國立臺灣大學文學院語言學研究所

碩士論文

Graduate Institute of Linguistics

College of Liberal Arts

National Taiwan University

Master Thesis

台灣華語自然語料中無聲捲舌音及齒音之聲學研究

An Acoustic Study on Voiceless Retroflex and Dental Sibilants in Taiwan Mandarin Spontaneous Speech

莊育穎

Yu-Ying Chuang

指導教授：馮怡蓁 博士

Advisor: Janice Fon, Ph.D.

中華民國 九十八 年 七 月

July, 2009

國立臺灣大學碩士學位論文
口試委員會審定書

台灣華語自然語料中無聲捲舌音及齒音之聲學研究
An Acoustic Study on Voiceless Retroflex and Dental Sibilants in Taiwan Mandarin Spontaneous Speech

本論文係莊育穎君（R95142001）在國立臺灣大學語言學研究所完成之碩士學位論文，於民國 98 年 7 月 17 日承下列考試委員審查通過及口試及格，特此證明

口試委員：

馮怡蓁（簽名）
（指導教授）

鄭靜宜

小笠原奈保美

誌謝

哆啦 A 夢有一個叫「任意門」的道具，可以帶我們到任何想去的地方。如果碩士學位是我欲前往之地，那這本碩士論文就是我的「門」。只不過，哆啦 A 夢的任意門只需輕轉門把，即可到達夢想之處，而我的這扇門，不僅不易開啟，走到門前的道路更是迂迴漫長且佈滿荊棘。

還記得三年前剛入學時，雄心壯志編織著美好遠景，這扇門就在伸手可及處。但是漸漸地，修課做研究佔據生活後，這扇門似乎不斷往後退，心理上及實際上皆離我越來越遠。等到真正著手進行論文的時候，猛一抬頭，才發現門已是在幾乎看不見的遠方，就好像是在那海拔四千公尺的玉山山頂上。幸運地，玉山雖不易攀爬，一路上有許多溫暖的手持續帶給我力量，支持我前進。

首先，要感謝台大語言所的所有師長，帶領我進入語言學研究的領域。尤其感謝我的指導教授馮怡蓁老師：謝謝老師是我的助力也是阻力，總是恐嚇中帶鼓勵，加油中挾威脅地提攜著我；謝謝老師看透且揪出了我做學問的劣根性，苦口婆心地讓我了解研究應有的方法及態度；謝謝老師耐心仔細地教導，不厭其煩地與我反覆討論，助我渡過瓶頸；謝謝老師啟發我對口語語料庫的興趣，指引我該前進的方向且讓我有往下個目標邁進的勇氣。

非常感謝我的兩位口試委員鄭靜宜老師及小笠原奈保美老師。謝謝兩位老師百忙之中不辭勞苦地參加口試，且給予我修改論文的指導及意見，使我在聲學量測及語音方面皆有更寬廣的認識及視野。另外，特別感謝中研院陳克健老師提供線上斷詞系統的使用，讓我能更有效率地完成論文。

感謝實驗室的大家。謝謝 Sally 學姊總是在一旁鼓勵我，替我加油打氣；謝謝 Shelly 學姊很有耐心地幫我聽語料及和我討論；謝謝 Chris 學姊貼心地聽我訴苦且每每提供有效的解決方法；謝謝盈潔和慧茹在語料的處理上幫了我許多忙。也感謝所上所有人好好的大家 — 乃欣學姊、芷誼學姊、文琦學姊、惠如學姊、季樺學姊、立心學姊、佩玥 — 謝謝大家在各方面支持著我完成論文。還要感謝一直和我在一起歡笑流淚的同學兼死黨們，Ruby、Angela、Susu、姚姚，有了大家的陪伴，這一路走來雖然艱辛卻不孤獨。

最後，我要感謝我的家人。謝謝阿公不斷地提醒我何時要畢業；謝謝爸爸媽媽無微不至的照顧、賠小心的呵護及風雨無阻的娃娃車接送；謝謝老哥總是給我繪圖設計上專業的建議；謝謝小貓喔逼適時地吵鬧替我趕走瞌睡蟲。大家的支持與關懷永遠是我最強而有力的後盾。

現在，終於我走到門前，手中緊握著鑰匙，到了要開啟這扇門的 moment 了。懷著感謝的心情，我將穿過這扇門，往另一扇更高更厚的門前進。

摘要

本研究旨在探討無聲捲舌音及齒音在台灣華語中的表現。過去的研究可以概括分為兩大類：以自然語料所做的社會語言學研究以及運用實驗材料所進行的聲學測量。由於研究材料和方法的限制，先前並未有從聲學角度探討自然語料的研究。有鑑於此，本研究以自然語料為研究材料，並且以聲學測量為研究方式來分析無聲捲舌音及齒音在台灣華語中的呈現。本研究共有十五位男女發音人，皆出生於台北或高雄。每位發音人參與約三十分鐘的訪談。在訪談語料中，所有的無聲捲舌音及齒音皆挑選出來並進行量測。本研究所採用的聲學測量是取頻譜的重心頻率值，且一併分析了四個因素，包括兩個社會因素：地區和性別，及兩個語言因素：重音程度及詞類。研究結果顯示，地區和性別皆有顯著影響。就性別因素而言，女生對於捲舌音和齒音的區分較為男生明顯。而就地區因素而言，男女生呈現不同的趨勢：台北女生的區分程度比高雄女生大，然而台北男生的區分程度卻比高雄男生小。至於兩個語言因素，結果亦發現兩者皆有顯著影響。在重音的情況下，捲舌音和齒音的區分程度較大；而相似的結果也出現在實詞的情況下。再者，研究結果顯示捲舌音化與齒音化這兩個音韻過程在台灣華語有著不同的功能。捲舌音化與社會因素(性別和地區)息息相關，而齒音化卻是與語言因素(重音程度和詞類)有相互關聯。進一步詮釋，本研究發現在台灣華語中，捲舌音化是呈現外在語言因素的不同；反之，齒音化能呈現語言內部的分類及差異。

關鍵詞：捲舌音，齒音，重音，詞類，自然語料，台灣華語

ABSTRACT

The present study investigated the realizations of voiceless retroflex and dental sibilants in Taiwan Mandarin. Past studies on this issue are mainly of two tracks – sociolinguistic studies on spontaneous speech and acoustic studies on experimental data. In this study we would like to examine sibilant realizations in spontaneous speech from the acoustic perspective. Fifteen speakers of both genders from two regions, Taipei and Kaohsiung, were recruited and each speaker contributed 30-minute-long speech data. All retroflex and dental sibilant tokens were labeled, and the centroid frequency of each sibilant was measured to determine its realization. Effects of four factors were looked into, including two social factors, region and gender, and two linguistic factors, prosodic prominence and word class. Results showed that both region and gender played determinant roles. Females generally made larger sibilant contrasts than males, but inconsistency was observed for cross-regional comparisons. While Taipei females distinguished sibilants better than Kaohsiung females, the opposite was observed for male speakers. As for the effects of linguistic factors, it was also found that sibilant realizations in Taiwan Mandarin were indeed subject to both prosodic prominence and word class. In particular, the strengthening effect was shown in linguistically prominent conditions – prosodically prominent and content word conditions, in which speakers tended to make greater sibilant distinctions. Our results further implicated distinctive functions for the processes of retroflexion and dentalization in Taiwan Mandarin. Retroflexion characterized speaker group discrepancies, while dentalization reflected different levels of linguistic prominence. In this regard, degrees of retroflexion were sensitive to extra-linguistic differences, whereas degrees of dentalization were sensitive to language-internal categorization.

Keywords: retroflex sibilant, dental sibilant, prosodic prominence, word class, spontaneous speech, Taiwan Mandarin

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1 INTRODUCTION

1.1 Linguistic background

With the spread of Chinese population, Mandarin has been brought to many areas around the world. Gradually, it develops into a number of dialects in different areas, due to factors such as geographical separation or contact with local languages. These dialects differ from Standard Beijing Mandarin, or *Putonghua*, in various linguistic aspects. Such a process of dialect formation is seen in many regions or areas where Chinese people populate.

Taiwan is of no exception. Before the arrival of Mandarin, Taiwan was essentially a multi-lingual society, owing to the composition of different ethnic groups. In the early twentieth century, a number of languages were spoken in Taiwan, including Min, Hakka, several indigenous languages, and even Japanese because of political occupation. In late 1940s, Mandarin officially came to Taiwan with the immigrants from Mainland China. At that time, in order to strengthen its authority, the government then implemented Mandarin-only policy, in which people were required to use Mandarin as the sole language in public occasions. Under such a social context, a lot of people in Taiwan started to learn and speak Mandarin, and eventually became bilinguals, or even trilinguals.

Among all the ethnic groups in Taiwan, the majority of people are of Min origin. To be specific, about 73.3% of people in Taiwan are Min people, 13% of them are Mainlanders,[1] 12% are Hakka, and 1.7% are Austro-Polynesian aborigines (S. Huang, 1993). A more recent demographic survey further revealed the slightly divergent compositions of ethnic groups in different regions (Yang, 2008). Take Taipei and Kaohsiung, the two metropolitan cities in Taiwan, for example. Figure 1.1 summarizes the distributions of the four major ethnic groups in these two cities. As can be seen, the general trend patterns similarly in both areas; that is, Min people have far bigger population than the other ethnic groups.

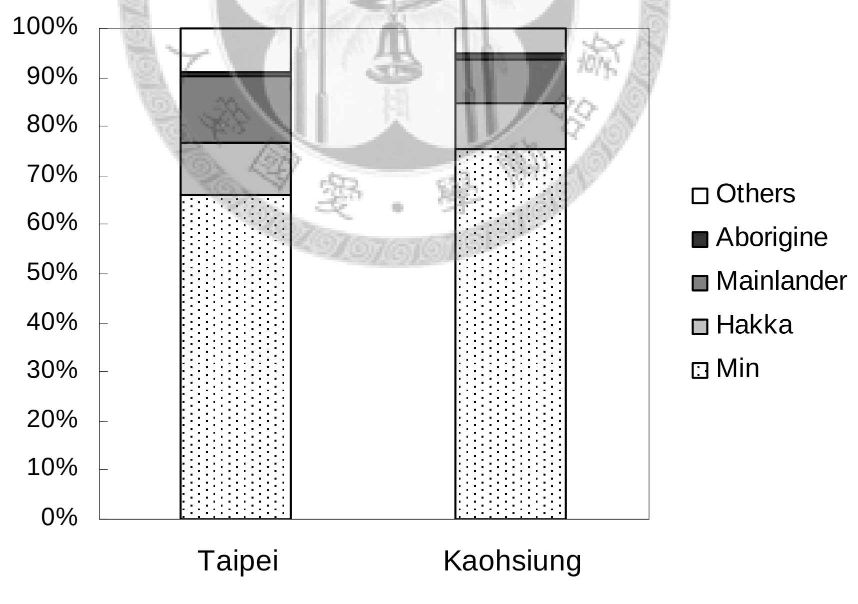


Figure 1.1 The demographic distributions of the four major ethnic groups (Min, Hakka, Mainlander, Aborigine) in Taipei and Kaohsiung (Yang, 2008).

[1] Mainlanders refer to those who came to Taiwan with the government from Mainland China after World War II.

Given the fact that a large number of people in Taiwan belong to the Min ethnic group, it is not surprising that Min is spoken much more frequently than the other local languages. In fact, about 80% of people in Taiwan are Mandarin-Min bilinguals (S. Huang, 1993). Regardless of the great number of Min speaking population in Taiwan, it should be noted that the actual usage of Mandarin and Min, nevertheless, was found to differ across regions in Taiwan (Ang, 1997). Generally speaking, Mandarin is more often spoken than Min in the Taipei area. Nevertheless, the usage of Min increases in cities southern to Taipei. In particular, the usage of Min even exceeds that of Mandarin in Southern Taiwan. It is a general trend that as one goes further south, the dominance of Min in the mixed code use is more frequently observed (Liao, 2000). These findings all convince us that Taiwan is by no means a linguistically homogeneous society, and the regional discrepancy in terms of Mandarin and Min usage characterizes one of such features.

1.2 Motivation

When Min native speakers learned to speak Mandarin, some linguistic features of Min would inevitably be carried over to their Mandarin speech. One of the most salient phonological features being identified at the early stage was the deretroflexion of retroflex sibilants. There are four retroflex sibilants in Mandarin, including three voiceless sibilants [tʂ], [tʂ^h], [ʂ], and one voiced sibilant [ʐ]. Specifically, deretroflexion depicts the process that voiceless retroflex sibilants [tʂ], [tʂ^h] and [ʂ], phones non-existent in Min, are substituted by voiceless dental sibilants [ts], [ts^h] and [s], phones existent in both Mandarin and Min. The substitution pattern for the voiced retroflex sibilant [ʐ], however, varies more, mainly due to the lack of a direct corresponding voiced non-retroflex sibilant in Mandarin and Min. Some common substituents include [z], [l] and [n] (Chan, 1984).

The realizations of retroflex sibilants, nevertheless, are actually much more complicated than mere substitution. Deretroflexion is so notorious a feature that in Chinese education, students in Taiwan are explicitly taught to learn the “standard” pronunciation. Starting from elementary schools, students are asked to pay extra attention to retroflex sounds when learning phonetic symbols[2]. It has always been highlighted that retroflex sounds are produced with the tongue curled up. Similar instructions can also be seen in the Mandarin phonetics textbook at college levels. Textbook demonstration of the “standard” retroflex articulation of [ʂ] was shown in Figure 1.2(a), which, in comparison with dental articulation of [s] in Figure 1.2(b), has a clearly curled-up tongue blade and further retracted place of articulation.

---

[2] The phonetic symbols here refer to *Zhuyinfuhao*, the sound transcription system officially used in Taiwan. Getting to know these symbols and making use of them are emphasized in the first two years of elementary school education.

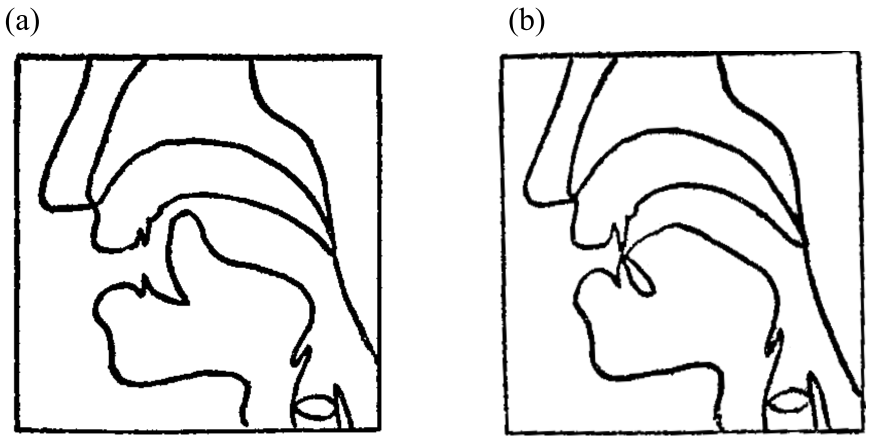


Figure 1.2 The production of (a) [ʂ] and (b) [s] from the Mandarin Phonetics Textbook (NTNU Mandarin phonetics committee, 2003).

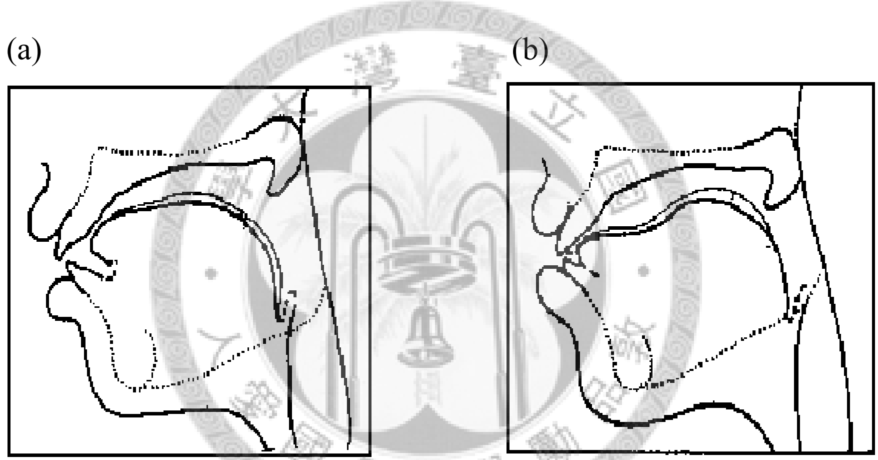


Figure 1.3 The X-ray slides of the production of (a) [ʂ] and (b) [s] from Ladefoged et al. (1984).

However, midsagittal X-ray data provided by Ladefoged et al. (1984) showed that even for Standard Beijing Mandarin, retroflex sounds are not produced with a curled-up tongue (see Figure 1.3). Instead, retroflex sibilants in Mandarin, just like dental sibilants, are produced with the upper surface of the tongue; in other words, the tongue is not curled up. In effect, the two sets of sibilants differ more crucially in terms of constriction position and tongue shape. With respect to such an official overcorrection of retroflex pronunciation in Taiwan Mandarin, the realizations of retroflex sibilants thus drew many researchers' attention, in terms of how and when Taiwan Mandarin speakers would use retroflex sibilants and also how the contrast between retroflex and dental sibilants was made.

Although a great number of studies have been conducted to investigate the realizations of retroflex sounds in Taiwan Mandarin from various perspectives, several gaps on this issue could still be observed. First of all, there is a gap of merging direction being studied. Because the substitution of retroflex sibilants with dental counterparts was first recognized as a salient feature, most studies focused on the deretroflexion process in Taiwan Mandarin (M.-C. Li, 1995; C. C. Lin, 1983; Rau & Li, 1994). Few, however, have paid attention to the realizations of dental sibilants. Although the general assumption is that dental sibilants are the unmarked segments and the process of turning marked into unmarked ones is linguistically universal, it is still interesting to investigate when and how such a process will be reversed. In the case of Taiwan Mandarin, the substitution of retroflex sibilants for dental ones has been observed from time to time (e.g., Chung, 2006). It is worthwhile to study the mechanism behind such a phenomenon.

Second, there is a gap of research materials and research methods. Early studies are mostly impressionistic, and results are often derived from perceptual observations

(Chan, 1984; Kubler, 1985; M.-C. Li, 1995; C. C. Lin, 1983; Rau & Li, 1994). However, sound perception is easily affected by various factors, such as ambient segments, individual voice quality, suprasegmental effects, etc. Later studies on this issue started to adopt a more objective way and acoustically measured retroflex and dental sibilants (Jeng, 2006; Tse, 1988, 1998). Nonetheless, the measurements are so far limited to experimental data. Under reading and citation conditions, subjects tend to be very aware of their own pronunciation, so retroflex and dental sibilants are usually found to be clearly distinguished. These totally different results thus create a mismatch between recent experimental studies and previous impressionistic ones on this issue.

Third, there is a gap of factors being examined. Even though the deretroflexion phenomenon has long been recognized and discussed, most studies are sociolinguistic research that centers on a number of extra-linguistic factors such as gender, social class, education level, etc. (M.-C. Li, 1995; C. C. Lin, 1983; Rau & Li, 1994). Few of them really focus on linguistic variables. Considering the fact that interview is the most frequent method for conducting sociolinguistic studies and also the fact that results are mostly derived from spontaneous speech in the interview, a lot of information regarding this issue will be masked if linguistic factors are not taken into consideration. For example, the copula verb *shi* is such a high-frequency word in Mandarin. Given its low semantic information and high frequency of use, it is predictable that its onset consonant will hardly be realized as the canonical retroflex sibilant [ʂ] in natural speech. Therefore, the results of simply averaging over the token numbers could possibly cause difficulty for interpretation.

Given the gaps noted above, in this study, we intend to investigate both dental and retroflex sibilants more thoroughly and completely, by acoustically measuring the sibilant realizations in spontaneous speech, with several linguistic factors being controlled for and looked into.

1.3 Aims of study

With regard to the gaps on retroflex and dental sibilants discussed above, we would like to investigate the realizations of both sibilants more thoroughly in the present study. Specifically, three research goals are tempted to be achieved.

Our first research goal is to adopt acoustic measurements to investigate the realizations of voiceless retroflex and dental sibilants in spontaneous speech. As mentioned previously, most of the past studies focused on the deretroflexion process. Moreover, due to limitation of research methods, a great number of sociolinguistic studies were based on impressionistic observations. Although several acoustic studies did appear later, they only examined experimental data. Therefore, in this study, we intend to objectively determine the issue of retroflex and dental sibilants in spontaneous speech by taking acoustic measurements. One thing to be noticed is that in order to adopt the same measurement to all sibilants, we exclude the voiced retroflex sibilant [ʐ] in this study mainly because it is no longer typically realized as a fricative in Taiwan Mandarin (Chan, 1984). Thus, the present study focuses on six voiceless retroflex and dental sibilants – [tʂ], [tʂ^h], [ʂ] and [ts], [ts^h], [s].

Second, social factors, gender and region, are of interest. Considering that gender is a common factor examined in many sociolinguistic studies, in this study we are particularly interested to see how the realizations of retroflex and dental sibilants are influenced by regional diversity, since it is a factor that has not yet been looked into on this issue. As aforementioned, Taipei differs from other regions in that the use of Mandarin dominates. On the contrary, the use of Min actually exceeds that of Mandarin in other southern cities. To examine regional differences, we thus select Taipei and Kaohsiung, the metropolitan city in Northern and Southern Taiwan, respectively. We would like to see whether different speaker groups have their specific realizations for retroflex and dental sibilants.

The third research goal is to see whether the realizations of retroflex and dental sibilants are subject to the influence of linguistic factors. In particular, we will examine one phonological factor – prosodic prominence, and one morpho-syntactic factor – word class. As commonly believed, prosodic prominence and word class in effect exhibit high correlation, *i.e.*, content words are more likely to receive prosodic prominence, while function words tend to be prosodically reduced. This leads to the assumption that prosodic prominence and content words signify the strengthening condition, whereas prosodic non-prominence and function words signify the reduction condition. However, exceptions can still be anticipated. For instance, function words do receive prominence under situations such as contrastive stress; on the other hand, content words are also very likely to be reduced to a great degree in the fast speech condition. In this regard, it is worthwhile to examine how the effect of prosodic prominence interact with that of word class, and it is predicted that different degrees of distinction between retroflex and dental sibilants can be observed in diverse conditions.

1.4 Significance

The significance of this study includes the following three aspects. First of all, although studying retroflex and dental sibilants in Taiwan Mandarin is not uncommon, or even seems to be out-dated, this study differs from previous ones in that it has better controls in both social and linguistic perspectives. A great number of social factors on this issue have been thoroughly examined. More often than not, the results, however, are hardly generalizable. Taiwan is in effect a heterogeneous society. Therefore, the lack of control regarding subject background, Min proficiency, etc. will introduce unwanted effects and thus make explanations difficult. Likewise, linguistic effects should not be neglected, especially when phonological variations are studied. Not all linguistic units are of equal status in a language, and our articulation will always reflect such linguistically internal discrepancies. Nonetheless, previous studies rarely, if not never, take into account the effects of linguistic factors. It is known that deretroflexion is a general phenomenon found in Mandarin spoken in Taiwan, but whether it is true in all speech contexts or exactly how this process interacts with other linguistic factors still remains unknown. Keeping these insufficiencies in mind, in this study we try our best to explore both social and linguistic factors in a more controlled way, in hope of obtaining more specific and less confounding results.

In addition, this study should be regarded as the first corpus-based study on the issue of retroflex and dental sibilants. Using data from a speech corpus has a number of advantages. For example, it is natural speech that is of interest. Natural speech is gaining attentions nowadays due to the fact that it better represents the authentic process of producing and perceiving language. Also, a speech corpus permits objective measurements. With the advancement of digital recording techniques, a lot of sound realizations can be acoustically determined, and thus, some flaws or inaccuracies of transcription-based analyses can be avoided. Most importantly, by using a speech corpus, a large number of data can be incorporated. A quantitative study has its merits, for it is much more representative and has higher generability. The present corpus-based study is thus considered pioneering in that it quantitatively and objectively looks into the realizations of retroflex and dental sibilants in spontaneous speech.

