45kUyXbzrNB

tell the story of what they saw immediately after watching each clip. The participants were encouraged to include as many details as possible. Since the animations did not feature any dialogue, they were able to freely describe what they believed to have happened in the cartoons. The story-telling process was also recorded with Edirol R-09.

### 4.4.2 Passage-reading tasks

Participants were asked to recite two passages for the passage-reading tasks, the first one presented in traditional Chinese characters and the other presented in simplified Chinese characters (see Appendix 4 and 5). These two passages were composed by the researcher in the attempt to include as many retroflex sibilants as possible but also using some alveolar sibilants while maintaining the meaningfulness of the passages. The participants were told to read the passages once before recording. Each passage with only 153 characters (traditional) and 159 characters (simplified), which could be recited within one minute at normal to slower reciting speed.

## 4.5 General procedure

The whole process of the experiment series comprised one survey, four production tasks, four perception tasks, and an interview. All participants completed a language background survey that asked questions such as where they were born, how many languages they spoke, and how often they read documents of simplified Chinese characters or heard Beijing Mandarin. A full list of questions in the original Mandarin survey is provided in Appendix 6 along with its translation in Appendix 7. After filling out the background survey, they proceeded to the four production tasks in the sequence of first watching and telling the stories of the cartoons followed by reading the passages presented in traditional Chinese characters and subsequently in simplified Chinese characters.

The perception study followed the production study with four tasks, three word-identification tasks and one speaker-identity-observation task. As previously mentioned, they completed the word identifications tasks in the order of traditional-character

stimulus, photographic stimulus, and finally simplified character stimulus. The last task of the whole experiment, the speaker-identity-observation task, is where the participants were divided into four groups according to the version of passage and speaker of the audio clip with which they were presented. The four combinations were traditional-character passage with Taiwan Mandarin audio clip, traditional-character passage with Beijing Mandarin audio clip, simplified-character passage with Beijing Mandarin audio clip, and simplified-character passage with Taiwan Mandarin audio clip. The series of experiments conducted for this dissertation concluded with general interview questions related to the speaker-identity-observation task and inquiries of further comments or opinions toward various aspects of this study, including but not limited to participants' awareness and ideologies of the usage of different written characters and sociophonetic variations of the retroflex sibilants in Mandarin Chinese.

# Chapter 5

## 5. Perception study

The perception study of this dissertation is consisted of two types of experiments: word-identification task and speaker-identity-observation task. The word-identification task is repeated three times but varies in the co-occurring stimuli, traditional Chinese characters, simplified Chinese characters, and photographs. This chapter begins with describing the method and process of data analysis of the word-identification tasks followed by explaining the findings from the analysis. It will then proceed to describing the results from the speaker-identity-observation task. The description of the results was focused on the participants' age and gender in order to compare the results from the perception study with those of the production study[8]. Finally, a summary of findings from the perception study is provided in the last section of this chapter.

### 5.1 Data analysis and results of word-identification tasks

All three word-identification tasks were composed of 18 target phonemes trials and 10 filler phoneme trials. Each of the 6 target phonemes, /tʂ/, /tʂ$^{h}$/, /ʂ/, /ts/, /ts$^{h}$/, and /s/, was tested three times, and thus 18 trials in total. The participants were asked to identify the word they heard during each trial. The responses from all participants were valid, and thus results from all 43 participants, 17 male and 26 female, were included in the analyses. A brief summary of the overall correct rate according to age and gender group is provided in Table 5.1. As shown in the table 14 participants, 3 male and 11 female, scored a 100 % correct rate for all trials. The lowest correct rate is between 60-69%, in which includes two male and one female participant from the oldest age group, 50 and older. Table 5.2 which only includes results from the target phonemes is also provided for comparison since there were two participants who each made one mistake in a filler trial,

[8] Greater difference in gender was found in the production study. Hence, to systematically compare the results between the perception and production study, gender difference was also discussed for the perception study even though it was not a major finding in the perception study of this dissertation.

but both received full marks in all the target trials.

| Participant's age and gender | Number of correct trials from all word-identification tasks including both target and filler phonemes<br>Total trials = 28*3 = 84 | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | 50-57 (60-69%) | 58-66 (70-79%) | 67-74 (80-89%) | 75-83 (90-99%) | 84 (100%) | Total |
| | 19 and under | male | | | | | 1 | 1 |
| | | female | | | | 1 | 1 | 2 |
| | 20-29 | male | | | | 1 | 1 | 2 |
| | | female | | | | 4 | 3 | 7 |
| | 30-39 | male | | 3 | | 3 | | 6 |
| | | female | | | | 1 | 3 | 4 |
| | 40-49 | male | | | | 3 | 1 | 4 |
| | | female | | | | 2 | 4 | 6 |
| | 50 and older | male | 2 | 1 | 1 | | | 4 |
| | | female | 1 (50) | 2 | | 4 | | 7 |
| | Total | male | 2 | 4 | 1 | 7 | 3 | 17 |
| | | female | 1 | 2 | 0 | 12 | 11 | 26 |

Table 5.1 Summarized result of all word-identification tasks

| Participant's age and gender | Number of correct trials of target phonemes from all word-identification tasks<br>Total trials = 18*3 = 54 | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | 27-31 (50-59%) | 32-36 (60-69%) | 37-42 (70-79%) | 43-47 (80-89%) | 48-53 (90-99%) | 54 (100%) | Total |
| | 19 and under | male | | | | | | 1 | 1 |
| | | female | | | | | 1 | 1 | 2 |
| | 20-29 | male | | | | | 1 | 1 | 2 |
| | | female | | | | | 3 | 4 | 7 |
| | 30-39 | male | 1 (30) | 2 | | 1 | 2 | | 6 |
| | | female | | | | | 1 | 3 | 4 |
| | 40-49 | male | | | | | 2 | 2 | 4 |
| | | female | | | | | 2 | 4 | 6 |
| | 50 and older | male | 1 (30) | 2 | 1 | | | | 4 |
| | | female | | 3 | | 1 | 3 | | 7 |
| | Total | male | 2 | 4 | 1 | 1 | 5 | 4 | 17 |
| | | female | 0 | 3 | 0 | 1 | 10 | 12 | 26 |

Table 5.2 Summarized result of target phonemes from all word-identification tasks

As seen in Table 5.2, 16 out of 43 participants (37.2%) made no mistakes on identifying the target phonemes and another 15 made less than six mistakes on all 54 target phoneme trials across the three word-identification task. These 32 participants are from the younger age groups and female participants showed a higher tendency of achieving a higher correct identification rate than male participants. 22 out of 26 female (85%) and 9 out of 17 male (53%) participants scored an over 90% correct rate when identifying retroflex and alveolar sibilants in the three word-identification tasks. The relative position of these participants in the relationship map is illustrated in Figure 5.1. The blocks in blue indicate full marks and green a 90% correct rate.

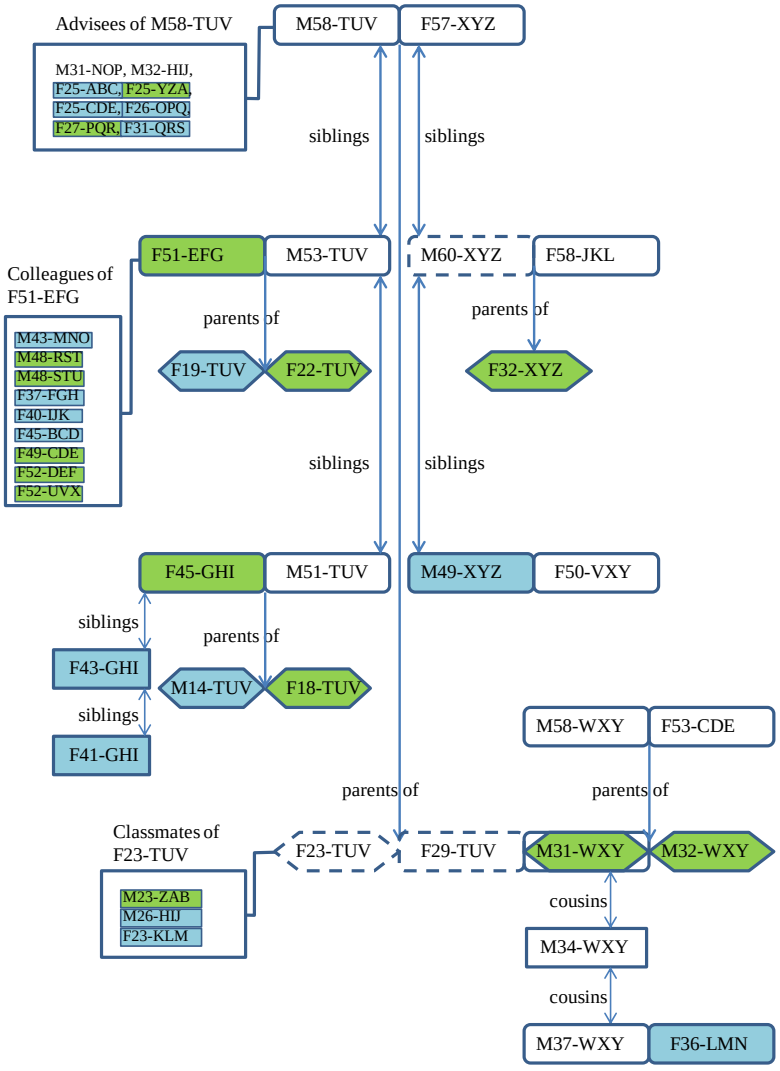



Figure 5.1 The 32 participants who received an over 90% correct rate on target phonemes

in all tasks

Two points are worth pointing out from this figure: (1) all participants under fifty years old scored an over 90% correct rate except for four male participants, and (2) only three female participants over fifty were highlighted in this figure. The four male participants under forty are M34-WXY, M37-WXY, M31-NOP, and M32-HIJ. The explanation for M34-WXY and M37-WXY is rather straight forward since they claimed in their language profile that they only use Mandarin for 50% and 20% of the time respectively. The other dominant language used is the Min dialect, in which retroflex consonants are not present in the sound inventory. Min is much more popular and frequently used for daily communications than Mandarin Chinese in the southern part of Taiwan, which is where they have lived for the majority of their lifetime, 32 years for both of them. M34-WXY said, of non-Mandarin usage, he uses Min for 40% of the time, while the other 10% is English and for work only. M37-WXY said Min accounts for 80% of the language spoken in everyday life and Mandarin is only employed when he meets people whom virtually speak no Min. Therefore, considering the facts that there are no retroflex consonants in the Min dialect, the frequency Min is used by these two, and their surrounding language environment, it is not unexpected that the word-identification tasks focused on differentiating retroflex and alveolar sibilants were reasonably difficult for them.

As for the other two male participants who are both members of the group of advisees of M58-TUV, their situation is somewhat similar with the WXY cousins, but their usage of Min in daily life is relatively lower. M31-NOP was born and raised in New Taipei City, a city in the northern part of Taipei where Mandarin is generally more frequently used than Min. In his language profile, he claimed that he speaks Min for about 35% of the time, while the other 65% is divided between Mandarin (50%) and English (15%). Even though 35% is less than 40% and a lot less than the 80% comparing with the WXY cousins and he lives in the northern part of Taiwan, it is already higher than his peers who typically use Min for less than 20% of the time and Mandarin for more than 70%. M32-HIJ, on the other hand, was born and raised in Kaohsiung, a city in the southern part of Taiwan. He lived there until nineteen years old, which he thereafter moved to Taipei. Although he wrote in his language profile that Min currently only

accounts for 10% of his daily language use, he did began speaking Min at age zero and still self-evaluated himself as speaking Min at the highest level of proficiency on a scale from 1 to 7, which was rarely claimed by other participants who do not currently use Min as frequently but also speak Min as a first language.

The three female participants who are over fifty years old but were still able to complete the three tasks with an over 90% correct rate are F51-EFG and her colleagues F52-DEF and F52-UVX. The three of them work at the Ministry of Science and Technology (MOST), a governmental institute which requires passing a civil service exam to receive an offer. This particular exam is well-known or even notorious for its low passing rate, and Mandarin Chinese is one of the tested subjects. Being able to pass the exam typically indicated that you have a higher basic knowledge of Mandarin Chinese, and generally spoke Mandarin at a close to the so-called ‘standard’ proficiency, by which the differences between retroflex and alveolar sibilants are expected to be realized in speech production and recognized in speech perception.

In addition to the working environment that contributed to their high rate of correctness in the word-identification tasks, their relatively low self-evaluated proficiency of Min and low frequency of use of Min should also be considered when rationalizing how their experiment results were comparably different from other participants who are fifty and older. All three of them were born and raised in the north or mid-cities of Taiwan. They all came from families that spoke little to no Min at home. F51-EFG claimed in her language profile that she started learning Min at age five, yet she only rated herself as speaking Min at level 1 proficiency. F52-DEF did not start using Min until eighteen and self-claimed to speak Min at level 3 proficiency. As for F52-UVX, though she did not learn Min until twenty-five, she evaluates herself as speaking Min at level 5 proficiency. This may be due to the fact that she is the one who uses Min the most frequently among the three. F52-UVX stated that she uses Min for about 10% of the time in her current daily life, which is already higher than the 1% of F52-DEF and 0% of F51-EFG. In sum, the three of them shared a similar language usage profile that is very different from the other four female participants who are also over fifty; therefore, their notably higher correct rate on the word-identification tasks is reasonably understandable.

With 31 participants out of a total of 43 scoring a higher than 90% correct

identification rate, it seems like neither the photographic stimuli, nor the orthographic stimuli had any or minimum effect on 72% of all participants. However, a closer look with a chi-square test showed that even though the photographic stimuli did not have a significant effect on the perception of Mandarin retroflex and alveolar sibilants, a difference in the orthographic stimuli did play a statistically significant role on affecting how Taiwan Mandarin speakers perceive these two series of consonants. The analysis and result from the word-identification task with photographic stimuli is provided in the next section 5.1.1 and following that is the section explaining how a variation in the type of Chinese character seen is associated with the consonant series perceived.

### 5.1.1 Photographic stimuli

The word-identification task with photographic stimuli used photos of three females representing different age groups, early twenties, late twenties, and early fifties. All three photos occurred once as stimulus for the three trials of each of the six target phonemes, 3 retroflex sibilants and their 3 corresponding alveolar sibilants. It was purposefully designed that the two words of the same minimal pair were stimulated by the same photo. Table 5.3 lays out how the photos were distributed throughout the tasks for target phonemes. The phonemes were coded from 1 to 6 and their three minimal pairs were coded from A to C for the ease of discussion and data sorting. 1A for example, was one of the trials for /tʂ/ and it represented the target word 知識 *zhīshí* ‘knowledge’; 2A indicated the its minimal pair or its corresponding alveolar sibilant was /ts/ and was represented by the word 姿勢 *zīshì* ‘posture’, which was tested alongside 1A. When either 1A or 2A was the target word of trial, the co-occurring background photo shown was the one of a female in her late twenties.

| Sound / Trial | Retroflex sibilants | | | Alveolar sibilants | | |
|---|---|---|---|---|---|---|
| | 1 /tʂ/ | 3 /tʂʰ/ | 5 /ʂ/ | 2 /ts/ | 4 /tsʰ/ | 6 /s/ |
| A | late 20s | early 50s | late 20s | late 20s | early 50s | late 20s |
| B | early 50s | early 20s | early 20s | early 50s | early 20s | early 20s |
| C | early 20s | late 20s | early 50s | early 20s | late 20s | early 50s |

Table 5.3 Two-way specification table of the photographic stimuli in the word-identification task

The purpose of the photo stimuli was to reproduce a similar setting as the one applied in Hay et al. (2006b), in which researchers investigated the perception of diphthongs /iə/ and /eə/ in New Zealand English. By measuring the participants' correct identification rate, it was found that the rates were affected by the photos. The result from Hay et al. (2006b) showed that participants were less likely to distinguish /iə/ and /eə/ when the photo represented a person belonging to the social group that is more likely to merge the diphthongs in reality. The hypothesis supporting the design of the current experiment is that participants would associate the merging of retroflex and alveolar sibilants with females of older age and vice versa. Therefore, with two photos representing females of a younger age, early and late twenties and only one photo representing a female of an older age, it was assumed that participants would show an overall higher tendency of identifying a retroflex-phoneme-target-sound with an actual retroflex-phoneme-target-word instead of a alveolar-phoneme-target-word.

Before explaining in details the result from each phoneme, a general overview of the correct rate is provided. Since the photo stimuli task was also conducted with traditional Chinese characters, it was considered as a treatment group to be compared to the control group, which was the word-identification task in traditional Chinese characters but with no additional stimuli. A side by side comparison of the correct rates from these two tasks according to age and gender is provided in the following Table 5.4 and Table 5.5. Participants generally achieved a higher correct identification rate with the photographic stimuli in comparison with the traditional-Chinese-character task. There were four more female participants (23) who made fewer than two mistakes in the photographic stimuli task compared to those in the traditional-Chinese-character task (19). While all participants scored an at least 50% correct rate with the photo stimuli, there were three participants, two male and one female, who scored below 50% with the traditional-Chinese-character only stimuli. There were four female participants from the group 50 and older who completed the task with less than two mistakes in the photo stimuli task. Among them, besides the three that were discussed in the previous section, F50-VXY also performed well in this task.

| Participant's age and gender | | Number of correct trials from word-identification task with photographic stimuli (target phonemes only) Total trials = 18 | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | 9-10 (50-59%) or under | 11-12 (60-69%) | 13-14 (70-79%) | 15-16 (80-89%) | 17 (90-99%) | 18 (100%) | Total |
| 19 and under | male | | | | | | 1 | 1 |
| | female | | | | | 1 | 1 | 2 |
| 20-29 | male | | | | | | 2 | 2 |
| | female | | | | | 3 | 4 | 7 |
| 30-39 | male | | 3 | | 1 | | 2 | 6 |
| | female | | | | | 1 | 3 | 4 |
| 40-49 | male | | | | | 2 | 2 | 4 |
| | female | | | | | | 6 | 6 |
| 50 and older | male | 2 (9) | 2 | | | | | 4 |
| | female | 1 (7) | 2 | | | 3 | 1 | 7 |
| Total | male | 2 | 5 | 0 | 1 | 2 | 7 | 17 |
| | female | 1 | 2 | 0 | 0 | 8 | 15 | 26 |

Table 5.4 Summarized result of word-identification task with photographic stimuli (target phonemes)

| Participant's age and gender | | Number of correct trials from word-identification task with traditional-Chinese-character stimuli (target phonemes only) Total trials = 18 | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | 9-10 (50-59%) | 11-12 (60-69%) | 13-14 (70-79%) | 15-16 (80-89%) | 17 (90-99%) | 18 (100%) | Total |
| 19 and under | male | | | | | | 1 | 1 |
| | female | | | | | 1 | 1 | 2 |
| 20-29 | male | | | | | 1 | 1 | 2 |
| | female | | | | 2 | | 5 | 7 |
| 30-39 | male | 1 (8) | 2 | | 1 | 1 | 1 | 6 |
| | female | | | | | 1 | 3 | 4 |
| 40-49 | male | | | | | 1 | 3 | 4 |
| | female | | | | | 2 | 4 | 6 |
| 50 and older | male | 1(8); 1 (10) | 2 | | | | | 4 |
| | female | 1 (8) | 2 | 1 | 1 | 1 | 1 | 7 |
| Total | male | 3 | 4 | 0 | 1 | 3 | 6 | 17 |
| | female | 1 | 2 | 1 | 3 | 5 | 14 | 26 |

Table 5.5 Summarized result of word-identification task with traditional Chinese

characters (target phonemes)

The photographic stimuli appeared to have an impact on the overall correct identification rate; however, it cannot be determined from these two tables whether the photographic stimuli and choice of sibilant phonemes were dependent variables. Therefore, a closer look into the data is necessary to observe the relation between the two variables. In the following Tables 5.6 (/tʂ/ and /ʦ/), 5.7 (/tʂ$^{h}$/ and /ʦ$^{h}$/), and 5.8 (/ʂ/ and /s/) itemized data of the responses from different age groups are provided. The age groups 19 and under, 20-29, and 40-49 are neglected from the tables shown here considering the fact that they contribute minimum to no variation to the final result since participants from these groups received almost perfect scores from all the tasks. However, a complete report of the responses is available in Appendices 8, 9, and 10.

| **Participant's age and gender** | **Number of *incorrect* responses of target phonemes pairs /tʂ/ and /ts/ from all three tasks** | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | 1A | 1B | 1C | /tʂ/ *zh* | 2A | 2B | 2C | /ʦ/ *z* |
| | | Age of person in photo | late 20s | early 50s | early 20s | | late 20s | early 50s | early 20s | |
| | | | 知識 *zhīshí* | 治理 *zhìlǐ* | 致意 *zhìyì* | Total | 姿勢 *zīshì* | 自理 *zìlǐ* | 恣意 *zìyì* | Total |
| | 30-39 | Male (6) | | | | | | | | |
| | | P | 1 | 1 | 0 | 2 | 1 | 2 | 3 | 6 |
| | | T | 4 | 1 | 0 | 5 | 0 | 2 | 2 | 4 |
| | | S | 3 | 2 | 0 | 5 | 0 | 3 | 3 | 6 |
| | | Female (4) | | | | | | | | |
| | | P | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | T | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | S | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | 50 and older | Male (4) | | | | | | | | |
| | | P | 2 | 1 | 0 | 3 | 2 | 2 | 4 | 8 |
| | | T | 4 | 2 | 0 | 6 | 1 | 1 | 1 | 3 |
| | | S | 3 | 1 | 0 | 4 | 3 | 0 | 0 | 3 |
| | | Female (7) | | | | | | | | |
| | | P | 2 | 1 | 1 | 4 | 2 | 1 | 3 | 6 |
| | | T | 4 | 2 | 0 | 6 | 2 | 1 | 1 | 4 |
| | | S | 0 | 1 | 3 | 4 | 2 | 1 | 1 | 4 |
| | Total | Male (17) | | | | | | | | |
| | | P | 4 | 2 | 0 | 6 | 3 | 4 | 8 | 15 |
| | | T | 8 | 3 | 0 | 11 | 2 | 3 | 3 | 8 |
| | | S | 6 | 3 | 0 | 9 | 4 | 3 | 3 | 10 |
| | | Female (26) | | | | | | | | |
| | | P | 4 | 1 | 1 | 6 | 3 | 1 | 3 | 7 |
| | | T | 7 | 2 | 2 | 11 | 2 | 1 | 1 | 4 |
| | | S | 1 | 1 | 3 | 5 | 2 | 2 | 2 | 6 |
| | | All (43) | | | | | | | | |
| | | P | 8 | 3 | 1 | 12 | 6 | 5 | 11 | 22 |
| | | T | 15 | 5 | 2 | 22 | 4 | 4 | 4 | 12 |
| | | S | 7 | 4 | 3 | 14 | 6 | 5 | 5 | 16 |

Table 5.6 Brief summarized result of responses from target phonemes pair /tʂ/ and /ʦ/

As illustrated in Table 5.6, it can be found that when the target phoneme was retroflex consonant /tʂ/, participants made a total of 22 incorrect responses during the traditional-Chinese-character task, yet only 12 incorrect when seeing the photographic stimuli. The trend was the same for both male and female participants. This result is in accordance with our hypothesis that participants would show an overall higher tendency of identifying a retroflex-phoneme-target-sound with an actual retroflex-phoneme-target-word instead of a alveolar-phoneme-target-word, because there were twice as more number of photos of younger people than older people, and thus less priming of retroflex-alveolar merger. However, when applying the Person's chi-square independency test, the responses were not sufficiently diversified to claim that there is significant association between the choice of phoneme category and the existence of photographic stimuli. $\chi^2$ (1, N = 258) = 3.387, which is insufficient to reject the null hypothesis at 3.84 ($\alpha$ = .05). Even by neglecting the 16 participants who made all correct responses and another 10 who made less than 2 incorrect responses, bringing N down to 102, $\chi^2$ (1, N = 102) = 3.753 is still insufficient to claim that the existence of photo stimuli had a significant effect on the speech perception of retroflex sibilants in Mandarin Chinese. When inspecting each individual trial, it did not seem that the person's age represented by the photo had the assumed effect either. However, with the limited data here, it would not be justifiable to directly make a valiant claim that the priming of age with photo stimuli had completely no effect on the perception of Mandarin retroflex sibilants.

