TONE MERGER IN GUANGZHOU CANTONESE

OU JINGHUA

M.PHIL

THE HONG KONG POLYTECHNIC UNIVERSITY

2012

The Hong Kong Polytechnic University

The Department of Chinese and Bilingual Studies

Tone Merger in Guangzhou Cantonese

Ou Jinghua

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Philosophy

June 2012

CERTIFICATE OF ORIGINALITY

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

Ou, Jinghua

ABSTRACT

Cantonese stands out from other Chinese dialects by having a complex tonal system. However, this complex tonal system has been reported to be in the process of merging. Although there has been a number of studies of the tone merger phenomenon in Hong Kong Cantonese (HKC), few studies have been conducted for Guangzhou Cantonese (GZC). Thus, this study conducted a comprehensive investigation into the variations and mergers of tones in contemporary GZC. Three experimental tasks, discrimination, identification and production, were administered to 75 subjects in three different age groups: young adults (20-25 yrs old); middle-aged (35-45 yrs old) and senior (over 50 yrs old). Forty-eight real syllables, generated by eight CV(C) roots, were used in all three tasks. An acoustic study was conducted to follow up the results of the production task.

In the discrimination task, AX paradigm was used to examine whether native GZC speakers can still contrast all the six tones in perception. The overall results showed that T3/T6 and T4/T6 were the two most readily confused pairs. However, the discrimination ability differed among the three age groups.

The identification task was administered to investigate the directions of tone mergers attested in the discrimination task. The stimuli were presented aurally to the participants who were asked to select a character from a list of six to represent the perceived syllable properly. Overall, the identification accuracy of T6 was significantly lower than those of the other tones. T6 was more often identified as T3 and also as T4. The error patterns differed among the three age groups.

In the production task, the stimuli were embedded into two sentence carriers and read aloud by the participants in order to investigate the production of the six tones. Overall results demonstrated that the production accuracies of T6 and T5 were significantly lower than those of the other tones and these two tones were mainly produced as T3. In addition, the accuracy of T3 was significantly low in the middle-aged group, whereas that of T2 and T3 were significantly low in the young adult group. As for the direction of errors, T3 was found to be produced as T6 and T2 as T5.

In all, two types of merger were identified: 1) T3/T6, a full-merger; and 2) T4/T6, a near-merger. The confusion between T3 and T5 was proposed as a case of alternative character readings.

The acoustic study was conducted on the T3/T6 full-merger to examine the phonetic realizations of the new level tone and to better understand the mechanism of this merger. The results showed that the two level tones have been merged into one category and realized as a mid-level tone. Also, our results suggested that different mechanisms of tone merger were adopted by different age groups.

The present study provided supporting evidence that tone merger is on-going in contemporary GZC. Tone mergers in the two varieties of Cantonese (HKC vs. GZC) were compared and implications for Cantonese as a whole were discussed.

(492 words)

ACKNOWLEDGEMENTS

Throughout the whole process of this study, there are many people who have given me help and support, without which I could not have finished this thesis. Therefore, I would like to take this opportunity to thank each of them.

First of all I would like to express my deepest gratitude to my supervisor, Dr. Roxana Fung, for her patience, encouragement and caring during the whole time. I really want to thank her for her invaluable supervision and support. She has always tried her best to make herself available for me whenever I have needed guidance in study or life, especially when I was in the middle of figuring out my future direction. I would have been lost without her. Also, I really want to thank her for giving me the opportunity to study in the LSA Summer School, which inspired me a lot and opened up a new world to me. I really enjoyed the time studying with and also learning from her.

I am grateful to my examiners, Dr. Lawrence NG, Professor ZHU Xiaonong and Dr. CHEUNG Kwan Hin, for their valuable comments and inspiring advice.

I would like to thank Mr. PENG Xiaofeng from Jinan University, who helped me so much in recruiting subjects in Guangzhou, and also provided me an office in the university to conduct the experiment. Also, I would also like to thank all my friends in Guangzhou who kindly helped "advertise" the experiment for me, especially GU Simin, WU Jingyao, OU Qiaoshan and ZHAO Zixun.

I am indebted to my dear colleagues in the Department of Chinese and Bilingual Studies for providing a stimulating and fun environment in which I was able to learn and grow, especially to JIN Jing, Echo KE and ZHANG Xue, for sharing my ups and downs during my life as a research student. I enjoyed the time “hanging out” and hunting for good food with them all. A special thanks goes to ZHANG Ling for lending her voice to me for the recording of the experiment.

I would also like to give my special thanks to Carmen Kung, for all the fun time during the ERP experiment period and all the sharing of her own experience with me.

Last, and most importantly, I wish to thank my family and friends for their support and concern, and especially my adorable niece who always sings to me, which brings me a lot of joy.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction

1.1 Introduction

Cantonese is spoken by the majority of people in Hong Kong, Guangdong and some cities of Guangxi in China. Cantonese spoken in Hong Kong and Guangzhou are generally treated as standard variants of this dialect. Standard Cantonese stands out from other Chinese dialects by having a complex tonal system. Generally speaking, there are six contrastive tones in Cantonese. The phonetic realizations of each tonal category in Guangzhou Cantonese (hereafter as GZC) are somewhat different from those in Hong Kong Cantonese (hereafter as HKC). In recent years, it has been reported that the phonemic tones are in the process of merging in Cantonese. However, aside from the work on HKC, little is known about the tone merger phenomenon in Cantonese communities outside Hong Kong. In addition, no systematic, in-depth study of the variations and mergers of tones in a broader Cantonese area of Mainland China has been published thus far. This study aimed first to provide a comprehensive investigation into the tone mergers in Cantonese spoken in Guangzhou, then to compare the merger phenomenon identified with that in HKC, and in the process, to achieve a better understanding of the evolution of tone in modern Cantonese.

There are seven sections in this chapter. Section 1.2 introduces the Guangzhou speech community. Section 1.3 offers an overview of the Cantonese tone system. Section 1.4 presents previous works on tonal confusion in Cantonese. Section 1.5 provides an overview of tone development in the history of the Chinese language.

Section 1.6 introduces the mechanisms of mergers, followed by an outline of the research aims of the present study in Section 1.7.

1.2 The Guangzhou speech community

Even though the two variants are mutually intelligible, the Cantonese spoken in the community of Guangzhou has some differences from the Cantonese spoken in Hong Kong mainly in pronunciation and lexis. This may be due to Guangzhou and Hong Kong having been separated for nearly 50 years, since the formation of the People's Republic of China. These two communities have experienced different social, linguistic and political influences. In Hong Kong, both the spoken and written forms of Cantonese have flourished. However, Cantonese is not the medium of education in Guangzhou, although it is used extensively as the local prestige dialect in the city. Since the 1980s, Putonghua (standard Mandarin) has been promoted vigorously in Guangzhou, as stipulated by the language reform and planning scheme. Putonghua has become the only spoken language that is government sanctioned for use in official settings, such as education, government administration and the mass media. Written Cantonese is suppressed and spoken Cantonese is discouraged officially. The Putonghua proficiency of the GZC speakers might be correlated positively with the education level. In this sense, the younger generation is more likely to have a better knowledge of and higher proficiency in Putonghua, as compared with their older counterparts. Indeed, the city has seen a great increase in the number of Cantonese-Mandarin biliterates during the past few decades. Apart from the strong influence of Putonghua, GZC is also affected by its more diverse dialectal neighborhood compared with HKC. There are ten administrative districts in Guangzhou: *Yuexiu* 越秀, *Liwan* 荔灣, *Dongshan* 東山, *Haizhu* 海珠, *Baiyun*

白雲, *Tianhe* 天河, *Huangpu* 黃埔, *Fangcun* 芳村, *Panyu* 番禺, *Huadu* 花都, and two counties - *Zengcheng* 增城 and *Conghua* 從化. Of the ten districts, *Yuexiu*, *Liwan*, *Dongshan* and *Haizhu* form the "old district", while *Baiyun*, *Tianhe, Huangpu* and *Fangcun* are the "new district" (Zhan, 2002, p. 7). The variety in the old district, called "*Xiguan* speech 西關話" was originally regarded as the representative of GZC, probably due to the early residence of Cantonese speaking ancestors in *Xiguan*. In the new districts and the two counties, the early inhabitants were primarily immigrants from the Hakka or Chaozhou dialectal regions (Li, 1994, p. 28-29). Therefore, there are quite considerable differences between the phonologies of the Cantonese spoken in the old and new districts. For instance, there are only eight contrastive tones in *Conghua* Cantonese, with the two rising tones of standard GZC merged into one (Zhan, 2002, p. 131).

1.3 Cantonese Tone system

Tones refer to syllabic-level F0 (fundamental frequency) patterns that are used to differentiate meaning in words composed of the same segmental make-ups (Bauer and Benedict, 1997). F0 changes of lexical tones are carried by the voiced segment of the syllable (Wang, 1967), and are an indispensable component in the phonological system of tone languages. Zhu (2012) pointed out that the variation of F0 alone may not be sufficient to differentiate among tones in a more complicated system, where more than one type of phonations is adopted. He proposed further that phonation also carried tonological information and should be incorporated into a tonal model. In Cantonese, voice quality was proposed as a possible cue for identification of T4 (Vance, 1976, 1977), since the lowest tone in a tonal inventory without contrastive phonation tends to be produced with creaky voice. This hypothesis was also supported by a study (Lam and Yu, 2010), where creaky voice was found to have a facilitating effect on the identification of T4, while those without creaks were more likely to be identified as T6. However, in the presence of F0, such information on voice quality may be taken as a secondary cue for distinguishing between tones. In sum, voice quality is a personal preference but not a distinctive feature in the production of Cantonese tone. One of the notations used widely in describing tone level in tone languages is the five-level system proposed by Chao (1947). Number '5' denotes the highest tone level while number '1' represents the lowest level. In this system, the tone contour is denoted by the movement between the tone level at the beginning and at the endpoint of the tone. Table 1.1 displays the Cantonese tone system. In general, there are six contrastive tones (T1 to T6) with three variants (T7 to T9) which are realized only on checked syllables of shorter duration with /-p/, /-t/ and /-k/ as codas. They are allotones of the three level tones (T1, T3 and T6).

T1 has two variants: high level and high falling. The most distinguishable phonetic difference between HKC accent and GZC accent also lies in the realization of T1. In old GZC, the alternation of pitch contour has semantic and morphological functions. For instance, some nouns can be derived from the related verbs of high falling tone by changing the contour to high level (Li, 1994, p. 76-77), for example /pou/ 煲, when pronounced as the high falling tone means "to cook", while it means "cooking utensil" with the high level variant. The distinction, however, may now be maintained only by the older speakers. As suggested by Shi (2004), most of the young speakers in

Guangzhou do not distinguish between high level and high falling tones nowadays and use them as free variations. In HKC, the high tone is realized mainly as high level, whereas the high falling variants are barely used in the population (So, 1996).

T4 may have two phonetic realizations, low falling or extra-low level. In fact, the low falling is used to depict the downdrifting effect, which is typically found in the lowest level tone of a language cross linguistically (see Yip, 2002, p. 23; Zhu, 2010, p. 277). Therefore, T4 has been taken as a level tone instead of a falling one in the present study.

Table 1.1

The Cantonese tone system

Tone number	Description	Chao's tone letter	Example
T1	High-level/High Falling	55/53	詩 /si/ 'poem'
T2	High-rising	25	史 /si/ 'history'
T3	Mid-level	33	試 /si/ 'try'
T4	Low-falling/ Extra-low level	21/11	時 /si/ 'time'
T5	Low-rising	13	市 /si/ 'city'
T6	Low-level	22	是 /si/ 'yes'
T7=T1	High-stopped	5	色 /sik/ 'color'
T8=T3	Mid-stopped	3	錫 /sik/ 'tin'
T9=T6	Low-stopped	2	食 /sik/ 'eat'

As shown above, the tonal system of Cantonese is quite complex. There are four level tones (T1, T3, T4 and T6) in the system, which differ only in relative F0 height, as illustrated by Figure 1.1. The pitch values of each tone have been normalized using the

T-value formula (see Section 3.2.3). The two rising tones (T2 and T5) share the same starting point and differ only in the end point. The acoustic similarity between the rising tones and among the level tones may impose difficulty on speakers for distinguishing and render the tones particularly susceptible to merging.

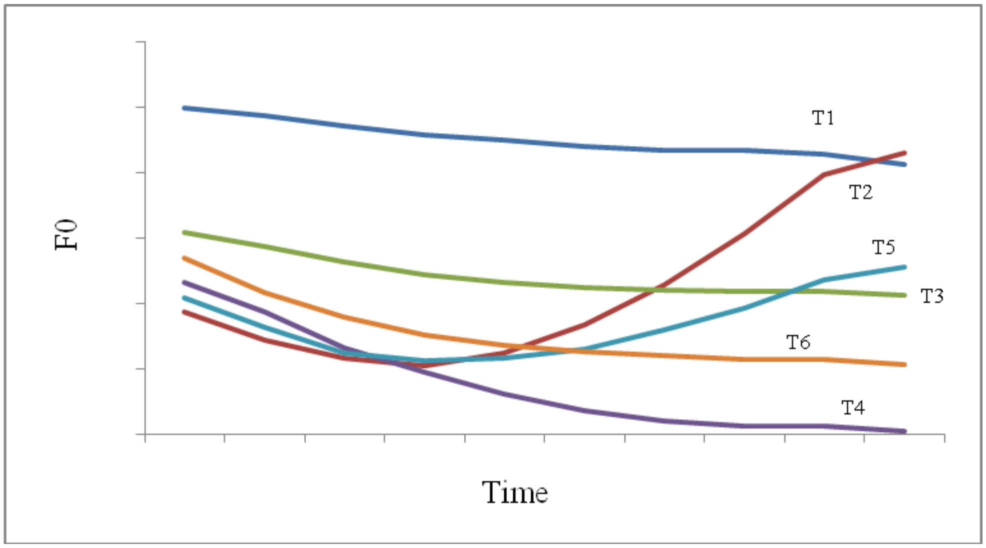


Figure 1.1. The F0 contours of the six tones in Cantonese

1.4 Tonal confusions reported in Cantonese

1.4.1 Tonal confusions reported in HKC

The phenomenon of tone merger in HKC has attracted considerable attention recently. There has been quite a number of studies reporting on tonal confusions by native speakers either in perception or production, although they were not set out primarily from the perspective of tone merger. In summary, four tone pairs were found in the literature to be readily confused by native Cantonese speakers: T3 (mid level) vs. T5 (low rising); T2 (high rising) vs. T5 (low rising); T3 (mid level) vs. T6 (low level); and T4 (extra low level) vs. T6 (low level).