Finally, results of this study can benefit the burgeoning field of information processing and signal recognition. In recent years, people start to get interested in human-machine interactions. Therefore, machines are designed to both produce and perceive speech. Since our study focuses on different sound realizations in distinct linguistic categories, our results, if applicable, can help machines make more precise predictions and better recognition of human spontaneous speech. Moreover, if different linguistic behaviors are indeed consistently found between genders and among dialects, social factors can also be set as separate parameters to enhance the overall processing speed and accuracy. In this way, speech recognition techniques will definitely be better improved and more widely applied in various fields in the upcoming future.

1.5 Organization

The organization of the following chapters is as followed. In Chapter 2, reviews of literature on retroflex and dental sibilants in Taiwan Mandarin, along with a number of relevant issues, are discussed. Chapter 3 introduces the methods for conducting this study, in terms of data collection, data processing, and sibilant measurements. Chapter 4 presents the statistical analyses and results. Discussion is provided in Chapter 5, and conclusion in Chapter 6.

CHAPTER 2 LITERATURE REVIEW

In this chapter, past studies on retroflex and dental sibilants in Taiwan Mandarin are discussed. Section 2.1 summarizes the overall phonological variation, resulting from contact with Min, in Taiwan Mandarin. Section 2.2 reviews studies of voiceless retroflex and dental sibilants, in terms of their articulation and acoustic correlates. Sociolinguistic studies are presented in Section 2.3, in which the effects of gender, region, and the phenomena of hypercorrection are particularly highlighted. Section 2.4 reviews studies on prosodic prominence, and the effect of word class is discussed in Section 2.5.

2.1 Phonological variations in Taiwan Mandarin

Given the fact that about 80% of the people in Taiwan are Mandarin-Min bilinguals (S. Huang, 1993), Min, among all the local languages in Taiwan, should therefore be considered to play the most crucial role in shaping Mandarin and propelling the emergence of such a Mandarin dialect in Taiwan. Indeed, according to early research, a great number of linguistic characteristics being identified are specific to Min-accented Mandarin (Cheng, 1985; Kubler, 1985). In terms of phonology, some sounds in Standard Beijing Mandarin were found to be altered, substituted, added, or deleted (Ing, 1984, 1985). It was even reported that Mandarin in Taiwan is actually developing in a trend towards the resemblance with Min, from which we should be able to see the great influence of Min (Dong, 1995).

For decades, phonological variations in Taiwan Mandarin have been consistently brought up and extensively studied. Commonly noted variations include the monophthongization of diphthongs, the merging of syllable-final nasals, the disappearance of neutral tone and several sandhi processes, etc. (Dong, 1995; Ing, 1984; Kubler, 1985). As for consonant variations, deretroflexion is by far the most frequently discussed topic. Most research attributes deretroflexion to the influence of Min, which has only dental sibilants but no retroflex counterparts (Y.-H. Lin, 1988).

Lin (1983) took the lead to systematically investigate the realizations of retroflex initials, followed by Chan (1984), who conducted a sociolinguistic study on [ʐ] in Taiwan, and Rau and Li (1994) and Li (1995), who examined the phonological variants of [tʂ], [tʂ^h] and [ʂ]. On contrary to what has been observed (e.g., Kubler, 1985, etc.), in all these studies, retroflex sibilants are still present in the speech of Taiwan Mandarin speakers, though the occurrence strongly depends on various social factors. In other words, instead of total deretroflexion as observed in early studies, speakers, at least for those in latter studies, still have the phonological representations of retroflex sibilants. As a matter of fact, total deretroflexion should be better considered as one of the salient features of Taiwanese Mandarin, which is a stigmatized variety of Mandarin that is heavily Min-accented and spoken mainly by the aged native Min speakers who acquire Mandarin as their second language (Hsu, 2006). Taiwan Mandarin, on the other hand, should refer to the variety of Mandarin, still influenced by Min but not stigmatized, and spoken by comparatively younger generation in Taiwan. In our study, it is Taiwan Mandarin that is of interest.

2.2 Voiceless retroflex and dental sibilants in Mandarin

2.2.1 Articualation

There are three pairs of voiceless retroflex and dental sibilants in Mandarin. Their romanization letters and corresponding phonetic symbols are shown in Table 2.1. According to the articulatory data of [ʂ] and [s] (Ladefoged & Wu, 1984), contradictory to the general assumption that a curled-up tongue is involved in the production of retroflex sibilants, [ʂ] in Mandarin, just like its dental counterpart [s], is also produced with the upper surface of the tongue approaching the alveolar region. As shown in Figure 1.3, [ʂ] and [s] are distinguished in the following three aspects. First, [ʂ] clearly has a further back constriction point than [s]: the constriction is roughly in the center region of the alveolar ridge for [ʂ], while it is anterior to the alveolar ridge for [s]. In addition, sublingual cavity, the space under the tongue, is present in the production of [ʂ], but not [s]. Finally, the tongue is slightly grooved when [s] is produced. Unfortunately the articulatory data for the other two affricate pairs are not available, but assuming that they only differ from the fricative pair in manner of articulation, we should be able to predict that the general articulation patterns resemble that of fricative sibilants.

Table 2.1 The three pairs of retroflex and dental sibilants in Mandarin.

Place of articulation	Sibilant type		
	Aspirated affricate	Fricative	Unaspirated affricate
Retroflex	*ch* [tʂ^{h}]	*sh* [ʂ]	*sh* [tʂ]
Dental	*c* [ts^{h}]	*s* [s]	*s* [ts]

2.2.2 Acoustic measurements

In search of acoustic cues for classifying fricatives, early studies made use of spectral analyses. Examination of power spectra has yielded several reliable measurements, inclusive of frication duration, amplitude, and spectral peaks (Behrens & Blumstein, 1988; Hughes & Halle, 1956; Manrique & Massone, 1981). A more recent and common method is spectral moments analysis, which applies statistical calculations to the power spectra (Forrest, Weismer, Milenkovic, & Dougall, 1988). These moments are mean/centroid (M1), standard deviation (M2), skewness (M3) and kurtosis (M4). In particular relation to sibilants, it has been indicated that M1 is a very robust cue (Jongman, Wayland, & Wong, 2000). Specifically, the mean or centroid frequency is negatively correlated with the length of the front cavity — the shorter the cavity anterior to the constriction position, the more energy fall on the high-frequency region in the spectra, thus resulting in higher M1 value.

In Mandarin, there are three series of sibilants: dental, retroflex, and alvelopalatal. Acoustic studies also started early in order to identify their spectral characteristics (e.g., Svantesson, 1986; Tse, 1988). Recently, Li (2008) measured these three fricative sibilants with spectral moments analyses. The moment values were calculated from the 40 ms window in the middle of the frication, which is considered to be the most steady portion of the frication noise. Results showed that centroid frequency (M1) alone is sufficient to separate retroflex and dental sibilants, for they contrast crucially in place of articulation.

As for Taiwan Mandarin, Tse (1998) studied retroflex and dental sibilants in experimental settings, in order to determine how merged each sibilant pair is in the speech of young people in Taiwan. The measurement he took was the lower limit of the energy concentration region in the power spectra. His acoustic results demonstrated that young people still distinguish retroflex and dental sibilants in their speech, and the degree of distinction interacts with gender effect. Moreover, he also found that the unaspirated affricate [tʂ] merges with its dental counterpart [ts] to a greater extent than the other two retroflex sibilants. Jeng (2006) also examined the realizations of retroflex and dental sibilants acoustically. She took four spectral moments, along with duration measurements. It was also found that centroid frequency could effectively distinguish retroflex and dental sibilants. In addition, the aspirated affricate pair [tʂ^{h}] and [ts^{h}] have particularly lower M1 values than the other two pairs, owing to the fact that frication diminishes during aspiration. In terms of duration, the frication of retroflex sibilants is in general shorter than that of dental sibilants. Another interesting finding in this study was that context plays a role. That is, retroflex sibilants become less retroflexed in a more natural (question-answer) context than in a less natural (reading) context.

Possibly due to the feature of deretroflexion in Taiwan Mandarin, the issue on the degree of retroflexion particularly arouses the interests of researchers in Taiwan. The relationship between acoustic parameters and the degree of retroflexion has been discussed with experimental manipulations. Tse (1989) once measured the frequency (Hz) and intensity (dB) of the second formants of the mid back unrounded vowel [ɯ], an allophone that only occurs following retroflex sibilants in Mandarin. Tse hypothesized that the more retroflex a sibilant is, the higher the frequency and the stronger the intensity of [ɯ]'s F2 would be. Although his results were far from satisfactory due to the small number of subjects, a slight trend compatible with his hypothesis was still observed. In addition to measuring the following vowels, sibilants are directly measured as well. Jeng (2006) asked subjects to perform four degrees of retroflexion – the most retroflexed, moderately retroflexed, slightly retroflexed, and non-retroflexed. Spectral analyses indicated that the former three retroflex versions differ from each other only to a small extent. On the other hand, the non-retroflexed version is significantly distinguished from those retroflex ones. Such a result further implies that there is a continuum between the realizations of retroflex and dental sibilants, in which acoustic measurements are able to authentically reflect sibilants' place of articulation along this continuum.

2.3 Sociolinguistic studies on retroflex sibilants in Taiwan Mandarin

According to Labov (1972), during the course of language change, there are three stages involved in the change of language forms, which are the *origin* of the change, the *propagation* of the change, and the *completion* of the change. It is in the second stage (propagation) that more and more people start to adopt new variant forms, and it is also in this stage that "social significance is inevitably associated with the variant and with its opposition to the older form" (p. 123). As a result, to study language variation, a great number of studies start with the investigation of variants in social contexts. Frequently examined social factors for the issue of retroflex and dental sibilants include gender, age, socio-economic status, education level, etc.

2.3.1 Gender and region

Gender is the most frequently examined factor in most previous sociolinguistic studies, and it also has a very salient effect. For example, in Lin's (1983) study on college students in Taiwan, gender is found to be the strongest factor, in which the correct usage of retroflex sibilants for females greatly outnumbers that for males. Subsequently, Li (1995) examined the realizations of [tʂ], [tʂ^{h}], [ʂ] in Taiwan Mandarin. Instead of using binary retroflex/non-retroflex distinction, she recognized the intermediate variants of the three retroflex sibilants which, according to her, are the realizations somewhere between retroflex sibilants and dental sibilants. Her results suggested that gender effectively interacts with speech contexts. To be specific, women retroflex more in the formal speech, but they actually substitute dental variants for retroflex sibilants more often than men in the casual speech. Li explained her finding with the well-established sociolinguistic theory. That is, in general, women are comparatively conservative speakers and tend to use more standard forms than men. However, in the course of sound change, usually women, instead of men, play the leading role (e.g. Labov, 1990). As a result, it is not difficult to understand why more vernacular forms are found in women's daily conversation.

Compared to gender, region is a much less studied factor on this issue. The only study that touches upon its effect is Lin's study (1983). In particular, Lin roughly divided his subjects into Taipei group and non-Taipei group, based on the subjects' residency prior to the age of 14. Broad categorization this might be, the effect is rather salient. Regardless of gender, when speaking Mandarin, native Min speakers from Taipei have higher percentage of producing correct retroflex sibilants than those from other regions. Moreover, discrepancies were shown even for native Mandarin speakers. That is, native Mandarin speakers from Taipei also have more correct retroflex production than native Mandarin speakers from other areas. This finding illustrates that region is indeed a potentially effective factor that deserves being pursuing in greater detail.

2.3.2 Hypercorrection

Labov (1972) studied the /r/ pronunciation of New York residents in relation to their social classes. Significantly, he found that the usage of rhotic /r/ (considered to be the prestigious form) of people from the lower middle class exceeds that of people from the upper middle class in formal conditions. A more detailed investigation was done in Labov's another study (1990), in which he discussed the complicated relationship between gender and social class. Particularly, he discovered that females constitute the high percentage of using rhotic /r/ of the lower middle class New Yorkers. Males are also doing the same, though on a smaller scale. The lower middle class, as well as females, are both categorized as linguistically insecure groups, therefore propelling them to use more standard forms. Such insecurity acts upon speakers' consciousness level and urges them to hypercorrect their speech in order to gain security.

The phenomenon noted above is parallel to the realizations of retroflex sibilants in Taiwan Mandarin. To be specific, most past studies, either sociolinguistic or acoustic, all showed a strong gender effect in terms of using retroflex sibilants. In addition, retroflex production is found more often in the formal context, in the speech of people of higher education level and higher income, etc. All these could serve as evidence for the fact that retroflex sounds are considered to be prestigious in Taiwan Mandarin. Therefore, people hypercorrect their pronunciation by using more retroflex sibilant tokens.

In early studies, while most attentions were directed to the detretroflexion process, researchers did notice that some non-retroflex tokens are pronounced as retroflexed one, especially in reading tasks (e.g., M.-C. Li, 1995; C. C. Lin, 1983). Most researchers explained this by stating that Taiwan Mandarin speakers have both dental and retroflex repertoire in their phonological systems, but they simply activate them wrongly during their productions. Chung (2006), however, held a rather different perspective. Instead of having wrong activation, Chung regarded retroflexion of dental sounds to be a hypercorrection process. Because retroflex sibilants have always been prestigious in

Taiwan Mandarin, to obtain authority and mark formality, people hypercorrect their speech to such an extent that they even mistakenly retroflex dental sibilants. The hypercorrection phenomenon unravels Taiwan Mandarin speakers' attitudes towards retroflex and dental sibilants, which should be facilitating in explaining how they are realized and contrasted in natural speech.

2.4 Prosodic prominence

Studies on prosodic prominence started early, due to the fact that it serves crucial linguistic functions. In English, for example, prominence at the word level contrasts lexical meaning (e.g., INcline *vs.* inCLINE); at the sentence level, locations of prosodic prominence enable speakers to convey different meanings or elements of focus (e.g., he HIT John *vs.* HE hit John *vs.* he hit JOHN). To understand what constitutes perceptual saliency of prosodic prominent elements, Lieberman (1960), as an first attempt, measured a number of dimensions on stressed syllables, in which he identified three acoustic correlates, including fundamental frequency, amplitude and duration. While these three cues should be more or less considered to be closely associated with suprasegmental features, more recent studies start to pay attention to the relationship between prosodic structure and fine-grained phonetic details. Specifically, researchers are interested in how prominence, defined at the suprasegmental level, influences phonetic features at the segmental level.

One of the early identified cues in the segmental level is vowel quality. In particular, vowels are articulated with greater gestural efforts in stressed or accented positions. This effect has been saliently verified both from the articulatory perspective (e.g., Beckman & Edwards, 1994; de Jong, 1995) and from the acoustic perspective (e.g., Cho, 2005; van Bergem, 1993). Similarly for consonants, it is discovered that in prosodically prominent conditions, greater distinction could be observed. For example, in terms of voicing, voiced and voiceless stops are better contrasted in the accented condition, which is reflected in a number of acoustic measurements such as VOT as in English (Cole, Kim, Choi, & Hasegawa-Johnson, 2007), or degree of prevoicing in Dutch (Cho & McQueen, 2005). Such a prosodic effect on segmental realizations is commonly referred to as prosodic strengthening, depicting the phenomenon that linguistic contrasts are maximized or maintained in prosodically stronger conditions. Since stressed and accented syllables are indeed in prosodically strong positions, it is conceivable that the strengthening effect could be generally observed for both vocalic and consonantal segments.

As opposed to prominence, reduction is a stress level that generally refers to a diminishing process of acoustic cues resulting from articulartory economy in prosodically non-prominent position. One common phenomenon of reduction is the assimilation effect. Specifically, in reduced conditions, segments usually become more similar to their surrounding segments and they are also more likely to lose their original distinctive phonetic features. van Bergem (1993), for instance, suggested that vowels in non-stressed, unaccented syllables move towards a position similar to the preceding and following consonants, instead of merely centralizing. As for consonants, the reduction process is basically comparable with that of vowels (van Son & Pols, 1999). In the perceptual aspect, Duez (1995), in a study of voiced stops in French spontaneous speech, indicated that prosodic prominence had an effect on consonant identification. In particular, voiced stops occurring in non-prominent syllables are significantly less successfully recognized than those in prominent syllables.

As for prosodic prominence in Mandarin, Chao (1968) proposed three levels of stress, including contrasting stress, weak stress, and normal stress. Accordingly, contrastive stress signifies the condition where speakers intend to contrast certain elements in the sentence. For instance, in the sentence *Bushi Huang xiansheng, shi WANG xiansheng* 'It's not Mr. Huang; it's Mr. Wang.', it can be seen that the speaker puts emphasis on the surname *Wang*. The syllable to be contrasted is thus said to possess contrastive stress. Contrastive stress is generally realized with a wider pitch range and longer duration, and usually with increased loudness. Weak stress is particularly associated with neutral tone syllables by Chao. A great number of grammatical suffixes in Mandarin (e.g., *-de* 'possessive marker') are of neutral tones. Moreover, neutral tones also have the function of distinguishing lexical meaning (e.g., *dong*^{1}*xi*^{1} 'east and west' *vs. dong*^{1}*xi*^{0} 'thing'). According to Chao, the name "neutral tone" is given because the original tonal range is "flattened to practically zero" (p.44), and the pitch height of the neutral tone syllable actually depends on the tone of its previous syllable. As opposed to contrastive stress, neutral tone syllables are relatively short in duration. Additionally, other identified acoustic features include low intensity, vowel centralization, and consonant weakening, as mentioned by Chao and in other researchers (e.g., Shi, 1994). As for normal stress, by Chao's definition, syllables that have neither contrastive stress nor weak stress belong to this category.

In addition to the three stratifications of stress, Chao (1968) further stated that stress in Mandarin is manifested primarily by pitch range enlargement and duration lengthening, and only secondarily by loudness. Later studies took Chao's idea and conducted a number of experiments to testify his observations. For example, in his acoustic study on sentence stress in Mandarin, Jin (1996) found that pitch and duration are truly the two most relevant correlates of sentence stress, with pitch ranked even higher than duration. The systemization of stress levels is found in Pan-Mandarin ToBI, developed by Peng, et al. (2007). Four levels of stress are identified. These four levels and the corresponding depictions are shown in Table 2.2.

Table 2.2 Relative levels of stress in Pan-Mandarin ToBI (Peng, et al., 2007).

Stress	Description
S0	syllable with lexical neutral tone
S1	syllable that has lost its lexical tonal specification (e.g., in a weakly-stressed position)
S2	syllable with substantial tone reduction (e.g. undershooting of tonal target with duration reduction)
S3	syllables with fully realized lexical tone

As can be seen, the stratification of stress in Pan-Mandarin ToBI is actually similar to that of Chao's (1968), except for dividing *weak stress* into two levels (S0 and S1), depending on whether the syllable is lexically specified as a neutral tone syllable. In addition, it is obvious that tonal realizations are taken as the sole criterion for identifying stress in Pan-Mandarin ToBI. Nonetheless, it should be noted that there are studies indicating that although pitch is an important cue for stress, it is not a necessary cue. For instance, Shen (1993) utilized natural speech and modified the acoustic parameters in order to see whether stress perception is harmed in lack of the F_0 cue. Results of perceptual experiments showed that without F_0 information, listeners are still able to identify stress locations. In this regard, it was concluded by Shen that in Mandarin, no one cue is indispensable; instead, stress prominence is marked by the integration of all relevant correlates.

2.5 Word class

The distinction of word class in our study refers to the contrast between content word and function word. By definition, content words, also termed open-classed words, are semantically informative elements, while function words, or close-classed words, are grammatically required elements. The common assumption is that prominence is usually with content word, whereas non-prominence is with function word. Nonetheless, it is still worth noticing that although the relationship between prominence level and word class is rather clear, this generalization is definitely not foolproof. As stated by Bolinger (1972), sentence accent is not determined by syntax, but information focus. As a result, even with syntactic category distinction, accent is still not predictable.

There are studies investigating the effect of word class on speech production and phonological variation. For example, van Bergem (1993) looked into Dutch vowel production in experimental settings. He found that word class, interacting with stress, is very influential in determining how vowels are realized. Specifically, for the same vowel, a stressed content word condition has more distinctive vowel realizations than an unstressed content word condition, which in turn has better realizations than function word condition. In studying /t/ and /d/ deletion in English spontaneous speech, Raymond et al. (2006) also pinpointed the influence of word class on such a phenomenon. In particular, for content words, deletion is more likely to occur in longer words; such an effect, on the contrary, is not found for function words.

While word class generally exerts an effect on speech production, it does not seem to play a crucial role in speech perception. For instance, Cutler and Foss (1977) examined how stress and word class affected sentence processing. They found that given the same stress level, there are no significant differences in terms of subjects' performance for content words and function words. In the identification test of French voiced stops, Duez (1995) did not see any effect of word class either. Voiced stops in content words, surprisingly, are not easier to be recognized than those in function words. Although the discrepancy between production and perception studies is indeed puzzling, and it does not seem to be satisfactorily explained yet, it still does not hurt our prediction that word class could have an effect on the realizations of retroflex and dental sibilants in Taiwan Mandarin. In this regard, the investigation of word class and its interaction with stress should be both interesting and worthwhile.

2.6 Summary

In this chapter, past studies on the realizations of retroflex and dental sibilants in Taiwan Mandarin were discussed, from both acoustic and sociolinguistic perspectives. In general, a gap can be observed; that is, acoustic studies focus on experimental data, whereas sociolinguistic studies most frequently investigate spontaneous speech. It is one of our main goals to fill this gap by acoustically studying the issue on retroflex and dental sibilants in spontaneous speech.

Previous research relevant to the four factors that are to be examined in this study was presented as well. For social factors, gender exerts a strong effect on the realizations of retroflex and dental sibilants, while the effect of region factor is much less examined and still remains unclear. As for linguistic factors, both factors of prosodic prominence and word class are active in speech production, but neither of them has been examined on the sibilant realizations in Taiwan Mandarin. As a result, in this study, we are going to investigate these four factors all together, and try to uncover interesting interactions among them.

CHAPTER 3 METHODS

This chapter introduces the methods for conducting the present study. Section 3.1 to Section 3.3 describes participants, data collection, and recording equipment in this study. Section 3.4 focuses on data processing procedure, in terms of transcription, word segmentation, and stress labeling. Section 3.5 illustrates the labeling and the acoustic measurements of retroflex and dental sibilant tokens.

3.1 Participants

This study utilized part of the speech data from the Mandarin-Min bilingual corpus constructed by Fon (2004). The data of fifteen speakers were selected for analyses in this study. These fifteen speakers, including both males and females, are all Mandarin-Min bilinguals from either Taipei City/County or Kaohsuing City/County. None of them lived or stayed in places other than Taipei or Kaohsiung for over six months prior to the age of eighteen. They were at the age of 20 to 35 years old at the time of recording. Each speaker's age is given in Appendix I. The numbers of speakers in each group are displayed in detail in Table 3.1.

As for language proficiency, each participant's Mandarin and Min recordings were listened to and judged by three additional bilinguals, who were proficient in both

Mandarin and Min. Each judge assigned a score to indicate the language proficiency for each participant, on a scale ranging from 1 (the least proficient) to 7 (the most proficient). The average proficiency scores for each speaker's Mandarin and Min are provided in Appendix I. Generally speaking, almost all speakers are proficient in Mandarin, but they do differ in Min proficiency levels. In order to compare Min proficiency among different speaker groups, each speaker's Min score was divided by his/her Mandarin score as a means of normalization. The average Min-to-Mandarin ratio for each speaker group is shown in Table 3.2.

Table 3.1 The number of subjects in each speaker group.

	Region	
Gender	Taipei	Kaohsiung
Male	4	4
Female	3	4

Table 3.2 The mean Min-to-Mandarin ratio for each speaker group.

	Region	
Gender	Taipei	Kaohsiung
Male	1.03	0.91
Female	0.79	0.78

3.2 Data collection

In the bilingual corpus of Fon (2004), the form of interview was adopted. Each speaker had two thirty-minute conversations in Mandarin and Min respectively with one interviewer. The interviewers themselves are also fluent Mandarin-Min bilinguals, who used either of the two languages solely to converse with the interviewees during the recording. The interviewer would ask open questions which were related to speakers' personal experiences or opinions on several issues. There were preset questions, but spontaneous questions were asked as well, with the goal to elicit maximum responses from the speakers. Frequently discussed topics included habits, movies and TV programs, food, traveling experiences, student or career lives, etc. In this study, only Mandarin data of the selected fifteen speakers were processed and analyzed. The recording length and discussed topics for each speaker are summarized in Appendix II.

3.3 Recording equipment

The recording equipment was SONY DAT recorder PCM-M1. Subjects wore a SHURE SM10A head-mounted microphone during the recording. Each recording was later digitalized and downsampled with the 22 kHz sampling rate, by using *Cool Edit Pro* (Version 2) software.