On the other hand, when the target phoneme was the alveolar consonant /ts/ participants only made 12 incorrect responses in the traditional-Chinese-character task, yet made a total of 22 incorrect responses when seeing the photographic stimuli, which is the opposite from that of the retroflex consonant /tʂ/. The trend was also the same for both male and female participants except that male participants made double the amount of mistakes compared to female participants. It may be inferred from the original hypothesis that when participants see photos of younger females, they have a higher chance to be primed with a retroflex response since younger people and females tend to make a greater distinction between retroflex and alveolar sibilants. This type of priming could be explained as a perceptual case of speakers hyper-correcting the non-retroflex

alveolar initials into their corresponding retroflex sounds because they are conscious about meeting the prescriptive standard of retroflection (Chung 2006).

Although the results from the target alveolar phoneme /ts/ were not statistically significant as a whole because female speakers did not respond with sufficient difference between the traditional-Chinese-character task and the photo stimuli task, the responses from male participants did pose a significant association between the existence of photo stimuli and the perceived consonant when hearing the alveolar sibilant /ts/. If the 4 male participants who received a 100% correct rate were excluded from the chi-square test along with another 5 male participants who made less than two mistakes, $\chi^2$ (1, N = 24) = 4.148, $p < .05$ is sufficient to reject the null hypothesis and claim that the two variables are dependent. However, since the chi-square test does not imply directionality, it could only be stated that male participants in this study, when seeing photographic stimuli, were estimated to be 3.4 times more likely to perceive the alveolar /ts/ as the retroflex /tʂ/ than when solely seeing traditional Chinese characters. The 95% confidence interval indicated a high degree of precision (1.0269 to 11.2571) in estimating the population odds ratio.

In summary, it was found with the retroflex /tʂ/ and alveolar /ts/ consonant pair that (1) participants generally tend to perceive /tʂ/ as /ts/ more often when they only saw the traditional Chinese characters than when seeing photographic stimuli; (2) they tended to perceive /ts/ as /tʂ/ more often when they saw the photographic stimuli than without; (3) the effect of the person's age represented by the photo still remains to be further studied; (4) male participants tended to perceive alveolar /ts/ as retroflex /tʂ/ much more often than female participants regardless of the stimuli; and finally (5) in the case of male participants who made three or more identification mistakes when perceiving the alveolar consonant /ts/, it could be said that there was a statistically significant association between the choice of phoneme category and the existence of photographic stimuli.

The results of the responses from the minimal pair /tʂ$^h$/ and /ʦ$^h$/ showed a similar trend with those from /tʂ/ and /ts/ except that the variation in total incorrect responses in relation to stimuli was not as substantial as that in the case of /tʂ/ and /ts/. As a result, only the first four points stated in the previous summary remain true for the pair /tʂ$^h$/ and /ʦ$^h$/. A brief summary of the responses is provided in the following Table 5.7.

| Participant's age and gender | | Number of *incorrect* responses of target phonemes pairs /tʂʰ/ and /tsʰ/ from all three tasks | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | 3A | 3B | 3C | /tʂʰ/ *ch* | 4A | 4B | 4C | /tsʰ/ *c* |
| | Age of person in photo | early 50s 延遲 *yánchí* | early 20s 池塘 *chítáng* | late 20s 赤字 *chìzì* | Total | early 50s 言詞 *yáncí* | early 20s 祠堂 *cítáng* | late 20s 刺字 *cìzì* | Total |
| 30-39 | Male (6) | | | | | | | | |
| | P | 2 | 0 | 0 | 2 | 2 | 2 | 3 | 7 |
| | T | 2 | 0 | 2 | 4 | 1 | 2 | 2 | 5 |
| | S | 1 | 1 | 0 | 2 | 0 | 3 | 2 | 5 |
| | Female (4) | | | | | | | | |
| | P | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | T | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | S | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 50 and older | Male (4) | | | | | | | | |
| | P | 2 | 2 | 1 | 5 | 3 | 2 | 3 | 8 |
| | T | 3 | 1 | 1 | 5 | 1 | 4 | 3 | 8 |
| | S | 0 | 1 | 1 | 2 | 2 | 3 | 3 | 8 |
| | Female (7) | | | | | | | | |
| | P | 2 | 1 | 0 | 3 | 1 | 1 | 3 | 5 |
| | T | 2 | 2 | 1 | 5 | 0 | 1 | 3 | 4 |
| | S | 1 | 1 | 2 | 4 | 1 | 0 | 4 | 5 |
| Total | Male (17) | | | | | | | | |
| | P | 4 | 2 | 1 | 7 | 5 | 4 | 6 | 15 |
| | T | 5 | 1 | 4 | 10 | 2 | 6 | 5 | 13 |
| | S | 1 | 2 | 1 | 4 | 2 | 6 | 5 | 13 |
| | Female (26) | | | | | | | | |
| | P | 2 | 1 | 0 | 3 | 2 | 1 | 3 | 6 |
| | T | 2 | 2 | 2 | 6 | 1 | 1 | 3 | 5 |
| | S | 1 | 1 | 2 | 4 | 1 | 0 | 6 | 7 |
| | All (43) | | | | | | | | |
| | P | 6 | 3 | 1 | 10 | 7 | 5 | 9 | 21 |
| | T | 7 | 3 | 6 | 16 | 3 | 7 | 8 | 18 |
| | S | 2 | 3 | 3 | 8 | 3 | 6 | 11 | 20 |

Table 5.7 Brief summarized result of responses from target phonemes pair /tʂʰ/ and /tsʰ/

The results from the only fricative pair /ʂ/ and /s/ among the three pairs differ from other two affricate pairs. The existence of photo stimuli did not seem to make an impact on the participants' responses when the target phoneme was the retroflex /ʂ/ for both male and female participants as shown in the following Table 5.8. As for the corresponding alveolar phoneme /s/, whether or not photo stimuli were presented still did not significantly influence how female participants perceived the alveolar consonant. The male participants seem to perceive /s/ as /ʂ/ more often in the absence of photographic stimuli, which is in the opposite trend from the results of the other two affricate pairs.

This finding may be due to the essential difference between a fricative and an affricate. When comparing the overall incorrect responses from all three word-

identification tasks across these six phonemes, participants made the most misjudgments on the pair /tʂ/ (48) and /ʦ/ (50), followed by the pair, /tʂ$^{h}$/ (34) and /ʦ$^{h}$/ (59), and made the least mistakes on the affricate pair /ʂ/ (39) and /s/ (33). It is also possible that the sample size was too small to reveal the true effects of the photo stimuli on the perception of /ʂ/ and /s/. Both of these assumptions are reasonable and require further study to verify.

| **Participant's age and gender** | **Number of *incorrect* responses of target phonemes pairs /ʂ/ and /s/ from all three tasks** | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | 5A | 5B | 5C | /ʂ/ *sh* | 6A | 6B | 6C | /s/ *s* |
| | | Age of person in photo | late 20s<br>詩人<br>*shīrén* | early 20s<br>使節<br>*shǐjiē* | early 50s<br>相識<br>*xiàngshí* | Total | late 20s<br>私人<br>*sīrén* | early 20s<br>死結<br>*sǐjié* | early 50s<br>相似<br>*xiàngsì* | Total |
| | 30-39 | Male (6) | | | | | | | | |
| | | P | 1 | 1 | 0 | 2 | 2 | 1 | 1 | 4 |
| | | T | 1 | 0 | 2 | 3 | 3 | 3 | 1 | 7 |
| | | S | 2 | 0 | 1 | 3 | 2 | 2 | 1 | 5 |
| | | Female (4) | | | | | | | | |
| | | P | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 |
| | | T | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 |
| | | S | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | 50 and older | Male (4) | | | | | | | | |
| | | P | 1 | 2 | 3 | 6 | 1 | 0 | 0 | 1 |
| | | T | 2 | 1 | 2 | 5 | 1 | 2 | 0 | 3 |
| | | S | 0 | 0 | 2 | 2 | 1 | 0 | 1 | 2 |
| | | Female (7) | | | | | | | | |
| | | P | 2 | 1 | 3 | 6 | 1 | 2 | 1 | 4 |
| | | T | 1 | 2 | 2 | 5 | 2 | 2 | 0 | 4 |
| | | S | 1 | 1 | 2 | 4 | 0 | 2 | 1 | 3 |
| | Total | Male (17) | | | | | | | | |
| | | P | 2 | 3 | 3 | 8 | 3 | 1 | 1 | 5 |
| | | T | 3 | 1 | 4 | 8 | 4 | 5 | 1 | 10 |
| | | S | 2 | 0 | 3 | 5 | 3 | 2 | 2 | 7 |
| | | Female (26) | | | | | | | | |
| | | P | 2 | 2 | 4 | 8 | 1 | 2 | 1 | 4 |
| | | T | 1 | 2 | 3 | 6 | 2 | 2 | 0 | 4 |
| | | S | 1 | 1 | 2 | 4 | 0 | 2 | 1 | 3 |
| | | All (43) | | | | | | | | |
| | | P | 4 | 5 | 7 | 16 | 4 | 3 | 2 | 9 |
| | | T | 4 | 3 | 7 | 14 | 6 | 7 | 1 | 14 |
| | | S | 3 | 1 | 5 | 9 | 3 | 4 | 3 | 10 |

Table 5.8 Brief summarized result of responses from target phonemes pair /ʂ/ and /s/

### 5.1.2 Orthographic stimuli

The purpose of the orthographic stimuli, using simplified Chinese characters as treatment to compare with the control group (traditional Chinese characters), was to examine whether the audio perceptual results of retroflex or alveolar sibilants were associated with the type of character visually perceived. A brief comparison of the overall correct

identification rate between these two types of characters is provided by contrasting Table 5.9 below with Table 5.5 from the previous section. On the surface, it seems that more participants received a 70% or higher correct identification rate when seeing simplified Chinese characters (38/43) than traditional Chinese characters (33/43).

| Participant's age and gender | | Number of correct trials from word-identification task with simplified-Chinese-character stimuli (target phonemes only) Total trials = 18 | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | 9-10 (50-59%) or under | 11-12 (60-69%) | 13-14 (70-79%) | 15-16 (80-89%) | 17 (90-99%) | 18 (100%) | Total |
| 19 and under | male | | | | | | 1 | 1 |
| | female | | | | 1 | | 1 | 2 |
| 20-29 | male | | | | | | 2 | 2 |
| | female | | | | 1 | | 6 | 7 |
| 30-39 | male | 1(8) | 1 | 1 | 1 | 1 | 1 | 6 |
| | female | | | | | | 4 | 4 |
| 40-49 | male | | | | | 1 | 3 | 4 |
| | female | | | | | | 6 | 6 |
| 50 and older | male | | 2 | 1 | 1 | | | 4 |
| | female | 1(8) | | 1 | 3 | 2 | | 7 |
| Total | male | 1 | 3 | 2 | 2 | 2 | 7 | 17 |
| | female | 1 | 0 | 3 | 4 | 4 | 14 | 26 |

Table 5.9 Summarized result of word-identification task with simplified-Chinese-character stimuli (target phonemes)

The other features of the two tables are rather similar to one another, including the number of people who scored an above 90% correct rate and the distribution of participants across the board in general. For an in depth investigation of how different varieties of written character have an effect on the perception of retroflex and alveolar sibilants in Mandarin Chinese, more detailed data is needed. Table 5.10, a compilation of the tables 5.6, 5.7, and 5.8, compares the total number of incorrect responses for each target phoneme. A common trend that can be easily identified is that more mistakes were made when seeing traditional Chinese characters while identifying retroflex sounds as well as seeing simplified characters while identifying alveolar sounds with /s/ being the only exception.

| **Number of *incorrect* responses of target phonemes pairs /tʂ/ and /ts/ from all three tasks** | | | | | | | | | **N = 18 * 3= 54** |
|---|---|---|---|---|---|---|---|---|---|
| | | 1A | 1B | 1C | /tʂ/ *zh* | 2A | 2B | 2C | /ts/ *z* |
| | | 知識 *zhīshí* | 治理 *zhìlǐ* | 致意 *zhìyì* | Total | 姿勢 *zīshì* | 自理 *zìlǐ* | 恣意 *zìyì* | Total |
| Total | Male (17) | | | | | | | | |
| | P | 4 | 2 | 0 | 6 | 3 | 4 | 8 | 15 |
| | T | 8 | 3 | 0 | 11 | 2 | 3 | 3 | 8 |
| | S | 6 | 3 | 0 | 9 | 4 | 3 | 3 | 10 |
| | Female (26) | | | | | | | | |
| | P | 4 | 1 | 1 | 6 | 3 | 1 | 3 | 7 |
| | T | 7 | 2 | 2 | 11 | 2 | 1 | 1 | 4 |
| | S | 1 | 1 | 3 | 5 | 2 | 2 | 2 | 6 |
| | All (43) | | | | | | | | |
| | P | 8 | 3 | 1 | 12 | 6 | 5 | 11 | 22 |
| | T | 15 | 5 | 2 | 22 | 4 | 4 | 4 | 12 |
| | S | 7 | 4 | 3 | 14 | 6 | 5 | 5 | 16 |
| **Number of *incorrect* responses of target phonemes pairs /tʂ$^{h}$/ and /ts$^{h}$/ from all three tasks** | | | | | | | | | **N = 18 * 3= 54** |
| | | 3A | 3B | 3C | /tʂ$^{h}$/ *ch* | 4A | 4B | 4C | /ts$^{h}$/ *c* |
| | | 延遲 *yánchí* | 池塘 *chítáng* | 赤字 *chìzì* | Total | 言詞 *yáncí* | 祠堂 *cítáng* | 刺字 *cìzì* | Total |
| Total | Male (17) | | | | | | | | |
| | P | 4 | 2 | 1 | 7 | 5 | 4 | 6 | 15 |
| | T | 5 | 1 | 4 | 10 | 2 | 6 | 5 | 13 |
| | S | 1 | 2 | 1 | 4 | 2 | 6 | 5 | 13 |
| | Female (26) | | | | | | | | |
| | P | 2 | 1 | 0 | 3 | 2 | 1 | 3 | 6 |
| | T | 2 | 2 | 2 | 6 | 1 | 1 | 3 | 5 |
| | S | 1 | 1 | 2 | 4 | 1 | 0 | 6 | 7 |
| | All (43) | | | | | | | | |
| | P | 6 | 3 | 1 | 10 | 7 | 5 | 9 | 21 |
| | T | 7 | 3 | 6 | 16 | 3 | 7 | 8 | 18 |
| | S | 2 | 3 | 3 | 8 | 3 | 6 | 11 | 20 |
| **Number of *incorrect* responses of target phonemes pairs /ʂ/ and /s/ from all three tasks** | | | | | | | | | **N = 18 * 3= 54** |
| | | 5A | 5B | 5C | /ʂ/ *sh* | 6A | 6B | 6C | /s/ *s* |
| | | 詩人 *shīrén* | 使節 *shǐjiē* | 相識 *xiàngshí* | Total | 私人 *sīrén* | 死結 *sǐjié* | 相似 *xiàngsì* | Total |
| Total | Male (17) | | | | | | | | |
| | P | 2 | 3 | 3 | 8 | 3 | 1 | 1 | 5 |
| | T | 3 | 1 | 4 | 8 | 4 | 5 | 1 | 10 |
| | S | 2 | 0 | 3 | 5 | 3 | 2 | 2 | 7 |
| | Female (26) | | | | | | | | |
| | P | 2 | 2 | 4 | 8 | 1 | 2 | 1 | 4 |
| | T | 1 | 2 | 3 | 6 | 2 | 2 | 0 | 4 |
| | S | 1 | 1 | 2 | 4 | 0 | 2 | 1 | 3 |
| | All (43) | | | | | | | | |
| | P | 4 | 5 | 7 | 16 | 4 | 3 | 2 | 9 |
| | T | 4 | 3 | 7 | 14 | 6 | 7 | 1 | 14 |
| | S | 3 | 1 | 5 | 9 | 3 | 4 | 3 | 10 |

Table 5.10 Compiled results of responses from target phonemes

As shown in Table 5.10, both male and female participants tended to perceive retroflex consonants as their corresponding alveolar phoneme when they were seeing traditional Chinese characters. In other words, they showed a higher tendency of correctly identifying a Mandarin retroflex consonant initial with the stimuli of simplified Chinese

characters. One noteworthy point is that the very few mistakes made by female participants from age groups 20-29 and 40-49 across all tasks were mainly made on identifying retroflex consonants as alveolars when they saw traditional Chinese characters, especially for the retroflex phoneme /tʂ/.

For each phoneme individually, no statistical significance was found for the association between type of character and the phoneme perceived. This is probably due to the smaller data sample size collected for this experiment. However, when considering the series of Mandarin retroflex consonants as a whole, it was found that the variation in orthography and variation of perceived consonants were in fact dependent variables. The Person's chi-square test of association revealed a statistically significant relationship between the identification of retroflex phonemes and the variety of written Chinese characters, $\chi^2$ (1, N = 387) = 5.95, p < .05. Participants of this experiment, when seeing simplified Chinese characters are estimated to be 1.83 times more likely to identify a retroflex sibilant with the actual retroflex sound than when seeing traditional Chinese characters. The 95% confidence interval indicated a high degree of precision (1.1151 to 2.8495) in estimating the population odds ratio.

The main contributors to this statistical significance are the male participants who are over fifty years old. This group alone showed a significant association between the identification of Mandarin retroflex consonants and the written Chinese character variable with chi-square test result $\chi^2$ (1, N = 36) = 4. Significance was not found when applying the Person's chi-square test of association to other age or gender groups.

If the 16 participants who made no mistakes were eliminated from the chi-square test on the retroflex phonemes as a whole, $\chi^2$ (1, N = 243) = 6.41, the p-value from the Fisher exact test would be even smaller. If the 10 participants who made only one or two mistakes on identifying target phonemes were also excluded from the chi-square test, $\chi^2$ (1, N = 153) = 7.13 would allow the data to reject the null hypothesis at a even more critical condition of $\alpha = .01$. Under this situation, the remaining participants who indeed responded differently with regard to the presented Chinese characters are estimated to be 1.99 times more likely to identify a retroflex sibilant with the actual retroflex sound instead of its corresponding alveolar sound when seeing simplified Chinese characters than traditional Chinese characters. Therefore, the findings from this perception study on

the association between orthography and perception of Mandarin retroflex sounds is in concordance with Tso's (2013) quantitative study of the interplay between variation in Chinese orthography and listeners' expectation.

The same statistical significance; however, could not be claimed for the series of alveolar consonant initials in Mandarin Chinese. The results from a chi-square test indicated that $\chi^2$ (1, N = 387) = 0.05, p > .05 for the alveolar phonemes, and could not reject the null hypothesis at 3.84 ($\alpha$ = .05). This result implies that the identification of Mandarin alveolar consonants is independent of the written Chinese character variable. Two factors could be accounted for such result, the limited difference in response made by the participants when presented with different characters and the responses for the phoneme /s/ showing an opposite trend from /ts/ and /ts$^h$/, which neutralized the already narrow difference even further.

To sum up, it was found from the results of the three word identification tasks that both photographic stimuli and orthographic stimuli influenced the participants' response to a certain extent even though 72% of all participants scored an over 90% correct identification rate across all three tasks. The participants responded to the stimuli with greater similarity between the affricate sibilant minimal pairs [tʂ, ts] and [tʂ$^h$, ts$^h$], but rather differently toward the fricative pair [ʂ, s]. For the photographic treatment task, (1) participants generally tend to perceive alveolar phonemes as their corresponding retroflex phonemes more often (with the exception of the [ʂ, s] pair) when they saw the photographic stimuli than without and vice versa; (2) the effect of the speaker's age represented by the photo still remains to be further studied; and (3) male participants seemed to perceive /s/ as /ʂ/ more often in the absence of photographic stimuli, which is in the opposite trend from the results of the other two affricate pairs. In the particular case of perceiving the alveolar consonant /ts/, a statistically significant association was found between the choice of phoneme category and the existence of photographic stimuli.

As for the orthographic stimuli, it was found with Person's chi-square test of association that there was a statistically significant relationship between the identification of retroflex phonemes and the variety of written Chinese characters. With a 95% confidence interval, the odds ratio estimated that participants, when seeing simplified Chinese characters, were at least 1.83 times more likely to identify a retroflex audio input

with the actual retroflex phoneme instead of its corresponding alveolar sound than when seeing traditional Chinese characters.

## 5.2 Speaker-identity-observation task

The goal of the speaker-identity-observation task was to examine whether participants' observation of the speaker's social identity would be affected by the discordance between two modes of social information input, audio and visual. In this section of the speech perception study, participants were divided into four groups according to the different combinations of the audio-visual stimuli: listening to Taiwan speaker and reading traditional characters, listening to Beijing speaker and reading simplified characters, listening to Taiwan speaker and reading simplified characters, and listening to Beijing speaker and reading traditional characters. Emphasis was on the latter two groups, in which there was conflict in the social identity represented by the audio and visual stimuli.

The distribution of participants across these four groups is provided in the following Table 11. As explained previously, due to particular interest of the study, the groups were not evenly divided.

| **Participant's age and gender** | **Number of participants in each audio-visual stimuli combination** | | | | | | **N =43** |
|---|---|---|---|---|---|---|---|
| | | | Audio: Taiwan Visual: traditional | Audio: Beijing Visual: simplified | Audio: Taiwan Visual: simplified | Audio: Beijing Visual: traditional | Total |
| | 19 and under | male | | | 1 | | 1 |
| | | female | | | 1 | 1 | 2 |
| | 20-29 | male | | | 1 | 1 | 2 |
| | | female | 1 | 1 | 2 | 3 | 7 |
| | 30-39 | male | 1 | 1 | 2 | 2 | 6 |
| | | female | 1 | 1 | 1 | 1 | 4 |
| | 40-49 | male | | | 2 | 2 | 4 |
| | | female | 1 | 1 | 2 | 2 | 6 |
| | 50 and older | male | | | 2 | 2 | 4 |
| | | female | 1 | 1 | 2 | 3 | 7 |
| | Total | male | 1 | 1 | 8 | 7 | 17 |
| | | female | 4 | 4 | 8 | 10 | 26 |

Table 5.11 Participant assignment for speaker-identity-observation task

The standard list of questions that the participants were asked includes an estimate of the speaker's age, profession or work place, and place of origin. They were also asked what they believed were the differences between the Taiwan Mandarin dialect and the Beijing Mandarin dialect, what they believed to be standard Mandarin, and whether they perceived themselves as speaking Mandarin to the their ideal stantard. The perceived aged of the two speakers who were both in their late twenties at the time of recording was ranged between 20 to 40 years old with an average of 30 years old for the Taiwan speaker and 28 years old for the Beijing speaker. Female participants generally assigned a younger perceived age to the speakers than male participants. The difference between male and female participants in the perceived age of the two speakers is presented in Figure 5.2.

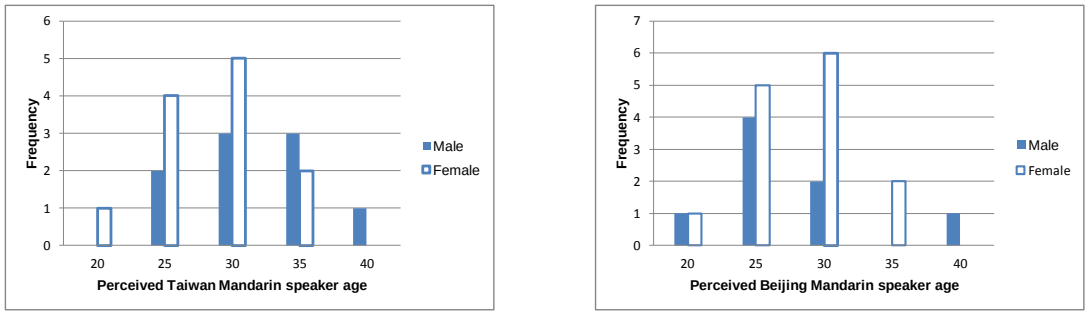



Figure 5.2 Perceived speaker age by male and female participants

The various professions of the speakers presumed by the participants include news reporter, teacher, scholar, student, office lady, and museum guide. These assumptions should be the result of the content recited by the speakers. Since the topic was on an English scholar's perspective on the opium war, it was logical to assume that the speaker was someone who had in depth knowledge on the historical event and were able to discuss the issue with clarity. Therefore, as expected, the majority of participants presumed that the speakers were either a news reporter (16) or a teacher (10). Details of the presumed professions of the speakers are presented in Figure 5.3. One point worth noticing is that there were close to equal number of participants who thought the Taiwan Mandarin speaker was either a news reporter or a teacher, while most of the participants who listened to the Beijing Mandarin speaker assumed that she was a news reporter.