Vance (1976) instructed participants to identify some target syllables embedded in three different sentence carriers in order to study the relationship between tone and intonation in Cantonese. He noted that his participants confused the low rising tone (T5) with the mid level tone (T3). Killingley (1985) followed up this issue by conducting a survey in Malayan Cantonese. Based on the data collected, Killingley argued that T3 and T5 were two allophonic variants of one toneme. In addition, she claimed that her proposal could also be extended to Hong Kong and Mainland Cantonese. However, it should be noted that Malaysia is a multilingual country, and the Cantonese spoken there is under a great influence from other languages or Chinese dialects such as Min. Killingley's proposal was refuted by native Cantonese linguists such as Cheung (1986). Instead, Cheung (1986) pointed out that the merger of these two tones in Malayan Cantonese might serve as an indication of the merger of T3 and T5 in mainstream Cantonese (i.e., HKC). Additionally, he proposed that the alternative pronunciations between T3 and T5 of some lexical items, such as '舍 /se/' and '蒜 /syn/', might be the sign of the very beginning of the merger of T3 and T5.

This issue has been picked up by Wong (2008) again. He provided a detailed examination of the morphemes that exhibit T3/T5 by comparing his own readings of the morphemes with the readings recorded in Wong Sek Ling's Cantonese Syllabary complied in 1941. The findings showed that the scenario was much more complicated in HKC. Some morphemes have changed from T3 to T5 or vice versa, whereas some morphemes have developed two alternative readings, T3 and T5, to differentiate lexical meanings. Instead of a phonological issue, as argued by Killingley (1985), it appears that the T3/T5 tonal confusion exhibits as a case of alternative character readings. Cheng

(2002 and 2003) followed up this tone change by asking 20 native speakers to read aloud a total of 314 minimal or near-minimal pairs of T3 or T5 in disyllabic forms. The results confirmed the bi-directional tonal confusion between T3 and T5, and showed that the major direction was T3 to T5. Similar findings were also attested in a production study carried out by Ng (2007). Taking the above studies together, the tonal confusion does not seem to be phonologically conditioned, but rather lexically conditioned. Different lexical items seem to favor different directions of the tonal confusion. For instance, the change from T5 to T3 mostly involved lexical items such as ‘臼 /k^{h}ɐu/’, ‘舅 /k^{h}ɐu/’, ‘似 /ts^{h}i/’ and ‘鱔 /sin/’; the change from T3 to T5 usually involved lexical items such as ‘愧 /k^{w}ɐi/’, ‘販 /fan/’ ‘潛 /ts^{h}im/’ and ‘蘊 /wɐn/’; and the bidirectional change between these two tones typically involved lexical items such as ‘試 /si/’, ‘舍 /se/’, ‘議 /ji/’ and ‘署 /tʃ^{h}y/’.

Unlike the aforementioned confusion between T3 and T5, the tonal confusions reported by the following studies were concerned with the phonemic status of the tonal categories. Fok (1974) administered various identification tasks to 511 listeners in order to explore the perceptual correlates of Cantonese tones. Natural speech and synthetic stimuli were used in this study. She found that T3 was readily confused with T6 and vice versa. Similar bidirectional confusions between T2 and T5 were also commonly found among her participants. Confusion between T4 and T6 was also attested. Among the six tones, T1 was the most resistant to confusion in this experiment.

Kei, Smyth, So, Lau and Capell (2002) studied the Cantonese tone production by 56 speakers acoustically in order to provide a normative profile of the standard tone production for speech therapists. The participants were instructed to read aloud 108 monosyllabic words (18 root syllables x 6 tones) and their productions were then normalized to z-scores. The authors found that T2/T5 confusion was the main source of production errors and different patterns of merging the two tones were observed. Some participants produced some T5 tokens as T2, and at the same time, some produced T2 tokens as T5. Intermediate forms were also attested. Some participants realized T2 and T5 tokens as a single tone with endpoint situated midway between T2 and T5. The findings from Fok (1974) and Kei et al. (2002) indicated that some HKC speakers might no longer distinguish some of the six tones, which laid the foundation of the study on tone merger in HKC.

Bauer, Cheung and Cheung (2003) followed on the findings of Kei et al. (2002) to investigate the variation and merger of the rising tones in HKC. Eight participants were asked to read aloud 122 monosyllabic stimuli and their productions were then analyzed acoustically. They reaffirmed the findings of the Kei et al. study in that one speaker merged T2 into T5, while another merged T5 into T2.

Yiu (2009) examined both the production and perceptual aspect of the suspected T2/T5 merger using 15 participants. The perception test contained two tasks: an identification task using 120 monosyllabic stimuli and a discrimination task using 20 bisyllabic stimuli. In the production task, the participants were asked to read aloud both the monosyllabic and bisyllabic stimuli. The productions were then examined acoustically. Mutual confusions of the two tones were attested in the production test; however, little evidence was found to support the confusion of T2/T5 in perception.

Mok and Wong (2010a and 2010b) administered both perception and production tasks to 27 participants. The participants were divided into two groups, the control and the potential mergers, according to the results of a preliminary screening. AX paradigm was employed for the discrimination task. The findings showed that the accuracy of the T2/T5 pair was the lowest in the control group, but remained relatively high in the potential merger group. As for the production task, 84 monosyllabic stimuli were embedded in sentence carriers for the participants to read aloud. Acoustic analysis was conducted to follow up the production of three potential merging tone pairs (T2/T5, T3/T6 and T4/T6). The F0 values at eight equally spaced points of each target syllable were elicited for t-test analysis to compare two tones in each of the three pairs. The results showed no significant differences for most of the measuring points between the two tones in each pair, suggesting that there was no tone merging for those potential pairs in production.

More conclusive evidence supporting the tone merger in the community of HKC was provided in a recent study conducted by Fung and Wong (2010a and 2010b). They administered discrimination and production tasks to 120 subjects of three age groups: 20 - 25; 35 - 45; and 50 - 58. It was found that the perception of the T2/T5 and T4/T6 contrasts was significantly lower than other tone pairs, but that of the T3/T5 contrast was well preserved. In the production task, the scores of T2, T3, T5, and T6 were significantly lower than those of T1 and T4. As for the direction of errors, different patterns were revealed in different age groups. In the oldest group, the T5 syllables were produced as T2, and the T6 syllables as T5. In the younger groups, bidirectional confusion was found between the T2/T5 syllables, and the T3/T6 syllables. The authors proposed further that T2/T5 was a full merger, T3/T6 a quasi-full-merger, T4/T6 a near-merger and T3/T5 a suspended merger in contemporary HKC.

1.4.2 Tonal confusions reported in GZC

While the tone merger in HKC was well documented, few studies have been conducted for GZC. To the best of my knowledge, the only work published thus far was the study by Li (2008). Using minimal pairs of T3 or T6 in disyllabic forms, she examined the perception of the T3/T6 pair among 22 young GZC speakers from the old districts *Liwan*, *Yuexiu*, *Dongshan* and *Haizhu*. The participants were asked to identify the disyllabic stimuli upon hearing the stimuli, and they could choose more than one answer from five given choices as long as they thought appropriate. The results displayed that quite a number of participants chose both the T3 and T6 simultaneously when a target tone of either T3 or T6 showed up. This finding suggested that the participants could not distinguish these two tones in perception. According to the scores in the perception task, two participants were subsequently chosen and asked to read aloud a list of monosyllabic stimuli of T3 or T6. However, the author did not provide any information about which monosyllabic stimuli were used. The tone productions were then examined by the F0 tractor in Praat. Only sketchy F0 traces were displayed without any further information. Nonetheless, the results of the F0 traces generally suggested that those who made more errors in the perception task tended to have more overlaps in F0 traces in their productions.

1.5 Tone changes in the history of Chinese language

The tone mergers in contemporary Cantonese could be the synchronic products of a diachronic tone change in the history of Chinese language. An overview of the historical development of tones in the Chinese language should be in order.

1.5.1 The evolution of tones in Chinese dialects

To review the evolution of tones in Chinese, it is necessary to examine the situation in Middle Chinese (abbreviated MC). MC was largely represented by the *Qieyun* and *Guanyun* rhyme tables, and it spanned approximately from AD 600 to AD 1200 (Norman, 1988, p. 24-25). In MC, four tonal categories, referred to by their traditional terms as *Ping* 平 (Level tone), *Shang* 上 (Rising tone), *Qu* 去 (Going tone) and *Ru* 入 (Entering tone) were recorded in the rhyming dictionaries. Norman (1988, p. 53) pointed out that the most important factor in the development of the tonal categories in MC to those modern dialects was the development of tonal register. He further elaborated that at some point in the history of Chinese, the primary four-way tonal contrast of MC was affected by the initial consonants: syllables with voiced initials began to be pronounced at a lower pitch than those which had voiceless initials. Consequently, each of the original tonal categories was split into two registers, *Ying* 陰 and *Yang* 陽. As long as the voicing of the initials persisted, such splits were merely allophonic; but when voicing was lost as a distinctive feature of a dialect's phonological system, these splits became phonemic. At this stage a new eight-contrast tonal system was created, namely *Yin Ping*, *Yin Shang*, *Yin Qu*, *Yin Ru*, *Yang Ping*, *Yang Shang*, *Yang Qu and Yang Ru*. This eight-tone system of MC is well preserved in Cantonese and the

Yin Ru tone has been further split into two sub-categories conditioned by vowel length, giving rise to nine categories (see Table 1.2). Cantonese is a very peculiar dialect with respect to tones since only a few have preserved the eight-way contrast faithfully; and more rarely, one of the categories has a further split (Norman 1988, p. 218).

Table 1.2

The Cantonese tone system from the historical perspective

<table>
<tr><td rowspan="2">Yin
register</td><td rowspan="2">Yin Ping
(T1)</td><td rowspan="2">Yin Shang
(T2)</td><td rowspan="2">Yin Qu
(T3)</td><td>Yin Ru
(upper)
(T7=T1)</td></tr>
<tr><td>Yin Ru
(lower)
(T8=T3)</td></tr>
<tr><td>Yang
register</td><td>Yang Ping
(T4)</td><td>Yang Shang
(T5)</td><td>Yang Qu
(T6)</td><td>Yang Ru
(T9=T6)</td></tr>
</table>

In most cases, modern dialects have merged one or more of the tonal categories after the tonal splits. Many Mandarin dialects show a tonal pattern similar to the four-tone system in Peking Mandarin (see Table 1.3). The *Ping* category has two registers; the *Yin Ping* category includes syllables that had voiceless initials in MC whereas the *Yang Ping* category includes syllables that had voiced initials in MC. With respect to the *Shang* category, no Mandarin dialects maintain the register distinction. Regarding the *Qu* tone, a register distinction in this category is relatively rare, found apparently in a few dialects spoken in the border region between the northern and central dialects. The MC *Ru* tone has evolved in many different ways: in some conservative dialects it is retained as a distinct category while, more typically, it has merged with other tonal categories.

Table 1.3

The tonal system of Peking Mandarin from the historical perspective

<table>
<tr><td>Yin Ping
(T1: Tone 55)</td><td rowspan="2">Shang
(T2: Tone 35)</td><td rowspan="2">Qu
(T3: Tone 214)</td></tr>
<tr><td>Yang Ping
(T4: Tone 51)</td></tr>
</table>

To investigate the tone merger in the northern dialects, Lien (1986) proposed two modes to account for the tone behavior during the historical development: vertical tone merger and horizontal tone merger. Horizontal merger refers to the coalescence among the four tonal categories *Ping*, *Shang*, *Qu* and *Ru*. Vertical merger refers to the coalescence between the *Yin* and *Yang* register within the same tonal category. Lien (1986) observed that *Qu* was a "merging category" (a target of tone merger), while *Ping*, *Shang* and *Ru* were "merged categories" (a source of tone merger) in the northern varieties. He pointed out that as a merging category, *Qu* tone never merged to other tone categories in a horizontal mode.

In both northern and southern dialects, the vertical merger of the two registers within the *Shang* category is more prevalent than that in the *Qu* category. Hence, dialects that have two *Qu* tones (*Yin Qu* and *Yang Qu*) but just one *Shang* tone are highly common. For instance, as recorded in Yuan (2001), *Suzhou*, *Changsha* and *Conghua* dialects exhibit the above tonal system. However, dialects that have two *Shang* tones (*Yin Shang* and *Yang Shang*) but just one *Qu* tone are less commonly found. Hence, it can be hypothesized that the vertical merger in the *Shang* tone should take place before that in the *Qu* tone.