3.4 Data processing procedure

3.4.1 Text transcription

Each recording was first transcribed in Chinese characters, and then romanized into *hanyupinyin*. Syllable labeling was done by using *Praat* computer software (Boersma &

Weenink, 2008). In each thirty-minute recording, there were about 7,500 syllables. The total syllable numbers contributed by each speaker are shown in Appendix II.

3.4.2 Word segmentation and word class labeling

The definition of a word in Chinese is not as clear-cut as that in most Western languages. Due to the influence of Chinese writing system, in which every syllable has its corresponding character, Chinese people often confuse the notion of word with that of syllable, considering that every syllable is a word. However, it has been shown that Modern Chinese consists of a great number of bisyllabic or multisyllabic words. Therefore, a word may be composed of one syllable, two syllables, or even three or more syllables. Accordingly, a word should better be defined as "a unit in the spoken language characterized by syntactic and semantic independence and integrity". (C. N. Li & Thompson, 2005, p. 13).

There have been debates over how to identify Chinese word boundaries. For example, certain grammatical morphemes (e.g., *-le* 'perfective aspect suffix') are hard to be classified, for some consider them to be separate words, whereas others regard them as part of the preceding words. Compound words (e.g., *zhongwenfenciguifan* 'Chinese segmentation standard') belong to an ambiguous category, in which they can be either treated as an independent unit or can be subdivided into several smaller units.

Furthermore, people also have different ideas towards the word boundaries of specific constructions (e.g., V-not-V) in Mandarin. To solve the word segmentation problem, in this study, we adopted the segmentation criteria developed by Academia Sinica, Taiwan, of which their definition of *word* is applicable with respect to both linguistics and information processing (C.-R. Huang, Chen, Chen, Wei, & Chan, 1996). For instance, according to their criteria, grammatical suffixes are regarded as separate morphemes. In addition, long compound words should be segmented; thus, *zhongwenfenciguifan* are subdivided into *zhongwen* 'Chinese' *fenci* 'segmentation' *guifan* 'standard'. Finally, V-not-V structure is segmented (e.g., *xihuan bu xihuan* 'like it or not').

The word segmentation of our corpus data first used the *Online Word Segmentation System* developed in the National Digital Archive Program, Taiwan (Chen, 1998). After the automatic text segmentation by the system, we hand-checked each segmented word and its corresponding labeled part of speech. The original part of speech coding of the online system was very specific. For instance, it distinguished different types of verb, such as transitive verb (e.g., *chi* 'eat'), intransitive verb (e.g., *zoulu* 'walk'), etc. However, our ultimate goal was just to classify all words into two major word classes – content word and function word; therefore, the broad labels such as "noun" or "verb" would be sufficient. In this regard, the part of speech coding was modified when we checked the segmentation data. All parts of speech identified in our study, with their corresponding word class categories, are shown in Table 3.3.

Table 3.3 The parts of speech and the corresponding word class categories[3].

Word class	Part of Speech	Example	
Content word	Adjective	*zhuyao*	'main'
	Adverb	*qishi*	'actually'
	Measure word	*zhong*	'kind'
	Noun	*zhuozi*	'desk'
	Numeral modifier	*henshao*	'few'
	Numeral	*san*	'three'
	Verb	*chifan*	'eat'
Function word	Aspect marker	*zhe*	'progressive marker'
	Conjunction	*suoyi*	'so'
	DE	*de*	'nominalizer'
	Discourse marker	*jiushishuo*[4]	'well'
	Determiner	*zhe*	'this'
	Negation	*bu*	'not'
	Preposition	*zai*	'at'
	Postposition	*zhilei*	'something like that'
	Pronoun	*wo*	'I'
	SHI	*shi*	'coverb'

3.4.3 Stress labeling

As noted previously, in Pan-Mandarin ToBI (Peng, et al., 2007), four levels of prosodic prominence are identified (see Table 2.2), and the criteria are generally based on tone solely, possibly due to the influence of Chao's observation that pitch is the

[3] There were several categories that did not belong to either content word class or function word class. These categories included "Filler," "Foreign word," and "Truncated word/sound."

[4] The Academia Sinica Corpus is a text-based corpus, and its data source is mainly from written materials. Therefore, the Online Segmentation System labels *jiushishuo* as an adverb, approximately equal to 'that is to say' in English. However, in spoken Mandarin, *jiushishuo* has been grammaticalized and became a discourse marker. In this regard, *jiushishuo* is considered as a discourse marker and thus categorized as function word by our standard.

primary cue for stress (1968). However, as noted earlier, stress correlates are more than only pitch. In addition, in spontaneous speech, usually the tonal contours are not realized clearly due to faster speech rate; thus, the total reliance on tonal realization will fall into difficulty. Owing to these limitations, in this study, we followed the basic stratification of Pan-Mandarin ToBI, but modified them by adding several other relevant variables for the ease of identification. The additional correlates included duration, amplitude, and segmental information, all of which are well-identified correlates of stress. The complete modified criteria are shown in Table 3.4.

Table 3.4 Modified criteria for levels of prosodic prominence.

Stress	Tone	Amplitude	Duration	Segmental info
S0	lexically neutral tone	soft	shortened	target neutralized
S1	loss of original tonal shape	soft	shortened	target neutralized
S2	default	default	default	default
S3	tone expanded/raised	loud	lengthened	target accurately reached

Examples of stress labeling are shown in Figure 3.1 and Figure 3.2. Figure 3.1 demonstrates the S1 and S2 conditions. As could be seen, for syllables labeled with S1, the tonal targets were not reached at all, and they could be simply considered as a string of lowering pitch contours. Furthermore, the spectrogram shows that the segments of S1 syllables were not really realized either. In general, a great degree of tonal and segmental reduction could be observed. On the other hand, Figure 3.2 shows the S2 and S3 conditions. Particularly for the syllable *la* ‘spicy,’ we could see that it had greater amplitude than the ambient syllables. In addition, compared with other Tone 4 syllables labeled with S2 in the same utterance, it was clear that the speaker indeed tried to put emphasis on that syllable, thus expanding the tonal range.

About 10% of the data were labeled independently by another labeler, who was previously trained with the criteria adopted in this study. The two labelers agreed upon 82% of the data, and the stress labels given by the two labelers were significantly correlated [Kendall's $W = .69$, $p < .001$]. Therefore, in this study, the original stress labels were adopted for analyses.

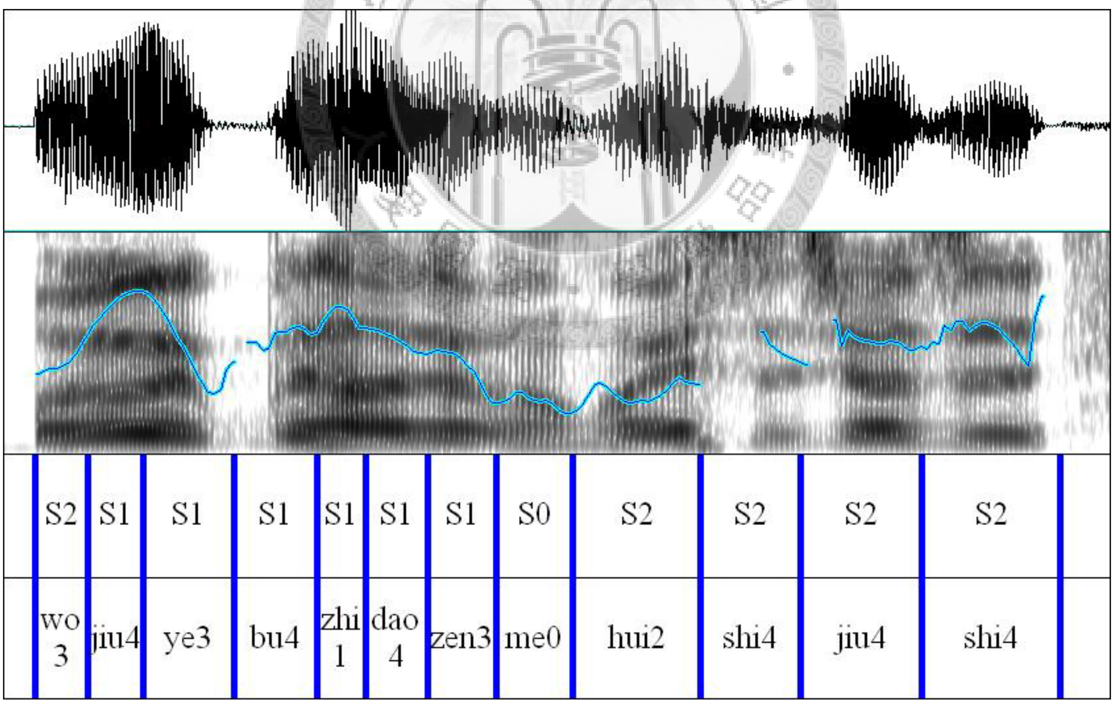


Figure 3.1 The waveform, spectrogram, tonal contour, stress labeling (tier 1) and syllable labeling (tier 2) of the utterance 'I did not know why, it's just…'.

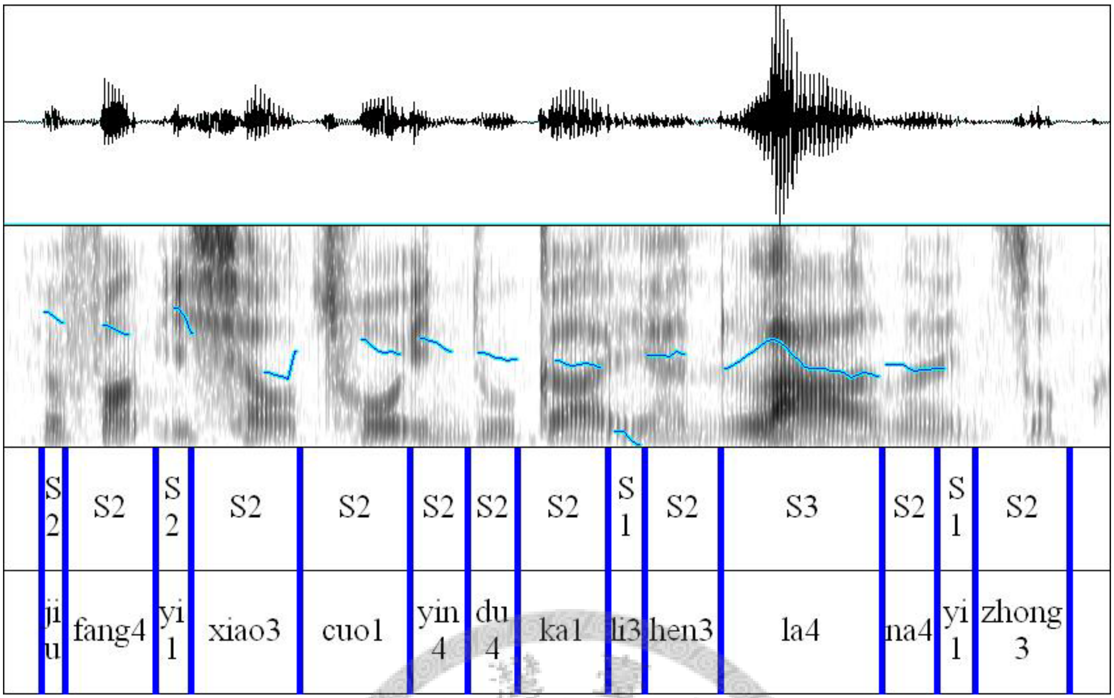


Figure 3.2 The waveform, spectrogram, tonal contour, stress labeling (tier 1) and syllable labeling (tier 2) of the utterance ‘There is a small amount of Indian curry, the kind that is very spicy.’.

3.5 Retroflex and dental sibilants

All retroflex sibilant ([tʂ], [tʂ^h], [ʂ]) and dental sibilant ([ts], [ts^h], [s]) tokens in the corpus were identified. The main frication part of each sibilant token was labeled. For affricates [tʂ]/[ts] and [tʂ^h]/[ts^h], the burst was excluded. In this study, we followed Li’s (2008) acoustic measurements on fricative sibilants, but modified in several aspects to meet our needs. In Li’s study, for each sibilant token, the spectrum of 40 ms Hamming window in the middle of the frication was taken. In addition, she filtered out frequency below 1000 Hz in order to eliminate low frequency noise. Then she measured all four spectral moment values plus one transition cue. In the present study, we also used Hamming window, and the analysis window was set at 5 ms. Instead of 40 ms, we took only 10 ms of the spectral slice, due to the fact that in spontaneous speech, the speech rate is usually faster, thus resulting in much shorter frication of sibilant tokens.

Moreover, to decide the cut-off frequency for the high-pass filter, we first examined the spectra of retroflex and dental sibilants in Taiwan Mandarin. Figure 3.3 presents ten males' and ten females' spectra of both retroflex and dental sibilants. The spectra were extracted from the sibilants of the syllable *shi* 'thing' and *si* 'in charge' uttered in isolation in experimental settings. As can be seen, dental sibilants had energy concentrated in the higher frequency range than retroflex sibilants. In addition, frequencies below 2000 Hz did not differ crucially for the two sibilants, and the spectral peaks in the range below 2000 Hz could be considered as frequency noise. To make our acoustic measurements focus on high frequency range, we thus filtered out frequency below 2000 Hz.

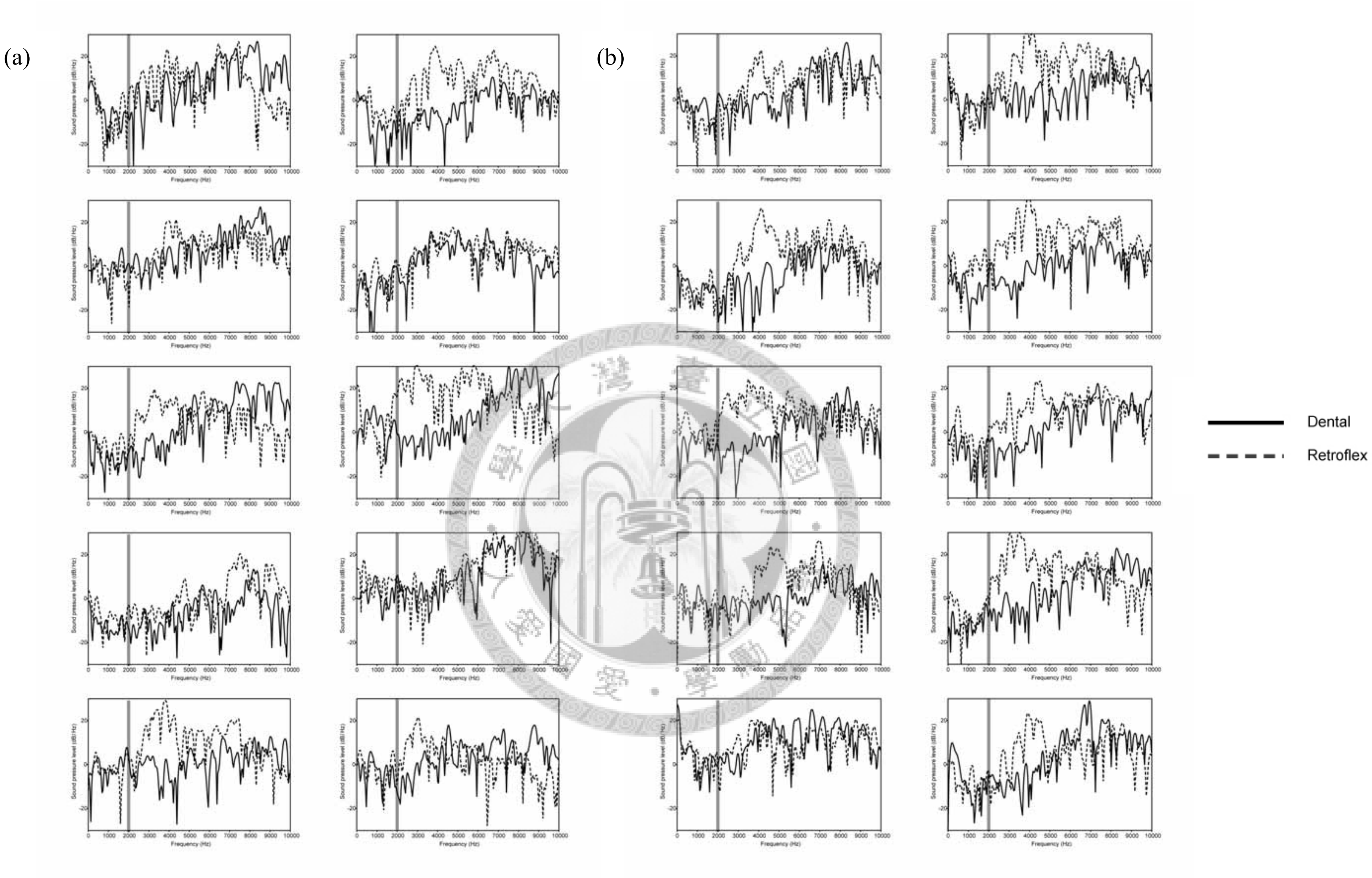


Figure 3.3 The spectra of retroflex and dental sibilants taken from the the syllable *shi* ‘thing’ and *si* ‘in charge’ respectively, produced by (a) 10 male and (b) 10 female speakers. The solid line in each spectrum indicates 2000 Hz.

Finally, we only took the first spectral moment measurement, *i.e.*, centroid frequency, in our study. Since the purpose of this study is not to search for acoustic cues that distinguished retroflex and dental sibilants, but to unravel the effects of linguistic factors on sibilant realizations, we thus chose to take the measurement of centroid frequency, which has been consistently found to be the most reliable variable in previous studies (e.g., Jeng, 2006; F. Li, 2008). The detailed measurement procedure is illustrated in Figure 3.4.

For the acoustic measurements we adopted, although 10 ms spectra taken in the middle of the frication was the general rule, we did some modifications when measuring the aspirated affricate [tʂ^{h}]. The aspiration portion usually carried the formant information of the following vowel; therefore, it was very likely that the middle 10 ms we took would fall on the aspiration part. To solve this problem, we took the middle 10 ms of the first one-third of the frication part, instead of the whole frication. A small pilot study on 120 [tʂ^{h}] and [ts^{h}] tokens showed that the aspiration started at on average the 32.88% of the frication. An example of the [ts^{h}] token was presented in Figure 3.5. It could be seen that the real sibilant frication part constituted about the first one-third part of the onset consonant [ts^{h}]. Therefore, we assumed the first one-third of the frication to have the least effect of aspiration.

(a)

(b)

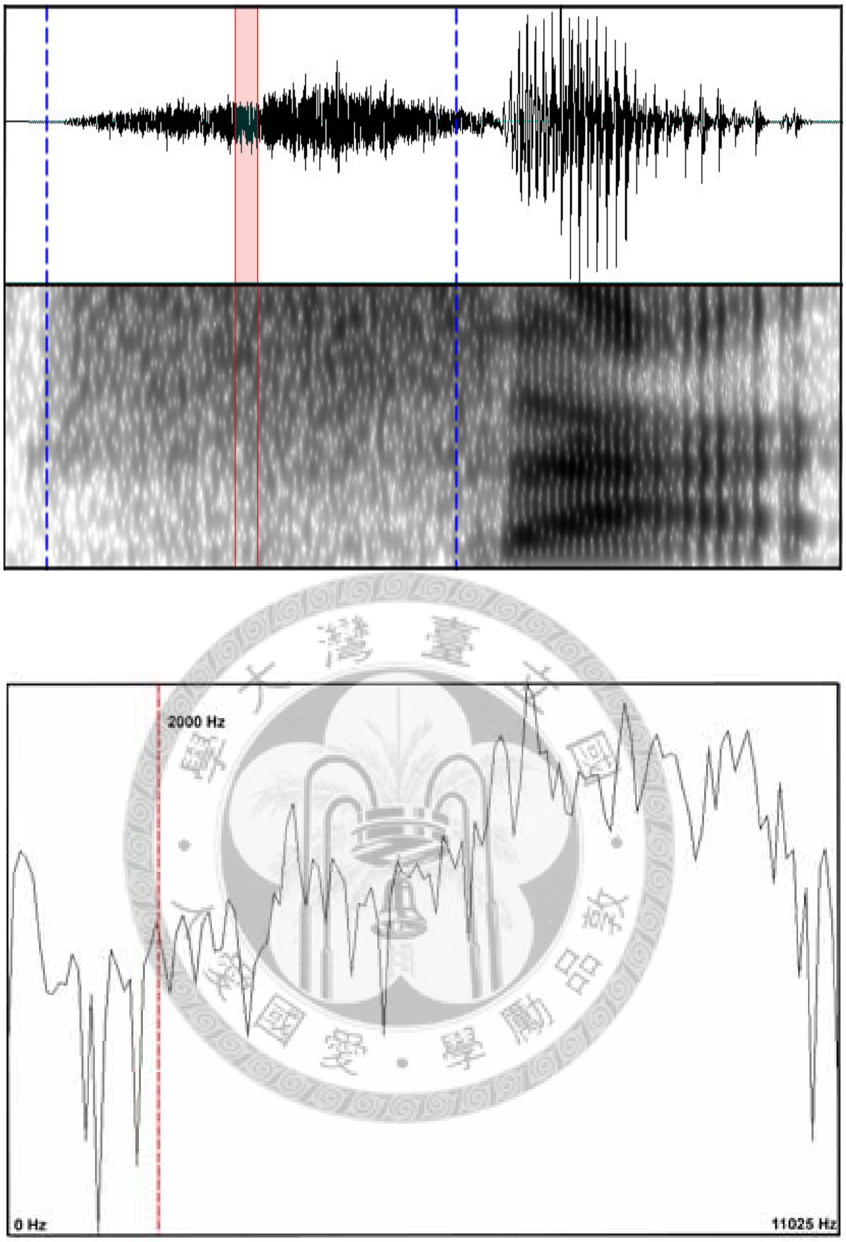


Figure 3.4 (a) The waveform and the spectrogram of the syllable *sai* 'game'. The dashed lines indicate the frication portion, where the solid lines indicate the 10 ms in the middle of the frication. (b) The spectrum extracted from the middle 10 ms of the frication. The centroid frequency was calculated from the range above 2000 Hz, indicated by the dashed line.

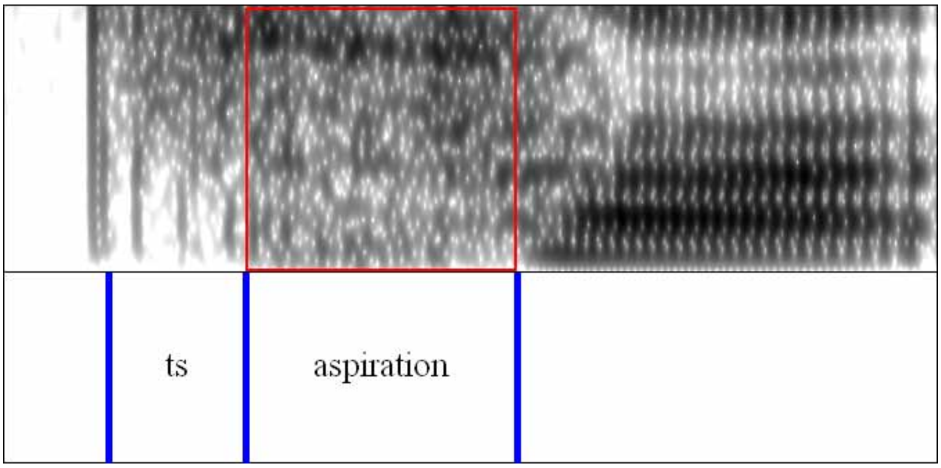


Figure 3.5 The spectrogram of the syllable *cai* 'vegetable,' with the frication and the aspiration parts being labeled.

CHAPTER 4 RESULTS

In this chapter, statistical results of the present study are presented. Section 4.1 summarizes the overall distribution of all sibilant tokens in our corpus data. Section 4.2 shows the analyses and results of canonically realized sibilant tokens. In Section 4.3, the effect of word class is examined. Section 4.4 investigates the interaction between the effects of stress and word class. Section 4.5 presents the analyses and results of substituted sibilant tokens, and the distribution of sibilant deleted tokens is shown in Section 4.6.

4.1 Overall distribution

The total number of voiceless retroflex and dental sibilant tokens in the corpus was 20,824, of which 16,258 of them were valid tokens and were thus included for acoustic analyses, whereas 4,566 of them were invalid tokens. The detailed distribution of both valid and invalid tokens is shown in Table 4.1.