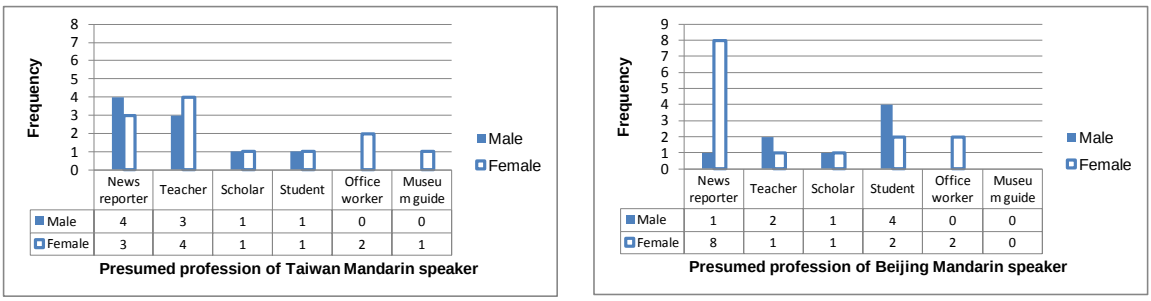



Figure 5.3 Presumed speakers professions by male and female participants

The results from asking the participants how old the speakers were and what were the presumed professions of them were anticipated. However, the question asking where they thought the speaker was from elicited responses well beyond expectation. The following Figure 5.4 is a summary of all the places mentioned by the participants and the perceived place of origin of the Taiwan Mandarin speaker. The possible places general include four Mandarin speaking regions in Asia: Taiwan, China, Hong Kong, and Singapore. A specific group of speakers, Hakka, was mentioned by some participants. 'Taiwan Hakka-speaking' in the figure refers to those whose first language may be Hakka or both Hakka and Taiwan Mandarin. While some participants only mentioned China, others made a further distinction between north and south of China and pointed out there was a difference in the 'accent' of these areas. What was surprising was the mentioning of Hong Kong and Singapore, which clearly were not foreseen responses for this question.

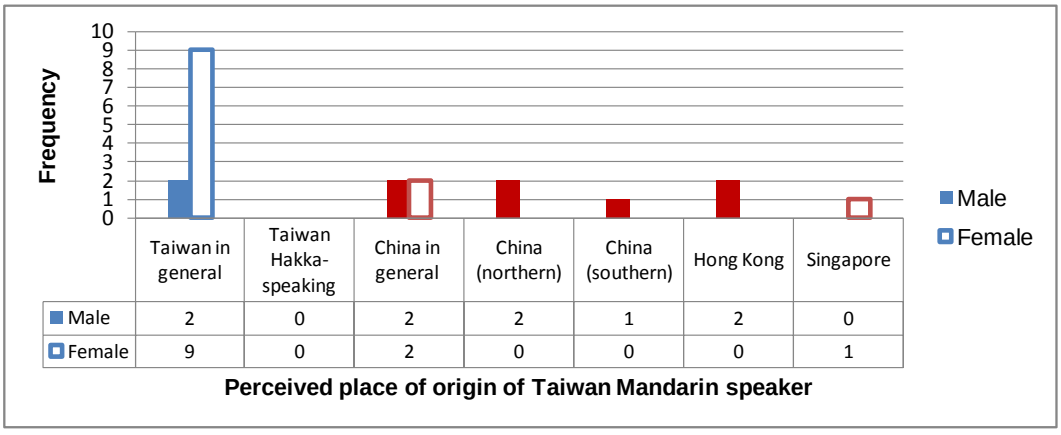



Figure 5.4 Perceived place of origin of Taiwan Mandarin speaker

As illustrated in the figure, while the majority of female participants who listened to the Taiwan Mandarin speaker accurately identified the speaker's place of origin, only

two male participants perceived the Taiwan Mandarin speaker as actually from Taiwan. These 21 participants include 16 (8 male and 8 female) who read the simplified-character passage and 5 (1 male and 4 female) who read the traditional-character passage. The participants who listened to the Taiwan Mandarin speaker are shown in Figure 5.5 with those who read the traditional-character passage highlighted in yellow, those who read the simplified-character passage and perceived the speaker as a Taiwanese marked in green, and those who read the simplified-character passage but thought the speaker was non-Taiwanese marked in red.

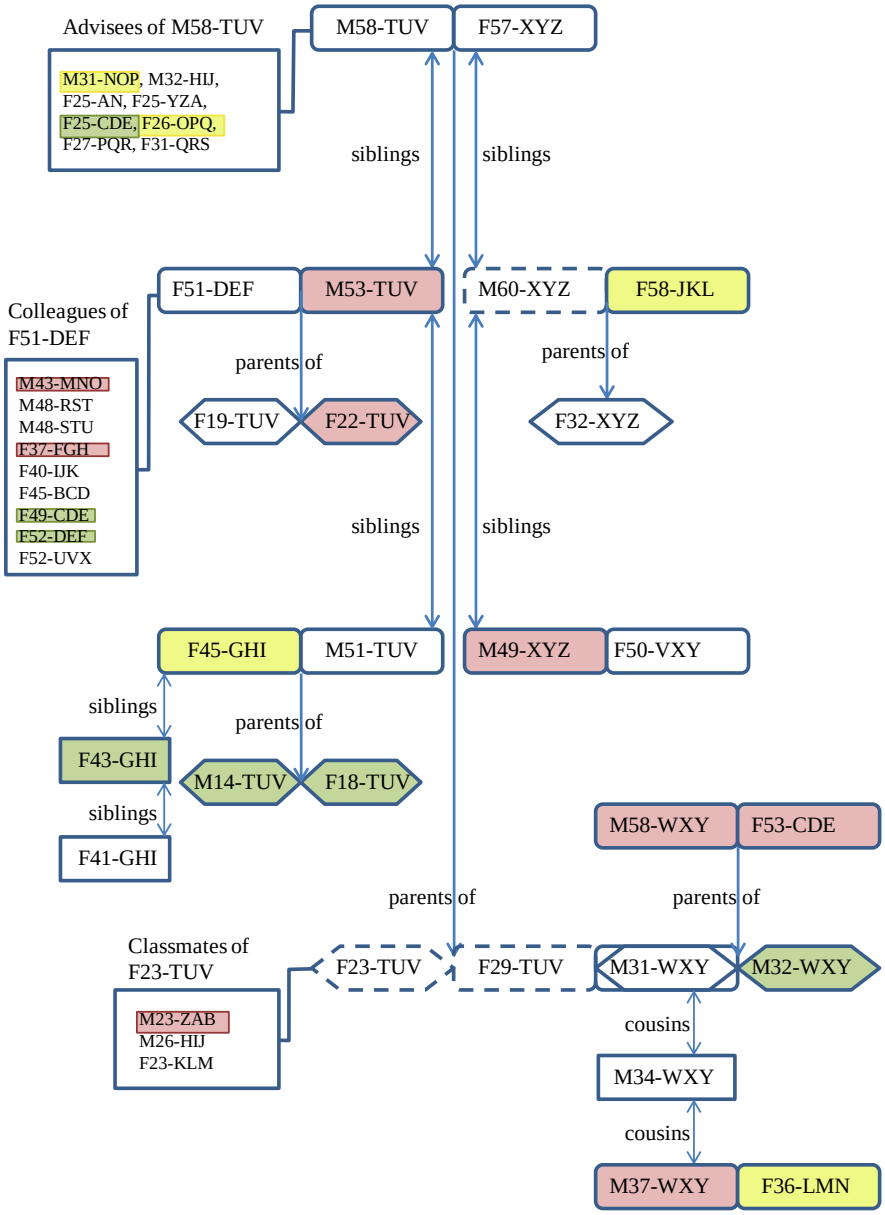



Figure 5.5 Participants who listened to the Taiwan Mandarin speaker

All female participants reading the traditional-character passage while listening to the Taiwan Mandarin speaker responded without doubt that the speaker was from Taiwan. One example is from the interview with F26-OPQ. The interviewer (represented by 'I' in the transcription), with experience from previous interviews, realized that Hong Kong and Singapore were also options in identifying the place of origin. Therefore, the areas Hong Kong and Singapore were intentionally provided as one of the options to choose from when asking about the speaker's place of origin. F26-OPQ's first response was 'Taiwan', and even after specifically asking if it were possible that the speaker was a Mainland China speaker, she still persisted with her original answer. Her reason was that the speaker did not have any accent. The Mandarin term she used was 口音 *kǒu yīn*, which the literal translation would be 'mouth sound' (IU 11). This term was frequently used by many participants to describe the difference in accent and it was used in alternation with 腔調 *qiāngdiào*, which literally means 'tune tone' but also translated as 'accent'.

| | | |
|---|---|---|
| 1 | I: | ..那你覺得她-- |
| 2 | | ..如果她是老師的話, |
| 3 | | ..你覺得她比較有可能是哪個地方的老師?, |
| 4 | | ..譬如說是台灣老師, |
| 5 | | ..香港老師, |
| 6 | | ..新加坡老師, |
| | | ..還是大陸的老師. |
| | | 'If you think she is a teacher, at what area do you think she teaches? For example, is she a teacher in Taiwan, Hong Kong, Singapore, or Mainland China?' |
| 7 | F26-OPQ: | ..Uh…台灣. |
| | | 'Uh, Taiwan.'_ |
| 8 | I: | ..台灣. |
| | | 'Taiwan.' |
| 9 | F26-OPQ: | ..Mm. |
| | | 'Mm' |
| 10 | I: | ..有沒有什麼特質讓你覺得她應該是台灣. |
| | | 'Are there any characteristics that make you believe it should be Taiwan?' |
| 11 | F26-OPQ: | …就沒有其他口音 ah, |
| | | '(She) does not have any other accent.' |
| 12 | I: | ..沒有其他口音. |

‘(She) does not have any other accent.’

13 F26-OPQ: ..對.
‘Yes’

14 I: ..有沒有可能是大陸那邊?.
‘Is it possible that she is from Mainland China?’

15 F26-OPQ: .. Uh …感覺比較像台灣 neh.
‘Mm…(I) feel like it’s more like Taiwan.’

The male participant, M31-NOP, who read the traditional-character passage while listening to the Taiwan Mandarin speaker assumed that she was from Hong Kong. He did not provide direct reasons as to why she sounded like a Hong Kong speaker, but instead stated that she did not sound similar a Taiwan speaker (IU 14). He added additional comments to Singapore and Mainland China Mandarin by claiming that there is English accent in Singapore Mandarin (IU 16-20) and Beijing accent in Mainland China Mandarin (IU 22-23). Since the Taiwan speaker had neither of these accents, and to M31-NOP did not sound like a Taiwan speaker, Hong Kong was assigned to her as the perceived place of origin by M31-NOP.

1 I: ..那…假使她是記者的話,
2 ..你覺得她比較有可能是香港記者,
3 ..台灣記者,
4 ..新加坡記者,
5 ..還是大陸記者?.
‘If she were a news reporter, would you think it is more likely that she a Hong Kong, Taiwan, Singapore, or Mainland China news reporter?’

6 M31-NOP: …嗯…感覺像香港.
‘Hmm, it feels like Hong Kong.’

7 I: ..香港.
‘Hong Kong.’

8 M31-NOP: ..Hunh.
‘Mm’

9 I: …嗯…有沒有什麼特質--
10 ..就講話的特質,
11 ..讓你覺得她是香港.
‘Are there any features in her speech that makes you feel it is Hong Kong?’

12 M31-NOP: …就是那個音跟台灣的那個--
13 ..講--
14 ..台灣人講國語不是很像,

15 ..然後…感覺新加坡或--
16 ..新加坡的會比較有那種--
17 ..英-
18 ..那種美-
19 ..美--
20 ..英語的那種腔調在裡面.
'Her accent is a little different from Taiwan Mandarin. And then (Mandarin) in Singapore would have some kind of English accent.'

21 I: ..Hunh.
'Huh'

22 M31-NOP: …然後大陸的話,
23 ..會比較有那種北京腔之類的.
'And then for Mainland (Mandarin), there would be some kind of Beijing accent.'

24 I: ..Huhhunh.
'Uh-huh'

25 M31-NOP: ..heN.\
'(sound of confirmation)'

M31-NOP was one of the participants introduced earlier at the beginning of this chapter, because he was one of the people under fifty but did not achieve a higher than 90% correct identification rate in the word-identification tasks. It is apparent from these two experiments that his ideal standard of Mandarin is different from others. Among the 24 mistakes out of a total of 54 trials across the three word-identification tasks, 5 of them were made on identifying retroflex sibilants as alveolar, while the other 19 were made by misjudging alveolar sibilants as retroflex. It can be inferred from this overtly noticeable pattern that he is unconfident in his ability to distinguish between retroflex and alveolar sibilants during perception, and thus he would rather hypercorrect the alveolar sibilants (19 out of 27) as retroflex. When extending this hypercorrection from word level to a complete passage, it is possible that he does not believe the speaker's Taiwan Mandarin is close enough to standard. As a result, even though he could not specify the reasons, he chose to believe that the speaker was from Hong Kong instead of Taiwan when the two other options were not plausible.

As for the participants who read simplified-characters (8 male and 8 female) while listening to the Taiwan Mandarin speaker, 9 of them (6 male and 3 female) did not perceive the speaker as Taiwanese. Among the three female speakers who did not

perceive her as a Taiwanese, two of them thought she was a Mainlander, and one believed she was from Singapore. F37-FGH, who thought the speaker was a mainlander, mentioned that her way of intonation and pronunciation was more like the 'unified standard'. The original term she used was 制式一點 *zhìshì yīdiǎn*, 'unified system a little' in literal translation (IU 22 and 24), and 'a little more towards the unified standard' in actual translation. She did not directly use the word 標準 *biāozhǔn* 'standard', which is a skilled way to avoid conceding a controversial opinion that there is one standard Mandarin dialect, or the dialect spoken by mainlanders are the so-called 'standard'.

| | | |
|---|---|---|
| 1 | I: | …那你覺得聽剛剛這位導讀的女生, |
| 2 | | …你覺得她是-- |
| 3 | | ..的口音像是-- |
| 4 | | ..台灣嗎?, |
| 5 | | ..還是香港, |
| 6 | | ..大陸, |
| 7 | | ..新加坡, |
| 8 | | ..你覺得她像是哪個區域來的女生.<br>'Where do you think the lady who recited this passage is most likely from? Taiwan, Hong Kong, Mainland, or Singapore?' |
| 9 | F37-FGH: | …我會覺得有點像大陸的.<br>'I feel like she is kind of like from Mainland ' |
| 10 | I: | ..有點像大陸的.<br>'Kind of like from China.' |
| 11 | F37-FGH: | ..對.<br>'Yes.' |
| 12 | I: | ..那…譬如說你是從, |
| 13 | | ..她的聲音特質哪個方面判斷, |
| 14 | | ..她比較像大陸的?.<br>'From which features, for example her voice quality, did you determine her as being from China?' |
| 15 | F37-FGH: | …Uh…一個是她的那個-- |
| 16 | | ..音調.<br>'One of them is from her accent.<br>' |
| 17 | I: | ..怎樣的音調?.<br>'What is it about her accent?' |
| 18 | F37-FGH: | ...Umm, |
| 19 | I: | ..可不可以[再詳細一點].<br>'Could (you) describe with more details?' |

| | | |
|---|---|---|
| 20 | F37-FGH: | [@@@@] |
| 21 | | …可能就是比較, |
| 22 | | …制式一點嗎. |
| | | ‘Probably a little bit more of the unified (standard).’ |
| 23 | I: | Uh-huh. |
| 24 | F37-FGH: | ..比較制式一點的音調, |
| 25 | | ..然後, |
| 26 | | …發音, |
| 27 | | ..感覺, |
| 28 | | …那種發音法, |
| 29 | | ..比較, |
| 30 | | ..比較有點像, |
| 31 | | ..大陸那邊的. |
| | | ‘The intonation is more like the unified (standard). The way of pronunciation is more like that from Mainland.’ |
| 32 | I: | Uh-huh. |

The female participant, F22-TUV, who perceived her as being from Singapore made interesting responses during the interview as shown in the following transcription. At the very beginning she sounded as if she was confident in her answer after thinking for a while (IU 5) that the speaker was from Singapore, but the turning point came when she was asked for the reasoning for her answer. At first she ruled out mainlanders claiming that all mainlanders retroflex their tongues. (F22-TUV did mention earlier in the interview that the speaker did not have clearly retroflexed consonants.) Subsequently, she then used a double negation saying that Taiwanese usually do *not not* retroflex their tongue (IU11), which is probably why she also dismissed the possibility of the speaker being Taiwanese. However, she then asked a rhetoric question of whether Hakka people retroflex their tongue (IU 15), and after two long pauses (IU 16 and 18), she decided that she believed that the speaker was from Taiwan but was a Hakka speaking person. Her response for the question was still documented as ‘Singapore’ since it was initial response without being interrupted, yet the unfolding of her thought process was truly worth mentioning.

| | | |
|---|---|---|
| 1 | I: | …你覺得她像是, |
| 2 | | ..台灣人香港人新加坡人大陸人, |
| 3 | | ..從哪裡來的, |
| 4 | | ..就是剛剛錄音那個人. |
| | | 'Where do you think she is from , the speaker of the recorded clip?' |
| 5 | F22-TUV: | … (1.5) 應該是新加坡人吧. |
| | | 'Should be from Singapore.' |
| 6 | I: | ..新加坡人. |
| | | 'A Singapore person.' |
| 7 | F22-TUV: | ..Hunh. |
| | | 'Huh' |
| 8 | I: | ..從哪裡覺得她是新加坡人?. |
| | | 'From what aspects did you determine that she is from Singapore?' |
| 9 | F22-TUV: | …因為卷舌大陸人會呀, |
| | | 'Because mainlanders retroflex their tongue.' |
| 10 | I: | ..Mm. |
| 11 | F22-TUV: | ..台灣人又不太會, |
| 12 | | …不捲舌. |
| | | 'Taiwanese usually do not not retroflex their tongue.' |
| 13 | I: | …所以你覺得她聽起來也不向台灣人. |
| | | 'So you don't think she sounds like Taiwanese.' |
| 14 | F22-TUV: | ..不太像 eh, |
| 15 | | ..eh 客家人會捲舌嗎?. |
| 16 | | …(2.0) Umm 我也不知道[eh]. |
| | | 'Not really; do Hakka people retroflex their tongue? Umm, I don't know either.' |
| 17 | I: | [猜]猜看. |
| | | 'Make a guess.' |
| 18 | F22-TUV: | …(1.8) [好像應該]是客家人吧, |
| | | 'Probably should be a Hakka person.' |
| 19 | I: | [沒關係]. |
| | | 'That's okay.' |
| 20 | | ..就台灣這邊的嗎?. |
| | | 'So from Taiwan?' |
| 21 | F22-TUV: | ..對, |
| | | 'Yes.' |
| 22 | I: | ..就台灣這邊的 |
| | | 'So, from Taiwan.' |
| 23 | F22-TUV: | ..Hunh. |
| | | 'Huh' |

Among the six male participants who did not correctly identify the Taiwanese speaker's place of origin when reading along the simplified-character passage, five of them thought she was from Mainland China, and the other thought she was from Hong Kong. Two of them specifically pointed out that she was from the northern areas of China, and said she was from the south. The prevalence of retroflex-alveolar neutralization is also known by most speakers of Taiwan Mandarin, and thus this response is understandable. One of the male participants, M43-MNO, who specifically stated that she was from the northern area of China, used the term 京片子 *jīng piānzi* 'Beijing accent'. 京 *jīng* is for 北京 *běijīng* 'Beijing' and the term is an informal way or slang for describing the true local accent of Beijing, which usually gives the strong impression of having clear retroflex consonants. When M43-MNO asked to explain why he thought the speaker was from mainland China, he first mentioned that she had more retroflex sounds (IU 14) and then talked about intonation, which he claimed that he could feel it but did not know how to explain it (IU 20 and 29). He eventually used term 京片子 *jīng piānzi* 'Beijing accent' to summarize his unexplainable feelings toward the difference in intonation.

| | | |
|---|---|---|
| 1 | I: | ...那你覺得她的, |
| 2 | | ..這樣子的唸法, |
| 3 | | ..像是台灣這邊的人, |
| 4 | | ..香港的人, |
| 5 | | ..新加坡的人, |
| 6 | | ..還是大陸那邊的人?. |
| | | 'According the way she read (the passage), where do you think she is from: Taiwan, Hong Kong, Singapore, or Mainland China?' |
| 7 | M43-MNO: | ..大陸的. |
| | | 'Mainland.' |
| 8 | I: | ..大陸的. |
| | | 'Mainland.' |
| 9 | M43-MNO: | ..heN. |
| 10 | I: | .. Uh 你覺得大陸那邊的人, |
| 11 | | 講話的特色, |
| 12 | | 譬如說, |
| 13 | | 她有什麼特色讓你覺得她應該是大陸那邊的人?. |
| | | 'What features does she have for you to believe that she is from mainland China?' |
| 14 | M43-MNO: | ...(1.5) Hmm 就捲舌音比較多. |

‘More retroflex sounds.’

15 I: ..捲舌音比較多.
‘More retroflex sounds.’

16 M43-MNO: ..對.
‘Yes.’

17 I: ...還有沒有別的特色,
18 ..讓你判別她是大陸人.
‘Are there other features that help you determine she is Chinese?’

19 M43-MNO: …Mm 音調吧.
‘Intonation.’
20 …可是音調我不太會講.
‘But I don’t really know how to explain intonation.’

21 I: ..Mm 可不可以稍微,
22 ..再敘述[一下].
‘Could you please elaborate a little more?’
23 M43-MNO: [比較]--
‘a little more’
24 ..這[感]覺--
‘It feels like’

25 I: [對]
‘Yes.’

26 M43-MNO: ..我知道,
27 ..我可以感覺,
28 ..但是我,
29 ..不會描述.
‘I know I can feel it, but I don’t know how to describe it.’

30 I: ..Okay.

31 M43-MNO: ..就京片子.
‘It’s the Beijing accent.’

The other participants who also thought the Taiwan speaker was a Beijing speaker gave similar responses and reasoning as F37-FGH and M43-MNO toward determining the speaker’s place of origin. Other comments on the speaker include (1) 她咬字比較清楚 *tā yǎozì bǐjiào qīngchǔ*, which literally means ‘her bite words is clearer’ and translates into ‘she has clearer articulation,’ and (2) 字正腔圓 *zì zhèng qiāng yuán*, which literally means ‘word upright accent round’ and translates into ‘clear articulation of words and overt intonation contour’. These were also the main features described by all participants during the interview when asked what the characteristics of Beijing Mandarin are.

As for the participants who still perceived the speaker as a Taiwan Mandarin speaker even though presented with a simplified-character passage made comments about her unclear pronunciation of retroflex sounds, her relatively flat intonation contour, and her not satisfyingly strong enough style of recitation in order to be considered from Mainland China. In sum, while the majority of female participants still perceived her as a Taiwan Mandarin speaker regardless of the type of character presented, male participants seemed to be affected by the simplified characters and perceived her as a Beijing Mandarin speaker.

In addition, each participant perceived the same features of the same speaker with great variation in degree. While the majority of male participants responded that she showed clear retroflection, the majority of female participants did not believe the retroflex initials were articulated clear enough when compared to Beijing Mandarin. There does not seem to be a strong correlation between the age of participant and the perceived speaker dialect, but probably only a difference in gender. Furthermore, the performance of the speaker-identity-observation task could not be predicted by the performance of the word-identification tasks. A high percentage of correct identification rate from the previous speech perception task did not entail a correct observation of speaker's place of origin in the speaker-identity-observation task and vice versa.

While 21 participants listened to the Taiwan Mandarin speaker recite the passage, the other 22 participants listened to the Beijing Mandarin speaker. As shown in figure 5.6, 5 of them (1 male and four female) were presented with a simplified-character passage, while the other 17 (7 male and 10 female) were presented with a traditional-character passage. Simplified-character-reading participants were highlighted in yellow, those who perceived the speaker as a Beijing Mandarin speaker were marked in green, and the others who thought the speaker was a non-Beijing Mandarin speaker were marked in red.