1.5.2 The historical tone change "*Zhuo Shang bian Qu*"

The vertical merger of the *Shang* tone in Chinese dialects may be motivated by a diachronic tone change, *Zhuo Shang bian Qu* 濁上變去. This sound change took place before the devoicing of consonants initials in MC. It refers to the migrating of *Shang* syllables with voiced obstruent initials to the *Qu* tone. Most northern dialects have been involved in this sound change. After the devoicing of the initial consonants, the *Shang* tone underwent the register split, giving rise to the *Yin Shang* and *Yang Shang* categories. Due to the previous migration of tone, the syllable inventory of the *Yang Shang* tone of many northern dialects is relatively small as compared with other categories, rendering this tone highly susceptible to merging. In Cantonese, the extent of *Zhuo Shang bian Qu* was smaller than that in the northern dialects. As documented in Luo (1997), two thirds of the *Yang Shang* syllables with MC voiced obstruent initials have undergone merging with the *Yang Qu* (T6) tone in Cantonese whereas the other one third of the words have remained in *Yang Shang* (T5). Chen and Newman (1984) suggested that the migration of *Yang Shang* syllables to the *Qu* category occurring in Cantonese might be due to either borrowing from the northern dialects, especially Mandarin; or development within Cantonese under the influence of the northern dialects. Hashimoto (1972, p. 44) characterizes those lexical items which have undergone this tone change, like '部 /pou^{22}/' and '杜 /tou^{22}/' as literary while those which have not, like '抱 /p^{h}ou^{13}/' and '肚 /t^{h}ou^{13}/' as colloquial. This, in a way, alludes to the traces of the northern influence in those literary items.

1.6 The mechanisms of mergers

After identifying the mergers, the subsequent question to explore is the route by which two tonemes become one: how individual words and tonemes targets move in relation to each other. Based on a vast number of studies on sound change, Labov (1994, p. 321) proposes three major mechanisms of sound mergers: 1) Merger-by-approximation; 2) Merger-by-transfer; and 3) Merger-by-expansion. Merger-by-approximation is a gradual and mutual process of approximation by phonetic realizations in which two phonemic categories lose their distinction. In this case, the merged phoneme might have an intermediate value between those of the original two. Merger-by-transfer refers to gradual unidirectional transfer of words from one phoneme to another. Merger-by-expansion is a process in which words of two phonemic categories transfer bidirectionally from one to the other. This mechanism results in a larger phonetic range for both, equivalent to the combined range of the two merged phonemes. The three mechanisms may operate at different rates, as suggested by Labov (1994, p. 323). Merger-by-transfer is the slowest; merger-by-approximation may take three or four generations; and merger-by-expansion appears to be completed in a single generation.

1.6.1 Dissociation of production and perception

Sound change involves the interplay of the language users' perceptions and productions. Most current studies of sound change have measured these two aspects of speech. Indeed, the role that speech production and perception play in sound change remains controversial. On one hand, it has been suggested that sound change is more speaker-based, which originated through the processes of articulatory reduction, simplification or variability (e.g., Halle, 1962). On the other hand, sound change has been described as more listener-based, as proposed by Ohala (1981), who described the "listener as a source of sound change". In fact, the relation of production and perception is quite complicated.

The production and perception of a language user can be perfectly matched. A good user may have good production and good perception, whereas a poor user may have poor production and poor perception. A production/perception mismatch is also possible in the case of a language user having good perception but poor production, since production involves the motor programming of the articulatory apparatus. How about the reverse? It has been reported that poor perception with good production is also attested during sound changes; that is, speakers maintain that two classes of sounds are 'the same' (non-distinctive) in perception, yet consistently differentiate them (distinctive) in production. This phenomenon was termed 'near-merger' by Labov, Yaeger, and Steiner (1972, Chapter 6). Since its introduction, "near-merger" has puzzled linguists because it challenges the prevalent models of phonological processing, and brings up suspicion in speakers' judgment about whether it can genuinely reflect their ability to discriminate speech sounds. According to the classic motor theory of speech perception (e.g., Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967), it is generally assumed that production and perception of speech share one common lexicon that contains phonological forms of lexical items in a language. On this account, good production must be derived from relevant phonological representations, which also support good perception. Therefore, good production but poor perception (i.e., near-merger), deriving from the same phonological forms is obviously counterintuitive.

Although this phenomenon is highly controversial, near-merger has frequently been attested as a transitional state in sound changes among many different languages in the literature (see Labov, 1994, p. 357-370; and references cited therein).

1.6.2 The apparent time hypothesis

A central theoretical hypothesis at the root of this work was the apparent time hypothesis (Labov, 1994, p. 43-72), which states that sound change proceeds through the various age levels can be projected into real time. To trace a linguistic variable through apparent time, it is advised to include participants that represent an age spectrum: from the older generation to the younger one. After setting up the appropriate age groups, the relationship between the ages of individuals and the value of the linguistic variable can then be examined. The distributions of linguistic variables from the oldest to the youngest can thus be projected as a progressive change taking place in that particular community. The present study adopted this apparent time approach in order to better capture the process of tone change in GZC.

1. 7 Research Aims

The study aims to provide a comprehensive investigation of the variations and mergers of tones in GZC through administering discrimination, identification and production tasks to 75 native speakers in order to answer the following research questions:

1. How many tonal categories exhibit variations or mergers in GZC?
2. What are the directions of the mergers?
3. What mechanisms are involved in the mergers?

4. Is age a factor in the tone mergers?
5. What are the similarities and differences between the mergers taking place in GZC and HKC, if any?
6. What implications can be drawn about the historical development of tonal system of Chinese?

Chapter 2

Experimental Study

2.1 Methods

2.1.1 Participants

Seventy-five native Guangzhou Cantonese speakers were recruited for this study. The participants were selected based on the following criteria: 1) They were born and raised in one of the four ‘old districts’ in GZ, namely *Yuexiu*, *Liwan*, *Dongshan* or *Haizhu*; 2) They had not left their places of birth for more than 12 consecutive months; 3) They had adopted Cantonese as the major communication means at home; 4) They had no known speech and hearing problems; 5) They had reading literacy in Cantonese of primary school level. Detailed information regarding each participant’s personal and family linguistic backgrounds were collected through a questionnaire (see Appendix I). The demographic information of all the participants is shown in Appendix II.

In order to examine the age factor in the tone mergers, the participants were recruited to form three age groups: the young adults (20-25 years); the middle-aged (35-45 years); and the seniors (above 50 years). The young adult group included 30 participants (Mean age = 22.6 years, S.D. = 1.57), matched for gender. The middle-aged group included 30 participants (Mean age = 38.3, S.D. = 4.02), also matched for gender. The senior group included 15 participants (Mean age = 55.73, S.D. = 4.61), with five males and ten females. It should be noted that the number of participants in the senior group did not match the other groups due to the difficulty of recruiting participants who were able to fulfill all the criteria for this age range.

2.1.2 Materials

Eight sets of syllables, derived from the roots /fɐn/, /jɐu/, /fu/, /si/, /jɐn/, /ji/, /jyn/ and /wɐi/, served as the target syllables. Each set consisted of six real syllables that contrasted only in tone, giving a total of 48 target syllables (see Table 2.1). The segmental structures were not controlled in the study, although different segmental structures, such as the initial consonants and the vowel height, can affect the intrinsic F0 of a syllable (Lehiste, 1970, Howie, 1974). An uncontrolled design was necessary for the study in order to include a wide variety of syllables, since there were great individual variations on which syllables or words of the same phonetic makeup merged among speakers. The linguistic materials, in fact, included all the possible syllables in GZC that were able to meet the following selection criteria: 1) The root chosen was able to derive real syllables for all six contrastive tones; 2) Each derived syllable was able to be represented by a character that was recognized readily by speakers with primary school education level; and 3) The character chosen did not involve alternative tone readings. For example, the syllable /se/ was excluded from the experiment since the syllable /se^{13}/ '社'has a tonal variation of /se^{33}/. One thing to note is that the characters selected for T5 (*Yang Shang* tone) were divided into two different groups in view of the historical sound change "*Zhuo Shang bian Qu* 濁上變去" in Mandarin. Accordingly, the first group (referred to as T5a in the following) involved the syllables "/fɐn/, /jɐu/, /fu/ and /si/", which have become the *Qu* tone in contemporary Mandarin. The second group (referred to as T5b in the following) involved the syllables "/jɐn/, /ji/, /jyn/ and /wɐi/" that remain as the *Shang* tone in contemporary Mandarin. These eight words were brought in deliberately to test the influence of Mandarin on Cantonese tone mergers. It was hypothesized that speakers with a stronger influence from Mandarin should have a stronger tendency to misperceive or mispronounce T5a words as T3/T6 (*Qu* tone) words.

Table 2.1

Eight sets of tonal contrasts

	/fɐn/	/jɐu/	/fu/	/si/	/jɐn/	/ji/	/jyn/	/wɐi/
T1 (55/ 53)	芬 ‘fragrance’	優 ‘excellent’	夫 ‘husband’	詩 ‘poem’	因 ‘reason’	衣 ‘clothes’	冤 ‘injustice’	威 ‘powerful’
T2 (25)	粉 ‘powder’	柚 ‘pomelo’	苦 ‘bitter’	史 ‘history’	忍 ‘bear’	椅 ‘chair’	丸 ‘meat ball’	委 ‘grieved’
T3 (33)	訓 ‘practice’	幼 ‘thin’	富 ‘rich’	嗜 ‘hobby’	印 ‘print’	意 ‘intention’	怨 ‘hatred’	畏 ‘fear’
T4 (21/ 11)	焚 ‘burn’	柔 ‘tender’	扶 ‘hold’	時 ‘time’	人 ‘person’	移 ‘move’	圓 ‘circle’	圍 ‘enclose’
T5 (13)	T5a				T5b			
	奮 ‘work hard’	誘 ‘seduce’	婦 ‘woman’	市 ‘city’	引 ‘attract’	耳 ‘ear’	軟 ‘soft’	偉 ‘great’
T6 (22)	份 ‘proportion’	右 ‘right’	父 ‘father’	事 ‘event’	孕 ‘pregnant’	二 ‘two’	願 ‘wish’	胃 ‘stomach’

All of the stimuli were pronounced by a native female GZC speaker, who was born and grew up in *Haizhu* district. Both the discrimination and the production test were conducted with the speaker. The results showed that the six-way tonal contrast was well maintained in her speech. The stimuli were recorded in a sound-attenuated recording studio in the Hong Kong Polytechnic University's Department of Chinese and Bilingual Studies. The accuracy of the stimuli was double-checked by two other native Cantonese speakers. The overall loudness of each of the stimuli was adjusted slightly by modifying the overall amplitude of each token as necessary.

2.1.3 Procedures

Production, discrimination and identification tasks were administered to the participants using a Visual Studio custom program running on an IBM X201computer. Information from the complete experiment was displayed visually on the computer screen and presented aurally to participants through the headphones before the experiment began. Instructions for each task were presented visually and aurally prior to the tryout of each task.

2.1.3.1 Discrimination task

The AX discrimination task was administered to the participants. Each tone was paired up with the other five tones to form the AB pairs, for example T1/T2, T1/T3, T1/T4, T1/T5 and T1/T6. Thus, each set of target syllables generated 21 tonal contrasts (15 AB pairs and 6 AA pairs), giving a total of 168 stimuli with 120 AB pairs and 48 AA pairs. This task was divided into three blocks of 56 pairs each with a short break in between. The order effect of each AB pair was counter-balanced by a random intra-pair sequence. In other words, whether a pair was AB or BA was decided randomly by the program. The intra-pair interval was set to 300 ms. Additionally, the order effect of the block was counter-balanced by the random selection of 56 pairs from the data pool of 168 pairs into each block. For each pair, the participants were instructed to click on the 'play' button on the computer screen once and to listen to the tone pair over the headphones. They were then asked to determine whether the tone pair presented was the same or different by clicking the appropriate button on the screen. The participants were allowed to listen repeatedly to the tone pair up to three times by clicking the 'replay' button. Once the participants selected a response, the program automatically provided a prompt for the next trial. There were three practice trials before the actual task. There was no time limit for this task.

2.1.3.2 Identification task

In this task, the 48 target syllables were embedded into a sentence carrier [nei^{55} kɔ^{33} tsi^{22} hɐi^{22} _] 'This is ___ character' produced by the same native speaker of GZC. The stimuli were drawn randomly by the computer program and presented aurally to the participants. Six Chinese characters, representing each of the six contrastive tones of the corresponding root syllable were presented visually on the computer screen. The participants were instructed to click on the 'play' button once and to listen to the stimulus over the headphones. Then they had to identify the perceived tone by clicking on the appropriate on-screen character. They were allowed to listen to each stimulus up to three times by clicking the 'replay' button. Once the participants selected a response on the character icon, the program automatically provided a prompt for the next trial. Before the task, there were three practice trials to acquaint participants with the procedure. No time limit was set for this task.

2.1.3.3 Production task

The 48 target syllables were embedded into two sentence carriers in different positions: sentence-internal position [ŋɔ^{13} ji$^{21/55}$ ka^{55} tʊk^{2} _ tsi^{22}] ‘I read the character _’ and sentence-final position [nei^{55} kɔ^{33} tsi^{22} hɐi^{22} _] ‘This is _ character’, giving a total of 96 stimuli. The sentence-internal carrier was employed to counterbalance the effect of intonation on the F0 pattern of the target word when imbedded in the sentence-final carrier. Moreover, the preceding and following syllables of the target syllables were both T6 (the low level tone) in the carriers. The stimuli were drawn randomly by the computer program and presented visually to the participants on the computer screen. For each trial, the participants were instructed first to click on the ‘recording’ button on the computer screen and then to read aloud the sentences at a normal speech rate. They were then asked to click on the ‘finish’ button and to move onto the next trial. Again, no time limit was set for this task. A training phase was conducted to ensure that the participants were familiar with all the target characters. They were allowed to go over all the words for as long as they wanted to. Each participant had one tryout before the task, thus adjustments, such as voice volume, could be made after listening to the recording. The recording was carried out with a Shure SM48S low-noise unidirectional microphone. A 15 cm mouth-to-microphone distance was maintained for all of the speakers. Each production of the stimuli was stored automatically as a separate wave file on the same IBM computer (with built-in 24 bit sound at a sampling rate of 44 kHz) for further analysis. The speech outputs were transcribed and assessed by the author, a native Cantonese speaker. Ten percent of the randomly selected speech outputs were rated a second time by the primary investigator and a second native Cantonese speaker. Some speech outputs were selected for acoustic analysis using Praat.