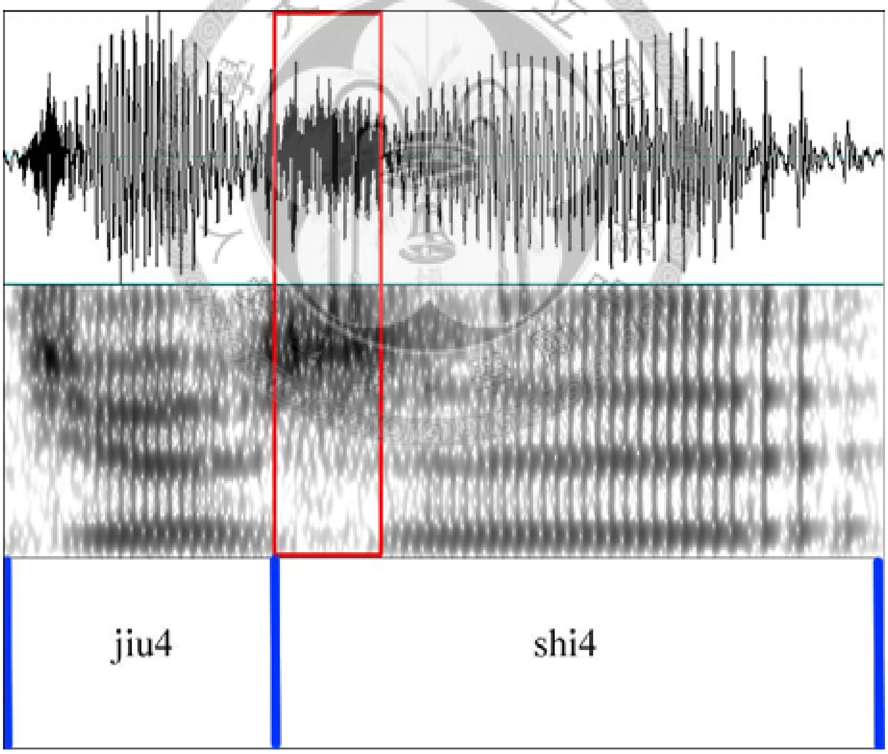
Table 4.1 The distribution of all sibilant tokens in the corpus.

Validity	Category	Token number	Percentage
Valid	Canonical		
	Voiceless	13,790	66.22%
	Voiced	2,309	11.09%
	Substituted		
	Voiceless	134	0.64%
	Voiced	25	0.12%
Invalid	deleted	3,160	15.17%
	soft/short	794	3.81%
	noise	232	1.11%
	truncated	176	0.85%
	wrong	146	0.70%
	extra-linguistic	56	0.27%
	glottalized	2	0.01%
Total		20,824	100.00%

4.1.1 Valid tokens

As shown in Table 4.1, among the valid sibilant tokens, there are two main categories – *canonical* and *substituted*. Each main category can further be divided into two smaller subsets – *voiceless* and *voiced*. The canonical category contained sibilant tokens that were canonically realized, *i.e.*, the sibilant being produced corresponded to its lexically specified sibilant category of that particular syllable. If the sibilant was realized as a voiceless sibilant, it belonged to the voiceless subset; on the other hand, if the sibilant was voiced, it belonged to the voiced subset. Although it is the voiceless sibilants that are investigated in this study, we also included voiced sibilant tokens for analyses, since the frequency below 2000 Hz of each sibilant token was filtered out, and it was the centroid frequency of each sibilant token that we measured. Therefore, the smaller frication amplitude resulting from voicing should not have caused problems for determining the place of articulation of the voiced sibilants. In the canonical category, 13,790 sibilants were voiceless and 2,309 of them were voiced. Figure 4.1 shows two realizations of the word *jiushi* ‘that is’. In Figure 4.1(a), the retroflex sibilant [ʂ] was voiceless, whereas in Figure 4.1(b), the sibilant was voiced, evidenced by the presence of the periodic waveform and the voice bar in the spectrogram.

(a)



(b)

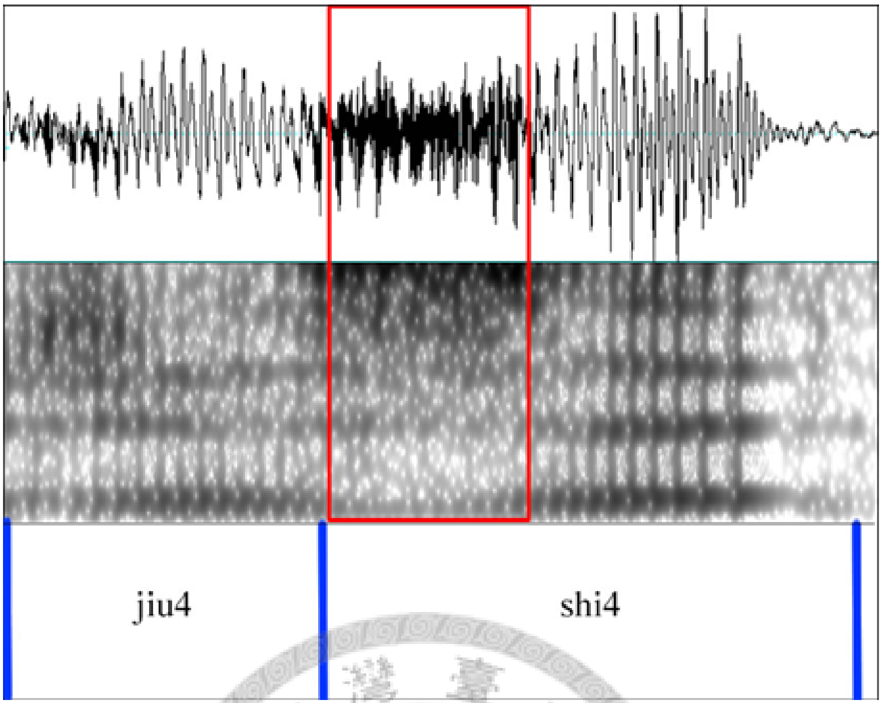


Figure 4.1 The waveform and the spectrogram of the word *jiushi* 'that is,' where the retroflex sibilant [ʂ̬] is voiceless in (a) and voiced in (b).

The substituted category consisted of sibilant tokens that were replaced by their sibilant counterparts and had already crossed the perceptual boundary; that is, for example, the dental fricative [s] was categorically perceived as [ʂ]. Nonetheless, sibilant tokens that were in the perceptual grey area were still labeled with their original sibilant categories. Only those that had obviously crossed the perceptual boundary were regarded as substituted tokens.

Substituted tokens were analyzed and discussed separately from canonically realized tokens due to the fact that different mechanisms were at play. For canonically realized sibilants, we focused on the degree of merging, which we defined two sibilants getting less different from each other. Thus, we would like to see the relationship between retroflex and dental sibilants that are correctly realized or in the perceptual grey areas. On the other hand, substitution was defined as one sibilant being perceptually identified as its counterpart, and its realization has already crossed the perceptual boundary between retroflex and dental sibilants. In this regard, substitution should be considered as extreme cases of merging, and we were concerned about when the extreme cases of merging would occur. In our present corpus-based data, totally 159 tokens were in the substituted category. Among them, 134 were voiceless, while 25 of them were voiced.

4.1.2 Invalid tokens

As for the invalid tokens, reasons why they were excluded for acoustic measurements are summarized in Table 4.1. The majority of the invalid tokens belonged to the category *deleted*, constituting almost 15% of all sibilant tokens. Since the sibilant frication was deleted, these tokens were not appropriate for our acoustic measurements. An example of sibilant deletion is shown in Figure 4.2, where the frication of [ʂ] was not present at all.

In addition, a number of the invalid tokens were categorized as *soft/short*, meaning that the frication was too soft or too short to be measured properly. Figure 4.3 demonstrates the soft situation. As can be seen, the fricative sibilant [ʂ] in the last syllable of the word *dagaishi* 'perhaps' was so soft that the frication of [ʂ] was barely visible from the waveform and the spectrogram. As for the sibilant duration, because we took 10 ms spectral slice of the sibilants, thus, if the sibilant duration labeled was shorter than 10 ms, the token was excluded from our analyses. However, for aspirated affricates, the 10 ms spectral slice was taken from the initial one-third of the sibilants, so the total duration of the aspirated affricate sibilants had to exceed 30 ms. Aspirated affricate tokens shorter than 30 ms were regarded as invalid tokens.

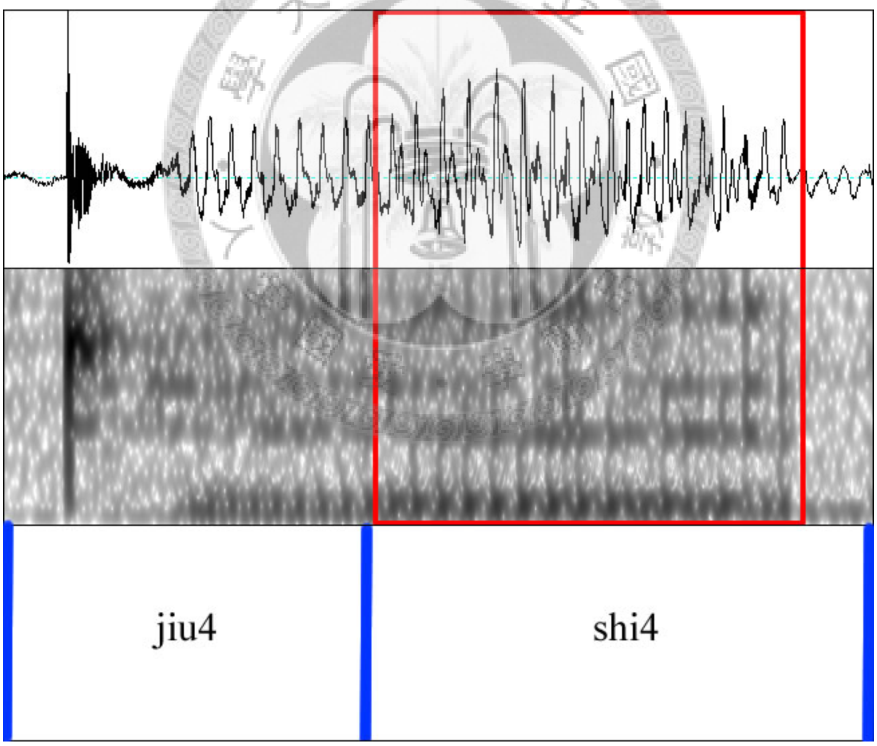


Figure 4.2 The waveform and the spectrogram of the word *jiushi* 'that is,' where the retroflex sibilant [ʂ] is deleted.

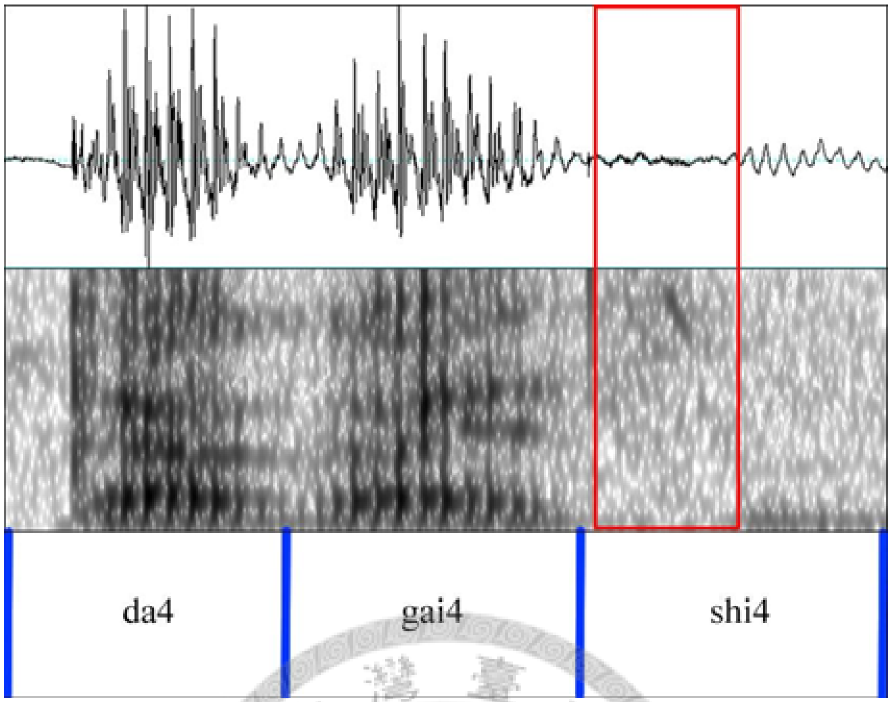


Figure 4.3 The waveform and the spectrogram of the word *dagaishi* ‘perhaps,’ where the retroflex sibilant [ʂ] is soft.

The *noise* category was excluded due to the interruption of background noise or microphone noise. Figure 4.4 shows how noise could affect the acoustical signals. Since noise would lead to false measurements, these tokens were discarded. For the *truncated* category, the production of syllables or sibilants were not complete, usually seen in disfluency or speech error conditions. For example, in Figure 4.5, it can be seen that for the word *tongche* ‘to commute,’ only the sibilant part of the second syllable was uttered, and the production of this syllable was incomplete. Although for most of the time, labelers were able to identify the truncated sound or syllable with contextual information, the speaker’s production was not complete after all. To avoid unwanted effects, these truncated tokens were not included for analyses.

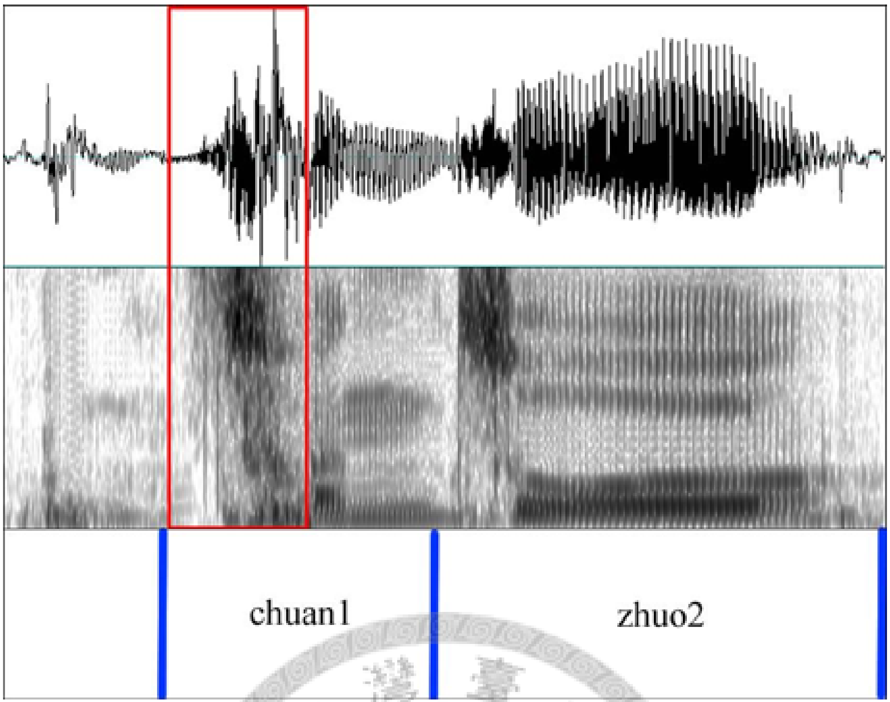


Figure 4.4 The waveform and the spectrogram of the word *chuanzhuo* 'clothing,' where the retroflex sibilant [tʂʰ] is affected by microphone noise.

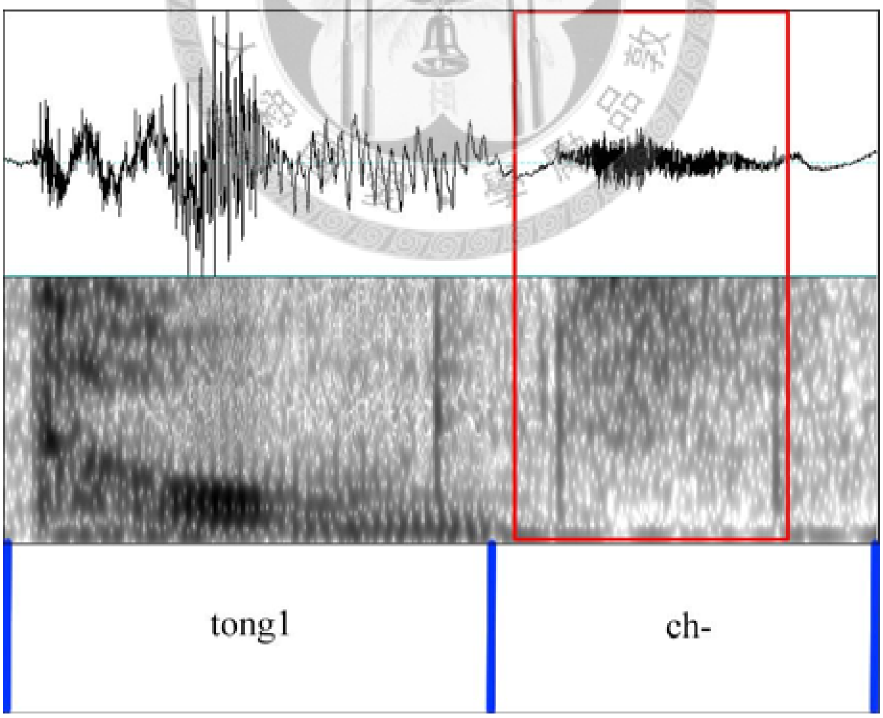


Figure 4.5 The waveform and the spectrogram of the word *tongche* 'to commute,' where the production of the second syllable is incomplete and only the retroflex sibilant [tʂʰ] is uttered.

In the *wrong* category were those sibilant tokens that were not correctly pronounced. For instance, as shown in Figure 4.6, the unaspirated affricate sibilant [ts] of the word *zai* 'at' was pronounced wrongly and sounded like the voiceless stop [t]. As can be seen, the frication portion of the sibilant was missing. The *extra-linguistic* category contained sibilant tokens overlapping with certain conditions such as speakers' laughing, stammering, yawning, or mimicking other people's speaking style during recording. Figure 4.7 demonstrates the sibilant production overlapping with the speaker's laughter. Finally, in the *glottalized* category we found two tokens whose sibilants were substituted by glottal stops. An example is provided in Figure 4.8.

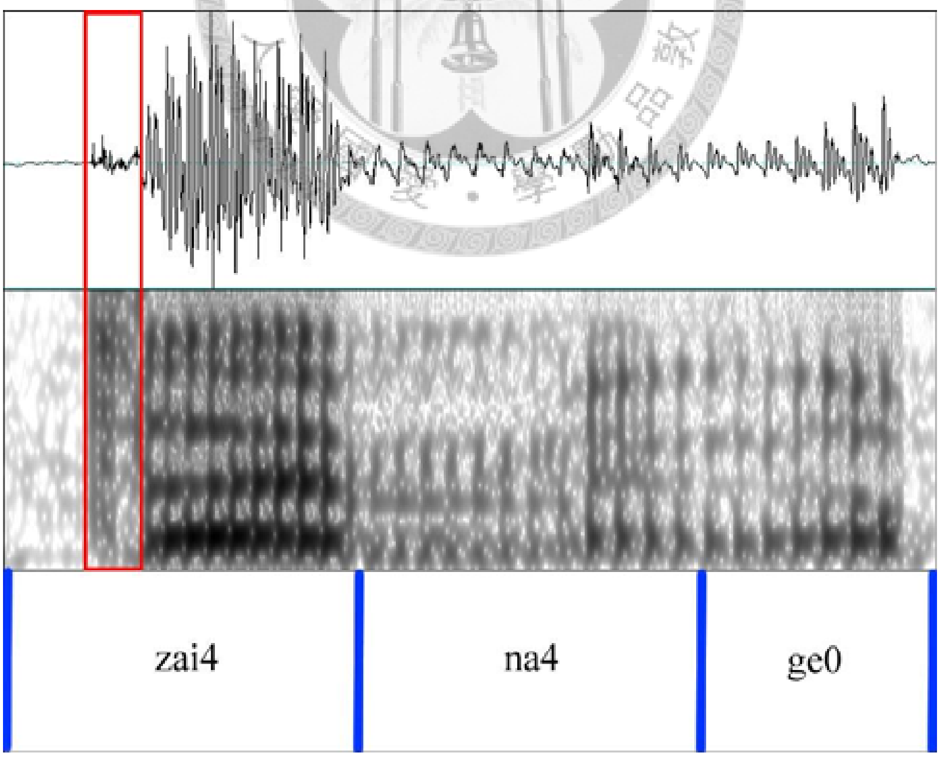


Figure 4.6 The waveform and the spectrogram of the phrase *zai nage* 'at that,' where the dental sibilant [ts] is pronounced wrongly and replaced by the voiceless stop [t].

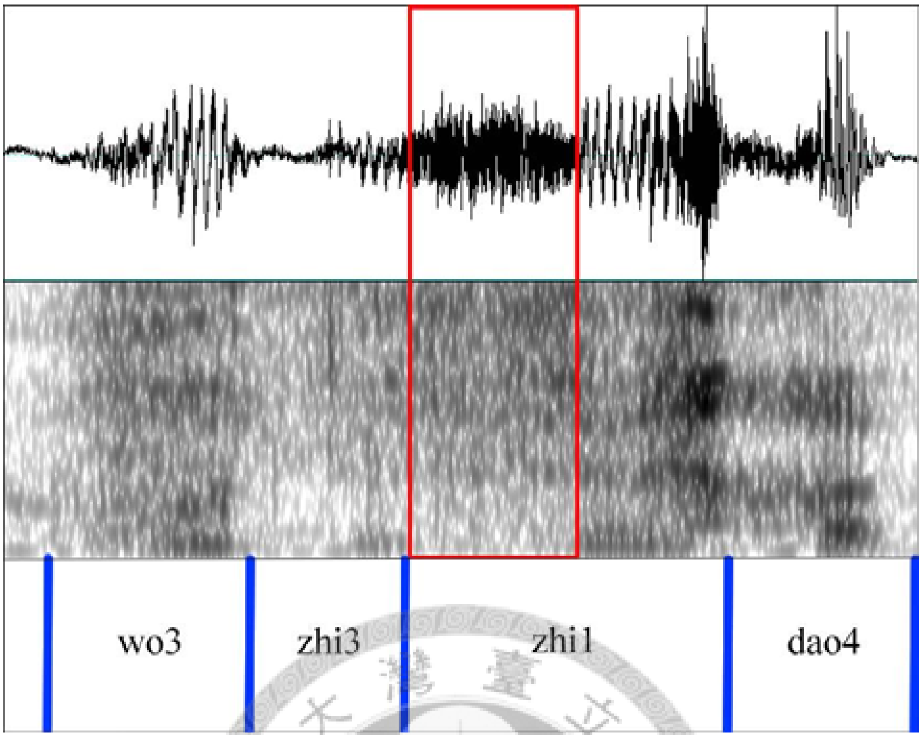


Figure 4.7 The waveform and the spectrogram of the phrase *wo zhi zhidao* 'I only know that…,' where the retroflex sibilant [tʂ] overlapped with the speaker's laughter.