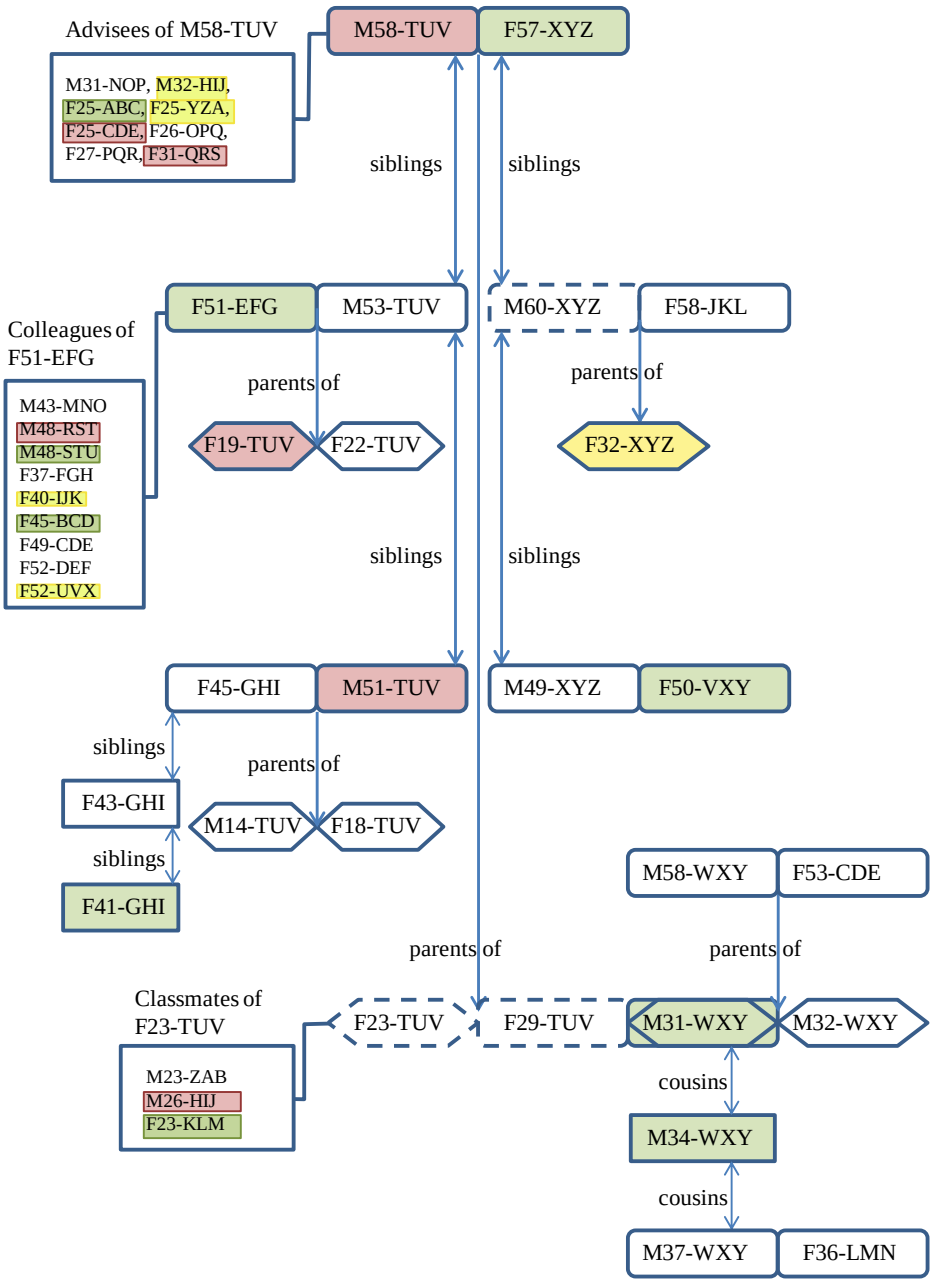



Figure 5.6 Participants who listened to the Beijing Mandarin speaker

As illustrated in Figure 5.6, a total of 10 participants (3 male and 7 female) who were given the traditional-character passage still perceived the speaker as a Beijing Mandarin speaker, while 7 participants (4 male and 3 female) believed that the speaker was a non-Beijing-Mandarin speaker. As for the 5 participants who were provided with the simplified-character passage, all 4 female participants perceived the speaker as from Mainland China, but the one male participant, M32-HIJ, thought the speaker was from

Hong Kong. M32-HIJ's performance on the word-identification task was discussed earlier in this chapter as he did not perform as well as his peers of the same age group. In line with M31-NOP who listened to the Taiwan Mandarin speaker and read a traditional-character passage, M32-HIJ also perceived the Beijing Mandarin speaker as originating from Hong Kong even though there was no conflict between the presented spoken dialect and written characters.

M32-HIJ stated that the Beijing Mandarin speaker sounded like neither from Beijing nor from Taiwan (IU 11-12). Hence, by using the exclusion method, he chose Hong Kong. This response was similar to that of F22-TUV who believed the Taiwan speaker was from Singapore. M32-HIJ further elaborated that the speaker did not have sufficient retroflex sounds to be similar to a Beijing Mandarin speaker (IU 21-22) and her accent was not similar to a Taiwan Mandarin accent (IU 46-48). As a result, even though he had rarely heard of Hong Kong people speak (IU 33-38), he still selected Hong Kong as the speaker was not a good representative of Beijing or Taiwan Mandarin to him.

| | | |
|---|---|---|
| 1 | I: | …如果你覺得她是學生的話, |
| 2 | | ..她比較有可能是大中華地區, |
| 3 | | ..哪個地方的學生. |
| | | 'If you think she is a student, where do you think she is from within the Asia pacific area?' |
| 4 | M32-HIJ : | …(2.0) eh, |
| 5 | | 香- |
| 6 | | 香港吧. |
| | | 'Hong Kong' |
| 7 | I: | ..香港. |
| 8 | | …(1.2) 有沒有什麼特質或特色, |
| 9 | | 讓你判斷她是香港. |
| | | 'Hong Kong. What are the features that lead you toward believing she is from Hong Kong?' |
| 10 | M32-HIJ : | …(1.5) eh, |
| 11 | | ..不太像中國, |
| 12 | | ..不太像台灣. |
| | | 'Not really like China; not really like Taiwan.' |
| 13 | I: | …Uh 怎麼說, |
| | | 'How come?' |
| 14 | | ..譬如[說], |
| | | 'For example' |

15 M32-HIJ : [捲]舌音,
16 ..然後,
17 ..也不太像台灣.
'The retroflex sounds. And does not sound like Taiwan.'

18 I: …捲舌音,
19 ..怎麼樣不太像台灣?.
'How does her retroflex sounds not similar to Taiwanese?'

20 M32-HIJ : …eh 沒有,
21 ..捲什麼音的話就不太像中國,
22 ..中國會比較多卷什麼音.
'No, her retroflex sounds are not similar to Chinese. Chinese have more retroflex sounds.'

23 I: Okay.

24 M32-HIJ : ..那感覺不太像,
25 ..一般台灣人.
'(She) does not seem like average Taiwanese.'

26 I: ..Okay.
27 ..那香港人--
28 ..她是一個很典型,
29 ..香港人的[特色嗎]?.
'About Hong Kong, is she a typical representative of Hong Kong features?'

30 M32-HIJ : [所以我]排除法,
31 ..就[是]--
'So I'm using the exclusion method.'

32 I: [Oh 排]除法.
'The exclusion method.'

33 M32-HIJ : ..就是--
34 ..Ah 選擇了香港,
35 ..因為我其實--
36 ..也沒有聽過什麼香港人--
'I chose Hong Kong because I've actually not heard of many Hong Kong people (speaking).'

37 I: Uh-[huh].

38 M32-HIJ : [講]過.
'Speaking'

39 I: …所以她的捲舌音不像台灣,
40 ..[但]--
'So her retroflex sounds are not like Taiwanese.'

41 M32-HIJ : [不]像--
42 ..不像--
43 ..捲舌音不像中國.

‘No, her retroflex sounds are not similar to Chinese.’

44 I: ..[不像]中國.
‘Not similar to Chinese.’

45 M32-HIJ : [然後],
46 ..腔調又不是很--
47 ..感覺不像,
48 ..台灣.
‘And then her accent; it doesn’t feel like Taiwanese.’

49 I: Okay.

As for the other two male participants who also perceived the Beijing Mandarin speaker as a Hong Kong person, M26-HIJ reasoned that although she did have retroflection in her articulation, the degree was not as strong[9] as a typical mainlander, and M51-TUV overtly pointed out there was a hint of Cantonese accent in her pronunciation. The other places of origin of this Beijing Mandarin speaker perceived by the participants include Mainland China (general, northern, and southern) and Taiwan (general and Hakka). The majority of female participants (11 out of 14) perceived the speaker as a Beijing Mandarin speaker regardless of the presented type of character, while only 3 out of 5 male participants considered the speaker to actually speak the Beijing Mandarin dialect. The number of participants who have selected each place of origin is presented in the following Figure 5.7.

[9] The original wording he used was:

| 雖然 | 捲舌音 | 有 | 發 | 出來， |
|---|---|---|---|---|
| suīrán | juǎnshéyīn | yǒu | fā | chūlái ， |
| although | retroflex.sound | EXIST | pronounce | out.come |

| 但 | 沒有 | 像 | 一般 | 大陸人 | 那麼 | 嚴重 |
|---|---|---|---|---|---|---|
| dàn | méiyǒu | xiàng | yībān | dàlùrén | name | yánzhòng |
| but | NEG.EXIST | similar | common | mainlander | so | serious |

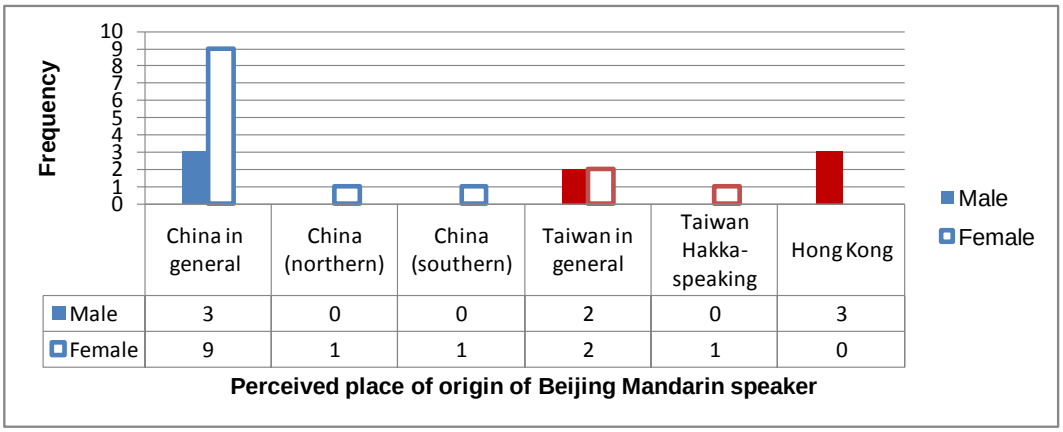



Figure 5.7 Perceived place of origin of Beijing Mandarin speaker

The female participant who thought the speaker was a Taiwan-Hakka speaking person is F22-TUV's sister, F19-TUV. It is unlikely that she overheard her sister's response for the questions because she was not present at the time F22-TUV was completing the experiment. As shown in the transcript below, F19-TUV brought up the group of Hakka people on her own (IU 13), which made us both laugh (IU 14-16) during the interview. She then decided that Taiwan was also a proper response, but it was still documented as Taiwan-Hakka speaking for special notice in this study. She was indecisive because (IU 5-8) she felt that the speaker did not sound similar to a Taiwanese, but on the other hand did not make the retroflex sounds clear enough to actually be a Beijing Mandarin speaker. She further explained that in her impression, the retroflex sounds were particularly emphasize in the Beijing Mandarin dialect (IU 30-32).

| | | |
|---|---|---|
| 1 | I: | …那你覺得她像是台灣這邊的人, |
| 2 | | ..還是香港人, |
| 3 | | ..新加坡人, |
| 4 | | ..還是大陸人. |
| | | 'Where do you think she is from: Taiwan, Hong Kong, Singapore or Mainland China?' |
| 5 | F19-TUV: | …(1.1)Hmm 應該不是台灣人. |
| 6 | | ..但是她又沒有大陸的捲舌音, |
| 7 | | ..就-- |
| 8 | | ..怪怪的. |
| | | 'I suppose she is not Taiwanese, but she does not have the retroflex sounds as a mainlander. There is something odd about it.' |
| 9 | I: | …如果一定要選一個地方, |
| 10 | | ..你覺得會-- |
| 11 | | ..你會比較選哪一個地方. |

‘If you must choose a place, where would you choose?’

12 F19-TUV: …Hmm,
13 …(1.0) 客家人好了,
‘I’ll go with Hakka.’

14 I: ..客家人 @@.
15 ..[那還是]台灣人 Ah.
‘Hakka, that is still Taiwanese.’

16 F19-TUV: [@@@]
17 ..好.
‘Okay.’

18 I: ..感覺像是台灣這邊的人.
‘Feels like she is from Taiwan.’

19 F19-TUV: ..好.
‘Okay.’

20 I: ..好.
21 ..那你覺得要--
22 ..所以就是因為她--
23 ..捲舌還不夠捲,
24 ..所以,
25 ..不像是大陸那邊的人.
‘So you think because her retroflex sounds are not retroflexed enough, so it is not likely that she is from China.’

26 F19-TUV: ..Hunh.
‘Huh’

27 I: ..所以大陸那邊的人,
28 ..捲舌要--
29 ..很捲.
‘So the people over there need to make really retroflexed retroflex sounds.’

30 F19-TUV: ..就是--
31 ..他們ㄓㄔㄕ那一類的話,
32 ..就會特別地強調.
‘Well, their *zhi chi shi* kind of sounds are specifically emphasized.’

33 I: ..Uh-huh.

The other participants who also perceived the speaker as a Taiwan Mandarin speaker responded with reasons similar to F19-TUV with emphasis that the retroflex series were not pronounce to the degree that would suffice for a Mainland China origin. Two of them raised the observation that it was possible for the speaker to be from the

southern part of China, where merger of the retroflex and alveolar consonants were also prevalent, they eventually chose Taiwan as their final answer to the question.

The participants who perceived the Beijing Mandarin speaker as originating from China provided interesting reasons as why they believed the speaker was a mainland despite the fact that they were presented with a traditional-character passage. Table 5.12 is a list of responses provide by these participants that are worthwhile of special notice.

| **Participant** | **Response** |
|---|---|
| F25-ABC | 她的捲舌比較多，台灣人唸不出來<br>'She has more retroflection, which cannot be produced by Taiwanese.' |
| M34-WXY | 就是大陸人講話，腔調、發音比較用心<br>'(It's obvious that) it is a mainlander speaking; (she) put more effort into the accent and pronunciation.' |
| F40-IJK | 大陸人，不過也有可能是台灣人很認真的念<br>'Mainland Chinese, but it is also possible that it was a Taiwanese reciting with effort.' |
| F41-GHI | 大陸南邊一點，沒有向北京腔那麼捲舌，也有可能是台灣人透過練習認真念<br>'More towards the southern part of China; not as retroflexed as the Beijing dialect; it is also possible that it was a Taiwanese practicing reading with effort.' |
| F45-MEI | 捲舌音重，講話語氣有種壓迫感<br>'Strong retroflex sounds and speaking in an aggressive style.' |
| F57-XYZ | 像台灣人認真的念，但還是大陸人<br>'Sounds like a Taiwanese reciting with effort, but still she is a mainlander.' |

Table 5.12 Impressive responses from Beijing-Mandarin-listening participants

Due to the fact that more participants perceived the speaker as a Beijing Mandarin speaker than a non-Beijing Mandarin speaker despite presented with a traditional-character passage, it may seem that the conflict between the dialect and version of character did not pose any effect on the participants. However, when examining their response in detail, traces of the effect can still be found. For instance, even though these

participants ultimately determined that the speaker was a mainlander, they still commented that there is still possibility the passage was recited by a Taiwanese, but ‘with effort’ as stated by F40-IJK, F41-GHI, and F57-XYZ. The perceived effort put into producing the retroflex sounds were mentioned by M34-WXY, yet F25-ABC directly concluded that Taiwanese simply could not achieve such degree of retroflection (even with extra effort). F45-BCD mentioned that the strong intensity of retroflection would result in an aggressive style of speech. In short, all participants were aware of the differences in speech production of retroflex sibilants between Taiwan Mandarin and Beijing Mandarin speakers. With the presence of a traditional-character passage, speech perception of an actual Beijing Mandarin speaker may lead to the impression that the same speech production could have been replicated by a Taiwan Mandarin speaker with additional effort.

In comparison with the perceived place of origin of the Taiwan Mandarin speaker, the Beijing Mandarin speaker’s dialect was more accurately identified by the participants. While less than half of the participants who listened to the Taiwan Mandarin speaker were able to identify her place of origin, the majority of participants from the Beijing Mandarin speaker group were able to correctly place her as a Mainland Chinese. However, the participants who performed well in this speaker-identity-observation task do not necessarily overlap with those who achieved a higher correct identification rate in the word-identification tasks and vice versa.

## 5.3 Summary of findings

In the two speech perception experiments conducted for this dissertation, it was found from the word identification tasks that 72% of all participants scored an over 90% correct identification rate across all tasks with the other 28% mostly consisting of participants over fifty years old. Regarding the variation among the three pairs of target sounds, it was observed that participants responded to the stimuli with greater similarity between the affricate sibilant minimal pairs [tʂ, ts] and [tʂ$^{h}$, ts$^{h}$], but rather differently toward the fricative pair [ʂ, s]. For the photographic stimuli test in particular, participants generally tend to perceive alveolar phonemes as their corresponding retroflex phonemes more often

(with the exception of the [ʂ, s] pair) with the presence of the photographic stimuli than without and vice versa. A statistical significance of association between the choice of phoneme category and the existence of photographic stimuli was only found in perceiving the alveolar consonant /ts/.

As for the orthographic stimuli, a statistically significant relationship between the identification of retroflex phonemes and the variety of written Chinese characters was found for all participants with a Person's chi-square test of association. With a 95% confidence interval, the odds ratio estimated that with the presence of simplified Chinese characters, participants were at least 1.83 times more likely to identify a retroflex audio stimulus with the actual retroflex phoneme instead of its corresponding alveolar sound than with the presence of traditional Chinese characters.

It was found from comparing the word-identification task and the speaker-identity observation task that performing well in one task did not entail achieving better results in the other. Although not statistically tested, a conflict between the speaker's dialect and presented characters at discourse level did pose an effect on how participants reacted towards speaker-identity observation. Among the 75% of speakers who were assigned to the audio-visual stimuli conflicting groups, half of them could not accurately identify the speaker's place of origin, in which the variation in orthography should be accounted for to a certain degree. Finally, it was confirmed that Taiwan Mandarin speakers in the study were consciously aware of the differences between the two dialects at both segmental and suprasegmental levels, yet oftentimes they found it difficult to explain in details exactly what the difference were at the suprasegmental level.

# Chapter 6

## 6. Production study

The production study was focused on the variations in the realization of retroflex sibilants in Mandarin Chinese by Taiwan Mandarin speakers. The variations in the corresponding alveolar sibilants were also examined to compare with the retroflex series. The goal was to examine whether variation in formality and orthography (Chinese characters) had an effect on the speech production of the retroflex sibilant series. Two types of experiments conducted for the production study, story-telling and passage-reading. The story-telling experiment was consisted of two stories that participants were asked to describe what they saw after watching each of the two 3-minute animations. The animations did not feature any dialogues or subtitles, and thus participants were not primed by any verbal stimuli. The passage-reading experiment included two approximately 160-Chinese-character passages, one in traditional Chinese characters and the other in simplified Chinese characters. All participants read both of them in the same order with the traditional passage preceding the simplified passage.

### 6.1 Process of data analysis

In order to systematically compare the variation in the production of retroflex and alveolar sibilants by all participants, an acoustic analysis of frication of the sibilants was applied. The spectrum measure utilized for calculation was the center of gravity (COG), because it is (1) “useful for measuring the frequency characteristics of aperiodic sounds in speech” (Styler 2017), and (2) the most robust and frequently used parameter to distinguish between retroflex and alveolar fricatives (Chang & Shih 2015). The length of the front cavity is in negative correlation with the value of COG. Therefore, a shorter front cavity, e.g., alveolar series resulted from a more anterior constriction would raise the frequency at which major energy is concentrated in the fricative spectrum, and thus a higher COG value.

Since the spectral measurement was focused on the frication interval, the first step was to separate the frication from the stop closure of the affricates. There are two affricates /tʂ/ and /tʂʰ/ in the retroflex series as well as the alveolar series /ts/, /tsʰ/ and one affricate of each series is aspirated. Therefore, additional separation of the aspiration from the frication noise was required for the aspirated affricates /tʂʰ/ and /tsʰ/. Following the method of Chang & Shih (2015:56), "the aspiration portion was distinguished from the frication based on two criteria: (A) the aspiration shows a formant pattern that is similar to that of the following vowel, and (B) the aspiration has a reduction of the intensity of high-frequency signals." All segmenting procedures were manually completed with Praat (Boersma & Weenink 2008), a tool for acoustic phonetic analysis. Praat screen shots illustrating how affricates were segmented are provided in the following with examples from participant F19-TUV. Figures 6.1 and 6.2 are the waveforms of the character 之 *zhi* which is unaspirated, while figures 6.3 and 6.4 are the waveforms of the characters 處 *chu* which is aspirated.