The whole experiment was implemented in the following order: production task, then identification task, and finally the discrimination task. The production was administered ahead of the two perception tasks to eliminate any priming effect. It took around 1.5 hours to complete the whole experiment.

2.2 Results

This section presents the respective results for the three tasks, in the order of discrimination, identification and production task. Within the presentation of the results for each task, the overall results will be shown first, followed by the findings for each age group.

2.2.1 Results on discrimination

The percent correct discrimination of each tone pair was calculated for each participant and then averaged first for each age group and then for all participants. The average percent correct discrimination for the 21 tone pairs across all participants were pooled first to check if there was any significant difference. The non-parametric Friedman's test was adopted for the discrimination task because ceiling effects and non-normal distribution were observed. In the presence of a significant overall main effect, follow-up pairwise comparisons were performed with the p-value adjusted using the Dunn-Bonferroni correction. Since age may have played a role in sound change, the data were then compared among the three age groups. If any significant age group effect was found, pair-wise comparisons of the percent correct discrimination for the 21 tone pairs were performed for each age group.

2.2.1.1 Overall results

Figure 2.1 shows the discrimination accuracy for the 21 tone pairs across all participants. The discrimination accuracy was significantly different among the tone pairs, χ^2 (20) = 311.2 and $p < .001$. The Dunn-Bonferroni pairwise test revealed that the percent correct discrimination for T3/T6 (M = 86.6%, S.D. = 19.1%) and T4/T6 (M = 89.7%, S.D. = 14.1%) were significantly lower than those of the other tone pairs, with over 95% correct (for T3/T6, adjusted $p < .05$ for all; for T4/T6, adjusted $p < .01$ for all). It should also be noted that the discrimination of T2/T5 had the third lowest percent correct (M = 93.8%, S.D. = 14.1%); although, the difference was not significant.

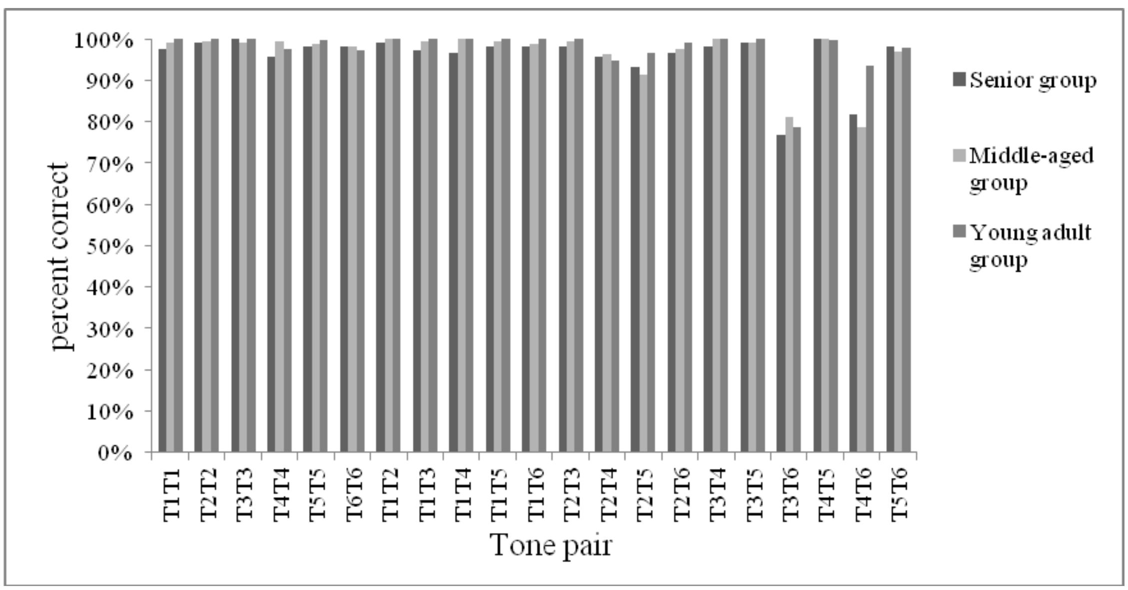


Figure 2.1. The average percent correct discrimination of the 21 tone pairs for the three age groups.

Gender and age may be two factors in tone discrimination. For gender, no significant difference was found between males and females in discriminating the 21 tonal contrasts. In order to examine the age-group effect on the discrimination of the tone pairs, the percent correct for the 21 tone pairs were compared across the three age groups using Mann-Whitney pair tests. Among the three most problematic tone pairs, T3/T6, T4/T6 and T2/T5, the percent correct discrimination of T3/T6 was found to be comparable across the three age groups, as illustrated in Figure 2.2. However, the percentage of correctly discriminating the T4/T6 contrast in the young adult group (M = 93.3%, S.D. = 9.7%) was significantly higher than that of the senior and the middle-aged group (M= 81.7%, S.D. = 18.8%; M = 78.5%, S.D. = 24.6%, respectively), adjusted $p < .05$ for both. Additionally, a marginal significance was found for the scores of the T2/T5 pair between the young adult and the middle-aged group, with the young adults better at discriminating this contrast than were the middle-aged speakers (adjusted $p = .075$). The findings on interaction of correct discrimination by age suggested that the speakers in each age group performed differently in discriminating the tonal contrasts, therefore, a separate analysis was carried out for the data for each age group to explore the individual group patterns.

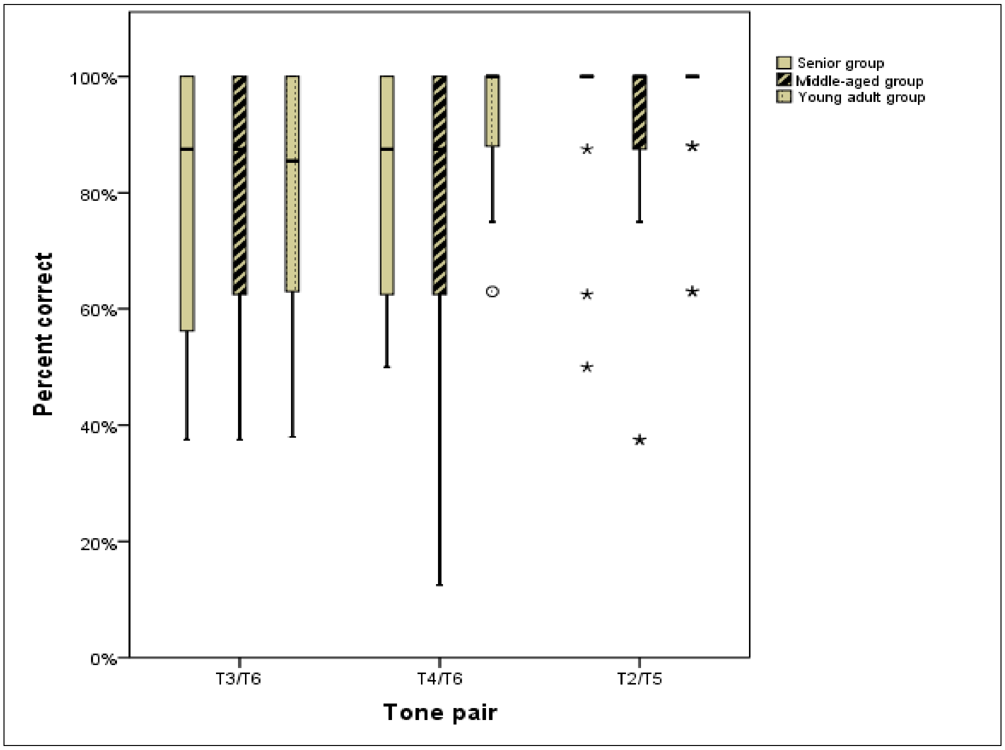


Figure 2.2. Boxplot of the percent correct for the discrimination of the T3/T6, T4/T6 and T2/T5 pairs for each age group. The top and bottom of the box refer to the limits within which the middle 50% of observations fall, with the horizontal line in the middle of the box representing the value of the median. The bottom and the top horizontal lines represent the lowest and the highest score respectively within one group.

2.2.1.2 Results for the senior group

Figure 2.3 demonstrates the average percent correct discrimination of the 21 tone pairs for the senior group. The percent correct discrimination of the T3/T6 (M = 76.7%, S.D. = 24.5%) and T4/T6 (M = 81.7%, S.D. = 18.8%) pairs had the lowest values, while that of the other tone pairs remained high (all above 95%). Discrimination accuracy was significantly different among the 21 tone pairs in the senior group (χ^2 (20) = 103.5 and p < .001). However, no significant difference was revealed by the pairwise comparisons.

Since there were 21 tonal contrasts, the critical level of significance after correction was extremely small. Besides, the sample size of the senior group was relatively small. Hence, a focused test was performed for this group instead. Among the accuracy means of the 21 tone pairs, only the three lowest, namely: T3/T6, T4/T6 and T2/T5, were compared with that of T1/T4 (M = 96.7%, S.D. = 7.4%). The tonal contrast T1/T4 was picked as the reference since it is the contrast between two extreme level tones in Cantonese. Wilcoxon tests showed that the percent correct discrimination of T4/T6 and T3/T6 were significantly lower (adjusted $p < .05$) than that of the other tone pair.

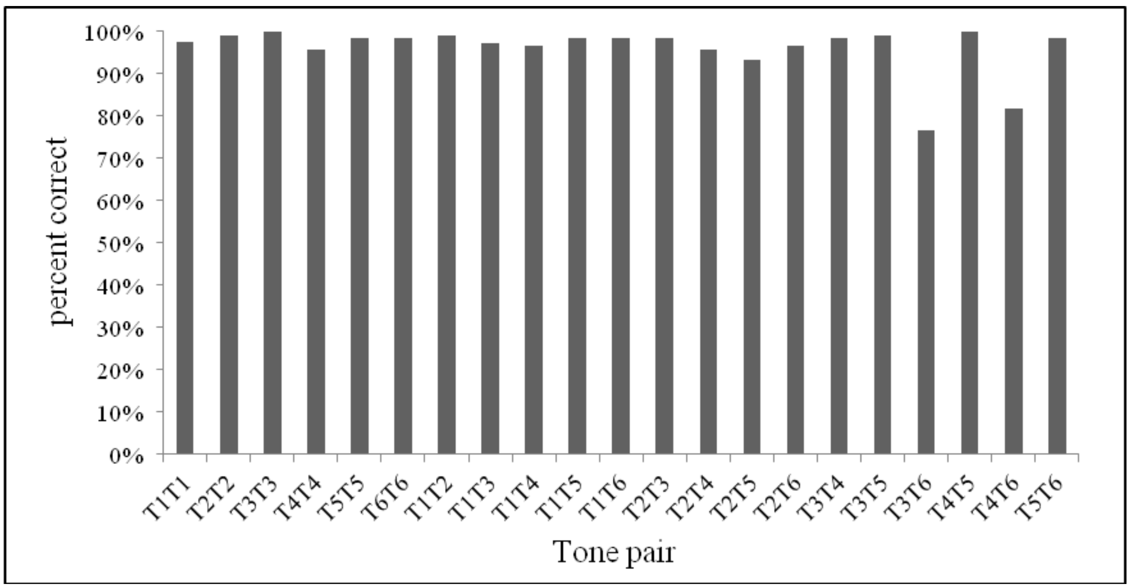


Figure 2.3. The average percent correct discrimination of the 21 tone pairs for the senior group.

2.2.1.3 Results for the middle-aged group

Figure 2.4 illustrates the discrimination accuracy of the 21 tone pairs for the middle-aged group. The percent correct in discriminating the T4/T6 (M = 78.8%, S.D. = 24.6%) and T3/T6 (M = 81.3%, S.D. = 19.6%) pairs were significantly lower than those of the other tone pairs (adjusted $p < .05$ for all). Unlike the senior group, the discrimination of T4/T6 presented a greater challenge to this age group than that of T3/T6.

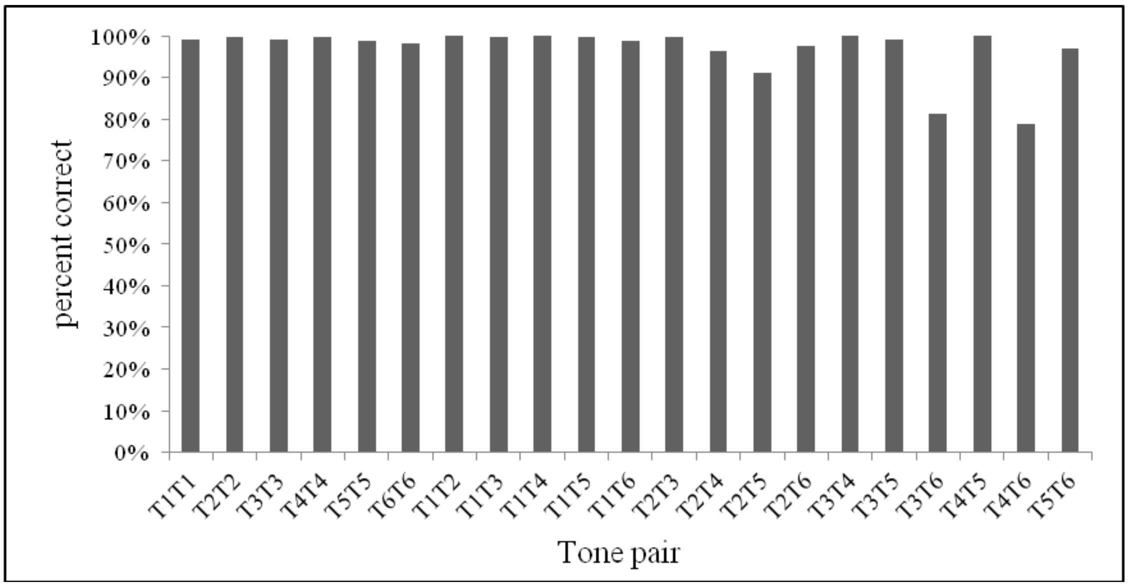


Figure 2.4. The average percent correct discrimination of the 21 tone pairs for the middle-aged group.