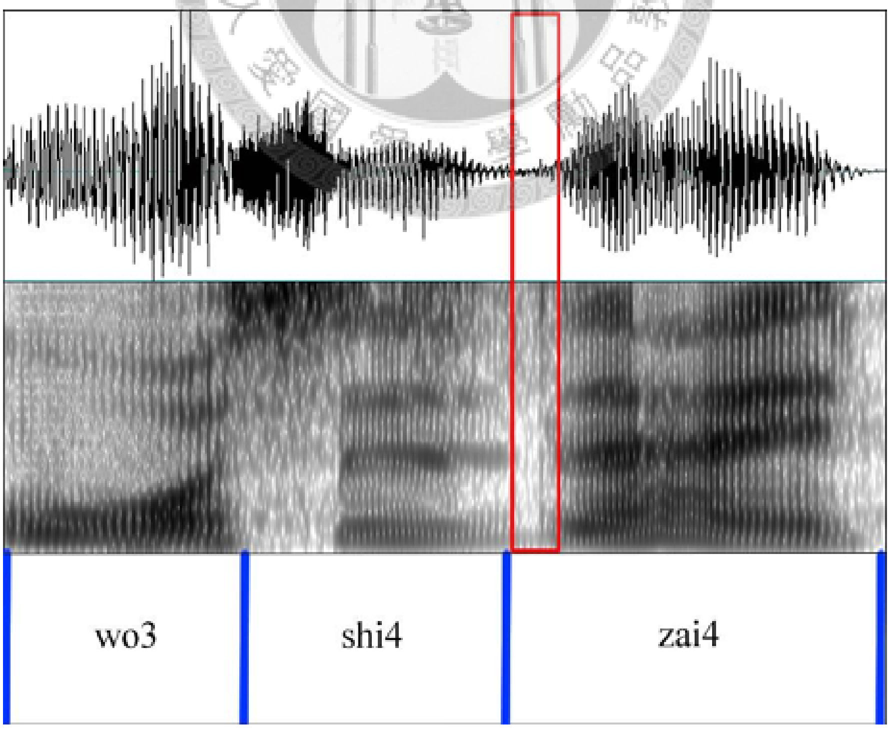


Figure 4.8 The waveform and the spectrogram of the phrase *wo shi zai* 'I was at…,' where the dental sibilant [ts] is replaced by a glottal stop.

4.2 Valid canonical sibilant tokens

Table 4.2 presented the detailed distribution of all valid canonical sibilant tokens, further subdivided by region (Taipei/Kaohsiung), gender (male/female), word class (content/function), stress level (S0/S1/S2/S3), sibilant type (aspirated affricate/ fricative/unaspirated affricate), and place (retroflex/dental). An additional factor we added into analyses was the vowel context, distinguishing whether the vowel following the sibilant was rounded or unrounded. As shown in a number of previous studies (e.g., Jeng, 2006), the following rounded vowel context lowers the frequency of previous sibilants due to coarticulation effects, and such lowering also effectively influences listeners' perception (Mann & Repp, 1980). Although vowel context is not the focus of the present study, its effect on voiceless sibilants, particularly for acoustic studies, is inevitable and should not be overlooked. Therefore, vowel context was taken as one factor in our analysis, and all sibilants were categorized according to their vowel contexts as well.

Table 4.2 The overall valid token distribution of (a) Taipei male (b) Taipei female (c) Kaohsiung male and (d) Kaohsiung female groups (R: retroflex; D: dental; c: aspirated affricate; s: fricative; z: unaspirated affricate).

(a)

Word class	Stress	Unrounded vowel context						Rounded vowel context						Total
		c		s		z		c		s		z		
		R	D	R	D	R	D	R	D	R	D	R	D	
Content	S0	0	0	0	0	0	33	0	0	0	0	0	0	33 (1%)
	S1	0	0	5	0	1	0	0	0	0	0	0	0	6 (0%)
	S2	415	130	894	115	371	194	86	33	282	75	172	139	2906 (62%)
	S3	61	22	131	28	57	23	13	8	24	17	35	26	445 (10%)
Function	S0	0	0	0	0	6	8	0	0	0	0	0	0	14 (0%)
	S1	0	0	14	0	4	0	0	0	0	1	0	0	19 (0%)
	S2	1	0	606	0	251	144	10	20	99	74	11	6	1222 (26%)
	S3	0	0	9	0	8	9	0	2	0	0	1	1	30 (1%)
Total		477 (10%)	152 (3%)	1659 (35%)	143 (3%)	698 (15%)	411 (9%)	109 (2%)	63 (1%)	405 (9%)	167 (4%)	219 (5%)	172 (4%)	4675 (100%)

(b)

Word class	Stress	Unrounded vowel context						Rounded vowel context						Total
		c		s		z		c		s		z		
		R	D	R	D	R	D	R	D	R	D	R	D	
Content	S0	0	0	0	0	0	17	0	0	0	0	0	0	17 (0%)
	S1	0	0	15	2	1	0	0	0	1	0	0	0	19 (1%)
	S2	112	48	494	43	237	168	63	27	169	37	123	141	1662 (45%)
	S3	62	26	138	26	85	51	32	14	44	10	52	14	554 (15%)
Function	S0	0	0	0	0	6	29	0	0	0	0	0	0	35 (1%)
	S1	0	0	13	0	3	1	0	0	0	0	0	0	17 (0%)
	S2	0	0	698	0	318	130	3	17	53	147	3	1	1370 (37%)
	S3	1	0	17	0	7	6	0	6	1	0	0	0	38 (1%)
Total		175 (5%)	74 (2%)	1375 (37%)	71 (2%)	657 (18%)	402 (11%)	98 (3%)	64 (2%)	268 (7%)	194 (5%)	178 (5%)	156 (4%)	3712 (100%)

(c)

Word class	Stress	Unrounded vowel context						Rounded vowel context						Total
		c		s		z		c		s		z		
		R	D	R	D	R	D	R	D	R	D	R	D	
Content	S0	0	0	0	0	0	37	0	0	0	0	0	0	37 (1%)
	S1	2	0	12	0	5	5	0	0	0	1	1	0	26 (1%)
	S2	222	66	580	59	271	217	99	30	195	67	184	120	2110 (52%)
	S3	88	25	132	40	57	35	40	9	38	21	69	35	589 (15%)
Function	S0	0	0	0	0	1	10	0	0	0	0	0	0	11 (0%)
	S1	0	0	22	0	10	0	0	0	0	5	0	0	37 (1%)
	S2	1	0	634	0	247	97	2	16	36	158	5	0	1196 (30%)
	S3	2	0	9	0	6	8	0	1	0	0	0	0	26 (1%)
Total		315 (8%)	91 (2%)	1389 (34%)	99 (2%)	597 (15%)	409 (10%)	141 (3%)	56 (1%)	269 (7%)	252 (6%)	259 (6%)	155 (4%)	4032 (100%)

(d)		Unrounded vowel context						Rounded vowel context						
		c		s		z		c		s		z		Total
Word class	Stress	R	D	R	D	R	D	R	D	R	D	R	D	
Content	S0	0	0	0	0	0	22	0	0	0	0	0	0	22 (1%)
	S1	1	0	17	0	6	3	0	0	1	0	0	0	28 (1%)
	S2	115	49	647	40	316	217	76	37	149	63	172	159	2040 (55%)
	S3	44	19	73	19	41	39	21	7	29	10	32	20	354 (10%)
Function	S0	0	0	0	0	3	19	0	0	0	0	0	0	22 (1%)
	S1	0	0	23	0	21	0	0	0	0	0	0	0	44 (1%)
	S2	0	0	705	0	226	68	9	6	49	76	3	2	1144 (31%)
	S3	0	1	11	0	14	0	0	0	0	0	0	0	26 (1%)
Total		160 (4%)	69 (2%)	1476 (40%)	59 (2%)	627 (17%)	368 (10%)	106 (3%)	50 (1%)	228 (6%)	149 (4%)	207 (6%)	181 (5%)	3680 (100%)

As can be seen from Table 4.2, the distribution of sibilant tokens was imbalanced. As a result, our analyses were restricted to certain categories when different factors were examined. In order to understand sibilant realizations as a whole, in our first analysis, we investigated the effects of *region*, *gender*, *stress*, *place*, *type* and *context*. *Word class*, unfortunately, could not be examined together because a lot of cells in the function word condition lacked sufficient data. Therefore, the first analysis focused on the content word condition. Furthermore, for the factor *stress*, not all categories could be analyzed due to insufficiency of sibilant tokens. Therefore, in this analysis, only S2 and S3 conditions were compared.

A six-way ANOVA was executed, with region (Taipei/Kaohsiung), gender (female/male), stress (S2/S3), place (retroflex/dental), type (c/s/z) and context (rounded/unrounded) all as between-subject factors. Results reported significant main effects of five factors [gender: $F(1, 10564) = 16.17$, $p < .001$; stress: $F(1, 10564) = 62.93$, $p < .001$; place: $F(1, 10564) = 546.15$, $p < .001$; type: $F(2, 10564) = 4.95$, $p < .01$; context: $F(1, 10564) = 1408.80$, $p < .001$]. Ten two-way interactions were significant, including region × gender [$F(1, 10564) = 54.27$, $p < .001$], region × stress [$F(1, 10564) = 5.02$, $p < .05$], region × place [$F(1, 10564) = 5.00$, $p < .05$], region × context [$F(1, 10564) = 27.20$, $p < .001$], gender × place [$F(1, 10564) = 139.74$, $p < .001$], gender × type [$F(2, 10564) = 27.20$, $p < .001$], stress × place [$F(1, 10564) = 27.72$, $p < .001$], place × type [$F(2, 10564) = 7.43$, $p < .001$], place × context [$F(1, 10564) = 27.72$, $p < .01$], type × context [$F(2, 10564) = 4.73$, $p < .01$]. Additionally, there were six significant three-way interactions [region × gender × place: $F(1, 10564) = 30.21$, $p < .001$; region × gender × type: $F(2, 10564) = 6.97$, $p < .001$; region × stress × place: $F(1, 10564) = 4.23$, $p < .05$; region × type × context: $F(2, 10564) = 3.70$, $p < .05$; gender × type × context: $F(2, 10564) = 6.56$, $p < .005$; place × type × context: $F(2, 10564) = 7.02$, $p < .001$] and three significant four-way interactions [region × gender × place × type: $F(2, 10564) = 6.49$, $p < .005$; gender × stress × place × type: $F(2, 10564) = 5.46$, $p < .005$; gender × place × type × context: $F(2, 10564) = 7.09$, $p < .001$]. Finally, one significant five-way interaction was reported: region × gender × stress × place × type [$F(2, 10564) = 3.45$, $p < .05$].

Figure 4.9 presents the centroid frequency of all canonically realized sibilant tokens in different conditions, and the error bars represented standard error. As can be seen, the stress effect was realized differently in three sibilant types. For aspirated affricate sibilants (c type), Kaohsiung speakers generally had better distinction of retroflex and dental sibilants in the S3 condition by realizing dental sibilants with higher frequency [$t(240) = -4.42$, $p < .001$]. Regardless of regional differences, stress effect was significantly shown in female speakers. Specifically, dental sibilants were realized higher in frequency as well [$t(225) = -4.47$, $p < .001$]. As for fricative sibilants (s type), female speakers differed in how the retroflex/dental contrast was made. Taipei female speakers did not make a distinction of retroflex and dental sibilants in terms of stress difference, whereas Kaohsiung female speakers contrasted retroflex and dental sibilants better in the S3 condition; post hoc independent *t* test reported that S3 dental sibilants were significantly realized higher than S2 ones [$t(130) = -3.33$, $p < .001$]. The stress patterns, on the other hand, were more similar for male speakers. In particular, dental and retroflex sibilants were both realized higher in centroid frequency in the S3 condition [dental: $t(420) = -4.99$, $p < .001$; retroflex: $t(2274) = -4.99$, $p < .05$], but the contrast made was larger. With regards to unaspirated affricate sibilants (z type), female speakers in general exhibited greater distinction in S3 than in S2. In particular, dental sibilants were realized higher and retroflex sibilants were realized lower in the S3 condition [dental: $t(372) = -3.02$, $p < .01$; retroflex: $t(565) = -4.46$, $p < .001$]. As for male speakers, the stress effect was not significantly shown.

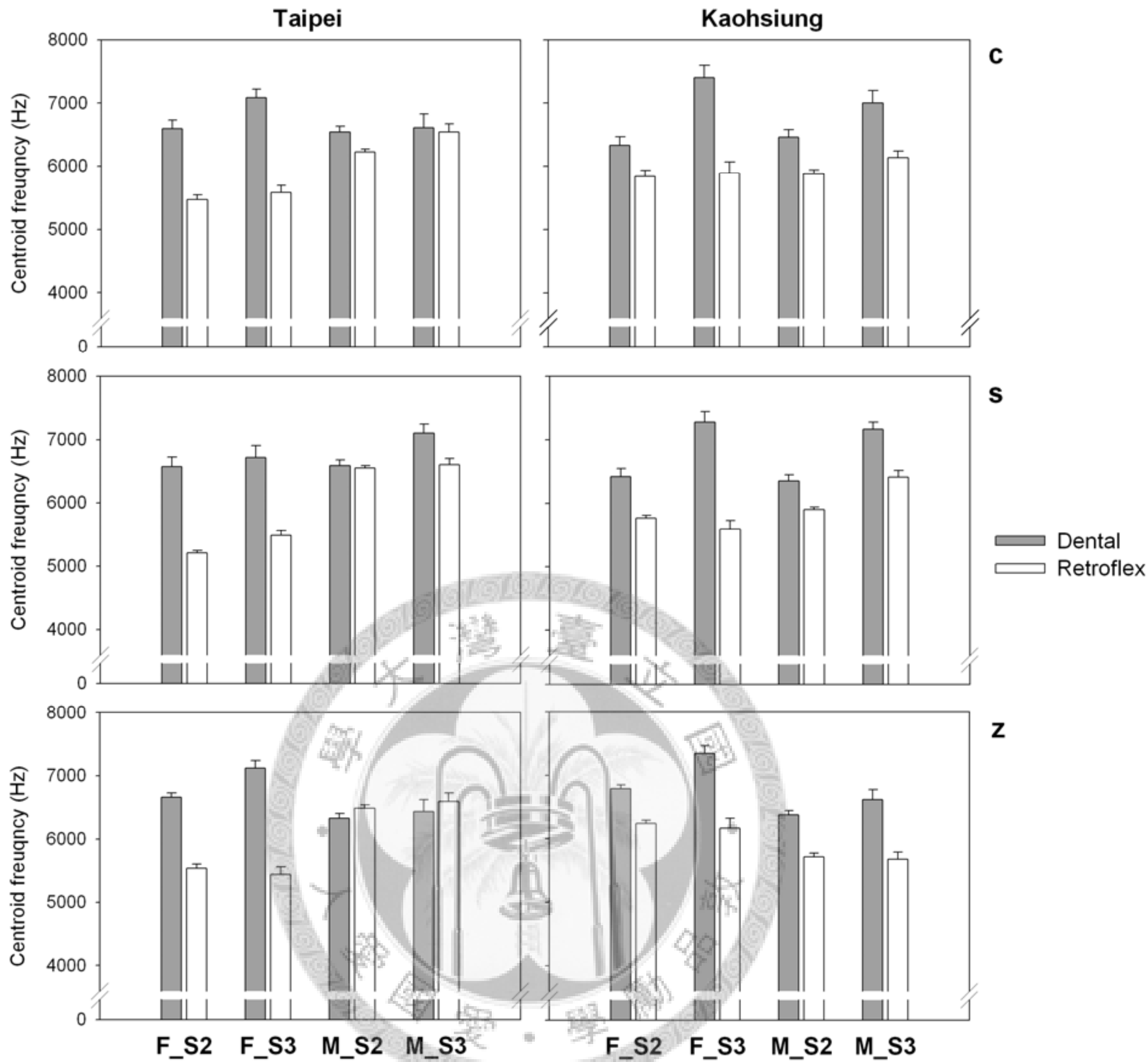


Figure 4.9 The mean centroid frequency of three sibilant types (c: aspirated affricate; s: fricative; z: unaspirated affricate) in the S2 and S3 conditions for speakers of both genders (F: females; M: males) from Taipei and Kaohsiung.

Cross-regional comparisons also revealed several interesting interaction effects. In particular, for female speakers, Taipei female speakers had significantly lower retroflex sibilants than Kaohsiung female speakers in the S2 condition [$t(2671) = -12.45$, $p < .001$], but these two groups did not differ in sibilant realizations in the S3 condition. The opposite pattern was observed for male speakers. Regardless of stress conditions, Kaohsiung male speakers did larger sibilant contrasts than Taipei male speakers by having lower retroflex sibilants in all three sibilant types [c: $t(1022) = -8.68$, $p < .001$; s: $t(2274) = -11.44$, $p < .001$; z: $t(1214) = -11.34$, $p < .001$]. The effect of sibilant type was particularly found for Taipei male speakers. Specifically, dental sibilants were significantly higher than retroflex ones for aspirated affricates (c type) [$t(766) = -3.26$, $p < .005$]; retroflex and dental sibilants did not differ significantly for fricatives (s type); retroflex were, nonetheless, slightly higher than dental sibilants for unaspirated affricates (z type) [$t(1015) = -1.98$, $p < .05$].

Moreover, gender differences were also found to effectively interact with the other factors. For Taipei speakers, across sibilant types, female speakers had both higher dental sibilants and lower retroflex sibilants than male speakers in the S2 condition [dental: $t(1148) = 2.52$, $p < .05$; retroflex: $t(3416) = -27.54$, $p < .001$]. As for Kaohsiung speakers, female speakers still made larger contrast, but both their retroflex and dental sibilants were higher in centroid frequency than male speakers' [dental: $t(1122) = 3.63$, $p < .001$; retroflex: $t(3024) = 2.07$, $p < .05$]. In the S3 condition, however, sibilant type came into play particularly for Taipei speakers. Females had only significantly lower retroflex sibilants than males for aspirated affricates (c type) and fricatives (s type) [c: $t(166) = -5.41$, $p < .001$; s: $t(335) = -8.85$, $p < .001$]. As for unaspirated affricates (z type), females made larger contrast than males by having both higher dental sibilants and lower retroflex sibilants [dental: $t(112) = 3.24$, $p < .005$; retroflex: $t(227) = -6.32$, $p < .001$]. On the other hand, no gender interaction was reported for Kaohsiung speakers in the S3 condition, illustrating that the sibilant contrast made by Kaohsiung female speakers did not differ significantly from that made by Kaohsiung male speakers.

In our six-way ANOVA analysis, vowel context was shown to have interactions with gender, place, and type. As shown in Figure 4.10, when followed by rounded vowels, the centroid frequency of preceding sibilants was significantly lowered. Female speakers showed significantly larger sibilant contrast in the unrounded condition than in the rounded condition for both aspirated affricates ($p = .054$) and fricatives ($p < .001$). The same trend was observed for male speakers as well (c: $p < .001$; s: $p < .005$). The lack of larger contrast in unrounded vowel context for unaspirated affricates (z type) was particularly due to the fact that in the unrounded vowel context, retroflex unaspirated affricates had higher centroid frequency than the other sibilant types, for both female speakers (c: $p < .005$; s: $p < .001$) and male speakers (c: $p < .001$; s: $p < .01$).

Moreover, the differences between retroflex and dental sibilants were significantly made by females for all sibilant types and in both vowel contexts. For male speakers, on the contrary, all sibilant contrasts were made except one condition. That is, no significant difference between retroflex and dental sibilants was reported for aspirated affricates (c type) in the rounded vowel context.

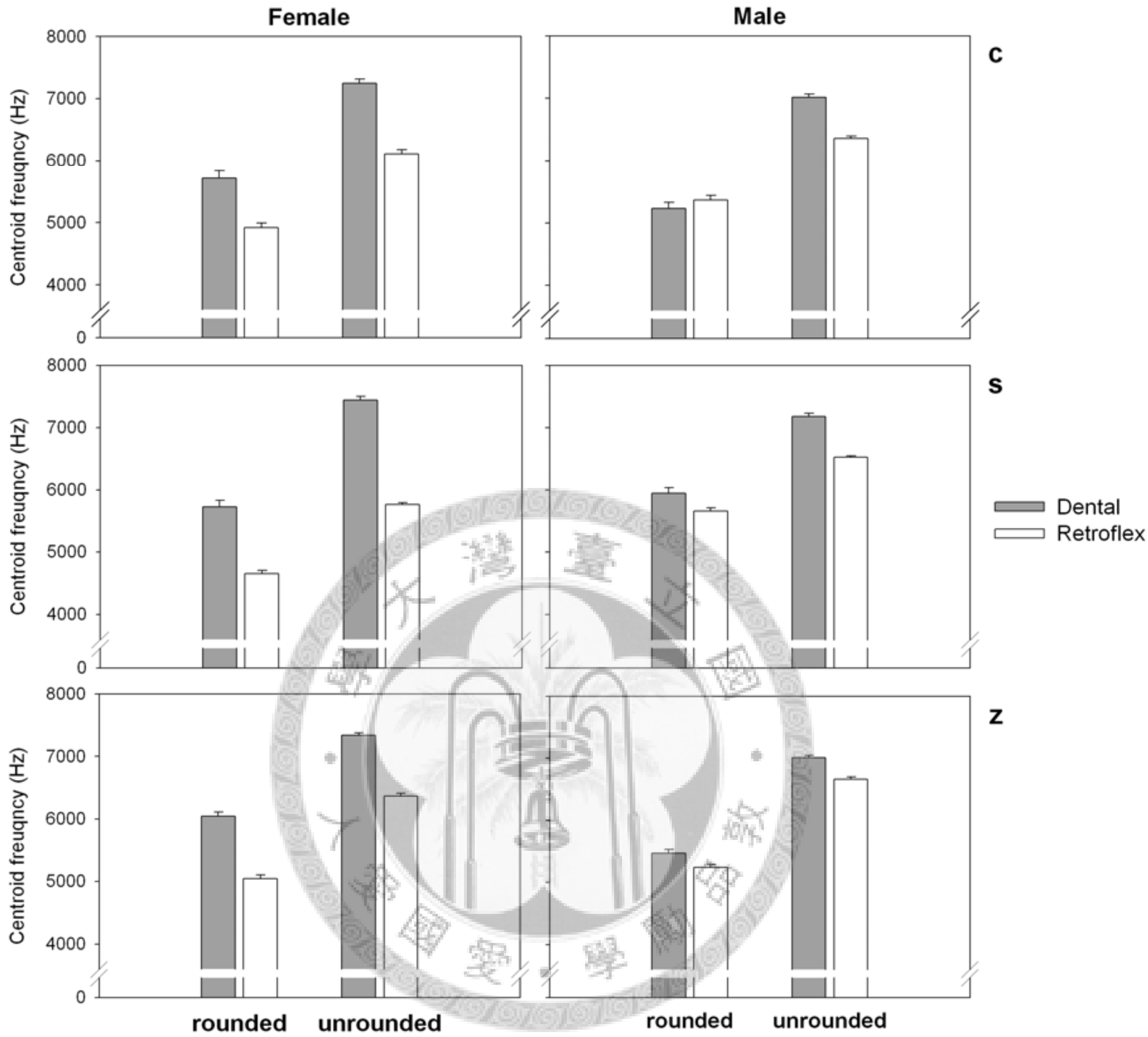


Figure 4.10 The mean centroid frequency of three sibilant types (c: aspirated affricate; s: fricative; z: unaspirated affricate) followed by rounded and unrounded vowels of both female and male speakers.

4.3 Word class

Our analyses so far were all limited to content word data, and the comparison between content words and function words was still not done yet. Therefore, in this section, the word class effect was of focus. Again, due to the limitation of data distribution in this study, we conducted two separate analyses.

We first analyzed the data of S2 unaspirated affricates (z type) in the unrounded vowel context. A four-way ANOVA with region (Taipei/Kaohsiung), gender (female/male), word class (content/function) and place (retroflex/dental) as between-subject factors was executed. Results showed significant main effects of all four factors [region: $F(1, 3456) = 11.78$, $p = .001$; gender: $F(1, 3456) = 4.75$, $p < .05$; word class: $F(1, 3456) = 23.52$, $p < .001$; place: $F(1, 3456) = 229.60$, $p < .001$]. Four two-way interactions were significant, including gender × place [$F(1, 3456) = 80.75$, $p < .001$], region × gender [$F(1, 3456) = 124.77$, $p < .001$], region × word class [$F(1, 3456) = 3.87$, $p < .05$] and place × word class [$F(1, 3456) = 4.48$, $p < .05$] . One three-way interaction was reported [region × gender × place: $F(1, 3456) = 68.77$, $p < .001$].

As shown in Figure 4.11, retroflex and dental sibilants were distinguished in both content word and function word conditions [content: $t(1989) = -12.61$, $p < .001$; function: $t(1479) = -7.98$, $p < .001$]. Moreover, for speakers from both regions, it was a general trend that both retroflex and dental sibilants in the content word condition were realized higher in frequency than those in the function word condition [retroflex: $t(2235) = 4.25$, $p < .001$; dental: $t(1233) = 5.48$, $p < .001$]. As could be observed, the distinction between retroflex and dental sibilants was made greater in the content word condition.