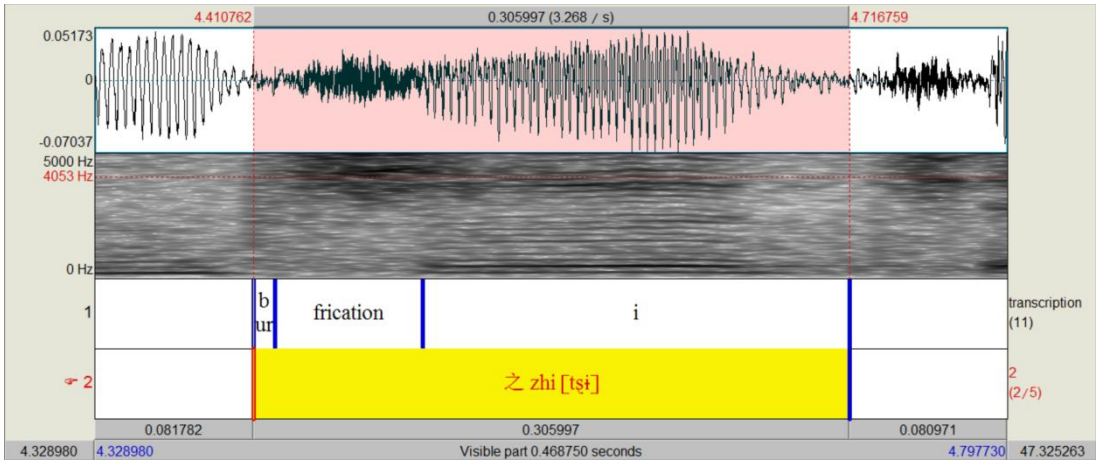



Figure 6.1 Waveform of the character 之 *zhi* in Praat (1)

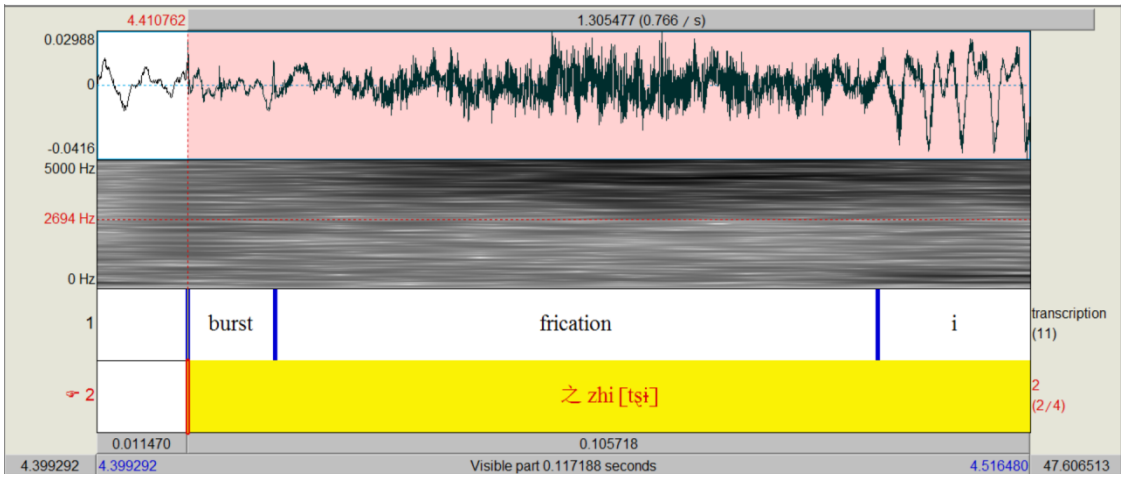



Figure 6.2 Waveform of the character 之 *zhi* in Praat (2)

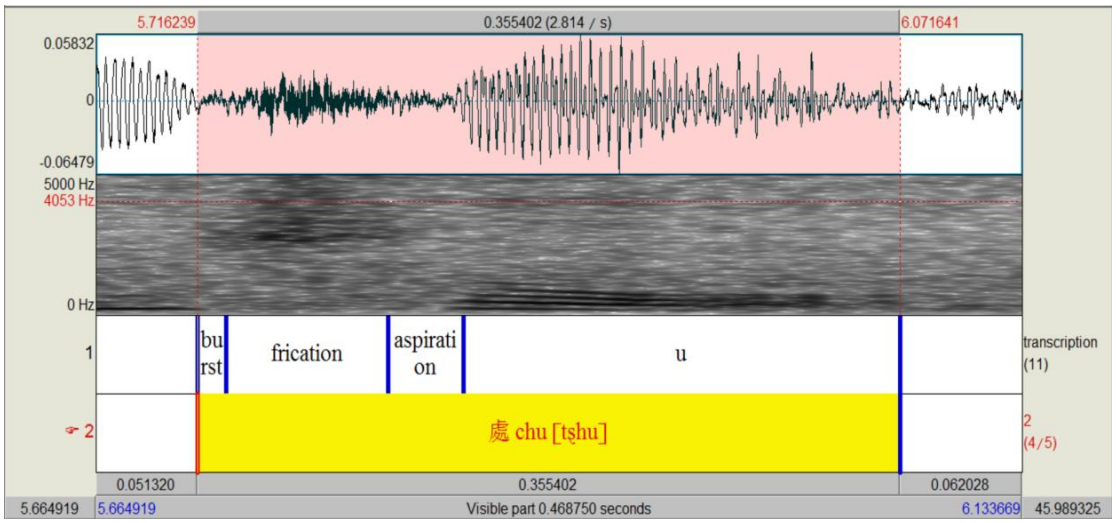



Figure 6.3 Waveform of the character 處 *chu* in Praat (1)

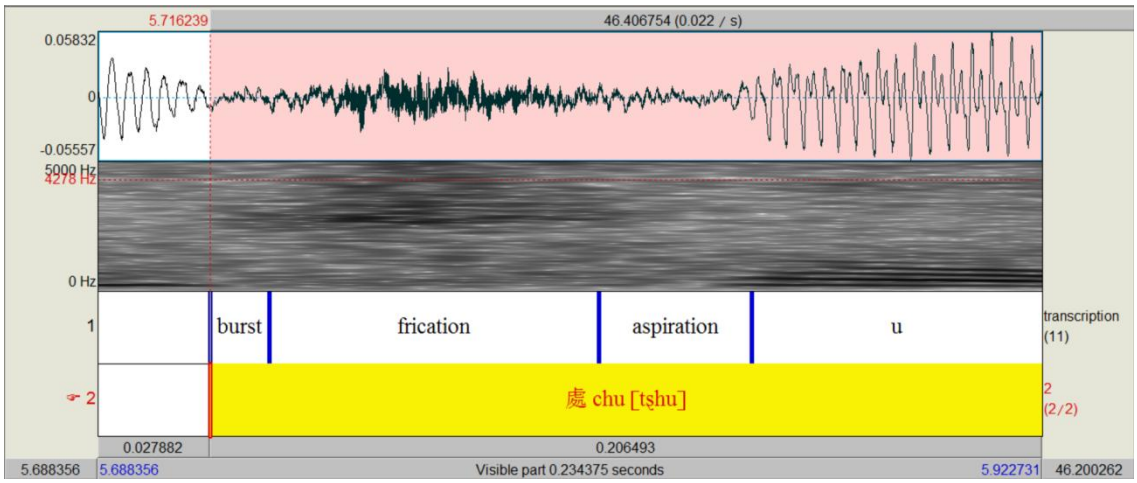



Figure 6.4 Waveform of the character 處 *chu* in Praat (2)

While the number of retroflex and alveolar tokens varied across participants in the story-telling tasks, there were a fixed number of tokens from the passage reading tasks. In the traditional-character passage reading task, there were 22 retroflex tokens and 11 alveolar tokens. As for the simplified-character passage reading task, there were 26 retroflex tokens and 12 alveolar tokens. In order to compare the tokens collected from the passage reading tasks to those of the story-telling tasks, the first 13 tokens of retroflex consonants and the first 6 tokens of alveolar consonants were extracted from each of the story-telling tasks for frication interval measurement. Therefore, the COG values were obtained from 74 retroflex sibilants and 35 alveolar sibilants for all participants. The spectral slice for measuring the COG of each token was taken from the middle 30 ms window of the total duration of frication noise.

A series of one-way repeated measures analysis of variation (ANOVA) was performed to compare the COG measures under different conditions. The mean values of the tokens were used to conduct the comparison. Hence each participant's tokens were summarized into six mean values according to formality, consonant series, and character conditions: mean COG of retroflex / alveolar consonants in casual speech, mean COG of retroflex / alveolar consonants in formal speech primed by traditional characters, and mean COG of retroflex / alveolar consonants in formal speech primed by simplified characters. Another set of measures that was utilized for comparison is the COG difference between retroflex and alveolar consonants under different conditions. In other

words, the relative distance in the place of articulation of retroflex and alveolar consonants during casual speech was compared to that of formal speech, and a similar comparison was made between the conditions of traditional character priming and simplified character priming.

## 6.2 Results from variation in formality

In order to examine whether the formality of speech had an effect on the production of retroflex sibilant in Mandarin Chinese, participants were first recorded while telling stories, and later recorded while reading a traditional-character passage, which served as examples of casual speech and formal speech respectively. The obtained COG measurements of the sibilant tokens were sorted and compared by age and gender.

An overall analysis of variance for repeated measures of tokens produced by all 17 male participants showed a significant difference between speech formality ($F(1,16) = 18.00$, $p < .05$). The mean COG of retroflex sibilants of all male participants was 7102.63 Hz during casual speech, which dropped to a mean of 6745.32 Hz in formal speech. A contrast on this difference was significant ($t(16) = 4.24$, $p < .05$). Using the standard deviation of contrast differences for each participant produced an effect size measure of $d = 1.03$, documenting the important effect of speech formality on the production of retroflex sibilants in Mandarin Chinese. The mean COG of retroflex productions seemed relatively higher than expected. Chang (2012) for example, arrived at a mean of 6224.1Hz (with a standard deviation of 1429.5) for all participants his study of Taiwan Mandarin retroflex sibilants. This relatively higher COG value was contributed by two groups of participants in this study, the male participants between 30-39 and participants over 50.

These two age groups included participants who had neutralized the retroflex sibilants and the alveolar consonants, and from the COG measurements, it can be concluded that they have substituted the retroflex phonemes with their alveolar counterparts. Table 6.1 provides a summary of the mean COG of each male participant under the two different formality conditions.

| | Mean COG of retroflex and alveolar sibilants | | | |
|---|---|---|---|---|
| | Casual speech (story-telling) | | Formal speech (traditional-character-passage reading) | |
| Participant | Retroflex sibilants | Alveolar sibilants | Retroflex sibilants | Alveolar sibilants |
| M14-TUV | 4911.54 | 10368.58 | 3975.41 | 9796.73 |
| M23-ZAB | 5151.50 | 9804.33 | 3821.55 | 9624.36 |
| M26-HIJ | 6655.46 | 8880.50 | 6104.18 | 8711.55 |
| Under 29 average | 5572.83 | 9684.47 | 4633.71 | 9377.55 |
| M31-NOP | 7916.38 | 7985.92 | 7856.86 | 8005.50 |
| M31-WXY | 6187.46 | 7919.92 | 5754.41 | 7713.91 |
| M32-HIJ | 8177.38 | 8442.08 | 7814.18 | 8316.56 |
| M32-WXY | 5495.46 | 7724.17 | 4855.14 | 7843.18 |
| M34-WXY | 7879.35 | 7741.25 | 7609.41 | 7655.00 |
| M37-WXY | 8577.88 | 8759.17 | 8590.55 | 8393.55 |
| 30-39 average | 7372.32 | 8095.42 | 7080.09 | 7987.95 |
| M43-MNO | 4664.46 | 7456.00 | 4338.64 | 7437.27 |
| M48-STU | 5485.62 | 7972.50 | 5251.50 | 8328.91 |
| M48-IJK | 8185.88 | 8269.80 | 7943.23 | 8220.36 |
| M49-XYZ | 7287.81 | 9071.86 | 7182.59 | 8929.73 |
| 40-49 average | 6405.94 | 8192.54 | 6178.99 | 8229.067 |
| M51-TUV | 8868.00 | 8916.17 | 8761.12 | 9071.73 |
| M53-TUV | 8892.65 | 8887.83 | 8688.36 | 9026.82 |
| M58-TUV | 7718.12 | 7783.67 | 7674.77 | 7850.00 |
| M58-WXY | 8689.69 | 8713.00 | 8448.59 | 8899.27 |
| 50 and over average | 8542.12 | 8575.17 | 8393.18 | 8711.96 |
| All male participants | 7102.63 | 8511.57 | 6745.32 | 8460.26 |

Table 6.1 Mean COG of male participants with variation in formality

It is illustrated in the table that the male participants over 50 years old in this experiment have virtually merged the retroflex sibilants with the alveolar sibilants. As for the 30-39 age group, due to the four participants (highlighted in gray) who have also merged the two sets of consonants, the group average COG was also relatively higher even though there were two other participants from this group who clearly made a distinction between retroflex and alveolar sibilants. However, a repeated measures ANOVA test still showed significant difference in the mean of COG under different formality conditions for both male participants between 30-39 ($F(1,5) = 8.72$, $p < .05$) and over 50 ($F(1,3) = 18.13$, $p < .05$). [10] All other age groups of male participants also produced retroflex sibilants with a significantly different mean value of COG.

[10] The age groups 19 and under and 20-29 were excluded from the tests as individual groups due to small sample sizes. However, when grouped together with the 30-39 age group, forming the group 39 and under, significant effect was found with $F(1,8) = 12.95$, $p < .05$.

A common language background of these participants highlighted in gray is that they all came from families that speak a home language other than Mandarin Chinese. M31-NOP and M32-HIJ have already been discussed previously stating that their home language was Min. M34-WXY and M37-WXY, who have spent the majority of their lives living in southern Taiwan also speak Min as their first language and it is also the language currently used most frequently. M58-WXY, who although has moved to Taipei since 18 also came from a family that spoke Min as the home language. He mentioned in his language background survey that even after living in Taipei for forty years now, he still felt more comfortable when people spoke with him in Min. As for the TUV brothers, the home language was the Shandong dialect of Mandarin. They currently use Taiwan Mandarin more often, yet when speaking with one another, the Shandong dialect is still preferred over Taiwan Mandarin. While Lin (1983) only claimed that speakers whose home language is Mandarin are more likely to retroflex than those whose home language is Min, the findings from this dissertation may be able to add to this statement that speakers whose home language is a Chinese dialect that is non-Taiwan Mandarin are more likely to neutralize the retroflex sibilants with alveolar sibilants than those who speak Taiwan Mandarin as a home language.

As for the mean difference between retroflex sibilants and alveolar sibilants under different formalities of speech, a significant effect was found for all the male participants as a whole. A repeated measures ANOVA showed that formality has an effect on the COG difference of these two consonant sets ($F(1,16) = 17.63$, $p < .05$). The mean COG difference of all male participants was 1408.95 Hz during casual speech production, which increased to a mean of 1714.94 Hz in formal speech production. A contrast on this difference was significant ($t(16) = 3.83$, $p < .05$). Using the standard deviation of contrast differences for each participant produced an effect size measure of $d = 0.93$, documenting the important effect of speech formality on the contrast between retroflex sibilants and alveolar sibilants in Mandarin Chinese for male participants of this study. What is worth noticing here is that this significance was mainly contributed by male participants of the 50 and older age group ($F(1,3) = 17.84$, $p < .05$ ) since no other age groups showed a significant effect.

A significant effect of speech formality on retroflex production was also found for

female participants of this experiment. An overall analysis of variance for repeated measures of tokens produced by all 26 female participants showed a significant difference between speech formality ($F(1,25) = 25.12$, $p < .05$). The mean COG of retroflex sibilants of all female participants was 5298.13 Hz during casual speech, which dropped to a mean of 5020.95 Hz in formal speech. A contrast on this difference was significant ($t(25) = 5.01$, $p < .05$). Using the standard deviation of contrast differences for each participant produced an effect size measure of $d = 1.42$, documenting the important effect of speech formality on the production of retroflex sibilants in Mandarin Chinese. This finding is in concordance with Lin (1983) and Rau & Li (1994) who both found that all speakers use significantly less retroflection in casual speech styles than in formal ones.

The mean values of COG found for female participants were similar to those of Chang (2012) since there were fewer female participants, compared to the number of male participants, who demonstrated a complete merger of retroflex and alveolar sibilants. Among the individual age groups, all groups except for the participants between 40-49 showed a significant effect of speech formality on the production of retroflex sibilants in Mandarin. The situation of the 40-49 age group will be discussed later with the results from the variation in orthography. There were three female participants who showed neutralizing of the two set of consonants were F50-VXY, F53-CDE, and, F58-JKL. Similar with the male participants who merged these sibilants, the three of them all spoke a home language that was not Taiwan Mandarin. F50-VXY and F58-JKL both came from Hakka-speaking families, and still currently live in Hsinchu County, Jhudong Township to be specific, which is an area where Hakka is used more frequently by locals than Mandarin. Like Min, there are no retroflex sibilants in the sound inventory of Hakka. Hence, the substitutions of alveolar consonant initials for retroflex ones were anticipated. F53-CDE, like her husband M58-WXY, came from a family that spoke Min as a home language. Even though she stated in her language background survey that she currently speaks Mandarin and Min for an equal amount of time, the effect of Min remains strong enough to motivate the merging of retroflex sibilants with alveolar sibilants.

The mean difference between COG of retroflex sibilants and their alveolar counterparts was also found to be significantly different for female participants under different formality conditions at ($F(1,25) = 39.88$, $p < .05$). The mean COG difference of

all female participants was 5073.16 Hz during casual speech production, which increased to a mean of 5347.25 Hz in formal speech production. These values are apparently higher than those of the male participants, which indicate that female speakers typically make a greater distinction between retroflex and alveolar sibilants. In addition, the distinction would be amplified when occurring in formal speech.

### 6.3 Results from variation in orthography

While the formality of speech showed an effect on the production of retroflex sibilants across gender and age group, the results from the variation in orthography indicated that whether participants were presented with traditional Chinese characters or simplified Chinese characters during speech production did not have an as strong effect as the formality of speech.

An overall analysis of variance for repeated measures of tokens produced by all 17 male participants showed a significant difference between written Chinese characters ($F(1,16) = 15.32$, $p < .05$). The mean COG of retroflex sibilants of all male participants was 6745.32 Hz when reading a traditional character passage, which dropped to a mean of 6596.55 Hz when reading a simplified character passage. A contrast on this difference was significant ($t(16) = 3.91$, $p < .05$). Using the standard deviation of contrast differences for each participant produced an effect size measure of $d = 0.95$, documenting the important effect of the type of written characters on the production of retroflex sibilants in Mandarin Chinese.

However, when examining the age groups of male participants in detail, it was found that the overall significant of variation in written character on the production of retroflex sibilants was solely contributed by the participants 29 and under at $F(1,2) = 132.58$, $p < .05$, and there were only three participants in this age group. None of the other individual age groups of male participants revealed a significant difference between seeing traditional and simplified characters. There was a general trend of lower COG values under the simplified-character condition, yet the difference was not great enough to yield a significant effect in a repeated measures ANOVA test. In comparison with the formality test, this result suggests that general male speakers of Taiwan Mandarin only

made particular effort in producing retroflex sibilants under conditions of formal speech, yet reading from a simplified-character passage did not provide a strong enough motivation for them to put in additional effort in the production of retroflex phonemes. The younger male participants of ages 29 and under did further distinguish their degree of retroflection between reading from a traditional-character passage and a simplified-character passage. However, since there were only three participants in this age group for this study, the result could not be generalized to all male speakers of Taiwan Mandarin.

As for the difference in mean COG value between retroflex and alveolar sibilants under different character conditions, no significant difference was found for any of the age groups, which indicated that the distinction made between retroflex and alveolar sibilants remained the same despite of the change in the perceived type of written Chinese characters.

Results from the COG values of the female participants on the other hand showed that female participants put slightly more effort into producing a retroflex consonant initial when presented with simplified characters than with traditional characters. An overall analysis of variance for repeated measures of tokens produced by all 26 female participants showed a significant difference between variants of Chinese characters ($F(1,25) = 17.25$, $p < .05$). The mean COG of retroflex sibilants of all female participants was 5020.95 Hz when reading a traditional-character passage, which dropped to a mean of 4875.44 Hz when reading a simplified-character passage. A contrast on this difference was significant ($t(25) = 4.15$, $p < .05$). Using the standard deviation of contrast differences for each participant produced an effect size measure of $d = 1.18$, documenting the important effect of the type of Chinese character on the production of retroflex sibilants in Mandarin Chinese. While the overall effect is significant across all female participants, there were only two individual age groups that that showed a significant effect of simplified characters on the production of retroflex sibilants: between 20-29 and 40-49. The younger female participants, similar with the younger male participants, seemed to further increase their degree of retroflection from reading a traditional-character passage to reading a simplified-character passage. It may be possible that the younger Taiwan Mandarin speakers are more sensitive to the differences between the represented dialects of these two types of Chinese characters.

As for the 40-49 age group, the result could be discussed in comparison with the variation in formality of speech. From the repeated measures ANOVA test on the effects of formality conditions of this group of female participants, no significance was found. However significance was found for the effect of simplified characters on the production of retroflex sibilants. It is reasonable to hypothesize that when they read the traditional-character passage after completing the story-telling tasks, they did not adjust their level of formality to a significantly different degree. However, when they were asked to read the following simplified-character passage, they were able to make a significant adjustment in the level of formality. Therefore, the significant effect achieved from reading the simplified Chinese characters may also suggest a modification in formality. Further research and a revised experiment design are required to finely distinguish between the formality effect and orthography effect that are both presumable present in reading the simplified-character passage.

The difference in mean COG value between retroflex and alveolar sibilants under different character conditions was found to be significant for all female participants at $F(1,25) = 4.96$, $p < .05$. However, when examining the individual age groups of female participants, it was found that this overall significance was solely contributed by the 20-29 age group, while all other age groups did not show a significant effect of character type on the difference of mean COG values between retroflex and alveolar sibilants. Hence, it can be concluded that only the younger female participants of this study tried to distinguish between retroflex and alveolar sibilants to a greater degree under the effect of variation in character type.

## 6.4 Summary of findings

The production study of this dissertation compared the speech production of Mandarin retroflex sibilants under two types of conditions, formality and character variation. Measurements of the center of gravity (COG), which measures the mean frequency of a spectrum, were obtained from the frication interval of all tokens. By comparing the mean COG of each participant under different conditions with a series of repeated measures ANOVA tests, it was generally found that the formality condition had a more apparent

effect on participants' production of retroflex sibilants than the character variation condition.

In more formal speech, passage reading in comparison to story-telling in this study, both male and female participants were found to increase their degree of retroflection as well as make a greater distinction between the retroflex consonant series and their alveolar counterparts. Also in more formal speech, but reading a passage in simplified characters rather than traditional characters did not cause significant effect to the production of retroflex sibilants or the distinction between to the two sets of sibilants. Even though an overall significance was found for both male and female participants, the effect was contributed by participants from younger age groups, while the other age groups did not show a significant difference between passage-reading in traditional and simplified characters.

# Chapter 7

## 7. General discussion and conclusions

This dissertation study took both a quantitative approach to discover statistically significant effects among variables from empirical data as well a qualitative approach to uncover participant's sociophonetic awareness of linguistic variation in Taiwan Mandarin with focus on the perception and production of retroflex sibilants in Mandarin Chinese. This chapter will first summarize the quantitative findings observed in the perception and production experiments, and will then discuss the quantitative results presented in the previous two chapters in relation with additional qualitative data that were acquired during the course of interview with the participants. Discussion topics include the correlation between speech perception and production of Mandarin retroflex sibilants under the priming of variation in written Chinese characters, the sociophonetics awareness of Taiwan Mandarin speakers, limitations of this study and suggested directions for future research.

Additional discussion on the cognitive processing of conflicting input information from two different modes is also provided with the framework of Paivio's (1971, 1986) dual-coding theory. Furthermore, the stylist variations found in this study is discussed under the framework of Eckert's (2012) three waves of variation studies. Finally, the implications of this dissertation will conclude the study.

### 7.1 Summary of the main findings

The role played by the variation in Chinese written characters was found to be much more important in the perception of retroflex sibilants than in the production of them. In the word-identification task of the perception study, a statistically significant relationship between the identification of retroflex phonemes and the variety of written Chinese characters was found for all participants with a Person's chi-square test of association. With a 95% confidence interval, the odds ratio estimated that with the presence of

simplified Chinese characters, participants were at least 1.83 times more likely to identify a retroflex audio stimulus with the actual retroflex phoneme instead of its corresponding alveolar sound than with the presence of traditional Chinese characters.

However, in the speech production study, minimal effect of variation in Chinese characters was found. Although an overall significance was found for both male and female participants, the effect was contributed by participants from younger age groups, while the other age groups did not show a significant difference between passage-reading in traditional and simplified characters. Due to the small sample size of each age group, it cannot be concluded that variation in Chinese characters has an effect on the production of Mandarin retroflex sibilants. Significant effects found from the production experiment were focused on the difference in formality of speech rather than variation in Chinese characters. It was found across age and gender that formal speech conditions would increase the degree of retroflection. However, the intensity of distinction between the production of retroflex and alveolar sibilants was only amplified for female participants and male participants over fifty years old.

Additional findings include behavioral differences between participants over and under fifty years old in this study. It was observed in both the perception and production experiments that participants younger than fifty years old regardless of gender (1) achieved a higher (around 90%) correct identification rate across all three word-identification tasks; and (2) showed a higher tendency of making clear distinctions between retroflex and alveolar sibilants during speech production. Both male and female participants older than fifty years old (1) scored lower correct identification rates in word-identification tasks; and (2) substituted more retroflex sibilants with alveolar phonemes.

Differences in relation to each pair of retroflex and alveolar phonemes were only observed in the speech perception study since they were not individually analyzed in the production study. From the word-identification task, it was found that participants responded to the stimuli with greater similarity between the affricate sibilant minimal pairs [tʂ, ts] and [tʂ$^{h}$, ts$^{h}$], but rather differently toward the fricative pair [ʂ, s]. For the photographic stimuli test in particular, participants generally tend to perceive alveolar phonemes as their corresponding retroflex phonemes more often (with the exception of

the [ʂ, s] pair) with the presence of the photographic stimuli than without and vice versa. This may be due to the essential differences between fricatives and affricates which is a potential topic of future research.

Finally, it was found from comparing the word-identification task and the speaker-identity observation task that performing well in one task did not entail achieving better results in the other. Among the 75% of speakers who were assigned to the audio-visual stimuli conflicting groups, half of them could not accurately identify the speaker's place of origin, in which the variation in orthography should be accounted for the confusion to a certain degree.