2.2.1.4 Results for the young adult group

Figure 2.5 displays the average percent correct discrimination of the 21 tone pairs for the young adult group. Only the percent correct in discriminating the T3/T6 (M = 78.7%, S.D. = 22%) pair was significantly lowered than those of the other tone pairs (adjusted $p < .05$ for all). Contrary to the results for the two older groups, the average percent correct in discriminating the T4/T6 pair was quite high (M = 93.3%) in this age group.

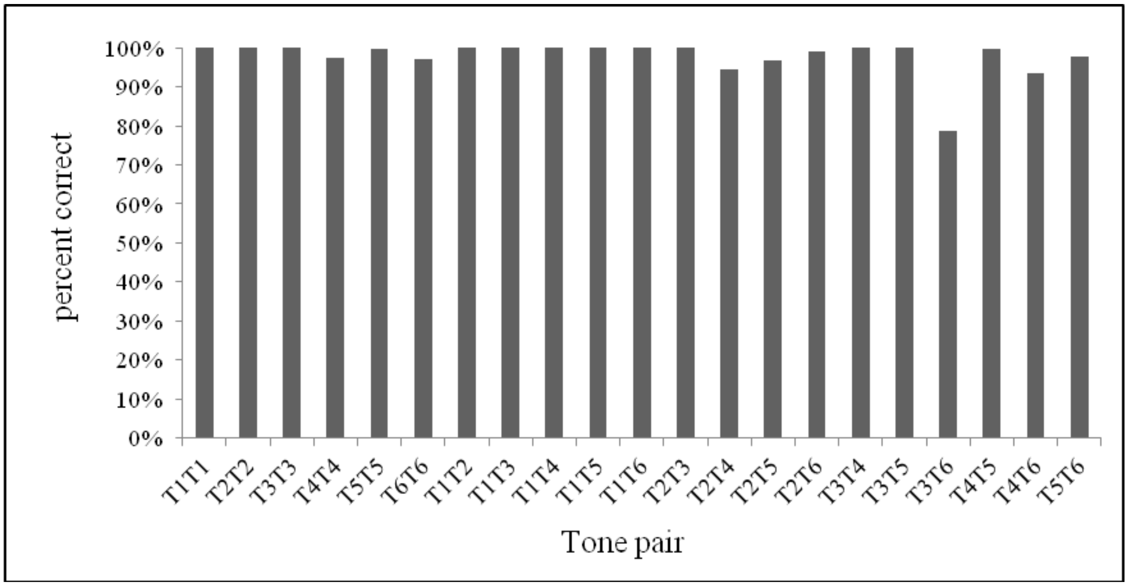


Figure 2.5. The average percent correct discrimination of the 21 tone pairs for the young adult group.

2.2.1.5 Summary of the results on discrimination

The results of the discrimination task are summarized as follows:

1. Overall, T3/T6 and T4/T6 were the most readily confused pairs;
2. The discrimination of T3/T6 was bad for all age groups;
3. The discrimination of T4/T6 was bad for the senior and the middle-aged groups, but quite good for the young adults;

4. The T2/T5 contrast was well maintained in all age groups. However, results also suggested a stronger tendency among the middle-aged participants to confuse this tone pair than did the other age groups.

2.2.2 Results on identification

The analysis of the results for this task was divided into two parts. First, the correct identification of each tone was calculated for each participant. Second, the identification pattern for each individual listener was constructed. Additionally, in order to test the hypotheses regarding the influence of Mandarin on Cantonese, the percentage of T5 misidentified as T3 or T6 was averaged for each of the two syllable sub-groups, T5a and T5b. An evaluation of the relationship between the syllable sub-group and the rate of misidentifying T5 as T3 was measured using Kendall's tau correlation. Group confusion matrices were compiled by summing the confusion matrices across participants. Non-parametric statistics were employed for comparisons of this data, as ceiling effects were observed and normal distribution could not be assumed.

2.2.2.1 Problematic tones

2.2.2.1.1 Overall results

Figure 2.6 shows the identification accuracy for the six tones for the three age groups. Tone type, gender and age may have been three factors in tone identification. For tone type, the identification accuracy of T6 (M = 64.9%, S. D. = 23.3%) was significantly lower than those of the other tones (adjusted $p < .001$ for all). The next two tones with relatively low identification accuracy were T3 (M = 86.7%, S.D. = 19.9%) and T5 (M = 89.5%, S.D. = 15.0%). No gender effects were found. With regard to age, the senior group made more identification errors than did the two younger groups for all six tones. In particular, the identification accuracy of T6 was significantly lower whereas that of T3 was marginally lower than the two younger groups (T6, adjusted $p < .001$ for all; T3, adjusted $p = .064$ and $p = .057$ respectively). In general, the results for the problematic tones for each age group were quite similar, henceforth, no separate analysis was carried out for any age group.

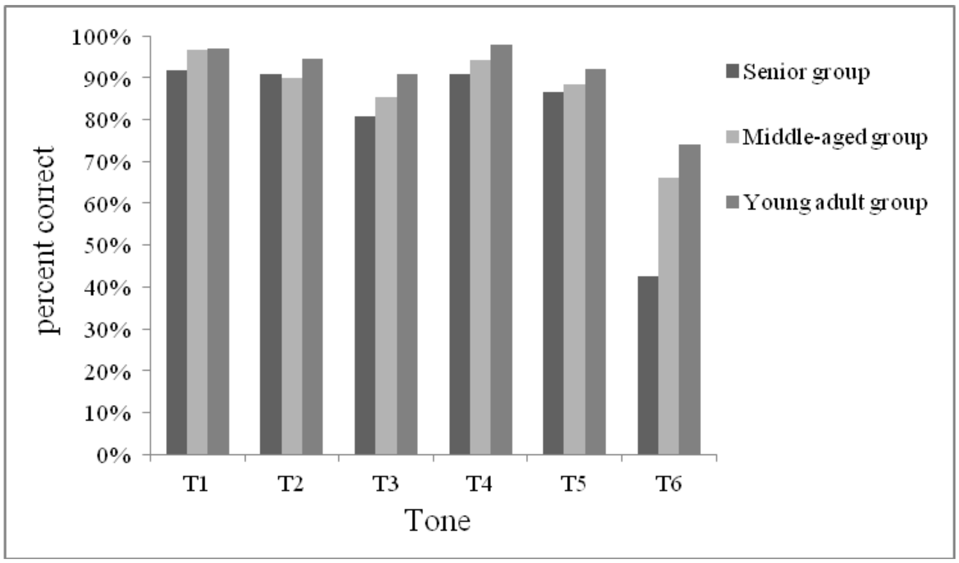


Figure 2.6. The average percent correct identification of the six tones for the three age groups.

2.2.2.2 Error patterns

2.2.2.2.1 Overall

Table 2.2 illustrates the confusion matrix for the six tones across all participants. The stimuli are arranged in rows whereas the responses are in columns. The numbers in a cell represents the percentage of responses being realized as that particular tone. For example, T2 was identified accurately 44 times out of 48 trials, resulting in an identification rate of 92%. Correct identifications are shown by the numbers appearing on the diagonal of the matrix (shaded grey). Percentages of identification errors between two tones are represented by the numbers in the non-shaded cells.

Table 2.2

Confusion matrix for the identification of the six tones across all participants with significant error rate asterisked

Target \ Perceived	T1	T2	T3	T4	T5	T6
T1	95.83	0.83	2.17	1.00	0.00	0.17
T2	0.00	92.00	0.67	2.50	3.83	1.00
T3	1.17	0.50	86.67	0.83	4.00	6.83*
T4	0.34	1.50	0.33	95.00	1.00	1.83
T5	0.16	5.17	2.67	1.17	89.50	1.33
T6	0.33	0.50	18.00*	12.00*	4.50	64.67

A statistical analysis was conducted to examine the error patterns for the three most readily misidentified tones further, namely: T6, T3, and T5.

Error pattern of T6: Most of the errors of T6 involved identifying the target tone as T3 and T4. The percentages of these two errors were significant (T6 as T3, adjusted $p < .001$ for all; T6 as T4, adjusted $p < .05$ for all), and the percentage of T6 as T3 (18%) was a bit higher than that of T6 as T4 (12%).

Error pattern of T3: T3 was significantly identified as T6 (adjusted $p < .05$ for all).

Error pattern of T5: None of the errors revealed in Table 2.2 was significant. Recall that there were two syllable sub-groups of T5: T5a and T5b. As explained in Section 1.5.2, it was hypothesized that T5a would be more readily identified as T3 or T6 if the Cantonese readings were influenced by the participants' Mandarin knowledge. However, the results showed that T5 was most readily misidentified as T2. Moreover, there was no significant difference between the error patterns of T5a (M = 97.3%) and T5b (M = 97.5%) among the three age groups. Hence, the identification rates of the two sub-groups were pooled together for all age groups, as described in the following sections.

In order to explore the age group effect on the identification patterns, the percentages of identification errors were compared across the three age groups using the Mann-Whitney pair test. It was found that the percentage of T6 misidentified as T4 in the senior group (M = 25.83%, S. D. = 16.7%) was significantly higher than that of the middle-aged group (M = 10%, S. D. = 12.5%) and the young adult group (M = 7.08%, S. D. = 9.1%), adjusted $p < .05$ for both. Additionally, marginal significance was found for the identification error of T6 as T3 between the young adult and the senior group. The senior speakers (M = 25%, S. D. = 16.4%) made more errors in misidentifying T6 as T3 than did the young adults (M = 14.52%, S. D. = 12.8%), adjusted $p = .06$. The error patterns of each age group are described in order.

2.2.2.2.2 Results for the senior group

Table 2.3 displays the confusion matrix of the senior group. The statistical analysis for the error patterns was centered on the three most readily misidentified tones: T6, T3, and T5.

Error pattern of T6: Most of the errors of T6 involved identifying the target tone as T4 and T3. The percentages of these two errors were significantly higher (both patterns, adjusted $p < .01$ for all) and almost equal (25.83% vs. 25%).

Error pattern of T3: No significant difference was found for the four error patterns. However, the percentage of T3 identified as T6 was the highest (10.83%) followed by that of T3 identified as T5 (5%).

Error pattern of T5: None of the errors revealed in Table 2.3 was significant. However, most of the errors of T5 involved identifying the target tone as T3 and T2; moreover, the percentages were equal (5.83% vs. 5%).

Table 2.3

Confusion matrix for the identification of the six tones in the senior group with significant error rate asterisked

Perceived / Target	T1	T2	T3	T4	T5	T6
T1	91.67	1.67	5.00	0.83	0.00	0.83
T2	0.00	90.83	1.67	0.00	6.67	0.83
T3	1.67	0.00	80.83	1.67	5.00	10.83
T4	0.00	4.12	0.00	90.83	0.88	4.17
T5	0.00	5.00	5.83	2.50	86.67	0.00
T6	0.00	1.67	25.00*	25.83*	5.00	42.50

2.2.2.2.3 Results for the middle-aged group

Table 2.4 presents the confusion matrix for the middle-aged group. The statistical analysis for error patterns was concentrated on the three most readily misidentified tones: T6, T3 and T5.

Error pattern of T6: Again, most of the errors of T6 involved identifying the target tone as T3 and T4. The percentages of these two types of error were significant (T6 identified as T3, adjusted $p < .01$ for all; T6 identified as T4, adjusted $p < .05$ for all). However, contrary to the case of the senior group, the percentage of T6 identified as T3 (17.92%) was considerably higher than that of T6 identified as T4 (10%) in this age group.

Error pattern of T3: No significant difference was found for the five error patterns. However, the percentage of T3 identified as T6 was the highest (6.25%) followed by that of T3 identified as T5 (4.5%).

Error pattern of T5: None of the errors revealed in Table 2.4 was significant. However, differed from that of the senior group, most of the errors of T5 involved identifying the target tone as T2 (5.42%) only.

Table 2.4

Confusion matrix for the identification of the six tones in the middle-aged group with significant error rate asterisked

Perceived / Target	T1	T2	T3	T4	T5	T6
T1	96.67	1.25	1.25	0.83	0.00	0.00
T2	0.00	90.00	0.42	4.58	4.17	0.83
T3	1.67	1.25	85.42	0.83	4.58	6.25
T4	0.42	1.66	0.42	94.17	1.25	2.08
T5	0.00	5.42	2.92	0.83	88.33	2.50
T6	0.83	0.00	17.92*	10.00*	5.00	66.25

2.2.2.2.4 Results for the young adult group

Table 2.5 shows the confusion matrix for the young adult group. Statistical analysis for error patterns was again concentrated on the three most readily identified tones: T6, T3 and T5.

Error pattern of T6: Again, most of the errors of T6 involved identifying the target tone as T3 (14.52%) and T4 (7.08%). However, only the error pattern of T6 identified as T3 was significantly higher (adjusted $p < .05$ for all).

Error pattern of T3: No significant difference was found for the four error patterns. Different from the two older groups, most of the errors of T3 involved identifying T3 as T6 only (5.42%).

Error pattern of T5: None of the errors revealed in Table 2.5 was significant. Similar to that of the middle-aged group, most of the errors of T5 involved identifying the target tone as T2 (5%) only.