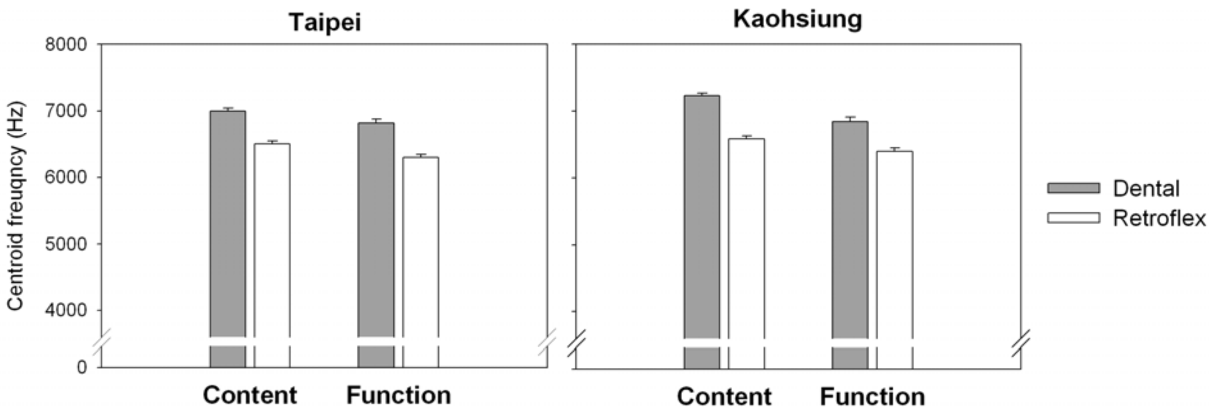


Figure 4.11 The mean centroid frequency of content and function word unaspirated affricates (z type) in unrounded vowel context of speakers from Taipei and Kaohsiung.

Our second analysis examined the data of S2 fricatives (s type) in the rounded vowel context. Figure 4.12 shows the realizations of retroflex and dental sibilants in the content word and function word conditions of the four speaker groups. As can be seen, particularly for Taipei male speakers, retroflex sibilants were realized a lot higher than dental sibilants in the function word condition. A closer examination of the distribution of Taipei males' sibilant tokens revealed that such a phenomenon actually resulted from speaker variability. Table 4.3 gives the token number and the percentage of S2 fricatives (s type) in the rounded vowel context contributed by each Taipei male speaker. Specifically, in the function word condition, retroflex sibilant tokens were contributed mainly by speaker CZX, while dental sibilants were contributed mostly by speakers HSK and YYS. Figure 4.13 shows the mean frequency range of each Taipei male speaker. As can be seen, the frequency ranges of CZX and JXW are about 2500 Hz higher than those of HSK and YYS. Because retroflex sibilants in the function word condition were mostly contributed by CZX, the mean centroid frequency was thus high. On the other hand, HSK and YYS contributed about 85% of dental sibilant tokens in the function word condition, thus leading to low centroid frequency. In this regard, when analyzing S2 fricatives (s type) in the rounded vowel context, we only included data of HSK and YYS in the function word condition for Taipei male speakers, owing to the fact that these two speakers had more comparable frequency ranges and the total retroflex and dental sibilant tokens of these two speakers were sufficient for analyses.

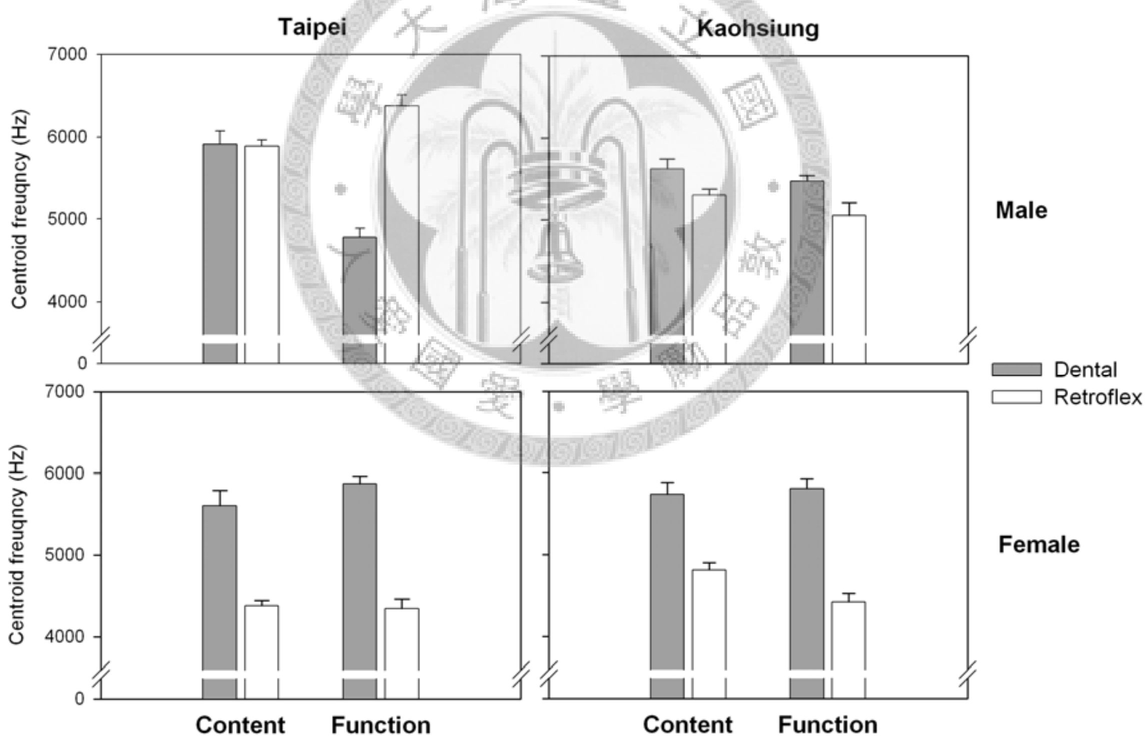


Figure 4.12 The mean centroid frequency of content and function word fricatives (s type) in the rounded vowel context of both female and male speakers from Taipei and Kaohsiung.

Table 4.3 The number and percentage (in paranthesis) of S2 fricative (s type) sibilant tokens in the rounded context contributed by each Taipei male speaker.

		Speaker			
Word class	Place	CZX	HSK	JXW	YYS
Content	Retroflex	108 (38%)	33 (12%)	65 (23%)	76 (27%)
	Dental	24 (32%)	6 (8%)	15 (20%)	30 (40%)
Function	Retroflex	64 (65%)	1 (1%)	13 (13%)	21 (21%)
	Dental	9 (12%)	30 (41%)	2 (3%)	33 (45%)

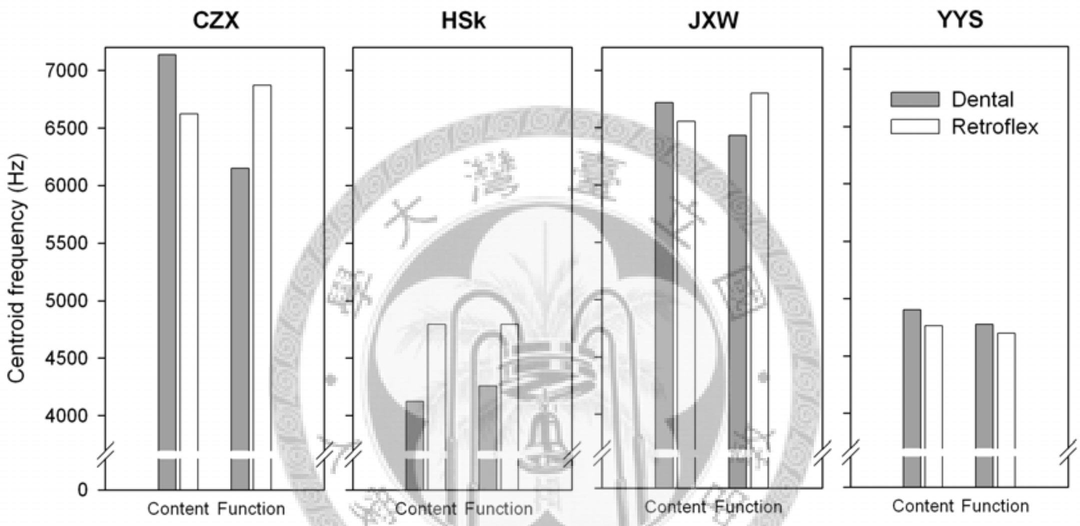


Figure 4.13 The mean centroid frequency of content and function word fricatives (s type) in the rounded vowel context of each Taipei male speaker.

Figure 4.14 presents the revised data after excluding speaker CZX and JXW from Taipei male group. A four-way ANOVA with region (Taipei/Kaohsiung), gender (female/male), word class (content/function) and place (retroflex/dental) as between-subject factors was carried out. The main effects of region and place were significant [region: $F(1, 1413) = 43.24$, $p < .001$; place: $F(1,1413) = 130.40$, $p < .001$]. There were two two-way interactions, including region × gender [$F(1, 1413) = 18.00$, $p < .001$] and gender × place [$F(1, 1413) = 82.80$, $p < .001$]. One significant three-way interaction was reported [region × gender × place: $F(1, 1413) = 7.56$, $p < .01$]. None of the four speaker groups showed significant effects of word class.

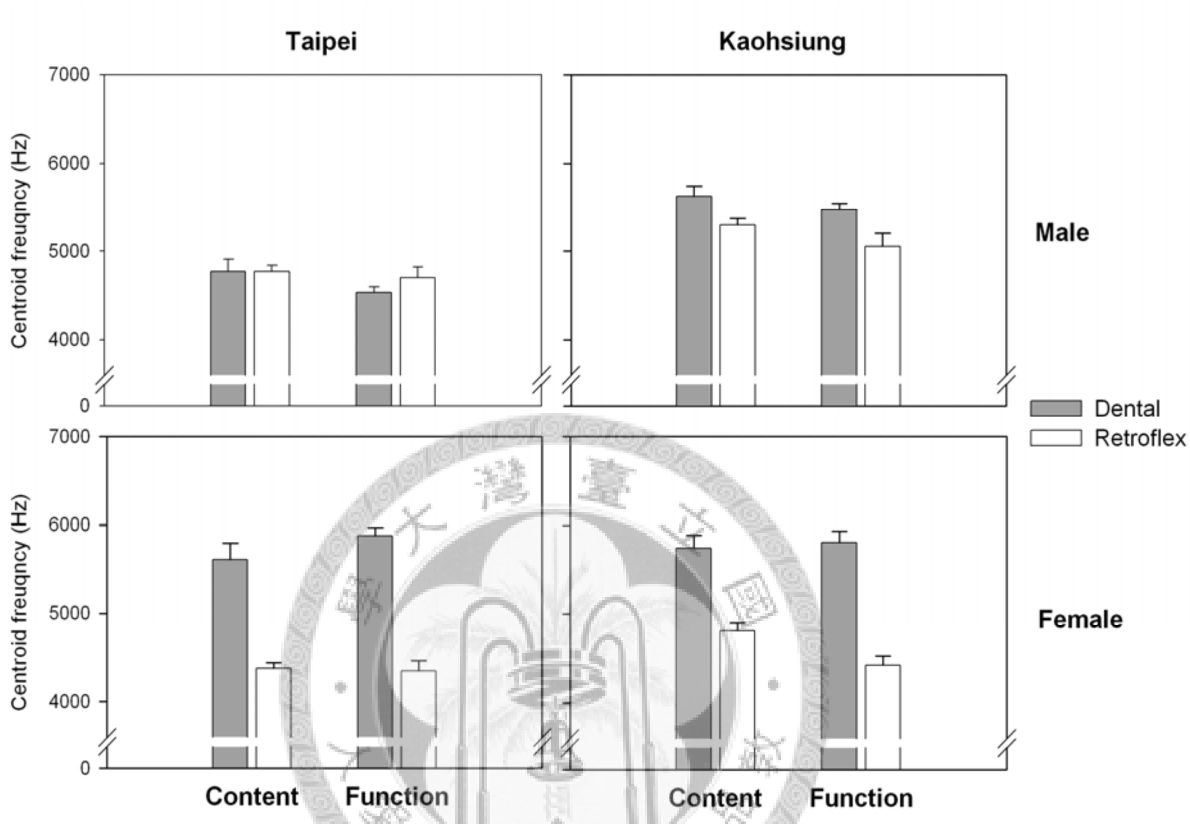


Figure 4.14 The mean centroid frequency of content and function word fricatives (s type) in the rounded vowel context, with speaker CZX and JXW excluded from Taipei male group.

4.4 The interaction between stress and word class

The effects of stress and word class were examined separately in the analyses so far. For example, when stress levels were compared, only content word data were included; in addition, when word classes were compared, only S2 data were put into analyses. In this regard, this section explores how stress and word class interact with each other, in order to obtain a better understanding of these two linguistic factors. Two analyses were conducted accordingly.

In the first analysis, S1, S2 and S3 fricatives (s type) in the unrounded vowel context of both content and function word classes were included. Due to the lack of dental sibilant tokens in these three stress conditions, only retroflex sibilants were examined. Figure 4.15 shows the realizations of retroflex fricatives in different stress levels and word classes. A four-way ANOVA with region (Taipei/Kaohsiung), gender (female/male), stress (S1/S2/S3) and word class (content/function) as between-subject factors was performed. Gender and stress were reported with significant main effects [gender: $F(1, 5875) = 81.84$, $p < .001$; stress: $F(2, 5875) = 5.02$, $p < .01$]. Two two-way interactions were also significant, including region × gender [$F(1, 5875) = 23.25$, $p < .001$] and gender × stress [$F(2, 5875) = 5.81$, $p < .005$].

As can be seen, for all three stress levels, male speakers had significantly higher centroid frequency than female speakers [S1: $t(119) = -2.63$, $p < .05$; S2: $t(5256) = -21.65$, $p < .001$; S3: $t(518) = -10.04$, $p < .001$]. In addition, while females did not show significant differences among the three stress levels, male speakers did make distinctions. In particular, post hoc Bonferroni tests showed that retroflex fricatives in the S1 and S2 conditions were realized significantly lower in centroid frequency than those in the S3 condition (S1: $p < .005$; S2: $p < .005$).

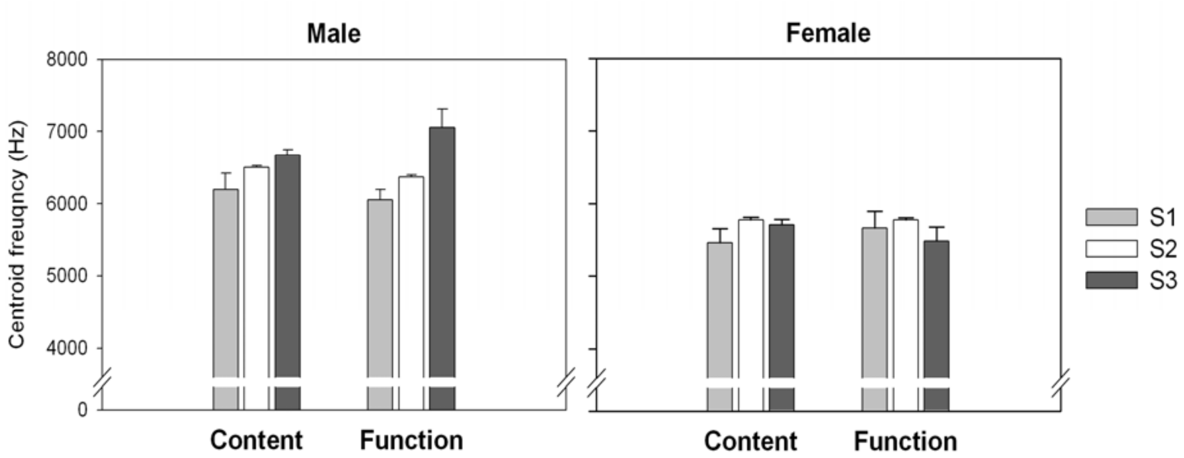


Figure 4.15 The mean centroid frequency of S1, S2 and S3 retroflex fricatives (s type) in the unrounded vowel context of the two word classes.

Our second analysis investigated S0 and S2 unaspirated affricate sibilants (z type) in the unrounded vowel context of both word classes. Due to the lack of retroflex unaspirated affricate tokens in the S0 condition, the comparison between S0 and S2 was limited to the data of dental sibilants only. A four-way ANOVA with region (Taipei/Kaohsiung), gender (female/male), stress (S0/S2) and word class (content/function) as between-subject factors was executed. Results showed significant main effects of three factors [region: $F(1, 1394) = 5.91$, $p < .05$; gender: $F(1, 1394) = 5.82$, $p < .05$; stress: $F(1, 1394) = 18.18$, $p < .001$]. Three two-way interactions were significant, including gender × stress [$F(1, 1394) = 7.67$, $p < .01$], region × gender [$F(1, 1394) = 6.75$, $p < .01$], stress × word class [$F(1, 1394) = 9.01$, $p < .005$].

Figure 4.16 shows how dental unaspirated affricates were realized in different stress and word class conditions. Post hoc independent *t* test reported that for both female and male speakers, in the content word condition, the centroid frequency of S0 and S2 dental unaspirated affricates were not significantly distinguished. However, S0 dental sibilants were significantly higher than S2 ones in the function word condition [$t(503) = 4.80$, $p < .001$]. As for the gender effect, female speakers did not distinguish S0 and S2 dental sibilants significantly, whereas male speakers realized dental sibilants higher in S0 than in S2 [$t(738) = 4.26$, $p < .001$]. A cross-gender comparison further revealed that female speakers realized dental unaspirated affricate sibilants higher than male speakers in the S2 condition [$t(1233) = 7.97$, $p < .001$], but not in the S0 condition.

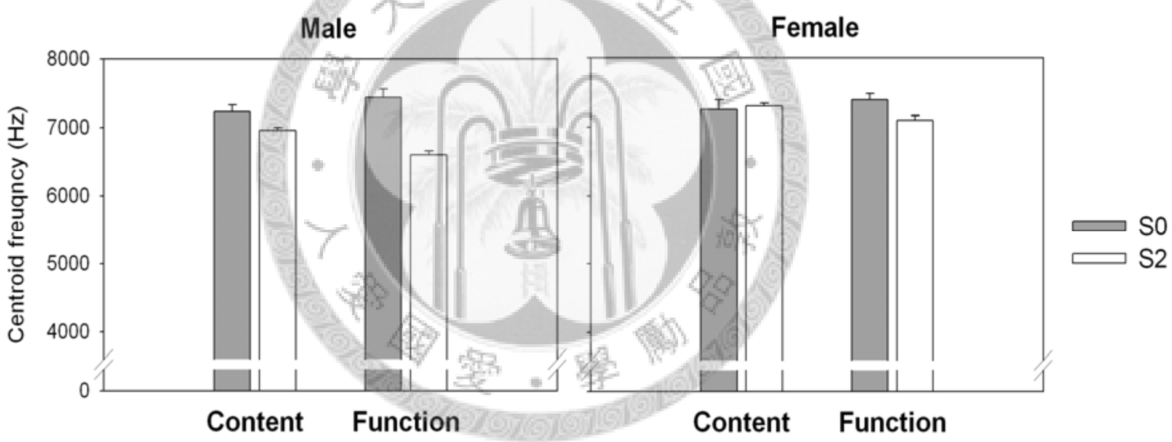


Figure 4.16 The mean centroid frequency of S0 and S2 dental unaspirated affricates (z type) in the unrounded vowel context of the two word classes.

4.5 Substituted valid sibilant tokens

Among all the valid data, 159 tokens were substituted by the other sibilant counterpart. In other words, retroflex sibilants were replaced by their dental counterparts, or vice versa. The distribution of these substituted tokens is shown in Table 4.4. As can be seen, substituted tokens mostly occurred more in the content word than in the function word condition. Moreover, usually it was in the S2 and S3 conditions that substitution was most likely to occur. Region and gender differences could be observed as well. In particular, Taipei males had quite a few retroflex tokens that were replaced by dental sibilant tokens, whereas Kaohsiung females tended to substitute retroflex sibilants for dental sibilants.

As shown in Table 4.4, the token number of substituted sibilants in each cell was too small for us to do statistical analyses. As a result, the statistical analysis of substituted tokens only utilized the data from Kaohsiung female speakers, who had relatively more dental sibilant tokens replaced by retroflex sibilants in the rounded vowel context. Furthermore, according to our previous analyses, the effects of sibilant type and word class were not significantly found for Kaohsiung female speakers, we thus combined all S2 substituted dental sibilant tokens in the rounded vowel context for further comparison. Figure 4.17 shows the mean centroid frequency of canonically realized retroflex and dental sibilants in the corresponding condition (Kaohsiung females' S2 sibilants in the rounded vowel context), and that of substituted dental sibilants. One-way ANOVA was adopted to test the frequency difference among the canonical dental sibilants, the canonical retroflex sibilants and the substituted dental sibilants. Significance was reported [$F(2, 837) = 77.76$, $p < .001$]. Post hoc Bonferroni test showed that substituted dental sibilants were significantly realized lower than canonical dental sibilants (p < .001) and also even lower than canonical retroflex sibilants (p < .001). It thus revealed that dental sibilants, once being substituted by retroflex sibilants, would become even more retroflex than canonical retroflex tokens.

Table 4.4 The overall substituted token distribution of (a) Taipei male (b) Taipei female (c) Kaohsiung male and (d) Kaohsiung female groups (R: retroflex; D: dental; c: aspirated affricate; s: fricative; z: unaspirated affricate). The categorization of R/D is provided in *Note* at the bottom of this table.

(a)

Word class	Stress	Unrounded vowel context						Rounded vowel context						Total
		c		s		z		c		s		z		
		R	D	R	D	R	D	R	D	R	D	R	D	
Content	S0	0	0	0	0	0	0	0	0	0	0	0	0	0 (0%)
	S1	0	0	0	0	0	0	0	0	0	0	0	0	0 (0%)
	S2	1	0	9	1	1	2	0	1	0	4	0	1	20 (53%)
	S3	0	0	10	1	0	0	1	1	0	1	0	0	14 (37%)
Function	S0	0	0	0	0	0	0	0	0	0	0	0	0	0 (0%)
	S1	0	0	0	0	0	0	0	0	0	0	0	0	0 (0%)
	S2	0	0	3	0	0	0	1	0	0	0	0	0	4 (11%)
	S3	0	0	0	0	0	0	0	0	0	0	0	0	0 (0%)
Total		1 (3%)	0 (0%)	22 (58%)	2 (5%)	1 (3%)	2 (5%)	2 (5%)	2 (5%)	0 (0%)	5 (13%)	0 (0%)	1 (3%)	38 (100%)

(b)

Word class	Stress	Unrounded vowel context						Rounded vowel context						Total
		c		s		z		c		s		z		
		R	D	R	D	R	D	R	D	R	D	R	D	
Content	S0	0	0	0	0	0	0	0	0	0	0	0	0	0 (0%)
	S1	0	0	0	0	0	0	0	0	0	0	0	0	0 (0%)
	S2	2	0	2	1	6	1	1	1	0	0	2	0	16 (53%)
	S3	0	0	0	0	1	1	3	0	3	1	1	0	10 (33%)
Function	S0	0	0	0	0	0	0	0	0	0	0	0	0	0 (0%)
	S1	0	0	0	0	0	0	0	0	0	0	0	0	0 (0%)
	S2	0	0	0	0	4	0	0	0	0	0	0	0	4 (13%)
	S3	0	0	0	0	0	0	0	0	0	0	0	0	0 (0%)
Total		2 (7%)	0 (0%)	2 (7%)	1 (3%)	11 (37%)	2 (7%)	4 (13%)	1 (3%)	3 (10%)	1 (3%)	3 (10%)	0 (0%)	30 (100%)

(c)

Word class	Stress	Unrounded vowel context						Rounded vowel context						Total
		c		s		z		c		s		z		
		R	D	R	D	R	D	R	D	R	D	R	D	
Content	S0	0	0	0	0	0	0	0	0	0	0	0	0	0 (0%)
	S1	0	0	0	0	0	0	0	0	0	0	0	0	0 (0%)
	S2	1	0	1	0	2	1	0	0	1	0	0	1	7 (54%)
	S3	0	0	1	0	0	0	0	0	0	0	0	0	1 (8%)
Function	S0	0	0	0	0	0	0	0	0	0	0	0	0	0 (0%)
	S1	0	0	0	0	0	0	0	0	0	0	0	0	0 (0%)
	S2	0	0	2	0	1	1	0	0	0	0	0	1	5 (38%)
	S3	0	0	0	0	0	0	0	0	0	0	0	0	0 (0%)
Total		1 (8%)	0 (0%)	4 (31%)	0 (0%)	3 (23%)	2 (15%)	0 (0%)	0 (0%)	1 (8%)	0 (0%)	0 (0%)	2 (15%)	13 (100%)

(d)

Word class	Stress	Unrounded vowel context						Rounded vowel context						Total
		c		s		z		c		s		z		
		R	D	R	D	R	D	R	D	R	D	R	D	
Content	S0	0	0	0	0	0	0	0	0	0	0	0	0	0 (0%)
	S1	0	0	0	0	0	0	0	0	0	0	0	0	0 (0%)
	S2	2	0	5	1	4	4	0	6	2	11	0	11	46 (59%)
	S3	0	0	8	0	0	0	1	1	1	3	0	3	17 (22%)
Function	S0	0	0	0	0	0	1	0	0	0	0	0	0	1 (1%)
	S1	0	0	0	0	0	0	0	0	0	1	0	0	1 (1%)
	S2	0	0	2	0	0	0	0	1	0	10	0	0	13 (17%)
	S3	0	0	0	0	0	0	0	0	0	0	0	0	0 (0%)
Total		2 (3%)	0 (0%)	15 (19%)	1 (1%)	4 (5%)	5 (6%)	1 (1%)	8 (10%)	3 (4%)	25 (32%)	0 (0%)	14 (18%)	78 (100%)

Note. The substituted sibilant tokens in this table were categorized according to their original places of articulation. For example, those in the "R" cells were retroflex sibilants, but pronounced as dental sibilants, and those in the "D" cells were dental sibilants that were replaced by retroflex sibilants.