### 7.1.1 Speech perception-production correlation of retroflex sibilants of Taiwan Mandarin speakers

From the results of the perception experiments and production experiments it was found that there was no static correlation between the perception and production of Mandarin retroflex and alveolar sibilants. The only correlation that could be confirmed is that speakers who could not distinguish between retroflex and alveolar sibilants in speech perception could not produce distinct versions of the two phonemes either. However, some speakers who did not make a distinction in the speech production of these two sets of consonants could still achieve a high correct identification rate in speech perception. This observation is discordant with other near-merger studies in which speakers constantly differentiated two classes of sounds in speech production, yet failed to distinguish them during speech perception. Due to the small sample size in this dissertation, confirmation of this observation requires further investigation .

One example of asymmetric performance is M48-RST, who only made two mistakes during the whole set of word-identification task, but did not clearly distinguish between retroflex and alveolar sibilants in speech production. The average COG of all his tokens of retroflex phonemes and alveolar phonemes respectively in casual, formal with traditional characters, and formal with simplified characters were (8185.88, 8224.80), (7943.23, 8220.36), and (7940.22, 8190.83). There is a slight difference of approximately 200 Hz between the retroflex and alveolar sibilant production, yet to the ear, they sound

as if they have merged.

M48-RST grew up in a Hakka speaking environment, and mentioned in his language background survey that he spoke Hakka and Mandarin each for 50% of the time. He does not speak Min. His reactions when first seeing simplified characters in the word-identification task was 如果我看不懂怎麼辦? *rúguǒ wǒ kàn bù dǒng zěnmebàn* 'what if I don't understand (the writing)?' However, when reading through his background survey, he claimed that he formerly read simplified characters almost every day for work, but now he rarely reads any simplified characters. It is difficult to determine whether his question was a real question or simply hedging his disappointment of seeing simplified characters. It turned out that he only made two mistakes in all three word-identification tasks, and there was one mistake each for the photographic stimuli task and traditional-character stimuli task, but scored a 100% correct identification rate in the simplified-character stimuli task. This result could probably serve as proof that the previous question was only a signal of his disappointment for seeing simplified characters.

The question to raise here is that whether it was possible that he could actually make a clear distinction between the production of retroflex and alveolar sibilants, yet deliberately chose not to during speech production. At first this did not seem possible since speakers in general experiment settings still tend to speak as they would out of experiment conditions. However, as more participants made comments signaling that they were upset about seeing simplified characters, it was hypothesized that intentionally not making a clear distinction between retroflex and alveolar sibilants in speech production may be a plausible cause to the asymmetry between the speech perception and production of retroflex consonants for some speakers of this study.

While it was more difficult to intentionally make incorrect identifications of words during the speech perception study, the result of the production study was rather controlled by the participants themselves. If they purposefully chose not to produce a clearly retroflexed retroflex phoneme or vice versa, the researcher conducting the experiment may not be able to detect the intentions unless the researcher is acquainted with the participant in person beyond the experiment settings.

There were others who also reacted in a more negative way when seeing simplified characters as part of the experiment tasks. M49-XYZ first said 怎麼是簡體

*zěnme shì jiǎntǐ* ‘why is it (in) simplified (characters)’during the word-identification task. During the second appearance of simplified characters in the passage reading task, he said again 怎麼又是簡體 *zěnme yòushì jiǎntǐ* ‘why is it (in) simplified (characters) again,’ which is a clear signal that he was disappointed with the presence of simplified characters. In his case though, he still produced distinct phonemes of retroflex consonants.

A counter example was found in F52-DEF’s comment when she first came across simplified characters in the perception tasks. She said 是簡體 eh *shì jiǎn tǐ* ‘it’s simplified characters!’ “eh” serves the function as a discourse marker in Taiwan Mandarin, which is generally used to show excitement when discovering unanticipated entities or concepts. Again, just before reading the simplified-character passage, she said 來認真念囉 *lái rènzhēn niàn luō* ‘(I’m) going to seriously read it’. This reaction, which is clearly different from those of M48-RST and M49-XYZ, may also be a signal that she made more effort in producing retroflex sibilants when reading the passage than she would have normally done outside of experiment settings.

In sum, the only correlation between the perception and production of retroflex sibilants found in this study was that Taiwan Mandarin speakers who could not distinguish between retroflex and alveolar sibilants in speech perception could not produce distinct versions of the two phonemes either. However, if they could make a clear distinction in the perception of the phonemes, the outcome of speech production would depend on the participants’ conscious choice of speech variety.

### 7.1.2 Sociophonetic awareness of Taiwan Mandarin speakers

The phenomenon mentioned in the discussion of the possible asymmetric correlation between the speech production and perception of Mandarin Chinese retroflex sibilants may be an indicator that speakers are aware of the fact that producing extensively retroflexed retroflex sibilant could be labeled as sounding like a Beijing Mandarin speaker, which is a generally disfavored comment for most Taiwan Mandarin speakers. Most participants when asking to describe the so-called standard Mandarin in their perspective would respond with answers such as, pronouncing the retroflex sounds clearly, making a distinction between [an] and [aŋ], producing clear intonation contour,

etc. However, when further asked if pronouncing the retroflex sibilants similar to that of a Beijing Mandarin speaker would meet their ideal standard Mandarin, the majority would say that that would be overdoing the retroflex pronunciation. Some mentioned that it is "one kind of standard," but just not the kind that Taiwan Mandarin speakers would appreciate, or even may make people feel uncomfortable.

M32-WXY replied in his answer that there should not be a Min accent in the ideal standard of Mandarin. The term he used was 台灣國語 *táiwān guóyǔ*, 'Taiwan Mandarin' in literal translation, but 'a Taiwanese Min accented Mandarin' would be a more understandable translation. M32-WXY and his brother M31-WXY What is worth mentioning is that M31-WXY and M32-WXY both produced well-distinct retroflex consonant initials during the speech production tasks as well as scored near perfect correct identification rates for the word-identification tasks in the speech perception study. Their parents M58-WXY and F53-CDE on the contrary scored lower than average in the word-identification tasks, and showed a merging of retroflex sibilants with the alveolar sibilants. During the interview, M32-WXY explained that although both parents spoke Min better than Mandarin, their home language was still Mandarin Chinese in hope that the next generation would not acquire the so-called 台灣國語 *táiwān guóyǔ*, 'Taiwan Mandarin' as explained in the above. This 台灣國語 *táiwān guóyǔ*, accent is a natural result from the contact of the two dialects, yet it is not an accent that represents the prestige or elite social class and far away from standard. Hence, it is not a positive comment to describe the accent of a person who typically speaks Min more often and better than Mandarin.

When participants were asked to describe the differences between Taiwan Mandarin and Beijing Mandarin, it was found that most of them were able to clearly point out segmental differences between the two dialects, while the segmental differences were generally referred to as 'accent'. In the speaker-identity-observation task, it was found that Taiwan Mandarin speakers have their own set of criteria to determine the so-called 'standard' Mandarin. There is no clear definition and probably little consensus among Taiwan Mandarin speakers to define 'standard' Mandarin since each person has their own ideal as to what degree of retroflection is not enough to be considered standard and to what degree would be considered as overdoing and being labeled as sounding like

a Beijing Mandarin speaker, let alone taking into consideration the many other distinguishing features between Taiwan Mandarin and Beijing Mandarin.

### 7.1.3 Limitations and future research

One limitation of this dissertation was not taking the differences in the vowel segments after the retroflex and alveolar sibilants into account in the analysis. The frequency of different occurrences was not controlled during the perception experiment design due to the fact that the frequency of usage of the whole word was prioritized over the distribution of vowels following the target consonant phonemes. As for production differences, one study that examined how vowel contexts (/a/, /i/, and /u/) affect the place contrast between retroflex and alveolar sibilant realizations briefly stated that "the /u/ context was where both Beijing and Taiwan Mandarin speakers produced a smaller spectral contrast, as coarticulation from the rounded vowel lowers the COG in alveolar realizations and in turn decreases the spectral distance between the two phonemes" (Chang and Shih 2015:61). A larger place contrast was found with the /i/ context[11], comparing to the /a/ and /u/ vowel context. In their study, the participants were asked to read the same passage in the corresponding written character of that participant's spoken dialect of Mandarin Chinese. In this experiment, however, the difficulty to control the vowel context was due to (1) participants read two different passages in two different character settings, in which a full control of the vowel context may result in creating awkward reading passages; and (2) participants were also asked to tell stories with liberty in word choice that could not possibly be identical across all participants for comparison.

Time limitations have also typically been a concern in experiment design. In order to collect natural data that reflects the speaker's sociophonetics knowledge and

[11] The vowel representation /i/ is used here for the convenience of discussion even though the actual retroflex and dental sibilants cannot be followed by the high front vowel. /i/ is realized as a unique apical segment in Mandarin Chinese when preceded by retroflex and alveolar sibilants. However they are realized with significant difference in F2 and F3 between co-occurrence with retroflex and alveolar consonant initials (Lee 2011). The phonetic and phonological properties of these vowel segments following the consonant series remain controversial as demonstrated with the great variation in the naming of these segments. They are commonly referred as apical vowels (Karlgren 1915-16), fricative vowels (Ladefoged & Maddieson 1996, Ao 1997), Syllabic fricatives (Wiese 1997, Duanmu 2000), and syllabic approximants (Lee & Zee 2003).

awareness of social variations, a whole series of tasks often require more time than general participants could offer. The average time each participant spent on completing the full set of tasks in this study was approximately fifty to eighty minutes depending on how willing the participants were to stay and elaborate on their thoughts toward the use of Taiwan and Beijing dialect as well as the variation in written characters.

The original design of the experiments included a dubbing task, by which the goal was to elicit natural spoken data from participants impersonating characters of seemingly different social economic backgrounds. We would be able to observe from their speech production of how they perceive the language variation among different social groupings. It was expected during experiment design that participants would naturally imitate or even exaggerate the style and features of speech of the impersonated character during the task execution. However, after a number of trials, it was found that not all participants were willing to virtually 'act out' the speech. Some were too shy to do so in the presence of an unfamiliar interviewer, others required a significant amount of time to think and plan their speech production, and still others simply gave the impression that the task was absurd and showed no interest in seriously completing the task. This production task was expected to yield interesting findings of speaker's sociophonetics knowledge. Hence, if the experiment design could be further revised, participants may be willing to cooperate in future research.

Another limitation of this dissertation that could be further investigated in future research is the correlation between speech perception and speech production, yet time would still be the main concern. Large amount of data from the same person needs to be collected and analyzed from the same participant at several intervals over a period of time. Ideally, participants would be asked to complete repeated tasks with different priming at several time points to avoid the priming effect of the previous task during the same experiment session. In addition, the passage reading tasks were rather short in duration to elicit not only a large amount of data, but also productions by participants who have truly immersed in the simplified character setting, rather than simply a short one minute passage for the sake of an experiment. In this dissertation study, participants were presented with both traditional and simplified characters during the same session. The best possible way to avoid priming effect was to either randomize the order of trials for

the word-identification tasks or to have them read different articles for the passage-reading task in two types of characters. The data would be more credible if all conditions were controlled to be the same setting except for the one independent variable subject to treatment tasks. A strict paired t-test could have been performed to precisely examine the effects of variation in Chinese characters. However, it was not feasible to ask participants to set aside time for multiple experiment sessions during the time of data collection for this study.

In addition, a replicated study of the current dissertation could be conducted with speakers of other dialects of Mandarin Chinese, such as Beijing Mandarin speakers and speakers from Hong Kong and Singapore as well as second generation Mandarin speakers in non-Mandarin dominant regions around the world. The prevalent usage of Mandarin in the Asia-Pacific area and the rather complex historical and political background associated with the usage of various Mandarin dialects as well as written characters imply that there are abundant cultural and ideological values behind the usage of Mandarin. Examples can be drawn from contemporary movies and TV[12] series depicting inter-regional relationships among speakers of different Mandarin dialects, in which linguistic attitudes and speakers' sociophonetic knowledge of the association between social factors and linguistic variations are exemplified through conversations in the films. Written text including fiction, drama, news, and historical documents would probably also provide additional in-depth information toward how the variation in written Chinese characters have changed speakers' attitude toward the Mandarin Chinese language.

## 7.2 Dual-coding theory and orthography

The perception study of this experiment was designed based on the hypothesis that orthography, Chinese characters in this study, has an effect on speech perception of retroflex and alveolar sibilants in Mandarin Chinese. It was assumed that the participants were conscious of the sociophonetics variations in speech production of Mandarin, and

[12] Two examples of such movies are *撒嬌女人最好命* '*Women Who Flirt*' and 失戀 33 天 '*Love is Not Blind*'.

that they were also aware that the different types of characters were associated with different social indexing, or indicating speakers of different dialects. The two auditory-visual combinations that received particular interest were cases where conflict occurred due to disagreement of the social indexing by the auditory and visual stimuli: the Taiwan Mandarin dialect with simplified Chinese characters and the Beijing Mandarin dialect with traditional Chinese characters. As shown in the findings, approximately half of the participants in these two combination settings were misguided by the written characters and incorrectly identified the speaker as a person of a different dialect. While there is no direct explanation as to how the conflicts were processed in the minds of the participants, what is definite is that there were two modes of input occurring approximately at the same time. One theory that is focused on the information processing of two modes of coding is the dual-coding theory (Paivio 1971, 1986)[13].

The dual-coding theory (Paivio 1971, 1986) was originally used to explain the effectiveness of imageability in predicting memory recall performance. The theory has been proven effective and gathered popularity in pedagogy research of all subjects, and in instructions design for classroom teaching particular has benefited the most from this theory. 'Dual-coding' indicates two different modes of information encoding and processing: verbal associations and visual imagery. Verbal associations are basically language systems which include auditory and speech, while visual imagery associations include information such as picture, sound, taste, imagination, and other nonverbal thoughts. Paivio's (1971) theory developed as he observed the influence of imagery on memory recall performance and concluded that the recall of concepts is processed differently according to the essential differences of the concepts: concrete concepts can be recalled by relying on both verbal and imagery associations, while abstract concepts rely primarily on verbal associations.

The theory was further explained in detail in Paivio (1986), stating that it has a hierarchical conceptual structure of four levels: the unit level, the sensorimotor subsystems level, the verbal/nonverbal symbolic systems level, and the cognitive system level. "The operative cognitive mechanisms are assumed to be at the unit level" (Paivio

[13] Other theories, the schema theory for instance, should also be capable of explaining the similar cognitive process with variation in the taken perspective and details of analysis.

1986:54) and thus activation of higher level systems require specific stimuli that are relevant in context or task. There are two types of units: *logogens* and *imagens*, which correspond to the relevant upper sensorimotor subsystems level and respectively to the verbal and nonverbal symbolic systems level. Finally, at the highest level is the cognitive system serving a symbolic or representational function.

Paivio (1986:54) suggested three types of processing:

(1) Representational: the direct activation of verbal or non-verbal representations.

(2) Referential: the activation of the verbal system by the nonverbal system.

(3) Associative: the activation of representations within the same verbal or nonverbal system.

These different modes of processing may serve as a plausible explanation to how conflicting input from auditory and visual stimuli was processed in the minds of the participants during experiments of this dissertation. The key factor was to determine how the visual input of Chinese characters was perceived by the participants. In conditions where there were no conflicts, the dual-code processing would be representational or referential as illustrated in Figure 7.1 and 7.2. In the control condition that traditional characters were presented with a Taiwan Mandarin speaker recording, both auditory and visual inputs were viewed as verbal associations. The social indexing of the two inputs agreed with one another as both pointed to a Taiwan Mandarin speaker.

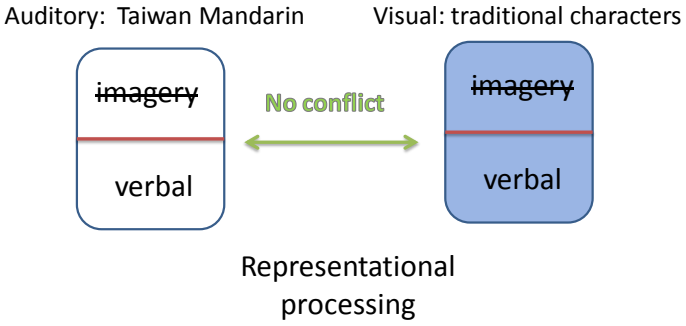



Figure 7.1 Representational processing when no conflict occurs in stimuli

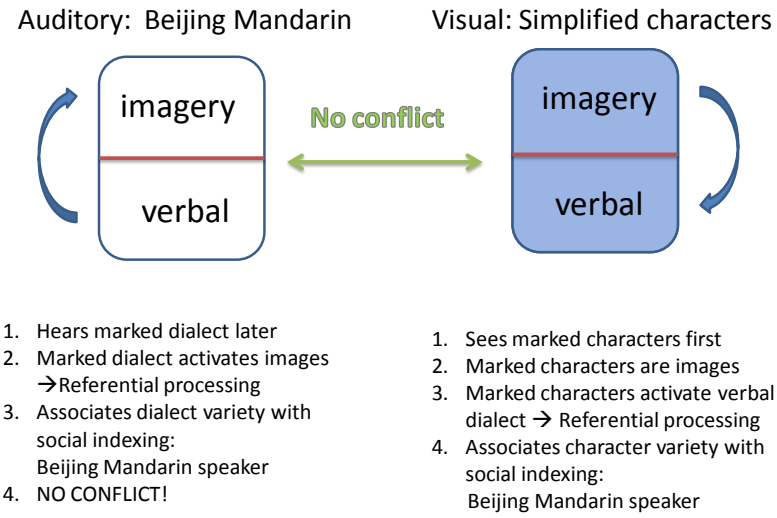



Figure 7.2 Referential processing when no conflict occurs in stimuli

As for the condition where simplified characters were presented with a Beijing Mandarin speaker recording, the process becomes slightly more complicated. It is hypothesized that the simplified characters would be first processed as imagery or simultaneously as both verbal and nonverbal information. This is possible because (1) comparing to traditional Chinese characters, simplified characters are the marked variety; (2) many participants claimed to not be used to reading simplified characters, in which cases these characters may be treated as blocks of images before processing them as verbal information; and (3) many participants mentioned during the interview that seeing simplified characters would make them think of the phonetic features of Beijing Mandarin, which could be perceived as an activation of the verbal system by the nonverbal system. All participants were presented with the transcription of the reading passage seconds before hearing the speaker reciting the passage. Hence, the processing of visual input should precede that of the auditory input. The auditory information of a Beijing Mandarin speaker, obviously verbal information, is also the marked variety to the participants of the experiment when compared to a Taiwan Mandarin speaker. Some participants also mentioned that simplified characters would come to mind when hearing the voice of a Beijing Mandarin speaker. Thus, it is also hypothesized that the marked dialect variety, the verbal information of the Beijing Mandarin dialect would activate the nonverbal system of the simplified character imagery. Since both the visual and auditory inputs indicate the identity of a Beijing Mandarin speaker, there generally is no conflict

between the two modes of information.

In conflicting situations, there is disagreement in the final identification of the speaker's dialect between the visual and auditory input. Figure 7.3 illustrates the dual coding process during which simplified characters were presented with a Taiwan Mandarin speaker recording. As previously mentioned, simplified characters are hypothesized to first be perceived as imagery rather than directly as verbal input. The soon activated verbal information of a Beijing Mandarin speaker would then be conflicting with the actual auditory stimuli of a Taiwan Mandarin speaker. Due to the occurring conflict in auditory expectations, it is hypothesized that the verbal information of the auditory input would also activate nonverbal imagery of traditional characters via referential processing. As a result, conflicts are present at both verbal and nonverbal modes, which in turn would create confusion for participants during the speaker-identity-observation task.

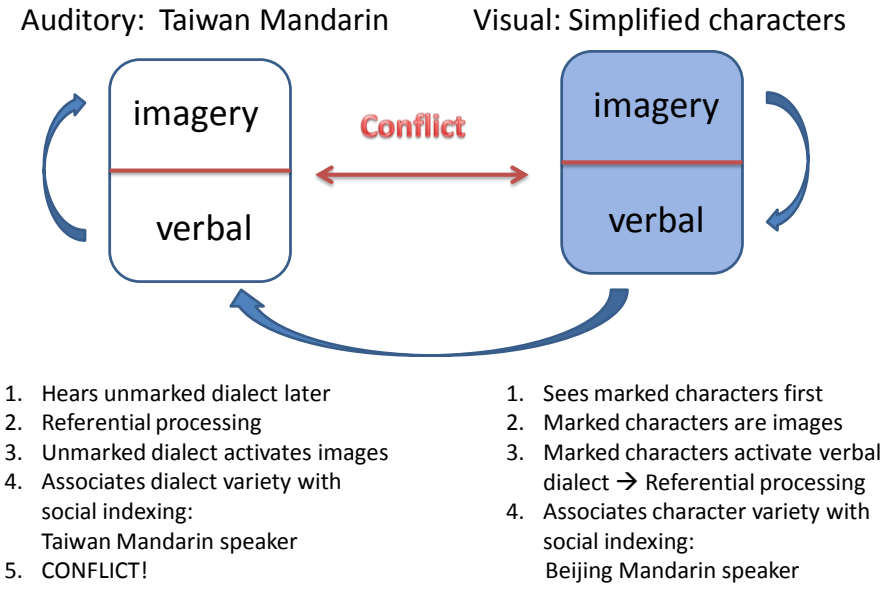



Figure 7.3 Marked characters initiates referential processing

A similar procedure of information processing is assumed to occur for the experiment setting of traditional characters presented with a Beijing Mandarin speaker recording as shown in Figure 7.4. The difference is that because the set of traditional characters is not the marked variety of characters, it may not necessarily activate the

nonverbal system at the beginning of information processing. If the referential processing were to occur, it is assumed to only take place after the participant has heard the auditory input of a Beijing Mandarin speaker, by which time the conflict has been initiated. The verbal input from traditional characters suggested a Taiwan Mandarin speaker, yet the provided recording was of a Beijing Mandarin speaker. Hence, conflict first occurs with the verbal system. As previously explained, participants have mentioned a connection between the hearing of Beijing Mandarin and the activation of simplified Chinese characters, which would result in the referential processing from verbal system to nonverbal system when viewing simplified characters as imagery instead of only verbal words. This activated information of simplified characters once again creates conflict in the imagery from visual input if it were activated by participants through referential processing.

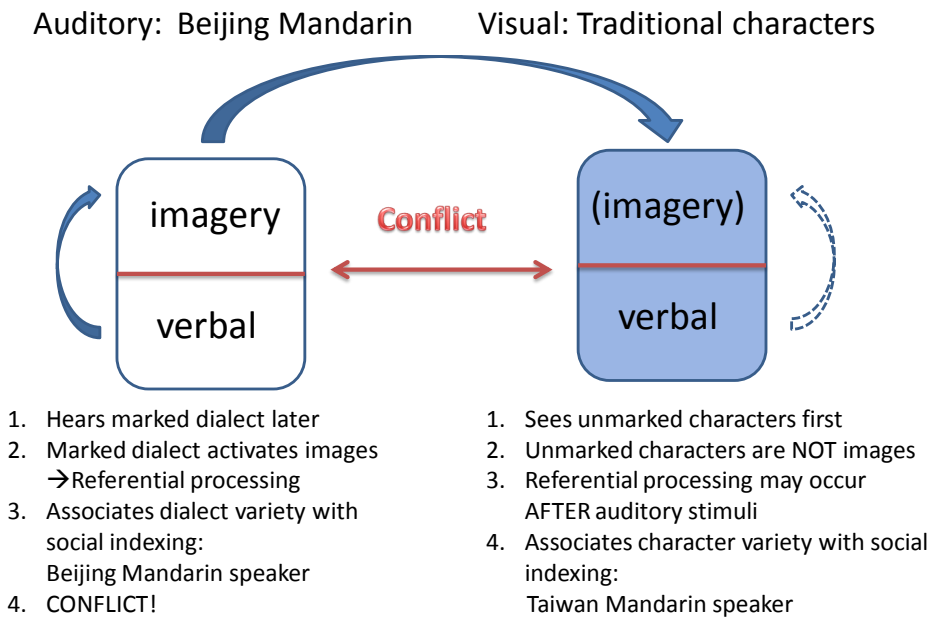



Figure 7.4 Unmarked traditional characters may initiate referential processing

In sum, how the Chinese characters are perceived by participants would influence the analysis of information processing of conflicting inputs. Whether Chinese characters are solely viewed as verbal input or both as verbal and nonverbal input would make a difference in applying the dual-coding theory to explain the conflict decoding process. It

is hypothesized here that for Taiwan Mandarin speakers, the marked variation of Chinese characters, simplified characters, are treated as nonverbal imagery that is capable of activating the relevant verbal information. The Beijing dialect, with the same logic, is also viewed as the marked variation of the two dialects, and thus would also activate relevant nonverbal information. However, what remains uncertain is that how participants arrived at the final decision of the speakers dialectal identify, and which input mode is a more signification determination factor to the ultimate decision. While the exact method for processing conflicting inputs of auditory and visual information remains to be further researched, the dual-coding theory may provide a reasonable hypothesis for preliminary explanation.