Table 2.5

Confusion matrix for the identification of the six tones in the young adult group with significant error rate asterisked

Perceived / Target	T1	T2	T3	T4	T5	T6
T1	97.08	0.00	1.67	1.25	0.00	0.00
T2	0.00	94.58	0.42	1.67	2.08	1.25
T3	0.41	0.00	90.83	0.42	2.92	5.42
T4	0.41	0.00	0.42	97.92	0.83	0.42
T5	0.43	5.00	0.83	0.83	92.08	0.83
T6	0.00	0.48	14.52*	7.08	3.75	74.17

2.2.2.3 Summary of the results on identification

The results of the identification task are summarized as follows:

1. Overall speaking, the identification rate of T6 was significantly lowered than that of the other tones;

2. Of the three age groups, the senior group committed the most identification errors followed by the middle-aged and the young adult group;

3. As for the directions of the errors, T6 was identified as T3 and T4 in the senior and middle-aged groups. It was identified as T3 only in the young adult group. In sum, the percentage of T6 identified as T4 increased with age;

4. Two other trends were noted. The first one was the relatively low identification rate of T3 in all age groups. It was most readily identified as T6, followed by T5;

5. The second one was the relatively low identification rate of T5 in all age groups. It was most readily identified as T2 and T3 in the senior and middle-aged groups. It was identified as T2 only in the young adult group. In sum, the percentage of T5 identified as T3 increased with age.

2.2.3 Results on production

The analysis of the production task results involved three parts. Primarily, the speech samples were assessed and the produced tones were transcribed by the investigator. A produced tone perceived as an intermediate between the target tone and a particular tone was marked as intermediate and was counted as an instance of production errors. Each participant's production score for each tone was calculated by the number of correctly produced trials over the total number of trials. Then the percent correct of each tone was measured by averaging the production score across all participants. Ten percent of the randomly selected speech outputs were rated a second time by the primary investigator and by a second native Cantonese speaker. Intra-rater and inter-rater reliability coefficients were derived using Pearson's correlation. Subsequently, in order to test the hypothesis regarding the influence of Mandarin on Cantonese, the percentage of T5 words produced as T3 or T6 words was averaged for each of the two syllable sub-groups, T5a and T5b. An evaluation of the relationship between the syllable sub-group and the percentage of T5 produced as T3 or T6 was measured using Kendall's tau correlation. Last, the pattern of production errors for each individual listener was constructed, with each response classified as either correct or incorrect according to the target tone. The intermediate forms of production were categorized into the nearest tone. For instance, a more T3-like intermediate production was categorized as a T3 realization. Group confusion matrices were compiled by summing the confusion matrices across participants. Non-parametric statistics were employed for this task as ceiling effects were observed and normal distribution could not be assumed.

2.2.3.1 Problematic tones

2.2.3.1.1 Overall results

Intra-rater and inter-rater reliability were calculated by repeating the evaluation for the production of eight randomly-selected speakers (10.7% of the data) by the primary investigator and the second rater. A good level of agreement was found between the judgments. Intra-rater reliability was $r = .97$ ($p < .001$) and intra-rater reliability was $r = .95$ ($p < .001$).

Figure 2.7 shows the production accuracy of the six tones for the three age groups. Tone type, gender and age may be three factors in tone production. Production accuracy was significantly different among the six tones, χ^2 (5) = 140.2 and $p < .001$. Dunn-Bonferroni pairwise comparisons revealed that the percent correct production of T6 (M = 75%, S. D. = 22.8%) and T5 (M = 79.4%, S. D. = 17.0 %) were significantly lower than those of the other four tones, which were over 89% (adjusted $p < .001$ for all). As for gender, no significant difference was found between males and females in production. A test for the age group effect on the production accuracy of the six tones showed marginal significance for T3 between the senior and the young adult group (adjusted $p = .067$).

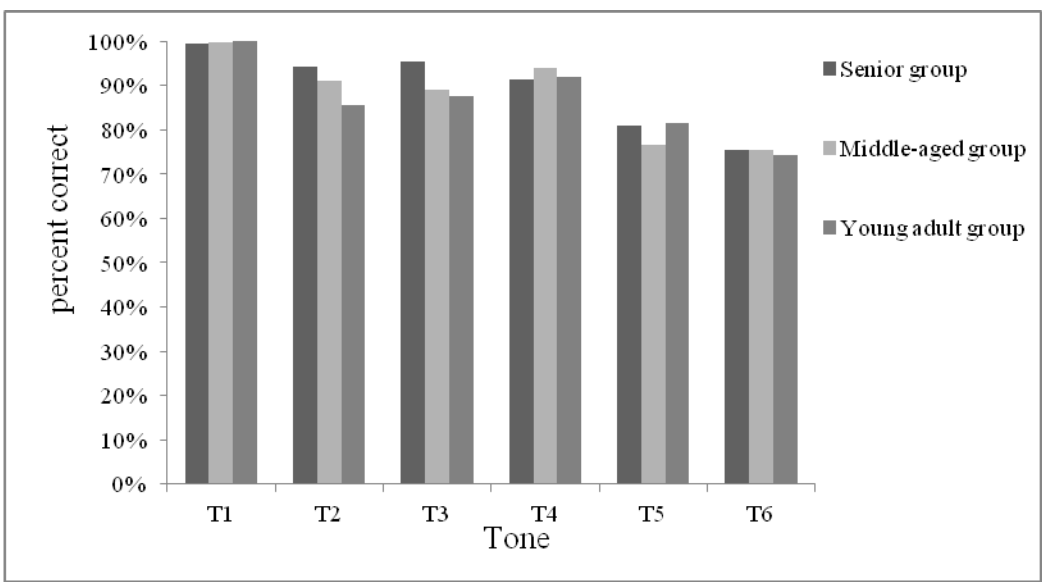


Figure 2.7. The average percent correct production of the six tones

for the three age groups.

Recall that there were two syllable sub-groups of T5. When examining the production errors of T5, most of the errors involved producing the target tone as T3 but not T6, as will be evident in the following confusion matrices. Therefore, the percentages of T5 produced as T3 were averaged for the T5a and T5b syllable sub-groups respectively within each age group, as shown in Figure 2.8. Correlation coefficients indicated that the percentage of T5 pronounced as T3 and the syllable sub-group correlated significantly, $r = .59$, $p < .001$. In contrast, age was not significantly correlated with the percentage of the error pattern T5 as T3, $r = -.136$, $p = .42$. Although there was no significant association between the age of participants and the extent to which they committed the production error, it should be pointed out that the percentage of T5a words produced as T3 in the senior group was notably lower than that of its younger counterparts (see Figure 2.8). In the following sections, calculation of the percentage of correct production for T5 will be separated for T5a and T5b syllable sub-groups, as well as the percentage of T5 words produced as T3 in the confusion matrices.

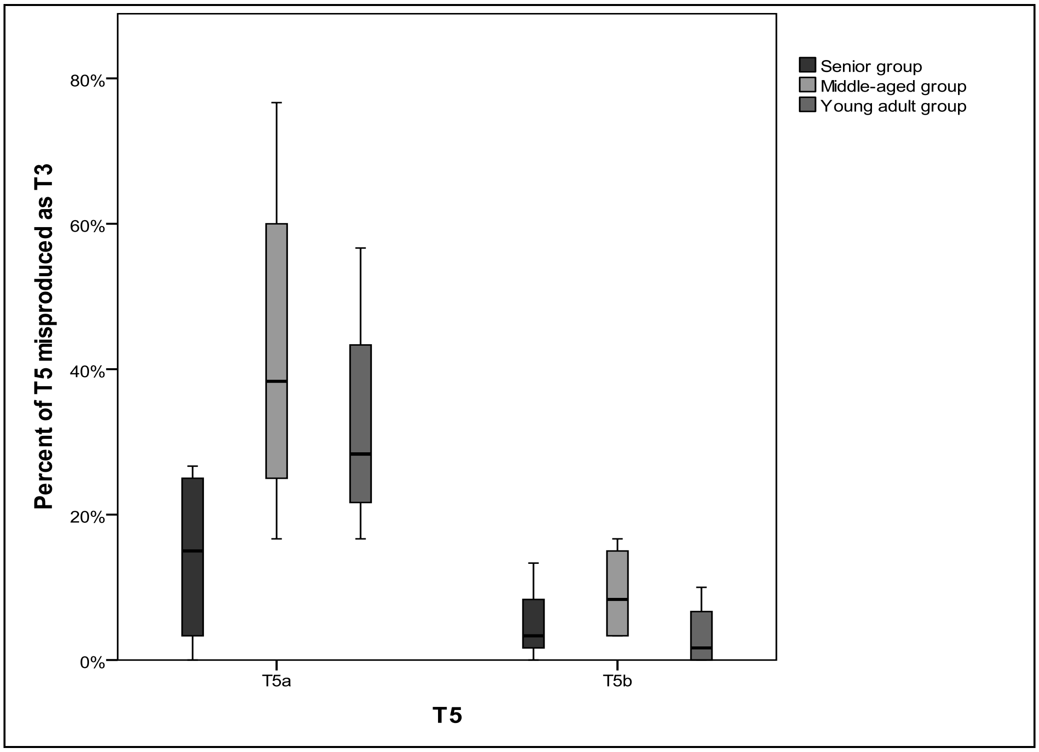


Figure 2.8. Boxplot of the percentages of T5 produced as T3 for the three age groups. The top and bottom of the box refer to the limits within which the middle 50% of observations fall, with the horizontal line in the middle of the box representing the value of the median. The bottom and the top horizontal lines represent the lowest and the highest score respectively within one group.

Figure 2.9 presents the average percent correct production of the T5a and T5b syllables, as well as those of the other five tones. Comparison among the seven tone groups indicated that the accuracies of T6, T5a were significantly lower (T6, adjusted $p < .05$ for four; T5a, adjusted $p < .001$ for one tone). Additionally, the accuracy of T3 was attested to be significantly lower than that of T1 (adjusted $p < .001$).

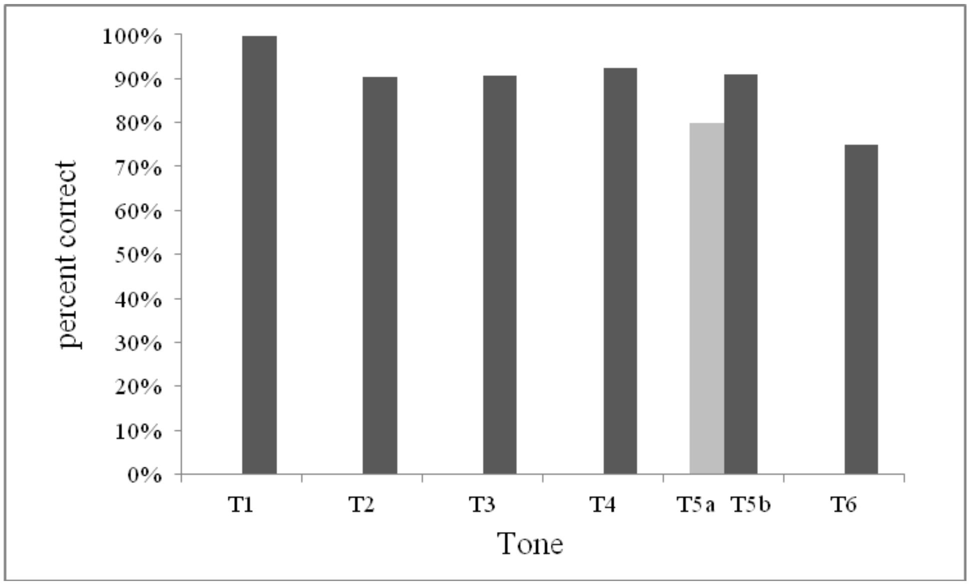


Figure 2.9. The average percent correct production of the six tones with that of T5 split by two syllable sub-groups for the three age groups.

2.2.3.1.2 Results for the senior group

Figure 2.10 presents the average percent correct production of the seven tone groups among the senior participants. The accuracies of T6 (M = 75.42%, S. D. = 22.6%) and T5a (M = 83.3%, S. D. = 14.7 %) were significantly lower than those of the other five tones, which remained fairly high, all above 90% (T6, adjusted $p < .01$ for all; T5a, adjusted $p < .05$ for all).

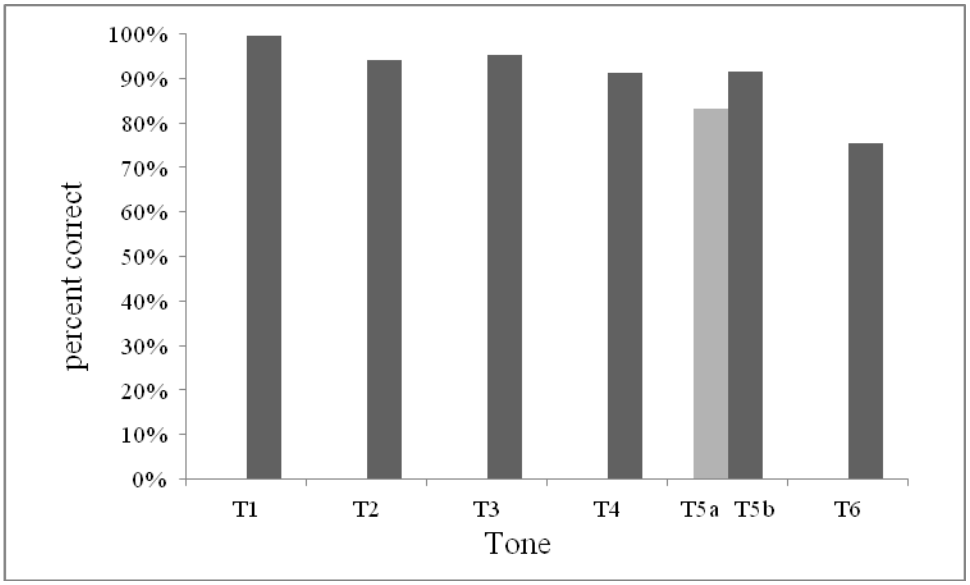


Figure 2.10. The average percent correct production of the six tones with T5 split by T5a and T5b syllable sub-groups for the senior group.

2.2.3.1.3 Results for the middle-aged group

The production accuracy of the seven tone groups among the middle-aged participants is shown in Figure 2.11. The accuracies of T6 (M = 75.42%, S. D. = 19.5%) and T5a (M = 76.7%, S. D. = 14.2%) were significantly lower than those of four other tones which remained fairly high, all above 90% (T6, adjusted $p < .05$ for all; T5a, adjusted $p < .01$ for all). Additionally, the accuracy of T3 was significantly lower than that of T1 (adjusted $p < .01$).