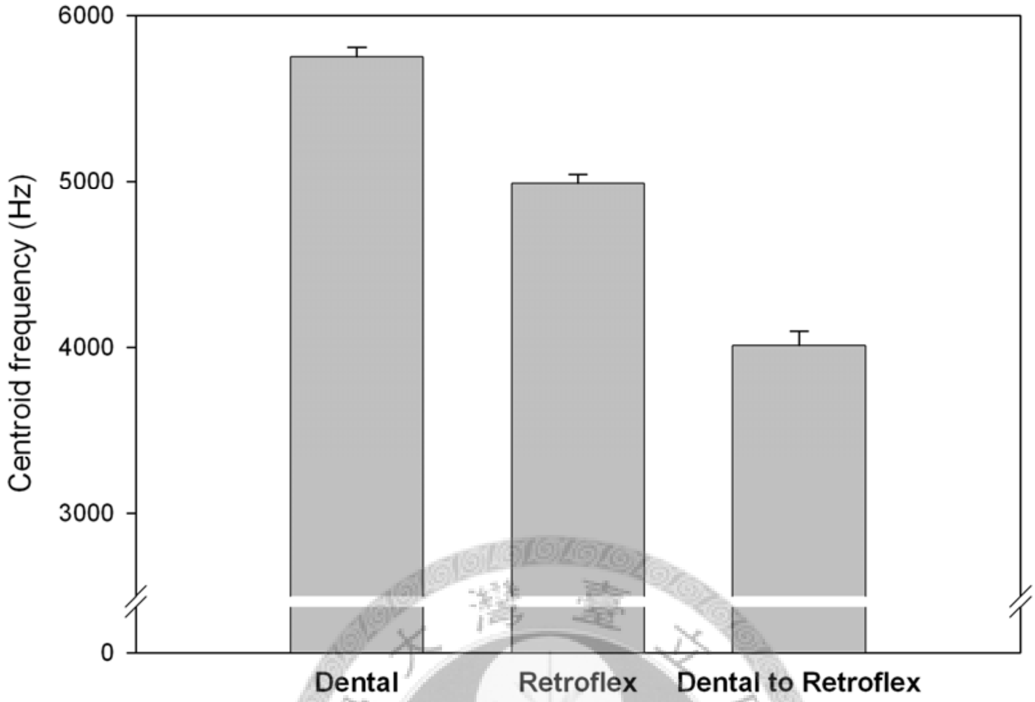


Figure 4.17 The mean centroid frequency of canonically realized dental/retroflex sibilants and substituted dental sibilants in rounded vowel context of Kaohsiung females.

4.6 Sibilant deleted tokens

Sibilant deletion was frequently seen in our spontaneous speech data. About 70% of the invalid tokens belonged to this category. Although acoustic measurements were not applicable, it was still worthwhile to look into its distribution in spontaneous speech data.

The distribution of these sibilant deleted tokens is summarized in Table 4.5, along with the deletion rate. The deletion rate was the proportion of deleted tokens to the total tokens in each cell. It was derived from the following formula:

$$\text{deletion rate} = \frac{N \text{ of deleted tokens}}{N \text{ of deleted tokens} + N \text{ of valid canonical tokens}}$$

For example, Table 4.2(a) shows that for S2 retroflex aspirated affricates (c type) in the content word condition, Taipei male speakers had 501 canonical valid tokens (415 in the unrounded context and 86 in the rounded context). As shown in Table4.5, Taipei male speakers had 7 deleted tokens in the same condition. The deletion rate of this condition would be 7/(7+501) = 0.013. Thus, about 1% of the S2 retroflex aspirated affricates in the content word condition were sibilant deleted tokens.

The deletion rate showed that for all groups of speakers, the occurrence of deleted tokens was particularly high in the S1 condition. This illustrated that for those tokens labeled with S1, most of their sibilants were deleted. It was further noticeable that unlike the stress effect, the word class effect, on the other hand, was not as salient. As we could see in Table 4.5, the deletion rate was not very different for the two word class categories. It thus seemed that stress had a stronger effect than word class in terms of determining sibilant realizations.

Table 4.5 The overall sibilant deleted token distribution and the deletion rate (in the parenthesis) of (a) Taipei male (b) Taipei female (c) Kaohsiung male and (d) Kaohsiung female groups (R: retroflex; D: dental; c: aspirated affricate; s: fricative; z: unaspirated affricate).

(a)

Word class	Stress	c		s		z		Total
		R	D	R	D	R	D	
Content	S0	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	11 (25%)	11 (25%)
	S1	14 (100%)	2 (100%)	37 (88%)	3 (100%)	53 (98%)	38 (100%)	147 (96%)
	S2	7 (1%)	9 (5%)	42 (3%)	2 (1%)	62 (10%)	57 (15%)	179 (6%)
	S3	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Function	S0	0 (0%)	0 (0%)	0 (0%)	0 (0%)	9 (60%)	0 (0%)	9 (39%)
	S1	0 (0%)	0 (0%)	158 (92%)	1 (50%)	25 (86%)	11 (100%)	195 (91%)
	S2	0 (0%)	1 (5%)	123 (15%)	0 (0%)	19 (7%)	42 (22%)	185 (13%)
	S3	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (9%)	1 (3%)
Total		21 (3%)	12 (5%)	360 (15%)	6 (2%)	168 (15%)	160 (22%)	727 (13%)

(b)

Word class	Stress	c		s		z		Total
		R	D	R	D	R	D	
Content	S0	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	2 (11%)	2 (11%)
	S1	26 (100%)	2 (100%)	26 (62%)	0 (0%)	34 (97%)	22 (100%)	110 (85%)
	S2	53 (23%)	2 (3%)	32 (65%)	1 (1%)	56 (13%)	31 (9%)	175 (10%)
	S3	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Function	S0	0 (0%)	0 (0%)	0 (0%)	0 (0%)	2 (25%)	0 (0%)	2 (5%)
	S1	0 (0%)	0 (0%)	78 (86%)	0 (0%)	50 (94%)	2 (67%)	130 (88%)
	S2	0 (0%)	0 (0%)	97 (11%)	0 (0%)	9 (3%)	26 (17%)	132 (9%)
	S3	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Total		79 (22%)	4 (3%)	233 (12%)	1 (0%)	151 (15%)	83 (13%)	551 (13%)

(c)

Word class	Stress	c R	c D	s R	s D	z R	z D	Total
Content	S0	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	11 (23%)	11 (23%)
	S1	21 (91%)	3 (100%)	107 (90%)	1 (50%)	83 (93%)	32 (86%)	247 (90%)
	S2	26 (7%)	4 (4%)	34 (4%)	2 (2%)	64 (12%)	44 (12%)	174 (8%)
	S3	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Function	S0	0 (0%)	0 (0%)	0 (0%)	0 (0%)	10 (91%)	3 (23%)	13 (54%)
	S1	0 (0%)	0 (0%)	168 (88%)	0 (0%)	58 (85%)	5 (100%)	231 (86%)
	S2	0 (0%)	0 (0%)	117 (15%)	0 (0%)	25 (9%)	64 (40%)	206 (15%)
	S3	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Total		47 (9%)	7 (5%)	426 (20%)	3 (1%)	240 (22%)	159 (22%)	882 (18%)

(d)

Word class	Stress	c R	c D	s R	s D	z R	z D	Total
Content	S0	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	7 (24%)	7 (24%)
	S1	10 (91%)	2 (100%)	106 (85%)	2 (100%)	86 (93%)	19 (86%)	225 (89%)
	S2	29 (13%)	5 (5%)	44 (5%)	0 (0%)	64 (12%)	59 (14%)	201 (9%)
	S3	1 (2%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (0%)
Function	S0	0 (0%)	0 (0%)	0 (0%)	0 (0%)	4 (57%)	1 (5%)	5 (19%)
	S1	0 (0%)	0 (0%)	305 (93%)	0 (0%)	45 (68%)	11 (100%)	361 (89%)
	S2	0 (0%)	0 (0%)	117 (13%)	1 (1%)	27 (11%)	57 (45%)	202 (15%)
	S3	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Total		40 (13%)	7 (6%)	572 (25%)	3 (1%)	226 (21%)	154 (22%)	1002 (21%)

CHAPTER 5 DISCUSSION

In this chapter, accounts for the results found in this study are provided. Section 5.1 summarizes the overall realizations of retroflex and dental sibilants identified in the present study. In section 5.2, the effects of social and linguistic factors on valid sibilant tokens, including canonically realized and substituted ones, are discussed. Section 5.3 focuses on linguistic factors in particular, looking into the interaction effects found between stress and word class. Explanations for the phonological processes of retroflexion and dentalization in Taiwan Mandarin are presented in Section 5.3.

5.1 Sibilant realizations in spontaneous speech

This study aimed to acoustically determine the realizations of retroflex and dental sibilants in Taiwan Mandarin spontaneous speech. Our results showed that both retroflex and dental sibilants were found to be canonically realized, substituted by the other sibilant counterparts, or deleted in spontaneous speech. In particular, the majority of sibilant tokens were realized as their lexically specified sibilant categories. The acoustic measurements revealed that retroflex and dental sibilants were generally distinguished in spontaneous speech. Such a result was comparable with previous research on experimental data (Jeng, 2006; Tse, 1998). In addition, we further found that sibilant realizations were subject to both social and linguistic factors, in which degrees of sibilant contrasts differed among speaker groups and linguistic conditions.

In our study, some sibilant tokens were categorized as being substituted by their sibilant counterparts. Specifically, it was found that sibilant substitution was bi-directional. That is, substitution of retroflex sibilants for dental ones and substitution of dental sibilants for retroflex ones were both existent in Taiwan Mandarin, and their occurrences were speaker-group dependent. Different mechanisms should be adopted to explain the two directions of substitution phenomenon.

Finally, sibilant deletion was also a common phenomenon found in spontaneous speech, in which about 15% of the sibilant tokens were not realized at all in our data. Although acoustic measurements were not applicable for these deletion tokens, we still found that sibilant deletion was tightly bound with linguistic factors, and it was more likely to occur in certain linguistic conditions.

5.2 The effects of social and linguistic factors

In the present study, results showed that social factors and linguistic factors closely interact with each other. Therefore, their effects were discussed together in this section. One thing to be noticed is that because the distribution of our corpus data was not balanced, the discussion had to be divided into several parts. In Section 5.2.1, we discuss stress realizations by different speaker groups. A relevant interaction effect found between sibilant type and vowel context is accounted for in Section 5.2.2. The effect of word class is discussed separately in Section 5.2.3.

5.2.1 Stress realizations by different speaker groups

In this study, it was found that stress exerted a salient effect in terms of sibilant realizations. In particular, larger sibilant contrasts were made in the stressed condition. Such a result was compatible with the effect of prosodic strengthening on segmental realizations, depicting that linguistic contrasts tend to be maximized or maintained in prosodically stronger conditions (e.g., Cho, 2005; Cho & McQueen, 2005; Cole, et al., 2007; de Jong, 1995). Nevertheless, despite the fact that larger contrasts were made in the S3 condition, speaker group differences were observed. Results showed that the realizations of stress actively interacted with the two social factors – region and gender. The specific discrepancies are summarized in Table 5.1.

As can be seen, the stress effect was observed for all speaker groups except Taipei males. S3 conditions were usually characterized by sibilants being realized towards a more dental direction. Cross-regional comparisons further showed that both Kaohsiung females and Kaohsiung males exhibited an even more obvious stress effect than Taipei speakers. It should be noticed that although the effect of stress was generally less shown in both Taipei female and male speakers, the weak effect of stress should be understood differently. Taipei females usually made larger sibilant contrasts than Kaohsiung females; therefore, the reason why they did not distinguish different stress levels might be due to the fact that they had reached the ceiling effect. In other words, Taipei females were very articulate even in the S2 condition. In this regard, the contrast between S2 and S3 was not easily revealed. On the other hand, Taipei males always had comparatively more merging sibilant realizations. Therefore, their lack of stress effect was more related to the fact that Taipei males generally did not distinguish their retroflex and dental sibilants successfully, regardless of stress levels.

Table 5.1 The degree of sibilant contrast made by different speaker groups.

Speaker group	Sibilant contrast	Realizations [a]
Taipei Male	No stress effect	
Taipei Female	S3 > S2	D (more dental)
Kaohsiung Male	S3 > S2	R & D (more dental)
Kaohsiung Female	S3 > S2	D (more dental)

[a.] The descriptions in this column indicate how the larger sibilant contrast was made. D stands for dental sibilants, while R stands for retroflex sibilants. For instance, the description for Taipei Female means that the larger contrast was made in the S3 condition by realizing dental sibilants toward a more dental direction.

Our results further showed that stress realizations were speaker gropu-dependent. Given the same stress levels, sibilant realizations differed among speaker groups. Table 5.2 summarizes the comparisons of sibilant contrast degrees with regard to the effects of gender and region. As can be seen in Table 5.2(a), females generally made larger sibilant contrasts than males. According to sociolinguistic theories, females are more conservative speakers and tend to use more standard forms, and the gender effect found in the present study showed similar patterns with previous sociolinguistic studies (M.-C. Li, 1995; C. C. Lin, 1983; Y.-H. Lin, 1988). Across stress levels, Taipei females consistently distinguished sibilants better by realizing retroflex sibilants in a more retroflexed way than Taipei males. For Kaohsiung males, on the other hand, female speakers showed larger sibilant contrasts, but both their retroflex and dental sibilants were realized towards dentalization. Such a result could be attributed to the fundamental physiological difference between females and males. It is known that females usually have shorter vocal tracts than males; therefore, it is expectable that their sibilant frequencies will altogether be higher. This explained why Kaohsiung females, though generally having larger sibilant contrasts, had both their retroflex and dental sibilants realized higher in frequencies than males.

Table 5.2 The comparisons of sibilant contrast degrees between (a) genders in the same region and (b) regions of the same gender.

(a)

Region	Stress	Sibilant contrast	Realizations [a]
Taipei	S2	Female > Male	R (more retroflex) D (more dental)
	S3	Female > Male	R (more retroflex)
Kaohsiung	S2	Female > Male	R & D (more dental)
	S3	No gender effect	

(b)

Gender	Stress	Sibilant contrast	Realizations [a]
Female	S2	Taipei > Kaohsiung	R (more retroflex)
	S3	No region effect	
Male	S2	Kaohsiung > Taipei	R (more retroflex)
	S3	Kaohsiung > Taipei	R (more retroflex)

[a.] The descriptions in this column indicate how the larger sibilant contrast was made. D stands for dental sibilants, while R stands for retroflex sibilants. For instance, the description for Taipei speakers in the S2 condition means that females made larger sibilant contrasts than males by realizing both retroflex sibilants and dental sibilants in more extreme directions.

The effect of region is shown in Table 5.2(b). As can be seen, the speaker groups that made greater sibilant distinctions all had retroflex sibilants being realized more retroflexed. Cross-regional comparisons revealed different patterns. For female speakers, Taipei females distinguished sibilants to a greater extent than Kaohsiung females in the S2 condition, whereas for male speakers, it was Kaohsiung males that made greater sibilant distinction, regardless of stress levels. This inconsistency was not expected. Our original prediction for the region effect was that since Kaohsiung speakers speak Min more frequently than Taipei speakers, if the frequency of using Min is indeed effective, the merging degree should be more salient for Kaohsiung speakers. While this was generally true for female speakers, male speakers did not follow this pattern. Examinations of male speakers' metadata showed that possibly there was another confounding factor involved, *i.e.*, Min proficiency. As shown in Table 3.2, while female speakers in the two regions had relatively comparable Min proficiency, Taipei males were on average more proficient in Min than Kaohsiung males. Previous studies have shown that Min proficiency have an effect on the realizations of retroflex sibilants. For example, Lin (1983) found that native Taiwanese speakers, proficient in Min, are more likely to mispronounce retroflex sibilants than native Mandarin speakers, non-proficient in Min. In other words, the more proficiently one speaks Min, the greater degree of sibilant merging would be expected. In the present study, since Taipei males showed more merging patterns than Kaohsiung males, it thus suggested that the effect of Min proficiency had overridden the effect of region, which is correlated with the frequency of using Min.

The sibilant merging patterns of Taipei males could be further illustrated by our substitution data. Substitution demonstrated the extreme cases for sibilant merging. In particular, it was found that the substitution of dental sibilants for retroflex ones, corresponded with the total deretroflexion process depicted in previous studies, was especially frequent for Taipei males. Such a result was compatible with the analysis of canonically realized sibilant tokens, in which Taipei males showed the most sibilant merging patterns, and in certain conditions, retroflex and dental sibilants were not even distinguished. Taken together, it could be concluded that for Taipei male speakers, sibilant merging was actively in progress. While most of previous studies usually roughly referred to this sibilant merging phenomenon as deretroflexion, our study further identified different realizations of sibilant merging. That is, for Taipei male speakers, not only did retroflex sibilants showed the tendency to become more similar to dental ones, but the extreme cases of total deretroflexion were seen from time to time.

While deretroflexion is generally the norm observed in Taiwan Mandarin, it was interesting to see that the other direction of substitution, *i.e.,* the substitution of retroflex sibilants for dental ones, was also found in our data. Particularly for Kaohsiung females, they had a number of instances of such substitution. Since this substitution direction was contradictory to what we generally observed in Taiwan Mandarin, we needed to explain its occurrence. Chung (2006) thought of such a phenomenon as hypercorrection. Given the fact that retroflexion is commonly considered to be prestigious, speakers would incorrectly substitute them for dental sibilants. Taking this into accounts, we should be able to see why Kaohsiung females hypercorrected so often. According to Labov (1990), in general, female speakers are more linguistically insecure, and thus they adopt the prestigious forms more often than male speakers. If we consider Taipei dialect as the standard dialect of Taiwan Mandarin, Kaohsiung speakers, especially females, would be eager to modify their speech in approach of the standard dialect. Therefore, it was not surprising that more hypercorrections were found in the speech of Kaohsiung females. Their attempts to adopt the standard form were also revealed in our canonically realized sibilant data. As discussed previously, while Kaohsiung females made smaller contrasts than Taipei females at the default stress level, this regional difference was however eliminated in the prosodically prominent condition. Taken together, it could be seen that Kaohsiung females were so sensitive to the standard performance that they were eager to approach the standard form. Such eagerness was further demonstrated by their mistaken usage of retroflexion in their speech.

5.2.2 Sibilant type and vowel context

When examining the sibilant realizations in response to stress by different speaker groups, our results further revealed an interesting interaction between sibilant type and the following vowel context. Accordingly, it had been shown that rounded vowels would lower the general frequencies of their preceding sibilants. In terms of the contrast between retroflex and dental sibilants, we further found that sibilants were better distinguished in the unrounded vowel context than in the rounded vowel context in this study. Moreover, it was found that frequencies were lowered to a greater extent for dental sibilants than for retroflex sibilants. Such a result corresponded to Jeng's (2006) finding that compared with retroflex sibilants, dental sibilants are much more affected by the roundedness of the following vowels.

The larger sibilant contrast in the unrounded vowel context could be accounted for from the production perspective. When followed by rounded vowels, the length preceding the constriction position in the vocal tract increases due to lip protrusion.

Such a production cause was reflected on the acoustic aspect, and thus we found the overall lowering of sibilant frequencies. Because the frequency range was suppressed to the lower frequency limit and because speakers needed to maintain the contrast between retroflex and dental sibilants within the compressed frequency range, the distinction would thus become smaller. For dental sibilants, their frequencies were in the high range when followed by unrounded vowels, so there was more room for lowering in the rounded vowel context. In this regard, it was conceivable that the frequencies of dental sibilants would be lowered to a greater extent than those of retroflex sibilants. The vowel context effect, therefore, would be more influential for dental sibilants than for retroflex sibilants.

Nevertheless, although it was a general trend that sibilant contrasts were made larger in the unrounded vowel context, such an effect was however absent for the unaspirated affricate sibilants. We thus suspected that the unaspirated affricate sibilant pair was possibly the most merged one among the three sibilant pairs. This speculation, in effect, has partially been evidenced by previous studies. For example, Tse (1998) found that unaspirated affricates are not distinguished by young male Taiwan Mandarin speakers, while aspirated affricates and fricatives are still well distinguished. Also, Jeng (2006) showed that the sibilant contrasts of affricate sibilant pairs were generally made smaller than that of the fricative pair. Tse (1998) suspected that the more merging pattern of unaspirated affricates might have something to do with the frication duration. Accordingly, the frication of unaspirated affricate sibilants was a lot shorter than that of aspirated affricate and fricative sibilants. His observation was proven both in Jeng's study (2006) and in our study. The mean frication duration for the three sibilant types is shown in Figure 5.1. As can be seen, unaspirated affricate sibilants (z type) had a lot shorter frication than the other, and the discrepancies were statistically significant for both sibilant categories [dental: $p < .001$; retroflex: $p < .001$]. Possibly owing to the short frication, the contrast between retroflex and dental sibilants could not be easily acquired, thus leading to a more merging pattern for unaspirated affricate sibilants.

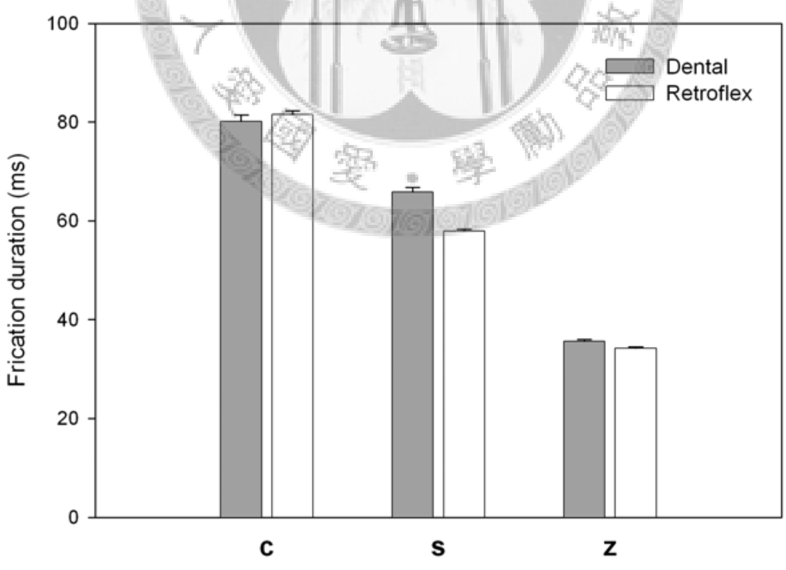


Figure 5.1 The mean frication duration of three sibilant types.