### 7.3 The third wave of variation studies

The study of variation and its association with social meaning has come a long way, and according to Eckert (2012) can be divided into three waves according to the focus of observation. The first wave in her categorization began with the Labovian variationist studies, addressing issues such as speech production across dialects, speech styles, and speech communities, etc. the characteristics of the first wave included focusing on (1) the socio-demographic features during observation and analysis instead of the within-speaker individual stylistic variations; (2) the arbitrary social classifications when making sense of variation rather than the dynamics of each speaker and their communities; and (3) the "pressures within the linguistic system" (Eckert 2012:90) when perceiving linguistic change instead of a speaker oriented point of view.

The second wave of variation studies, which occurred shortly after the beginning of the first wave, utilized an ethnographic approach in analyzing and explaining observations of linguistic variations with a shift of focus from the standard to the vernacular. Signature works of this wave include Milroy's (1980) social network theory and Eckert's (1989, 2000) "jocks" and "burnouts" ethnographic study of high school social categories and their networks. With the ethnographic approach, focus was given to more local categories rather than the sociologist's primary categories; therefore, providing more concrete connections between the speaker and observed variations.

The third wave of variation studies made a breakthrough of moving beyond the obvious static social categories to the speaker's immediately related social experience and stylistic choice of variant realization. One example study is that of Zhang's (2005, 2008) analysis of 'yuppies' in Beijing, wealthy young financial professionals who work in foreign-owned corporations. In order to contrast themselves with state-owned corporations, they developed a speech style that included various features of nonmainland Mandarin, which in particular were indices of Hong Kong or Taiwan speakers. They also stylistically avoided marked features of Beijing Mandarin to exclude themselves from the local social network and advanced to a more appreciated elite social status. As Eckert (2012:96) summarized "third-wave studies often begin with styles, seeking out what makes them distinctive, in an attempt to fill out the kinds of resources and meanings that give language its social life." Hence, the third wave of variation studies are more focused on speaker's self-oriented stylistic enregisterment in linguistic variation.

Although there was not a systematic observation of the association between variation based on stylistic choice and certain communities of speakers, a preliminary observation was that Taiwan Mandarin speakers were not only aware that simplified Chinese characters were associated with the Beijing Mandarin accent, but many also made an attempt to show their preferences toward this dialect. Those who probably disfavored the dialect often explicitly made negative comments about the presence of simplified characters in the tasks. On the other hand, those who may have favored or at least were not disappointed about the presence of simplified characters gave positive comments toward the unexpected presentation of simplified characters. It is however, unclear and remains to be further studied whether this approval or disapproval of the simplified characters were shown in their speech perception or speech production.

It is hypothesized that showing preference for character because of it social indexing is rather difficult in speech perception tasks, yet rather simple and easy to control during speech production tasks. For future research, the correlation between the preference of Chinese written characters and speech production of Beijing Mandarin featured phonemes, such as the retroflex consonants, could be further studied to see if there were stylistic choices made in the speech production based on preference for written characters.

## 7.4 Implications and conclusion

It was demonstrated via a series of speech perception and production experiments that the variation in Chinese written characters showed an effect on the speech perception of retroflex sibilants of Taiwan Mandarin speakers. The effect of characters on speech production was not as straightforward as that in perception. From the data collected in this study, minimal effect was found; however, when taking the speaker's attitude towards different varieties of characters into consideration, there may be an interesting effect that awaits to be discovered in future research.

It was found through the interview with participants of this study that Taiwan Mandarin speakers are fully aware of the variation in the production of retroflex sibilants. They are also aware of the association between simplified characters and the Beijing Mandarin dialect and this association was activated during the speech perception and production experiments of this dissertation.

This study adds to the finding of the research in sociophonetics variations that an asymmetry in speech production and speech perception may be a deliberate choice of the speaker instead of a result of unconscious perception and production of speech. In addition, using varieties of orthography as conditional stimuli creates the question as whether this type of stimuli should be considered as verbal or nonverbal. In the current study, it was proposed that the marked variant of the characters, which is the variant less familiar to the speaker, should be considered as nonverbal while the unmarked variant should be considered as verbal stimuli. The cognitive processing of two modes of input in this situation could be explained with the dual coding theory. Finally, this dissertation also shows that the abundant cultural and ideological values associated with the usage of Chinese written characters and spoken dialects are potential topics of future research.

**References**

Ao, Xiaoping B. 1997. *Phonetics and Phonology of Nantong Chinese*. Ph. D. Dissertation, The Ohio State University.

Bao, Zhiming. 1996. The syllable of Chinese. *Journal of Chinese Linguistics*, 24.2: 312-353.

Baran, Dominika Marta. 2007. *Linguistic Practice, Social Identity, and Ideology: Mandarin Variation in a Taipei County High School*. Ph.D. Dissertation, Harvard University.

Beebe, Leslie. 1981. Social and situational factors affecting communicative strategy of dialect code-switching. *International Journal of the Sociology of Language*, *32*, 139-149. Thakerar, J., & Giles, H. (1981). They are—so they speak: Noncontent speech stereotypes. *Language and Communication*.1: 251-256.

Boersma, Paul. and David Weenink. 2008. *Praat: Doing phonetics by computer* (v. 5.0.50) Amsterdam: Institute of Phonetic Sciences.

Campbell-Kibler, Kathryn. 2007. Accent, (ING), and the social logic of listener perceptions. *American Speech* 82: 32-64.

Cao, Wen. 2002. *Hanyu yuyin jiaocheng [Teaching manual of Chinese phonetics]*. Beijing: Yuyan Wenhua Daxue chubanshe.

Chang, Yung-Hsiang. and Chilin Shih. 2015. Place contrast enhancement: The case of the alveolar and retroflex sibilant production in two dialects of Mandarin. *Journal of Phonetics* 50: 52-66.

Chang, Yung-Hsiang. 2012. Variability in cross-dialectal production and perception of contrasting phonemes: the case of the alveolar-retroflex contrast in Beijing and Taiwan Mandarin. Ph.D Dissertation, University of Illinois at Urbana-Champaign.

Chao, Yuen-Ren. 1968. *A Grammar of Spoken Chinese*. Berkeley: University of California Press.

Chao, Yuen-Ren. 1948. *Mandarin Primer*. Cambridge: Harvard University Press.

Chao, Yuen-Ren. 1930. A system of tone letters. *Le maitre phonetique*.45: 24-27.

Chen, Ping. 1999. *Modern Chinese: History and sociolinguistics*. Cambridge: Cambridge University Press.

Cheng, Chin-Chuan. 1973. *A Synchronic Phonology of Mandarin Chinese*. The Hague: Mouton.

Cheng, Robert L. 1968. Tone sandhi in Taiwanese. *Linguistics, 41*, 19-42.

Cheng, Robert L. 1966. Mandarin Phonological Structure. *Journal of Linguistics 2.2:* 135–262.

Chung, Karen Steffen. 2006. Hypercorrection in Taiwan Mandarin. *Journal of Asian Pacific Communication* 16.2: 197-214.

Clopper, Cynthia G., Janet B. Pierrehumbert, and Terrin N. Tamati. 2008. Lexical bias in cross-dialect word recognition in noise. Paper presented at *Laboratory Phonology* 11.

Clopper, Cynthia G. and David B. Pisoni. 2004. Homebodies and army brats: Some effects of early linguistic experience and residential history on dialect categorization. *Language Variation and Change* 16: 31-48.

Cooper, Robert L.1975. Introduction to language attitudes II. *International Journal of the Sociology of Language* 6: 5-9

Drager, Katie. 2011. Speaker age and vowel perception. *Language and Speech* 54.1: 99-121.

Drager, Katie. 2005. The influence of social characteristics on speech perception. MA thesis, University of Canterbury.

Duanmu, San. 2007. The *phonology of Standard Chinese* (2 ed). New York: Oxford University Press.

Duanmu, San. 2000. The *phonology of Standard Chinese*. New York: Oxford University Press.

Duanmu, San. 1990. *A formal study of syllable, tone, stress, and domain in Chinese languages*. PhD. Dissertation, MIT.

Eckert, Penelope. 2012. Three Waves of Variation Study: The Emergence of Meaning in the Study of Sociolinguistic Variation. *Annual Review of Anthropology*, 41: 87-100.

Eckert, Penelope. 2000. *Linguistic Variation as Social Practice*. Oxford: Blackwell.

Eckert, Penelope. 1989. *Jocks and Burnouts: Social Categories and Identity in the High School*. New York: Teach. Coll. Press.

Graff, David, William Labov and Wendell A. Harris. 1986. Testing listeners' reactions to phonological markers of ethnic identity: A new method for sociolinguistic research.

In: D. Sankoff (ed.) *Diversity and Diachrony.* Benjamins, pp. 45-58.

Feifel, Karl-Eugen 1994. *Language attitudes in Taiwan: A social evaluation of language in social change.* Taipei, Taiwan: Crane.

Flanigan, Beverly Olson and Franklin Paul Norris. 2000. Cross-dialect comprehension as evidence for boundary mapping: Perceptions of the speech of Southeastern Ohio. *Language Variation and Change* 12: 175-201.

Fon, Janice, and Wen-Yu. Chiang. 1999. What does Chao have to say about tones? -- a case study of Taiwan Mandarin. *Journal of Chinese Linguistics* 27: 15-37.

Foulkes, Paul, Gerard J. Docherty, Ghada Khattab, and Malcah Yaeger-Dror. 2010. Sound judgments: perception of indexical features in children's speech . In Dennis R. Preston and Nancy A. Niedzielski (eds.) *A reader in sociophonetics*. Berlin: Mouton de Gruyter.

Fridland, Valerie, Kathryn Bartlett and Roger Kreuz. 2004. *Do you hear what I hear?* Experimental measurements of the perceptual salience of acoustically manipulated vowel by Southern speakers in Memphis, TN. *LVC* 16: 1-16.

Fridland, Valerie and Sohei Okamoto. 2009. The influence of regional vowel shifts on front vowel perception. Paper presented at *NWAV* 38. Ottawa: University of Ottawa.

Hay, Jennifer, Aaron Nolan and Katie Drager. 2006a. From Fush to Feesh: Exemplar Priming in Speech Perception. *The Linguistic Review* 23: 351-79.

Hay, Jennifer, Paul Warren and Katie Drager. 2006b. Factors influencing speech perception in the context of a merger-in-progress. *Journal of Phonetics* 34.4: 458–484.

Hay, Jennifer and Drager, Katie. 2010. Stuffed toys and speech perception. *Linguistics* 48.4: 865-892.

Huang, Shuanfan. 1993. *Yuyan, shehui, yu zuqun yishi: Taiwan yuyan shehuixue de yanjiu. [Language, society, and ethnic identity: A sociolinguistic study on Taiwan]*. Taipei, Taiwan: Crane.

Janson, Tore. 1986. Sound change in perception: An experiment. In J. Ohala & J. Jaeger (Eds.), *Experimental phonology* (pp. 253-260). Orlando, FL: Academic Press.

Janson, Tore and Richard Schulman. 1983. Non-distinctive features and their use. *Journal of Linguistics* 19: 321-336.

Koops, Christian. 2011. *Local Sociophonetic Knowledge in Speech Perception*. Ph.D Dissertation, Rice University.

Koops, Christian, Elizabeth Gentry, and Andrew Pantos. 2008. The effect of perceived speaker age on the perception of PIN and PEN vowels in Houston, Texas. *University of Pennsylvania Working papers in Linguistics* 34.2: 93-101.

Kubler, Cornelius. 1986. The Influence of Southern Min on the Mandarin of Taiwan. *Anthropological Linguistics* 27(2): 156-176.

Kubler, Cornelius. 1979. Some Differences between Taiwan Mandarin and 'Textbook Mandarin.' *Journal of the Chinese Language Teachers Association* 14.3: 27-40.

Hamann, S. 2002. Postalveolar fricatives in Slavic languages as retroflexes. In S. Baauw, M. Huiskes & M. Shoorlemme (Eds.), *OTS Yearbook 2002* (pp. 105-127). Utrecht: Utrecht Institute of Linguistics.

Hamann, Silke and Susanne Fuchs. 2010. Retroflexion of Voiced Stops: Databfrom Dhao, Thulung, Afar and German. *Language and Speech,* 53.2: 181-216.

Hamilton, Philip J. 1996. *Phonetic constraints and markedness in the phonotactics of Australian Aboriginal languages*. Toronto: University of Toronto.

Heniz, John M. and Kenneth N. Stevens. 1961. On the properties of voiceless fricative consonant. *Journal of Acoustical Society of America* 33:589-596.

Howie, John M. 1976. Acoustical studies of Mandarin vowels and tones. Cambridge: Cambridge University Press.

Hsueh, Frank. 1986. *Beijing yinxi jiexi [Analysis of Beijing sound system]*. Beijing: Beijing Yuyan Xueyuan Chubanshe.

Huang, James C.T., Y.H. Audrey Li, and Andrew Simpson. 2014. *The Handbook of Chinese Linguistics*. Oxford: Blackwell Publishing.

Huang, Shuanfan. 1993. *Yuyan, shehui, yu zuqun yishi: Taiwan yuyan shehuixue de yanjiu. [Language, society, and ethnic identity: A sociolinguistic study on Taiwan]*. Taipei, Taiwan: Crane.

Huang, Yi-Hsuan, E-Chin Wu, and Janice Fon. 2012. The effect of Min proficiency on production and perception of tones in Taiwan Mandarin. *Proceedings of Speech Prosody 2012*.

Hughes, George W. and Morris Halle. 1956. Spectral properties of fricative consonants.

*Journal of Acoustical Society of America 28*: 303-310.

Jeng, Jing-Yi. 2006. The acoustic spectral characteristics of retroflexed fricatives and affricates in Taiwan Mandarin. *Journal of Humanistic Studies* 40.1: 27-48.

Jongman, Allard, Ratree Wayland, and Serena Wong. 2000. Acoustic characteristics of English fricatives. *Journal of the Acoustical Society of America* 108.3: 1252-1263.

Karlgren, Bernhard. 1915-16. *Études sur la phonologie chinoise*. Uppsala: K. W. Appelberg.

Labov, William. 2001. *The principles of language change. Vol. II: Social Factors*. Oxford: Basil Blackwell.

Labov, William. 1994. *The principles of language change*. *Vol. I: Internal Factors*. Oxford: Basil Blackwell.

Labov, William. 1966. The social stratification of English in New York City. Washington, DC: Center for Applied Linguistics.

Labov, William and Sharon Ash. 1997. Understanding Birmingham. In: C. Bernstein, T. Nunnally, and R. Sabino (eds.) *Language Variety in the South revisited*. Tuscaloosa: University of Alabama Press, pp. 508-573.

Labov, William, Mark Karen, and Corey Miller. 1991. Near-mergers and the suspension of phonemic contrast. *Language Variation and Change* 3: 33-74.

Labov, William, Malcah Yaeger, and Richard Steiner. 1972. *A Quantitative Study of Sound Change in Progress*. 2 Vols. Philadelphia: U.S. Regional Survey.

Lambert, Wallace E., R. C. Hodgson, R. C. Gardner, and S. Fillenbaum. 1960. Evaluational reactions to spoken languages. *Journal of Abnormal and Social Psychology* 60.1: 44-51.

Ladefoged, Peter. 2005. *A course in phonetics*. Boston: Wadsworth Publishing.

Ladefoged, Peter and Ian. Maddieson. 1996. *The sounds of the world's languages*. Oxford: Blackwell.

Ladefoged, Peter and Z. Wu. 1984. Places of Articulation: An Investigation of Pekinese Fricatives and Affricates. *Journal of Phonetics* 12: 267-289.

Lee, Chao-Yang, Yu Zhang, and Ximing Li. 2014. Acoustic characteristics of Mandarin voiceless fricatives. *Journal of Chinese Linguistics* 43:151-170.

Lee, Sang-Im. 2011. *An articulatory and acoustic investigation of Mandarin apical*

*vowels*. Poster presented at the 162nd meeting of the Acoustical Society of America.

Lee, Wai-Sum and Eric Zee. 2003. Illustrations of the IPA: Standard Chinese (Beijing). *Journal of the International Phonetic Association* 33.1: 109-112.

Levon, Erez. 2007. Sexuality in context: Variation and the sociolinguistic perception of identity. *Language in Society* 36: 533-554.

Levon, Erez. 2006. Hearing 'gay': Prosody, interpretation and the affective judgments of men's speech. *American Speech* 81: 56-78.

Li, Fangfang, Jan Edwards, and Mary Beckman. 2007. Spectral measures for sibilant fricatives of English, Japanese, and Mandarin Chinese. In *Proceedings of the XVIth international congress of phonetic sciences*.

Li, Jessica Min-Chieh. 1995. A Sociolinguistic Variation Study of Chinese Retroflex Initials /tʂ/, /tʂh/ and /ʂ/ in Taiwan. Province University, Taiwan: MA Dissertation.

Li, Charles N. and Sandra A. Thompson. 1981. *Mandarin Chinese: A Functional Reference Grammar*. Berkeley: University of California Press.

Li. Wen-Chao. 1999. A diachronically-motivated segmental phonology of Mandarin Chinese. New York: Peter Lang

Lin, Chun Chung. 1983. A Sociolinguistic Study of the use of the Retroflex Sounds in Mandarin in College Students in Taiwan. *Bulletin of the College of Arts and Letter, National Central University* 1:1-15.

Lin, Yen-Hwei. 2007. *The sounds of Chinese*. Cambridge: Cambridge University Press.

Lin, Yen-Hwei. 1989. *Autosegmental treatment of segmental processes in Chinese phonology*. Ph.D. Dissertation, University of Texas, Austin.

Luo, Changpei and Jun Wang. 1981. *Putong yuyinxue gangyao [Outline of general phonetics]*. Beijing: Shangwu Yinshuguan.

McGowan, Kevin B. 2011. The Role of Socioindexical Expectation in Speech Perception. PhD dissertation, University of Michigan.

Milroy, Lesley. 1980. *Language and Social Networks*. Oxford: Blackwell.

Niedzielski, Nancy. 1999. The effect of social information on the perception of sociolinguistic variables. *Journal of Language and Social Psychology* 18: 62-85.

Niedzielski, Nancy. 1997. The effect of social information on the phonetic perception of sociolinguistic variables. PhD dissertation, University of California, Santa Barbara.

Norman, Jerry. 1988. *Chinese*. Cambridge: Cambridge University Press.

Paivio, Allan. 1986. *Mental representations: a dual coding approach*. Oxford. England: Oxford University Press.

Paivio, Allan. 1971. *Imagery and verbal processes*. New York: Holt, Rinehart and Winston.

Plichta, Bartłomiej and Dennis R. Preston. 2005. The /ay/s have it: the perception of /ay/ as a North-South stereotype in US English. *Acta Linguistica Hafniensia* 37: 107-130.

Plichta, Bartłomiej and Brad Rakerd. 2010. Perceptions of /a/-–fronting across two Michigan dialects. In: D. Preston and N. Niedzielski (Eds.), *A Reader In Sociophonetics*. DeGruyter, pp. 223-239.

Pulleyblank, Edwin G. 1984. *Middle Chinese: a study in historical phonology*. Vancouver: Univeristy of British Columbia Press.

Rau, Der-hwa, and Ming-chieh Li. 1994. Phonological variation study of Chinese retroflexed initials (ts) (tsh) (s). Proceedings of the 4th International Conference on Teaching Chinese as a foreign Language, 345-366.

Shih, Chilin. 1995. Study of vowel variations for a Mandarin speech synthesizer. *Proceedings of Eurospeech '95*.

Stainton, Michael. 1999. The politics of Taiwan aboriginal origins. In M.A. Rubinstein (Ed.), *Taiwan: A new history*. New York: M.E. Sharpes, Inc, pp. 27-44.

Staum, Laura. 2008 Experimental investigations of sociolinguistic knowledge. Ph.D. Dissertation, Stanford University.

Sternberg, Robert J. 2003. Cognitive theory (3rd ed.). Belmont, CA: Thomson Wadsworth.

Strand, Elizabeth A. 1999. Uncovering the role of gender stereotypes in speech perception. *Journal of Language and Social Psychology* 18.1: 86-99.

Strand, Elizabeth A., and Keith Johnson. 1996. Gradient and visual speaker normalization in the perception of fricatives. In: D. Gibbon (Ed.) *Natural language processing and speech technology: Results of the 3rd KONVENS Conference, Bielefeld, October 1996*. Berlin: Mouton, pp. 14-26.

Styler, Will. 2017. *Using Praat for Linguistic Research* (v 1.7). Retrieved from http://savethevowels.org/praat/UsingPraatforLinguisticResearchLatest.pdf.

Su, Hsi-Yao. 2005. *Language Styling and Switching in Speech and Online Contexts: Identity and Language Ideologies in Taiwan.* Ph.D. Dissertation, University of Texas at Austin.

Svantesson, Jan-Olof. 1986. Acoustic analysis of Chinese fricatives and affricates. *Journal of Chinese Linguistics* 14.1: 53-70.

Thakerar, Jitendra, & Howard Giles. 1981. They are—so they speak: Noncontent speech stereotypes. *Language and Communication* 1: 251-256.

Thomas, Erik R. 2000. Spectral differences in /ai/ offsets conditioned by voicing of the following consonant. *Journal of Phonetics* 28: 1-25.

Thomas, Erik R., Jeffery Reaser. 2004. Delimiting perceptual cues used for ethnic labeling of African American and European American voices. J*ournal of Sociolinguistics*: 54-87.

Tse, John Kwock-Ping. 1998. *Tai$^{2}$-wan$^{1}$ di$^{4}$qu$^{1}$ nian$^{2}$quin$^{1}$ren$^{2}$* ㄓㄞˊ *yu* ㄚㄘㄣ *zhen$^{1}$de bu$^{4}$fen$^{1}$ma?* [Do the Young People of Taiwan Really Not Distinguish Between *zh-*, *ch-*, *sh-*, and *z-*, *c-*, *s-*?] *The World of Chinese Language* 90: 1-7.

Tso, Ru-Ping Ruby. 2013. Chinese Characters and Speech Perception: The Interplay Between Orthography and Listener Expectation. M.A. qualifying paper, Rice University.

Walker, Abby. 2007. The effect of phentic detail on perceived speaker age and social class. Proceedings of ICPhS XVI, Saarbrücken, 6-10 August 2007, pp. 1453-1456.

Wan, I-Ping and Jeri J. Jaeger. 2003. *The Phonological Representation of Taiwan Mandarin Vowels: A Psycholinguistic Study*. *Journal of East Asian Linguistics*, 12.3: 205-257.

Wei, Jennifer. 1994. To *–Er* is to Err: A Case of Code-switching in Standard Mandarin. *Indiana East Asian Working paper Series on Language and Politics in Modern China*, Paper #3, January.

Williams, Angie, Peter Garrett, and Nikolas Coupland. 1999. Dialect recognition. In: Dennis R. Preston (ed.) *Handbook of perceptual dialectology*. Vol.1 Philadelphia: Benjamins, pp. 345-358.

Willis, Clodius. 1972. Perception of vowel phonemes in Fort Erie, Ontario, Canada, and

Buffalo, New York: An application of synthetic vowel categorization tests to dialectology. *Journal of Speech and Hearing Research* 15: 246-255.

Wu, Zongji. 1992. *Xiandai hanyu yuyin gaiyao [Outline of Modern Chinese phonetics]*. Beijing: Huayu Jiaoxue Chubanshe.

Wu, Yuwen. 1994. *Mandarin Segmental Phonology*, PhD dissertation, University of Toronto.

Xu, Shirong. 1980. *Putonghua Yuyin Zhishi [Basics of Putonghua pronunciation]*. Beijing: Wenzi Gaige Chubanshe.