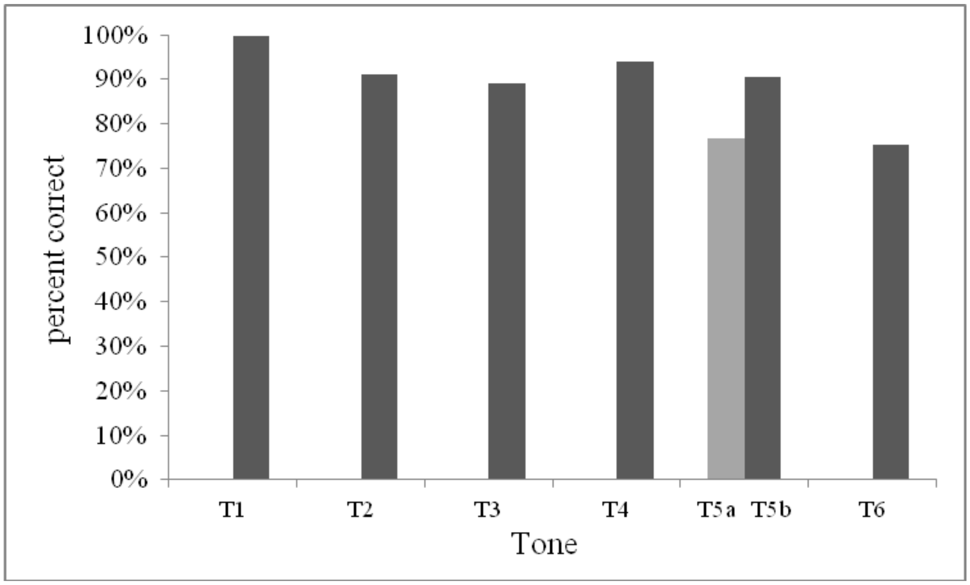


Figure 2.11. The average percent correct production of the six tones with T5 split by T5a and T5b syllable sub-groups for the middle-aged group.

2.2.3.1.4 Results for the young adult group

The production accuracy of the seven tone groups among the young adult participants is shown in Figure 2.12. The accuracies of T6 (M = 74.38%, S. D. = 26.5%) and T5a (M = 80%, S. D. = 15.9%) were significantly lower than those of three other tones which remained fairly high, all above 90% (both T6 and T5a, adjusted $p < .05$ for all), In addition, the accuracies of T2 (M = 85.63%, S. D. = 21.4%) and T3 (M = 87.5%, S. D. = 12.8%) were significantly lower than that of T1 (both T2 and T3, adjusted $p < .01$).

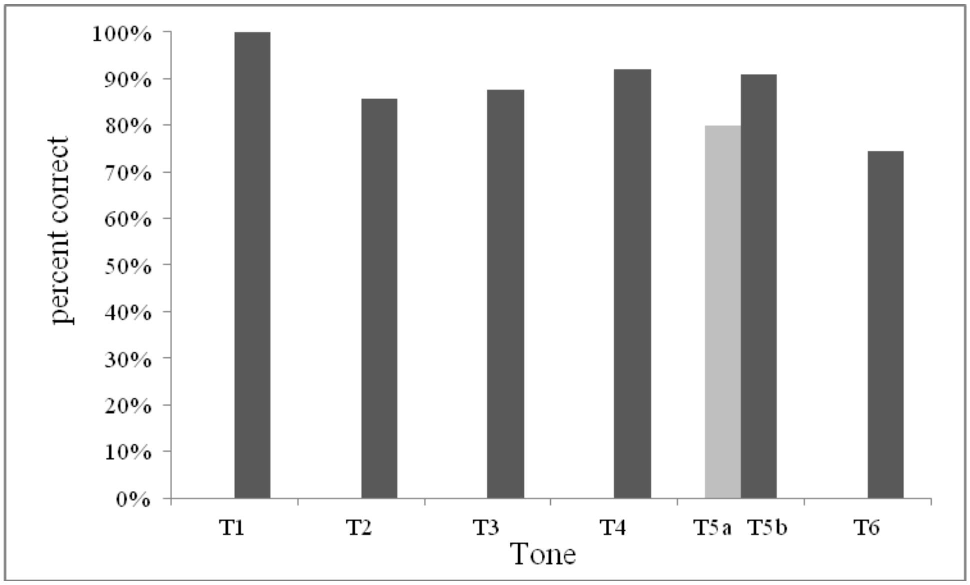


Figure 2.12. The average percent correct production of the six tones with T5 split by T5a and T5b syllable sub-groups for the young adult group.

2.2.3.2 Error patterns

2.2.3.2.1 Overall

The confusion matrix for the six tones across all participants is illustrated in Table 2.6. The target tones are arranged in rows and the actual productions in columns. The number in each cell represents the percentage of the productions being realized as that particular tone. For example, T3 was produced accurately 84 times out of 96 trials, resulting in a percentage of 87.5%. Percentages of correct production are shown by the numbers appearing on the diagonal of the matrix (shaded grey), while percentages of production errors are represented by the numbers in the non-shaded cells. The cells in the row of T5 as the target tone were divided into three sub-cells: overall percentages when T5a and T5b syllables were combined (the upper rows in the cells refer), percentages for T5a syllables (the left columns in the lower rows refer) and percentages for T5b syllables (the right columns in the lower rows refer).

Table 2.6

Confusion matrix for the production of the six tones across all participants with significant error rate asterisked

<table>
<tr><th>Produced / Target</th><th>T1</th><th colspan="2">T2</th><th colspan="2">T3</th><th colspan="2">T4</th><th colspan="2">T5</th><th colspan="2">T6</th></tr>
<tr><td>T1</td><td>99.83</td><td colspan="2">0.00</td><td colspan="2">0.17</td><td colspan="2">0.00</td><td colspan="2">0.00</td><td colspan="2">0.00</td></tr>
<tr><td>T2</td><td>0.00</td><td colspan="2">89.50</td><td colspan="2">0.42</td><td colspan="2">0.50</td><td colspan="2">9.08</td><td colspan="2">0.50</td></tr>
<tr><td>T3</td><td>0.08</td><td colspan="2">0.00</td><td colspan="2">89.67</td><td colspan="2">0.00</td><td colspan="2">1.67</td><td colspan="2">8.58*</td></tr>
<tr><td>T4</td><td>0.17</td><td colspan="2">0.00</td><td colspan="2">0.33</td><td colspan="2">92.67</td><td colspan="2">2.33</td><td colspan="2">4.50</td></tr>
<tr><td rowspan="2">T5</td><td rowspan="2">0.00</td><td colspan="2">2.66</td><td colspan="2">15.67</td><td colspan="2">0.75</td><td colspan="2">79.42</td><td colspan="2">1.50</td></tr>
<tr><td>0.83</td><td>3.06</td><td>17.22*</td><td>3.75</td><td>0.56</td><td>0.42</td><td>80.00</td><td>90.96</td><td>1.39</td><td>1.81</td></tr>
<tr><td>T6</td><td>0.50</td><td colspan="2">0.50</td><td colspan="2">17.50*</td><td colspan="2">4.25</td><td colspan="2">2.25</td><td colspan="2">75.00</td></tr>
</table>

Statistical analysis was conducted to further examine the error patterns of the three most readily misproduced tones, namely: T6, T5a and T3.

Error pattern of T6: Among the five production errors of T6, the percentage of T6 produced as T3 was significantly higher than those of the other errors (adjusted $p < .001$ for all).

Error pattern of T5a: At first glance, most T5 were misproduced as T3. On closer examination, the low production rate was caused mainly by T5a. The percentage of T5a produced as T3 was the highest and significantly different from that of the other errors of T5a (adjusted $p < .001$ for all).

Error pattern of T3: The percentage of T3 produced as T6 was significantly higher than that of the other errors (adjusted $p < .001$ for all).

In order to explore the age group effect on the production errors, the percentages of production errors were compared across the three age groups. The production error of T3 produced as T6 in the senior group (M = 2.5%, S. D. = 4.5%) was significantly lower than that in the middle-aged group (M = 10%, S. D. = 13.7%) and that in the young adult group (M = 10.21%, S. D. = 11.1%), adjusted $p < .01$ for both. The error patterns of each age group are described in order.

2.2.3.2.2 Results for the senior group

The confusion matrix for the production of the six tones among the senior participants is illustrated in Table 2.7. A statistical analysis of the error patterns was conducted for the two most readily misproduced tones, namely: T6 and T5a.

Error pattern of T6: Most of the errors of T6 involved producing the target tone as T3 and the percentage of T6 produced as T3 was significantly higher than those of the other errors (adjusted $p < .01$ for all).

Error pattern of T5a: The percentage of T5a produced as T3 was the highest and significantly different from those of the other errors of T5a (adjusted $p < .05$ for all).

Table 2.7

Confusion matrix for production of the six tones in the senior group with significant error rate asterisked

<table>
<tr><th>Produced
Target</th><th>T1</th><th colspan="2">T2</th><th colspan="2">T3</th><th colspan="2">T4</th><th colspan="2">T5</th><th colspan="2">T6</th></tr>
<tr><td>T1</td><td>99.59</td><td colspan="2">0.00</td><td colspan="2">0.41</td><td colspan="2">0.00</td><td colspan="2">0.00</td><td colspan="2">0.00</td></tr>
<tr><td>T2</td><td>0.00</td><td colspan="2">94.17</td><td colspan="2">0.00</td><td colspan="2">1.25</td><td colspan="2">4.58</td><td colspan="2">0.00</td></tr>
<tr><td>T3</td><td>0.00</td><td colspan="2">0.00</td><td colspan="2">95.42</td><td colspan="2">0.00</td><td colspan="2">2.08</td><td colspan="2">2.50</td></tr>
<tr><td>T4</td><td>0.42</td><td colspan="2">0.00</td><td colspan="2">0.83</td><td colspan="2">91.25</td><td colspan="2">2.50</td><td colspan="2">5.00</td></tr>
<tr><td rowspan="2">T5</td><td rowspan="2">0.00</td><td colspan="2">2.50</td><td colspan="2">15.42</td><td colspan="2">0.42</td><td colspan="2">80.83</td><td colspan="2">0.83</td></tr>
<tr><td>0.83</td><td>2.47</td><td>14.21*</td><td>5.00</td><td>0.83</td><td>0.00</td><td>83.30</td><td>91.70</td><td>0.83</td><td>0.83</td></tr>
<tr><td>T6</td><td>0.00</td><td colspan="2">0.00</td><td colspan="2">16.25*</td><td colspan="2">5.83</td><td colspan="2">2.50</td><td colspan="2">75.42</td></tr>
</table>

2.2.3.2.3 Results for the middle-aged group

The confusion matrix for the production of the six tones among the middle-aged participants is illustrated in Table 2.8. A statistical analysis of the error patterns was conducted for the three most readily misproduced tones, namely: T6, T5a and T3.

Error pattern of T6: The percentage of T6 produced as T3 was significantly higher than that of the other errors (adjusted $p < .01$ for all).

Error pattern of T5a: T5a was significantly produced as T3 (adjusted $p < .001$ for all).

Error pattern of T3: T3 was significantly produced as T6 (adjusted $p < .05$ for all).

It should be noted that the percentage of T2 produced as T5 was significantly different from that of one other error (adjusted $p < .05$), although the accuracy of T2 was not significantly lower.

Table 2.8

Confusion matrix for the production of the six tones in the middle-aged group with significant error rate asterisked

<table>
<tr><th>Produced / Target</th><th>T1</th><th colspan="2">T2</th><th colspan="2">T3</th><th colspan="2">T4</th><th colspan="2">T5</th><th colspan="2">T6</th></tr>
<tr><td>T1</td><td>99.79</td><td colspan="2">0.00</td><td colspan="2">0.21</td><td colspan="2">0.00</td><td colspan="2">0.00</td><td colspan="2">0.00</td></tr>
<tr><td>T2</td><td>0.00</td><td colspan="2">91.04</td><td colspan="2">1.04</td><td colspan="2">0.42</td><td colspan="2">6.88</td><td colspan="2">0.62</td></tr>
<tr><td>T3</td><td>0.00</td><td colspan="2">0.00</td><td colspan="2">88.96</td><td colspan="2">0.00</td><td colspan="2">1.04</td><td colspan="2">10.00*</td></tr>
<tr><td>T4</td><td>0.20</td><td colspan="2">0.00</td><td colspan="2">0.21</td><td colspan="2">93.96</td><td colspan="2">1.88</td><td colspan="2">3.75</td></tr>
<tr><td rowspan="2">T5</td><td rowspan="2">0.00</td><td colspan="2">1.67</td><td colspan="2">18.75</td><td colspan="2">1.03</td><td colspan="2">76.67</td><td colspan="2">1.88</td></tr>
<tr><td>0.83</td><td>1.67</td><td>21.22*</td><td>4.58</td><td>0.42</td><td>1.27</td><td>76.70</td><td>90.40</td><td>0.83</td><td>2.08</td></tr>
<tr><td>T6</td><td>0.00</td><td colspan="2">0.00</td><td colspan="2">19.79*</td><td colspan="2">2.50</td><td colspan="2">2.29</td><td colspan="2">75.42</td></tr>
</table>

2.2.3.2.4 Results for the young adult group

The confusion matrix for the production of the six tones in the young adult group has been compiled in Table 2.9. A statistical analysis of the error patterns was conducted for the four most readily misproduced tones, namely: T6, T5a, T2 and T3.

Error pattern of T6: T6 was significantly produced as T3 (adjusted $p < .05$ for all).

Error pattern of T5a: T5a were significantly produced as T3 (adjusted $p < .001$ for all). Different from those of T5a, most of the errors of T5b involved producing the target tone as T2.