However, it should be noticed that the long duration of aspirated affricate sibilants (c type) in effect resulted from the aspiration portion following the frication. As aforementioned, our pilot examination on aspirated affricates showed that frication constituted only the first one-third portion of the total duration. If we applied such a criterion to compute the actual frication duration for aspirated affricates, the frication duration of aspirated affricates was actually as short as, or even shorter than, that of unaspirated affricates. Nevertheless, aspirated affricates were not as merged as unaspirated affricates. One possible explanation was that the cue for sibilants' place of articulation existed in not only the frication part but the aspiration portion as well. When looking for acoustic parameters that distinguish sibilants, Li (2008) found that the second formant values measured in the onset positions of the vowels reliably indicate preceding sibilant categories. Alone the same line, since aspiration usually carries formants of the following vowels, it was very likely that the place information of the preceding sibilants was also embedded in the aspiration parts. Although so far no direct evidence is available, this inference should be able to help explain why aspirated affricates, also having short frication duration, were not shown to be as merged as unaspirated affricates.

Given the fact that unaspirated affricates were potentially the most merged sibilant pair in Mandarin, the lack of vowel context effect for their sibilant contrasts was thus conceivable. As discussed previously, unrounded vowel context was an easier condition to make sibilant contrasts, since the frequency range was not suppressed. However, because unaspirated affricate sibilants generally showed greater merging patterns themselves; thus, large contrasts could not be made even in the easier vowel context.

5.2.3 The effect of word class

In our analyses, we examined the effect of word class in two conditions – one was the unaspirated affricates in the unrounded vowel context, and the other was the fricatives in the rounded vowel context. The effect of word class was indeed observed. In particular, it was found that both retroflex and dental sibilants were realized towards a more dental direction in the content word condition than in the function word condition; the sibilant distinction was still made greater in the content word condition. This finding was parallel to the general finding that content words, semantically more informative, are usually articulated more clearly than function words (e.g.,van Bergem, 1993). In this regard, retroflex and dental sibilants were thus distinguished better. Nonetheless, it was noticeable that no significant word class effect was reported in the analysis of fricative sibilants. Such a discrepancy might be due to the influence of vowel contexts. As noted previously, the frequency range was compressed in the rounded vowel context. Therefore, it was possible that the word class effect was not salient enough to be revealed in the rounded vowel context.

5.3 The interaction between stress and word class

In the present study, the interaction between stress and word class was looked into. To include sufficient sibilant tokens for symmetrical analyses, we had no choice but to examine either dental sibilants or retroflex sibilants only. Two analyses were conducted accordingly, and in these two analyses, we were able to investigate the sibilant realizations in the two prosodically reduced stress levels, *i.e.*, S0 and S1, respectively.

Our first analysis focused on the realizations of S1, S2 and S3 retroflex fricatives in both content word and function word conditions. In particular, stress effect was found significant. Retroflex sibilants were realized towards a more dentalized direction in the stressed condition. Such a result corresponded well with our finding regarding the sibilant realizations in previous analyses. That is, sibilants were realized more dental in the S3 condition. The effect of word class, however, was not reported to be significant in this analysis.

The second analysis looked into the realizations of S0 and S2 dental unaspirated affricates. It was found that the effect of word class was revealed in the S2 condition. Specifically, dental sibilants were more dental in the content word condition than in the function word condition. Comparable with our previous findings of the word class effect, here we also observed that sibilants in the content word conditions had greater degree of dentalization than those in the function word conditions. Moreover, in this analysis, the effect of stress was revealed as well. Specifically, dental sibilants were realized more dental in the S0 condition than in the S2 condition. This, however, contradicted with our previous finding in which dentalization was more saliently shown in a more stressed condition. A closer examination of the S0 dental unaspirated affricate data revealed that among 132 sibilant tokens, 128 of them were of the syllable *zi* ‘noun suffix’. Since this syllable constituted the great majority of the S0 dental unaspirated affricate category, it was possible that the greater degree of dentalization resulted from lexically specific realization of this syllable. Figure 5.2 shows the mean centroid frequency of dental unaspirated affricates of the syllable *zi* in three stress levels and two word class conditions. Statistical analyses did not report significance for either the effect of stress or that of word class, though there seemed to be a trend that S3 dental unaspirated affricates were realized a lot more dental than S0 and S2 ones. Since *zi* was the only sibilant we had in the S0 category, its idiosyncratic realization should be able to account for the contradictory pattern we observed in terms of the relationship between stress and the degree of dentalization.

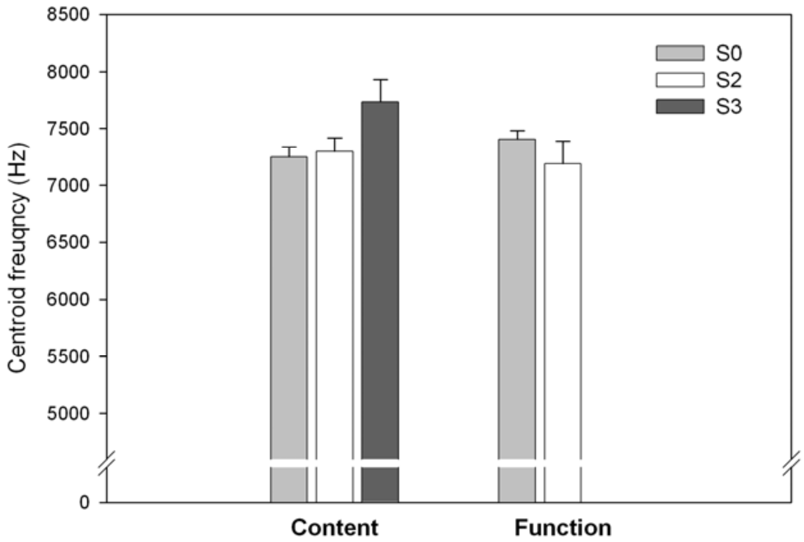


Figure 5.2 The mean centroid frequency of the dental unaspirated affricate of the syllable *zi* in different stress levels and word class conditions.

Sibilant deletion data could also reflect how the two linguistic effects act upon sibilant realizations. In particular, deletion rate was high in the S1 condition. Since S1 is prosodically reduced condition, it was understandable that speakers tend not to make efforts to achieve clear articulation. Thus, deletion should be considered as an extreme case of reduction, in which the whole sibilant segment was deleted. Considering the effect of word class, on the other hand, we found that it was not as salient a factor as stress, especially when it was examined along with stress. With respect to our deletion data, it could be seen that regardless of word class conditions, sibilants were very likely to be deleted in the S1 condition. In this regard, it seemed to suggest that when these two linguistic factors were investigated together, it was the effect of stress that was stronger and more easily detected.

5.4 Retroflexion and dentalization in Taiwan Mandarin

Our examination in the present study showed interesting sibilant realizations in relation to social and linguistic factors. Specifically, when we looked into gender or region differences, it was more often the case that retroflexion characterized the effects of these two social factors. On the other hand, when we investigated the two linguistic factors, dentalization usually characterized the effects of stress and word class. These observations seemed to implicate different statuses of retroflexion and dentalization in Taiwan Mandarin.

Cross-regional and cross-gender comparisons showed that social differences were crucially determined by retroflex sibilants. That is, larger sibilant contrasts usually resulted from greater degree of retroflexion. This might be due to the emphasis of retroflexion in the education system of Taiwan. Retroflexion is generally conceived as the prestigious form that marks formality; therefore, it is the social consensus that whoever speaks “standard” Mandarin should have proper retroflexion in his or her speech. Such a social expectation thus acts upon the stereotypes of gender and region. Speakers who are expected to speak more standard Mandarin should show better retroflexion in his or her speech. In this regard, since females are more often regarded as conservative speakers and the Mandarin spoken in Taipei is generally considered to be more standard, it is socially expected that they should do retroflexion to a greater extent.

As a result, in this study, we observed that the effects of social factors were shown by degrees of retroflexion.

Dentalization, on the other hand, represented the effects of linguistic factors. In particular, sibilants were realized towards a more dental direction in prosodically prominent conditions. Also it was found that sibilants were more dentalized in the content word condition than in the function word condition. It thus seemed that the place-of-articulation phonetic feature [dental] was enhanced in linguistically more prominent conditions. Possibly for Mandarin-Min bilinguals, dental sibilants, existent in both Mandarin and Min, were more active phonemes for them so that they utilized dental sibilants to make linguistic contrasts. Moreover, from the perspective of language universals, dental sibilants are much more unmarked than retroflex sibilants, considering their frequency of occurrence in the world's languages. For example, the UPSID-PC database (Maddieson & Precoda, 1990) reports 143 languages to have dental/alveolar sibilants, while only 29 languages are found to have retroflex sibilants. Given the unmarkedness of dental sibilants, it would be easier for speakers to contrast linguistic levels by realizing different degrees of dentalization.

Another interesting observation was that although greater degrees of dentalization were overall shown in linguistically prominent conditions, not only dental sibilants became more dentalized, but even retroflex sibilants showed the tendency to be more dentalized in stressed conditions. One might wonder whether lexical confusion would be created if retroflex sibilants also became more dentalized. A similar situation is found in the production of Dutch stops /t/ and /d/. In syllable-initial position, /t/ is pronounced as unaspirated, thus resulting in short VOT. It was found by Cho et al. (2005) that once /t/ is prosodically strengthened, its VOT became even shorter, which makes /t/ likely to be confused with /d/. In this regard, Cho et al. (2005) discovered that to avoid lexical confusion, the strengthening effect is compensated by the other feature enhancement, that is, the degree of prevoicing. /d/ showed greater degree of prevoicing in prosodically strong conditions, which could help maintain the contrast between /t/ and /d/. The finding of Dutch stops could thus help explain the prosodic strengthening effect of retroflex and dental sibilants in Taiwan Mandarin. Although the enhancement of the feature [dental] could possibly endanger the contrast between retroflex and dental sibilants, we suspected that probably their contrast would be maintained or maximized in other articulatory aspects, such as the shape of tongue. Due to the fact that the acoustic measurement we adopted in this study, the centroid frequency, was accurate in characterizing differences in place of articulation, other acoustic measurements, such as skewness or kurtosis in the spectral moments analysis, might be able to reveal the strengthening effect in other articulatory aspects.

In summary, our findings suggested that retroflex and dental sibilants functioned differently in Taiwan Mandarin. Retroflexion usually marked formality, and the degree of retroflexion was stable across speaker groups; on the other hand, dentalization characterized linguistic prominence, and the degree of dentalization was flexible in terms of signaling stress differences. In this regard, within the same speaker group, stress and word class differences were exhibited by dentalization; on the other hand, within the same stress level or word class condition, regional and gender differences were demonstrated by retroflexion.

CHAPTER 6 CONCLUSION

The current study has demonstrated that retroflex and dental sibilants in Taiwan Mandarin were generally distinguished in spontaneous speech. However, their realizations were influenced a number of factors. From the social perspective, we found that female speakers distinguished retroflex and dental sibilants better than male speakers. The regional effect, however, was different between genders. Specifically, Taipei females made larger sibilant contrasts than Kaohsiung females; on the contrary, Taipei males merged retroflex and dental sibilants to a greater extent than Kaohsiung males. It was suspected that the factor of Min proficiency was confounded. Linguistically, prosodic prominence and word class also played significant roles in affecting sibilant realizations. In prosodically prominent conditions, larger sibilant contrasts were made, whereas in reduced conditions, sibilants exhibited high tendency to be deleted. As for word class, it was found that given the same stress level, sibilants in the content word condition were distinguished better than those in the function word condition. When stress and word class were examined together, nonetheless, it was more often the case that the effect of stress was more saliently shown.

In addition to the general realizations of retroflex and dental sibilants in Taiwan Mandarin, effects regarding vowel context and sibilant type were discovered. It was found that the degree of sibilant distinction differed in rounded and unrounded vowel contexts. In particular, retroflex and dental sibilants were better contrasted when the following vowels were unrounded. In the rounded vowel context, because the frequency range was suppressed to the lower limit, the distinction was thus smaller. Moreover, among the three pairs of retroflex and dental sibilants, we observed that unaspirated affricates had a more merging pattern than fricatives and aspirated affricates. It was speculated that frication duration might play an important role in affecting the degree of sibilant merging.

Finally, our results implicated different functions for the phonological processes of retroflexion and dentalization in Taiwan Mandarin. Degrees of retroflexion was more stable and insensitive to linguistic differences, but they characterized speaker group differences, both in terms of region and gender. On the other hand, degrees of dentalization did not differ much across speaker groups, but they were comparatively more flexible and active in reflecting linguistic differences, both in terms of stress and word class. In this regard, it should be concluded that retroflexion marked formality, which was more related to the outer speech context, whereas dentalization marked linguistic prominence, which was more related to language-internal categorization.

In the present corpus-based study, analyses were limited and a number of phenomena regarding the realizations of retroflex and dental sibilants could not be well testified, due to the imbalanced data distribution. With larger amount of spontaneous speech data provided in the future, more linguistic aspects regarding the realizations of retroflex and dental sibilants in Taiwan Mandarin can therefore be better and more completely studied.

REFERENCES

Ang, U. (1997). 1997 Taiwan gonggong changsuo shiyong yuyan diaocha [A survey on language use in public occasions in Taiwan, 1997]. In C. Tung (Ed.), *Taiwan yuyan fazhan xueshu yantaohui lunwenji [Proceedings of the Conference on Language Development in Taiwan]* (pp. 83-100). Hsinchu: Quanmin Bookstore.

Beckman, M. E., & Edwards, J. (1994). Articulatory evidence for differentiating stress categories. In P. A. Keating (Ed.), *Papers in Laboratory Phonology III: Phonological structure and phonetic form* (pp. 7-33).

Behrens, S. J., & Blumstein, S. E. (1988). Acoustic characteristics of English voiceles fricatives: A descriptive analysis. *Journal of Phonetics, 16*, 295-298.

Boersma, P., & Weenink, D. (2008). Praat: doing Phonetics by Computer (Version 5.0.22). Amsterdam.

Bolinger, D. (1972). Accent is predictable (If you're a mind-reader). *Language, 48*(3), 633-644.

Chan, H. (1984). *The phonetic development of Mandarin /ʐ/ in Taiwan: a sociolinguistic study.* Unpublished MA. thesis, Fu Jen Catholic University.

Chao, Y.-R. (1968). *A Grammar of Spoken Chinese*. Berkeley and Los Angeles: University of California Press.

Chen, K.-J. (1998). Online Word Segmentation system (http://ckipsvr.iis.sinica.edu.tw/) Retrieved April, 2009

Cheng, R. L. (1985). A comparison of Taiwanese, Taiwan Mandarin, and Peking Mandarin. *Language, 61*(2), 352-377.

Cho, T. (2005). Prosodic strengthening and featural enhancement: Evidence from acoustic and articulatory realizations of /ɑ, i/ in English. *Journal of Acoustical Society of America, 117*(6), 3867-3878.

Cho, T., & McQueen, J. M. (2005). Prosodic influences on consonant production in Dutch: Effects of prosodic boundaries, phrasal accent and lexical stress. *Journal of Phonetics, 33*, 121-157.

Chung, K. S. (2006). Hypercorrection in Taiwan Mandarin. *Journal of Asian Pacific Communication, 16*(2), 197-214.

Cole, J., Kim, H., Choi, H., & Hasegawa-Johnson, M. (2007). Prosodic effects on acoustic cues to stop voicing and place of articulation: Evidence from Radio News speech. *Journal of Phonetics, 35*, 180-209.

Cutler, A., & Foss, D. J. (1977). On the role of sentence stress in sentence processing. *Language and Speech, 29*, 1-10.

de Jong, K. J. (1995). The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation. *Journal of Acoustical Society of America, 97*(1), 491-504.

Dong, Z.-s. (1995). Taiwan hanyu fanyan yingxiangxia de ruogan "Guoyu" shengmu bianti chugao [First draft of several initial variants in Mandarin under the influence of Chinese dialects in Taiwan]. *Yuwen Xuebao, 2*, 1-28.

Duez, D. (1995). On spontaneous French speech: Aspects of the reduction and contextual assimilation of voiced stops. *Journal of Phonetics, 23*, 407-427.

Fon, J. (2004). *A preliminary construction of Taiwan Southern Min spontaneous speech corpus* (Technical report No. NSC-92-2411-H-003-050-). Taipei: National Science Council.

Forrest, K., Weismer, G., Milenkovic, P., & Dougall, R. N. (1988). Statistical analysis of word-initial voiceless obstruents: Preliminary data. *Journal of Acoustical Society of America, 84*(1), 115-123.

Hsu, H.-j. (2006). *Some aspects of phonological leveling in Taiwan Mandarin.* Unpublished Ph.D. dissertation, National Taiwan Normal University, Taipei.

Huang, C.-R., Chen, K.-J., Chen, F.-Y., Wei, W.-C., & Chan, L.-l. (1996). *A segmentation standard for Chinese information processing: Design criteria and content*: Academia Sinica.

Huang, S. (1993). *Yuyan, shehui yu zuqun yishi – Taiwan yuyan shehuixue de yanjiu [Language, society, and ethnic identity – Studies in language sociology in Taiwan]*. Taipei: Crane Publisher.

Hughes, G. W., & Halle, M. (1956). Spectral properties of fricative consonants. *Journal of Acoustical Society of America, 28*(2), 303-310.

Ing, R. O. (1984). Youguan huayu fayin de yixie wenti [Issues on the pronunciations of Mandarin]. *The World of Chinese Language, 35*, 6-16.

Ing, R. O. (1985). Guoyu fayin zai Taiwan: muqian qushi yu yiban cuowu zhi tantao [Taiwan Mandarin pronunciation: a discussion on current trend and common mistakes]. *Teaching and Research, 7*, 113-125.

Jeng, J.-Y. (2006). The acoustic spectral characteristics of retroflexed fricatives and affricates in Taiwan Mandarin. *Journal of Humanistic Studies, 40*(1), 27-48.

Jin, S. (1996). *An acoustic study of sentence stress in Mandarin Chinese.* Unpublished doctoral dissertation, The Ohio State University.

Jongman, A., Wayland, R., & Wong, S. (2000). Acoustic characteristics of English fricatives. *Journal of Acoustical Society of America, 108*(3), 1252-1263.

Kubler, C. C. (1985). The influence of Southern Min on the Mandarin of Taiwan. *Anthropological Linguistics, 27*(2), 156-176.

Labov, W. (1972). Hypercorrection by the lower middle class as a factor in linguistic change. In W. Labov (Ed.), *Sociolinguistic patterns* (pp. 122-142). Philadelphia: University of Pennsylvania Press.

Labov, W. (1990). The intersection of sex and social class in the course of linguistic change. *Language Variation and Change, 2*, 205-254.

Ladefoged, P., & Wu, Z. (1984). Places of articulation: An investigation of Pekingese fricatives and affricates. *Journal of Phonetics, 12*, 267-278.

Li, C. N., & Thompson, S. A. (2005). *Mandarin Chinese: A Functional Reference Grammar*. Taipei: The Crane Publishing.

Li, F. (2008). *The phonetic development of voiceless sibilant fricatives in English, Japanese and Mandarin Chinese.* Unpublished doctoral dissertation, the Ohio State University.

Li, M.-C. (1995). *A sociolinguistic variation study of Chinese retroflex initials /tʂ/, /tʂʰ/ and /ʂ/ in Taiwan.* Unpublished MA. thesis, Providence University, Taichung.

Liao, C.-C. (2000). Changing dominant language use and ethnic equality in Taiwan since 1987. *International Journal of Sociogy of Language, 143*, 165-182.

Lieberman, P. (1960). Some acoustic correlates of word stress in American English. *Journal of Acoustical Society of America, 32*(4), 451-454.

Lin, C. C. (1983). A sociolinguistic study of the use of the retroflex sounds in Mandarin in college studennts in Taiwan. *Bulletin, College of Arts and Letters, National Central University*, 1-15.

Lin, Y.-H. (1988). Consonant variation in Taiwan Mandarin. In K. Ferrara, B. Brown, K. walters & J. Baugh (Eds.), *Linguistic change and contact (Proceedings of the 16th Annual New Ways of Analyzing Variation in Language Conference), Texas Linguistics Forum 30* (pp. 200-208). Austin: University of Texas, Department of Linguistics.

Maddieson, I., & Precoda, K. (1990). UPSID-PC. The UCLA Phonological Segment Inventory Database (http://www.linguistics.ucla.edu/faciliti/sales/software.htm).

Mann, V. A., & Repp, B. H. (1980). Influence of vocalic context on perception of the [s]-[ʃ] distinction. *Perception & Psychophysics, 28*(3), 213-228.

Manrique, A. M. B. d., & Massone, M. I. (1981). Acoustic analysis of perception of Spanish fricative consonants. *Journal of Acoustical Society of America, 69*(4), 1145-1153.

NTNU Mandarin phonetics committee (2003). *Mandarin Phonetics*. Taipei: Cheng-Chung Book Co.

Peng, S., Chan, M. K. M., Tseng, C., Huang, T., Lee, O. J., & Beckman, M. E. (2007). Toward a Pan-Mandarin system for prosodic transcription. In S.-A. Jun (Ed.), *Prosodic Typology: The Phonology of Intonation and Phrasing*. Oxford: Oxford University Press.

Rau, D.-h., & Li, M.-C. (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.

Raymond, W. D., Dautricourt, R., & Hume, E. (2006). Word-initial /t, d/ deletion in spontaneous speech: Modeling the effects of extra-linguistic, lexical, and phonological factors. *Language Variation and Change, 18*, 55-97.

Shen, X.-N. S. (1993). Relative duration as a perceptual cue to stress in Mandarin. *Language and Speech, 36*(4), 415-433.

Shi, F. (1994). Suprasegmentals of Chinese from various aspects. In F. Shi & R. Liao (Eds.), *Yu3yin1 cong2gao3 [Papers in Phonetics]* (pp. 215-225). Beijing: Beijing Language and Culture University Press.

Svantesson, J.-O. (1986). Acoustic analysis of Chinese fricatives and affricates. *Journal of Chinese Linguistics, 14*(1), 53-70.

Tse, J. K.-P. (1988). A spectrographic analysis of the coronal affricates and fricatives in Mandarin Chinese. *Studies in English Literature and Linguistics, 14*, 149-199.

Tse, J. K.-P. (1989). Degrees of retroflexion and the 'empty vowel' in Mandarin Chinese. *Studies in English Literature and Linguistics, 15*, 95-100.

Tse, J. K.-P. (1998). Taiwan diqu nianqingren ㄓㄔㄕ yu ㄗㄘㄙ zhende bufen 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.

van Bergem, D. R. (1993). Acoustic vowel reduction as a function of sentence accent, word stress, and word class. *Speech Communication, 12*, 1-23.

van Son, R. J. J. H., & Pols, L. C. W. (1999). An acoustic description of consonant reduction. *Speech Communication, 28*, 125-140.

Yang, W.-S. (2008). *Quanguo kejia renkou jichu ziliao diaocha yanjiu [A survey on National Hakka population]*. Taipei: Council of Hakka Affairs, Executive Yuan, Taiwan.

APPENDIX I

Region/Gender	Speaker	Age	Mandarin proficiency	Min proficiency
Taipei/Male	CZX	24	7	6
	HSK	33	7	6
	JXW	35	5	7
	YYS	28	6	6
Taipei/Female	LRL	22	7	4
	XHR	31	7	7
	XHY	32	7	6
Kaohsiung/Male	GWH	21	7	4
	KCZX	23	6	7
	LJH	24	7	5
	LZW	27	6	7
Kaohsiung/Female	CJH	22	7	5
	CN	20	7	4
	GYX	22	6	5
	WJL	35	7	7

APPENDIX II

Region/Gender	Speaker	Recording duration	Syllable number	Discussed topics
Taipei/Male	CZX	30'10"	8,641	future plan, life experience, family, historical stories
	HSK	33'09"	8,285	food, traveling experience
	JXW	30'27"	5,379	traveling experience, study abroad, TV program, singer, pet
	YYS	31'53"	7,676	school life, career, marriage, dating experience
Taipei/Female	LRL	30'20"	6,671	Min usage, hobby, school life, part-time job, traveling experience
	XHR	30'53"	8,111	Min usage, childhood, history, marriage
	XHY	32'13"	8,243	education issue, food, TV program, career life
Kaohsiung/Male	GWH	30'14"	6,486	school life, political activity, food, hobby
	KCZX	30'37"	7,082	playing band, school life, future plan
	LJH	30'37"	7,550	education issue, culture, school life, singing in a choir
	LZW	30'32"	7,044	car repairing, part-time job, internet, TV program, education issue
Kaohsiung/Female	CJH	30'43"	4,823	food, sports, part-time job, music, Taipei/Kaohsiung comparison
	CN	30'37"	7,151	movie, TV program, school life, dorm life
	GYX	35'08"	8,635	family, school life, health, future plan
	WJL	31'19"	8,054	study abroad, teaching experience, TV program, education issue, marriage