Zhang, Qing. 2008. Rhotacization and the "Beijing Smooth Operator": the social meaning of a linguistic variable. *Journal of Sociolinguistic*, 12: 201-222.

Zhang, Qing. 2005. A Chinese yuppie in Beijing: phonological variation and the construction of a new professional identity. *Language in Society*, 34.3:431-466.

**Appendix 1 List of interview questions for speaker-identity-observation task**

1. 請問你覺得講者念得好聽嗎? 'How does the speaker's recitation sound to you?'

2. 請問你記得鴉片戰爭事件的起因嗎? 'Do you remember the initial cause of The Opium War?'

3. 請問你大約多久會閱讀一次簡體字的文件或字幕?

'Approximately how often do you read documents or subtitles in simplified Chinese characters?'

4. 請問你覺得講者幾歲? 'How old do you think the speaker is?'

5. 請問你覺得講者的職業為何? 'What occupation do you think the speaker holds?'

6. 請問你覺得講者有多高? 'How tall do you think the speaker is'

7. 請問你覺得講者在哪裡長大?如何判斷? 'Where do you think the speaker is from and how do you decide?'

8. 請問你覺得講者的中文發音標準嗎? 'Do you think the speaker's Mandarin is standard?'

8a. 如回答標準則問，有沒有比你的小學國語老師還要標準?

'If yes, do you think it is even more standard than your *Guóyǔ* teacher from elementary school?'

8b. 如回答不標準則問，有沒有哪一位公眾人物是你覺得發音標準的?

'If no, are there any public figures that you believe speak standard Mandarin?'

8c. 你覺得標準的中文發音有哪些特質?

'Which do you think are the features of standard Mandarin?'

8d. 你覺得你自己的中文發音標準嗎?

'Do you perceive yourself as speaking standard Mandarin?'

9. 請問你覺得講者有多重? 'How much do you think the speaker weighs?'

10. 請問你覺得講者還會說其他語言嗎?如果會的話是哪些?

'Do you think the speaker speaks other languages? If yes, what other languages does she speak?'

**Appendix 2 News excerpt in simplified Chinese characters for speaker-identity-observation task**

## 鸦片战争特别报导：英国学者蓝诗玲的看法

蓝诗玲（Julia Lovell）发现，在 1920 年代以前的中国历史教育中，鸦片战争只是晚清政府疲于应付的诸多问题之一，而中国人对待这场战争的态度，也不是爱国主义这一种，而后来的论述为什么发生了变化？

1840 年，鸦片战争正式爆发。在中国历史教科书中，此役乃中国近代史的起点。战争最终失败了，不平等条约的签定，开启了中国的百年国耻。中国沦为半殖民地、半封建社会，中国人民为了争取民族独立与民主，展开了反抗“帝国主义及其走狗”的斗争，最终，历史选择了社会主义。

硝烟散尽 170 余年，每当这场战争被提及，国人心里都好像还会痛，“割地赔款”、“丧权辱国”等教科书里的说法已成为习惯性的表达。即便是曾经被清廷割让的香港回归祖国，也未必就彻底解开了这个历史的疙瘩。

1997 年，香港回归不久，22 岁的英国人蓝诗玲第一次踏足中国，到南京大学学习中文与历史。在这个中英两国签订《南京条约》的地方，她看了谢晋导演的电影《鸦片战争》，影片里全是脸谱化的形象——残忍、阴险、好色的帝国主义者，以及正义的、英勇抗战的中国人，把近代中国的屈辱演得淋漓尽致。

与此同时，在南京大学的课堂上，她的历史教授执意提醒这位英国公民——她的祖国曾在中国为毒品而打仗，这是一段极不道德的过去。

她还参观了《南京条约》史料陈列馆，见到了普通中国人的愤怒。这一切的体验，让她震惊于那场战争竟给中国留下了如此巨大的创伤。她说：“那是我第一次亲身体验到当代中国与其历史之间的密切关系。”在此之前，她只是在英国选择学习中文，上过一节讨论鸦片战争的课，在整个小学到大学，她对此一无所知。对她的国家而言，在帝国主义向全球狂飙突进的背景下，一百多年前与东方古国的这场冲突，相对于英国对印度或非洲的剥削，只是一个小插曲。

从那时起，蓝诗玲就决定探寻与还原一个尽量真实、全面的鸦片战争。

**英国学者看鸦片战争：强调国耻比较有中国特色**

2011-11-22 07:30 南方人物周刊

http://cul.cn.yahoo.com/ypen/20111122/712626.html

**Appendix 3 News excerpt in traditional Chinese characters for speaker-identity-observation task**

## 鴉片戰爭特別報導：英國學者藍詩玲的看法

藍詩玲（Julia Lovell）發現，在 1920 年代以前的中國歷史教育中，鴉片戰爭只是晚清政府疲於應付的諸多問題之一，而中國人對待這場戰爭的態度，也不是愛國主義這一種，而後來的論述為什麼發生了變化？

1840 年，鴉片戰爭正式爆發。在中國歷史教科書中，此役乃中國近代史的起點。戰爭最終失敗了，不平等條約的簽定，開啟了中國的百年國恥。中國淪為半殖民地、半封建社會，中國人民為了爭取民族獨立與民主，展開了反抗“帝國主義及其走狗”的鬥爭，最終，歷史選擇了社會主義。

硝煙散盡 170 餘年，每當這場戰爭被提及，國人心裡都好像還會痛，“割地賠款”、“喪權辱國”等教科書裡的說法已成為習慣性的表達。即便是曾經被清廷割讓的香港回歸祖國，也未必就徹底解開了這個歷史的疙瘩。

1997 年，香港回歸不久，22 歲的英國人藍詩玲第一次踏足中國，到南京大學學習中文與歷史。在這個中英兩國簽訂《南京條約》的地方，她看了謝晉導演的電影《鴉片戰爭》，影片裡全是臉譜化的形象——殘忍、陰險、好色的帝國主義者，以及正義的、英勇抗戰的中國人，把近代中國的屈辱演得淋漓盡致。

與此同時，在南京大學的課堂上，她的歷史教授執意提醒這位英國公民——她的祖國曾在中國為毒品而打仗，這是一段極不道德的過去。

她還參觀了《南京條約》史料陳列館，見到了普通中國人的憤怒。這一切的體驗，讓她震驚於那場戰爭竟給中國留下了如此巨大的創傷。她說：“那是我第一次親身體驗到當代中國與其歷史之間的密切關係。”在此之前，她只是在英國選擇學習中文，上過一節討論鴉片戰爭的課，在整個小學到大學，她對此一無所知。對她的國家而言，在帝國主義向全球狂飆突進的背景下，一百多年前與東方古國的這場衝突，相對於英國對印度或非洲的剝削，只是一個小插曲。

從那時起，藍詩玲就決定探尋與還原一個儘量真實、全面的鴉片戰爭。

**英國學者看鴉片戰爭：強調國恥比較有中國特色**

2011-11-22 07:30 南方人物周刊

http://cul.cn.yahoo.com/ypen/20111122/712626.html

**Appendix 4 Article for passage-reading task in traditional Chinese characters**

人與人之間的相處，需要長時間的經營。
rén yǔ rén zhī jiān de xiàng chù，xū yào zhǎng shí jiān de jīng yíng。

對於剛認識的人，一般人比較不會在言詞間透露自己私人的事，
duì yú gāng rèn shí de rén，yī bān rén bǐ jiào bù huì zài yán cí jiān tòu lù zì jǐ sī rén de shì，

反而會較為保守，並且不恣意表現個人的思緒。
fǎn ér huì jiào wéi bǎo shǒu，bìng qiě bù zì yì biǎo xiàn gè rén de sī xù。

如果是在長輩面前，說話會特別小心謹慎，
rú guǒ shì zài zhǎng bèi miàn qián，shuō huà huì tè bié xiǎo xīn jǐn shèn，

回答問題時，常常因為想呈現最好的一面，而可能會稍有延遲。
huídá wèn tíshí，cháng cháng yīn wéi xiǎng chéng xiàn zuì hǎo de yī miàn，ér kě néng huì shāo yǒu yán chí。

如果是平輩，生長環境與學經歷又相似，則很容易變成朋友，
rú guǒ shì píng bèi，shēng zhǎng huán jìng yǔ xué jīng lì yòu xiàng sì，zé hěn róng yì biàn chéng péng yǒu，

互通有無，做知識的交流
hù tōng yǒu wú，zuò zhī shí de jiāo liú。

**Appendix 5 Article for passage-reading task in simplified Chinese characters**

做为一国元首，治理好一个国家真的很不容易，
zuò wéi yī guó yuán shǒu，zhì lǐ hǎo yī gè guó jiā zhēn de hěn bù róng yì，

不论内政外交，都要能够面面俱到，还要时时体察民情。
bù lùn nèi zhèng wài jiāo，dōu yào néng gòu miàn miàn jù dào，hái yào shí shí tǐ chá mín qíng。

内政财务方面要懂得开源节流，避免出现财政赤字，
nèi zhèng cái wù fāng miàn yào dǒng dé kāi yuán jiē liú，bì miǎn chū xiàn cái zhèng chì zì，

外交方面除了需要出国访查，也要接待外国使节。
wài jiāo fāng miàn chú le xū yào chū guó fǎng chá，yě yào jiē dài wài guó shǐ jiē。

除此之外还要了解民情，与民亲近，
chú cǐ zhī wài hái yào le jiě mín qíng，yǔ mín qīn jìn，

公开露面并且向大家挥手致意时，
gōng kāi lù miàn bìng qiě xiàng dà jiā huī shǒu zhì yì shí，

虽不须自理妆容，因为有专业随行人员打理，
suī bù xū zì lǐ zhuāng róng，yīn wéi yǒu zhuān yè suí háng rén yuán dǎ lǐ，

但还是要随时注意自己的姿势与言词。
dàn hái shì yào suí shí zhù yì zì jǐ de zī shì yǔ yán cí。

**Appendix 6 Language Background Survey**

1. 我到今年 (2013) 的 12/31 那天會是_____歲。

2. 我是 □ 男生 □女生。

3. 我會說_____種語言。

4. 請列舉上題的語言，並標示學習的時間與熟稔的程度:

(1) _____ (語言名稱); _____歲 到 _____歲 (學習時間)

熟稔程度: 只會一點點＝1 2 3 4 5 6 7＝母語的程度
(請圈選)

(2) _____ (語言名稱); _____歲 到 _____歲 (學習時間)

熟稔程度: 只會一點點＝1 2 3 4 5 6 7＝母語的程度
(請圈選)

(3) _____ (語言名稱); _____歲 到 _____歲 (學習時間)

熟稔程度: 只會一點點＝1 2 3 4 5 6 7＝母語的程度
(請圈選)

(4) _____ (語言名稱); _____歲 到 _____歲 (學習時間)

熟稔程度: 只會一點點＝1 2 3 4 5 6 7＝母語的程度
(請圈選)

(5) _____ (語言名稱); _____歲 到 _____歲 (學習時間)

熟稔程度: 只會一點點＝1 2 3 4 5 6 7＝母語的程度
(請圈選)

(6) _____ (語言名稱); _____歲 到 _____歲 (學習時間)

熟稔程度: 只會一點點＝1　　2　　3　　4　　5　　6　　7＝母語的程度
(請圈選)

5.

我目前比較常使用的語言有：（請列舉語言使用機會的百分比，**總和應為 100%**）

_____ (語言名稱); _____ (%)　_____ (語言名稱); _____ (%)

_____ (語言名稱); _____ (%)　_____ (語言名稱); _____ (%)

_____ (語言名稱); _____ (%)　_____ (語言名稱); _____ (%)

6. 我出生的城市是_____。

7. 請**依序列舉**出生之後有住過**六個月以上**的城市，並標示居住的時間:

(1) _____ (城市名稱); _____歲 到 _____歲 (居住時間)

(2) _____ (城市名稱); _____歲 到 _____歲 (居住時間)

(3) _____ (城市名稱); _____歲 到 _____歲 (居住時間)

(4) _____ (城市名稱); _____歲 到 _____歲 (居住時間)

(5) _____ (城市名稱); _____歲 到 _____歲 (居住時間)

(6) _____ (城市名稱); _____歲 到 _____歲 (居住時間)

(7) _____ (城市名稱); _____歲 到 _____歲 (居住時間)

(8) _____ (城市名稱); _____歲 到 _____歲 (居住時間)

(9) _____ (城市名稱); _____歲 到 _____歲 (居住時間)

(10) _____ (城市名稱); _____歲 到 _____歲 (居住時間)

8. 我平時閱讀簡體字的頻率是：

☐ 從未看過簡體字　　☐年閱讀五次以下

☐ 半年閱讀五次以下　　☐個月會閱讀至少一次

☐ 每周會閱讀至少一次　　☐天會閱讀至少一次

9. 我平時聽到大陸人講中文的頻率是：(在影視媒體上聽到的也算)

☐ 從未聽過大陸人講話　　☐年聽到五次以下

☐ 半年聽到五次以下　　☐個月聽到至少一次

☐ 每周聽到至少一次　　☐天聽到至少一次

10. 我在分辨台灣人和大陸人講的中文時，我可以達到的準確度為：

完全分辨不出來＝1　2　3　4　5　6　7＝一定可以分辨得出來
(請圈選) ←——————————→

11. 請列舉三項你認為是台灣人的中文和大陸人的中文發音上的差異：

(1)

(2)

(3)

**Appendix 7 English translation of the language background survey**

1. How old will you be by December 31th, 2013?

2. Are you female or male?

3. How many languages do you speak?

   (This includes all the dialects of Mandarin Chinese. The participants were told that even if they could only speak a little of that language, they could still count it in and specify the proficiency in the following question.)

4. Please list the languages that you speak and provide the learning period for each language. In addition, on a scale of one to seven with seven being the proficiency of a native speaker, choose your self-evaluated proficiency of that language.

5. Based on all the languages you have listed above, please state the frequency of usage of each language. With a total of 100%, please provide the percentage of the usage frequency of each language. The final number should add up to 100.

6. Which city (of what country) were you born in?

7. If you have moved away from the city you were born in, please list all the cities that you have lived in for more than six consecutive months. Please also provide the age span of which you stayed in each listed city.

8. How often do you read simplified Chinese characters?
   This includes newspapers, novels, magazines, web-pages, product manuals, subtitles of movies, etc.
   Please choose one from the following six options: never, less than five times per year, less than five times per half a year, at least once a month, at least once a week, or at least once a day.

9. How often to you hear Beijing Mandarin Chinese?
   This includes face-to-face conversations, radio broadcast, movies, talk shows, soap operas, online media, etc.
   Please choose one from the following six options: never, less than five times per year, less than five times per half a year, at least once a month, at least once a week, or at least once a day.

10. On a scale from one to seven with seven being 100%, how accurate do you think you

can distinguish between the Taiwan Mandarin dialect and the Beijing Mandarin dialect?

11. Please list three features that you believe are the most notable differences between the Taiwan Mandarin dialect and the Beijing Mandarin dialect.

## Appendix 8 Summarized result of responses from target phonemes pairs /tʂ/ and /ts/

| Participant's age and gender | | Number of *incorrect* responses of target phonemes pairs /tʂ/ and /ts/ from all three tasks | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | 1A | 1B | 1C | /tʂ/ *zh* | 2A | 2B | 2C | /ts/ *z* |
| | | 知識 *zhīshí* | 治理 *zhìlǐ* | 致意 *zhìyì* | Total | 姿勢 *zīshì* | 自理 *zìlǐ* | 恣意 *zìyì* | Total |
| 19 and under | Male (1) | | | | | | | | |
| | P | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | T | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | S | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | Female (2) | | | | | | | | |
| | P | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| | T | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 |
| | S | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| 20-29 | Male (2) | | | | | | | | |
| | P | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | T | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | S | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | Female (7) | | | | | | | | |
| | P | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 1 |
| | T | 2 | 0 | 1 | 3 | 0 | 0 | 0 | 0 |
| | S | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 2 |
| 30-39 | Male (6) | | | | | | | | |
| | P | 1 | 1 | 0 | 2 | 1 | 2 | 3 | 6 |
| | T | 4 | 1 | 0 | 5 | 0 | 2 | 2 | 4 |
| | S | 3 | 2 | 0 | 5 | 0 | 3 | 3 | 6 |
| | Female (4) | | | | | | | | |
| | P | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | T | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | S | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 40-49 | Male (4) | | | | | | | | |
| | P | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 1 |
| | T | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 |
| | S | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 |
| | Female (6) | | | | | | | | |
| | P | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | T | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| | S | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 50 and older | Male (4) | | | | | | | | |
| | P | 2 | 1 | 0 | 3 | 2 | 2 | 4 | 8 |
| | T | 4 | 2 | 0 | 6 | 1 | 1 | 1 | 3 |
| | S | 3 | 1 | 0 | 4 | 3 | 0 | 0 | 3 |
| | Female (7) | | | | | | | | |
| | P | 2 | 1 | 1 | 4 | 2 | 1 | 3 | 6 |
| | T | 4 | 2 | 0 | 6 | 2 | 1 | 1 | 4 |
| | S | 0 | 1 | 3 | 4 | 2 | 1 | 1 | 4 |
| Total | Male (17) | | | | | | | | |
| | P | 4 | 2 | 0 | 6 | 3 | 4 | 8 | 15 |
| | T | 8 | 3 | 0 | 11 | 2 | 3 | 3 | 8 |
| | S | 6 | 3 | 0 | 9 | 4 | 3 | 3 | 10 |
| | Female (26) | | | | | | | | |
| | P | 4 | 1 | 1 | 6 | 3 | 1 | 3 | 7 |
| | T | 7 | 2 | 2 | 11 | 2 | 1 | 1 | 4 |
| | S | 1 | 1 | 3 | 5 | 2 | 2 | 2 | 6 |
| | All (43) | | | | | | | | |
| | P | 8 | 3 | 1 | 12 | 6 | 5 | 11 | 22 |
| | T | 15 | 5 | 2 | 22 | 4 | 4 | 4 | 12 |
| | S | 7 | 4 | 3 | 14 | 6 | 5 | 5 | 16 |

**Appendix 9 Summarized result of responses from target phonemes pairs /tʂh/ and /tsh/**

| Participant's age and gender | Number of *incorrect* responses of target phonemes pairs /tʂʰ/ and /tsʰ/ from all three tasks | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | 3A | 3B | 3C | /tʂʰ/ *ch* | 4A | 4B | 4C | /tsʰ/ *c* |
| | | | 延遲 *yánchí* | 池塘 *chítáng* | 赤字 *chìzì* | Total | 言詞 *yáncí* | 祠堂 *cítáng* | 刺字 *cìzì* | Total |
| | 19 and under | Male (1) | | | | | | | | |
| | | P | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | T | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | S | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Female (2) | | | | | | | | |
| | | P | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | T | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | S | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | 20-29 | Male (2) | | | | | | | | |
| | | P | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | T | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 1 |
| | | S | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Female (7) | | | | | | | | |
| | | P | | 0 | 0 | 0 | 1 | 0 | 0 | 1 |
| | | T | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 |
| | | S | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| | 30-39 | Male (6) | | | | | | | | |
| | | P | 2 | 0 | 0 | 2 | 2 | 2 | 3 | 7 |
| | | T | 2 | 0 | 2 | 4 | 1 | 2 | 2 | 5 |
| | | S | 1 | 1 | 0 | 2 | 0 | 3 | 2 | 5 |
| | | Female (4) | | | | | | | | |
| | | P | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | T | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | S | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | 40-49 | Male (4) | | | | | | | | |
| | | P | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | T | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | S | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Female (6) | | | | | | | | |
| | | P | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | T | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 |
| | | S | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | 50 and older | Male (4) | | | | | | | | |
| | | P | 2 | 2 | 1 | 5 | 3 | 2 | 3 | 8 |
| | | T | 3 | 1 | 1 | 5 | 1 | 4 | 3 | 8 |
| | | S | 0 | 1 | 1 | 2 | 2 | 3 | 3 | 8 |
| | | Female (7) | | | | | | | | |
| | | P | 2 | 1 | 0 | 3 | 1 | 1 | 3 | 5 |
| | | T | 2 | 2 | 1 | 5 | 0 | 1 | 3 | 4 |
| | | S | 1 | 1 | 2 | 4 | 1 | 0 | 4 | 5 |
| | Total | Male (17) | | | | | | | | |
| | | P | 4 | 2 | 1 | 7 | 5 | 4 | 6 | 15 |
| | | T | 5 | 1 | 4 | 10 | 2 | 6 | 5 | 13 |
| | | S | 1 | 2 | 1 | 4 | 2 | 6 | 5 | 13 |
| | | Female (26) | | | | | | | | |
| | | P | 2 | 1 | 0 | 3 | 2 | 1 | 3 | 6 |
| | | T | 2 | 2 | 2 | 6 | 1 | 1 | 3 | 5 |
| | | S | 1 | 1 | 2 | 4 | 1 | 0 | 6 | 7 |
| | | All (43) | | | | | | | | |
| | | P | 6 | 3 | 1 | 10 | 7 | 5 | 9 | 21 |
| | | T | 7 | 3 | 6 | 16 | 3 | 7 | 8 | 18 |
| | | S | 2 | 3 | 3 | 8 | 3 | 6 | 11 | 20 |

## Appendix 10 Summarized result of responses from target phonemes pairs /ʂ/ and /s/

| Participant's age and gender | | | Number of *incorrect* responses of target phonemes pairs /ʂ/ and /s/ from all three tasks | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | 5A | 5B | 5C | /ʂ/ *sh* | 6A | 6B | 6C | /s/ *s* |
| | | | 詩人 *shīrén* | 使節 *shǐjiē* | 相識 *xiàngshí* | Total | 私人 *sīrén* | 死結 *sǐjié* | 相似 *xiàngsì* | Total |
| 19 and under | Male (1) | | | | | | | | | |
| | P | | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | T | | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | S | | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | Female (2) | | | | | | | | | |
| | P | | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | T | | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | S | | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 20-29 | Male (2) | | | | | | | | | |
| | P | | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | T | | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | S | | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | Female (7) | | | | | | | | | |
| | P | | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 |
| | T | | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | S | | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 30-39 | Male (6) | | | | | | | | | |
| | P | | 1 | 1 | 0 | 2 | 2 | 1 | 1 | 4 |
| | T | | 1 | 0 | 2 | 3 | 3 | 3 | 1 | 7 |
| | S | | 2 | 0 | 1 | 3 | 2 | 2 | 1 | 5 |
| | Female (4) | | | | | | | | | |
| | P | | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 |
| | T | | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 |
| | S | | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 40-49 | Male (4) | | | | | | | | | |
| | P | | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | T | | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | S | | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | Female (6) | | | | | | | | | |
| | P | | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | T | | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | S | | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 50 and older | Male (4) | | | | | | | | | |
| | P | | 1 | 2 | 3 | 6 | 1 | 0 | 0 | 1 |
| | T | | 2 | 1 | 2 | 5 | 1 | 2 | 0 | 3 |
| | S | | 0 | 0 | 2 | 2 | 1 | 0 | 1 | 2 |
| | Female (7) | | | | | | | | | |
| | P | | 2 | 1 | 3 | 6 | 1 | 2 | 1 | 4 |
| | T | | 1 | 2 | 2 | 5 | 2 | 2 | 0 | 4 |
| | S | | 1 | 1 | 2 | 4 | 0 | 2 | 1 | 3 |
| Total | Male (17) | | | | | | | | | |
| | P | | 2 | 3 | 3 | 8 | 3 | 1 | 1 | 5 |
| | T | | 3 | 1 | 4 | 8 | 4 | 5 | 1 | 10 |
| | S | | 2 | 0 | 3 | 5 | 3 | 2 | 2 | 7 |
| | Female (26) | | | | | | | | | |
| | P | | 2 | 2 | 4 | 8 | 1 | 2 | 1 | 4 |
| | T | | 1 | 2 | 3 | 6 | 2 | 2 | 0 | 4 |
| | S | | 1 | 1 | 2 | 4 | 0 | 2 | 1 | 3 |
| | All (43) | | | | | | | | | |
| | P | | 4 | 5 | 7 | 16 | 4 | 3 | 2 | 9 |
| | T | | 4 | 3 | 7 | 14 | 6 | 7 | 1 | 14 |
| | S | | 3 | 1 | 5 | 9 | 3 | 4 | 3 | 10 |