Error pattern of T2: The percentage of T2 produced as T5 was significantly higher than those of the other errors (adjusted $p < .001$ for all).

Error pattern of T3: The percentage of T3 produced as T6 was significantly higher than those of the other errors (adjusted $p < .05$ for all).

Table 2.9

Confusion matrix for the production of the six tones in the young adult group with significant error rate asterisked

<table>
<tr><th>Produced / Target</th><th>T1</th><th colspan="2">T2</th><th colspan="2">T3</th><th colspan="2">T4</th><th colspan="2">T5</th><th colspan="2">T6</th></tr>
<tr><td>T1</td><td>100.00</td><td colspan="2">0.00</td><td colspan="2">0.00</td><td colspan="2">0.00</td><td colspan="2">0.00</td><td colspan="2">0.00</td></tr>
<tr><td>T2</td><td>0.00</td><td colspan="2">85.63</td><td colspan="2">0.00</td><td colspan="2">0.20</td><td colspan="2">13.54*</td><td colspan="2">0.63</td></tr>
<tr><td>T3</td><td>0.21</td><td colspan="2">0.00</td><td colspan="2">87.50</td><td colspan="2">0.00</td><td colspan="2">2.08</td><td colspan="2">10.21*</td></tr>
<tr><td>T4</td><td>0.00</td><td colspan="2">0.00</td><td colspan="2">0.28</td><td colspan="2">92.01</td><td colspan="2">2.71</td><td colspan="2">5.00</td></tr>
<tr><td rowspan="2">T5</td><td rowspan="2">0.00</td><td colspan="2">3.75</td><td colspan="2">12.71*</td><td colspan="2">0.62</td><td colspan="2">81.46</td><td colspan="2">1.46</td></tr>
<tr><td>0.83</td><td>5.00</td><td>16.25*</td><td>1.67</td><td>0.42</td><td>0.00</td><td>80.00</td><td>90.83</td><td>2.50</td><td>2.50</td></tr>
<tr><td>T6</td><td>1.25</td><td colspan="2">1.25</td><td colspan="2">15.83*</td><td colspan="2">5.21</td><td colspan="2">2.08</td><td colspan="2">74.38</td></tr>
</table>

2.2.3.3 Summary of the results on production

The results for the production task are summarized as follows:

1. Overall, the production accuracy of T6, T5a and T3 was significantly low;

2. The accuracy of T6 was comparable among the three age groups. The accuracy of T5 was the lowest in the middle-aged group. The accuracy of T3 increased with age;

3. As for the directions of errors, T6 was most readily produced as T3. T5a was also most readily produced as T3. However, T3 was most readily produced as T6;

4. A trend was noted regarding T2. The production accuracy of T2 was well preserved in the senior group. However the accuracy decreased in the two younger groups. It was most readily produced as T5 in these two groups.

2.3 Summary of the whole experimental study

2.3.1 The perception tasks

In this experimental study, two perception tasks, discrimination and identification, were conducted to investigate the perception of the six tones among the GZC speakers.

2.3.1.1 The discrimination task

Twenty one tonal contrasts were presented to 75 native GZC speakers in the discrimination task to examine whether the GZC speakers could still contrast the six tones in perception. T3/T6 and T4/T6 were the most readily confused pairs. However, the discrimination ability differed among the three age groups. Specifically, the discrimination of T3/T6 was bad for all age groups, while that of T4/T6 was bad for the senior and the middle-aged groups, but was good for the young adult group. The T2/T5 contrast was well maintained in all three age groups. Nonetheless, the results for the middle-aged group suggested a stronger tendency to confuse this tone pair than did the other two age groups.

2.3.1.2 The identification task

The identification task was administered to investigate the directions of the tone mergers attested in the discrimination task. Eight sets of target syllables contrasting in tone (thus 48 stimuli) were presented to the participants who were asked to identify the syllables upon hearing the stimuli by choosing one corresponding character. It was found that the identification accuracy of T6 was significantly lower than the other tones. The identification errors increased with age.

As for the directions of misidentification, T6 was more often identified as T4 but also as T3 by the senior group. On the contrary, the tone was more often identified as T3 but also as T4 by the middle-aged group. However, it was identified as T3 only by the young adults.

Taking the results of the discrimination and identification tasks together, it is manifest that the loss of contrasts between T6/T3 and T6/T4 was caused by the erosion of T6. T6 was merged to T4 and T3 in the older groups but was merged only to T3 in the young adult group.

2.3.2 The production task

In this task, eight sets of stimuli were embedded into two sentence carriers (thus 96 stimuli) and read aloud by our participants in order to investigate the production of the six Cantonese tones. The tone pairs that caused confusion in production were different from those in perception. In all age groups, the production accuracy of T6 and T5a were significantly lower than the other tones. In addition, the accuracy of T3 was also significantly low in the middle-aged group, whereas those of T2 and T3 were significantly low in the young adult group.

As for the direction of errors, T6 was mainly produced as T3 in all age groups. In the two younger groups, T3 was also found to be produced as T6. T5a was also found to be produced as T3 among all age groups, and T2 was found to be produced as T5 among the younger participants.

2.3.3 Dissociation between the perception and the production

As shown in the above, production was found to be disassociated from perception. Such asymmetry may be found at the individual level, that is, a participant may manage to maintain the difference between two tones in production, but fail in perception or vice versa. In some instances, the asymmetry may be found at the community level. For example, the production of T2 was difficult (third lowest correct rate among the six tones) and was readily confused with T5 by the young adults, who performed quite well in distinguishing these two tones in perception (above 95% correct). Another more interesting yet controversial example revealed in the present study is the T4/T6 pair. In the discrimination task, the older speakers failed to discriminate this tone pair in perception, but were able to distinguish it in production, exhibiting a near-merger phenomenon. The issue of near-merger will be discussed further in Chapter 4.

2.3.4 Types of mergers identified

Based on the findings of the discrimination and production tasks in the experimental study, the following types of (non)-mergers were identified, as summarized in Table 2.10. Three terms were adopted, “Good”, “Poor” and “Bad”. “Good” refers to the situation in which the score of a tone pair is not significantly low in an experimental task; “Poor” refers to the situation where the score of a tone pair is not significantly low but it is next to the lowest; “Bad” refers to the situation where the score of a tone pair is significantly lower than the others in an experimental task.

Table 2.10

Summary of the tone mergers identified in contemporary GZC

Tonal contrast	Senior group		Middle-aged group		Young adult group		Status
	Discrimination	Production	Discrimination	Production	Discrimination	Production	
T3/T6	Bad	Bad	Bad	Bad	Bad	Bad	Full-merger
T4/T6	Bad	Good	Bad	Good	Good	Good	Near-merger
T3/T5b	Good	Good	Good	Good	Good	Good	Well preserved
T3/T5a	Good	Bad	Good	Bad	Good	Bad	Alternate character readings
T2/T5	Good	Good	Good	Good	Good	Poor	Potential merger

2.3.4.1 T3/T6: Full-merger

All three age groups failed to distinguish the T3/T6 contrast in perception and their performance was concluded to be "bad", as shown in Table 2.10. The production performance of the three age groups was also "bad". Taken together, the phonemic contrast between T3 and T6 was neutralized in both perception and production, and these two tonal categories can be proposed as being fully merged in a sub-community of contemporary GZC.

The subsequent question would be the direction of the merger. In perception, the identification accuracy of T3 was much higher than T6, and the percentage of T6 identified as T3 was much higher than the reverse. It can be concluded that T6 was merged to T3 unidirectionally in perception.

In production, the correct rate of T3 was substantially higher than T6, and the percentage of T6 produced as T3 was much higher than the reverse. It can be concluded that T6 was merged to T3 unidirectionally in production. That being said, it was observed that the percentage of T3 produced as T6 increased with decreasing age. That is, the merger tended to be bidirectional among the younger participants.

2.3.4.2 T4/T6: Near-merger

T4/T6 exhibited as a case of near-merger in the middle-aged and the senior group, where speakers consistently distinguished these two tonal categories in production, yet failed to perceive the difference between them. For these speakers, the rate of T6 identified as T4 was much higher than the reverse. Thus, T6 was merged to T4 in perception.

Unlike their older counterparts, the young adult speakers managed to maintain the tonal contrast of T4/T6 in both production and perception.

2.3.4.3 T3/T5: Alternative character readings

The GZC speakers maintained the distinction of this tonal contrast in perception, while they consistently produced T5a as T3. Contrary to the situation of T5a, T5b was distinguished from T3 by the three age groups. Therefore, rather than a phonemic merger between these two tonal categories in production, it seems more plausible that the confusion between T3 and T5 is a case of alternate readings for characters.

2.3.4.4 T2/T5: Potential merger

This tonal contrast was well preserved among the GZC speakers in general, except that "poor" performance was concluded for the young adult speakers by virtue of the relatively low production accuracy of T2 (the third lowest), and T2 was most readily produced as T5 in the young adult group. It is highly plausible that the young adult speakers may have started to collapse this tonal contrast, first from production with T2 merged to T5.

Chapter 3

Acoustic Study

3.1 Introduction

In the experimental study, the tone pair T3/T6 was identified as a full-merger according to the results of the perception and production tasks. In the production task, the ratings and transcriptions of the speech samples were based merely on the perception of the raters. Intermediate forms for this tone pair were categorized binarily into either T3-like or T6-like to facilitate statistical analysis. Thus, the results for the production tasks alone do not provide us a full picture of the phonetic realization of the merged tones. The following questions remain unanswered. Is there one phonetic form or two variants of the newly evolved tonal category? What is the phonetic form? Is it a mid-level or low-level tone? What is the mechanism of the merger? In order to answer these questions, the data collected in the production task were followed up by an acoustic study with the following objectives: 1) To examine the phonetic realizations of the new level tone; and 2) To better understand the mechanism of the level tone merger in GZC.

3.2 Methods

3.2.1 Participants

The data collection for this acoustic study was based on the speech samples produced by 12 participants, two mergers and two non-mergers from each age group. The participants were selected based on their production and discrimination scores in the experimental study. The mergers were participants who had failed to maintain the T6/T3 contrast in both production and perceptions tasks, and the non-mergers were participants who had maintained all the six tonal contrast in both tasks. The backgrounds of the participants are displayed in Table 3.1.

Table 3.1

Backgrounds of the selected participants

				Percent correct in production			Percent correct in discrimination	
	Subject No.	age	sex	T3	T6	Mean	T3/T6 pair	Overall
Non-mergers								
Senior group	54	67	M	100	100	100	100	96
	34	58	F	100	100	100	100	97
Middle-aged group	70	35	M	100	100	100	100	99
	52	40	F	100	100	100	100	99
Young adult group	20	20	M	100	100	100	100	100
	16	21	F	100	100	100	100	100
mergers								
Senior group	68	56	M	75	18.8	46.9	62.5	84
	25	57	M	70.5	62.5	66.5	37.5	87
Middle-aged group	11	43	M	68.7	56.2	62.5	50	88
	29	43	F	56.2	37.5	46.8	62.5	89
Young adult group	69	24	M	75	68.7	71.8	50	85
	65	24	M	73.5	56.3	64.9	37.5	89

The selection criteria for the mergers were not very straightforward due to the dissociation between the perception and production that was commonly attested among the participants, as discussed in section 2.3.3. Since the acoustic study focused on the production side, the mean production score of T3 and T6 was taken as the primary indicator, the mean discrimination score of this tone pair as the secondary indicator, and the mean score of the overall performance in the discrimination task as the tertiary indicator. Consequently, the non-merger group was comprised of participants who scored 100% correct in the production and also 100% correct in the perception of the T3/T6 contrast. Their overall scores in discrimination were no less than 96%. The merger group was comprised of participants whose mean scores of the production of T3 and T6 were equal to or less than 75%. The scores of their perceptions of the T3/T6 contrast were less than 65% and their overall performance scores less than 90%.

3.2.2 Materials

Of the eight syllables (/fɐn/, /jɐn/, /fu/, /si/, /jɐn/, /ji/, /jyn/ and /wɐi/) in the experimental study, there were three types of syllable structures, CV, CVV and CVN. Three CV syllables were selected to control the composition of rimes. In the CV structure, the tone can be realized more consistently over the rime without any possible perturbation resulting from the changing of the vowel (e.g., in the CVV structure) or changing from vowel to nasal consonants (e.g., in the CVN structure). Furthermore, the adoption of the CV structure in studying the realization of tone has been a common practice in the field. Therefore, only speech samples of three CV roots (/si/, /fu/ and /ji/) were included in the acoustic study. In total, 432 target syllables (3 syllables x 2 carrier positions x 6 tonal contrasts x 12 participants) were analyzed.

3.2.3 Procedures

The pitch trajectory of the target syllable was analyzed using Praat. The start and end of the vocalic segment of the syllable were selected manually. The onset of the tone was marked by the start of vocalic modality, and the offset by the final glottal pulse. F0 (Hz) of the vocalic segment at 10 equidistant time points was measured via a Praat script. Manual measurements from the waveform were used when the pitch tracking was distorted by non-modal voice, such as creaky voice. The raw data had to be normalized to reduce individual variations. Among the various normalization methods, The T-value proposed by Shi (1986) was adopted in the present study for the following two reasons: 1) The resulting values from this formula correspond to the traditional five-level transcription system for Chinese tones devised by Chao (1947). Thus, the pitch value of the new level tone(s) can be compared straightforwardly with the canonical tones; 2) Unlike those of z-scores, the resulting values from this normalization method are positive, thus making subsequent statistical analysis easier. Therefore, F0 values were normalized by converting to T-value using the following formula (Shi, 1986),

$$T = [(lg\ P_i - lg\ P_{min})/(lg\ P_{max} - lg\ P_{min})] \times 5$$

where P_i was the F0 value at each time point i, P_{max} and P_{min} referred to the maximum and minimum F0 value across six tones of one target syllable.

The mean T-values of the ten time points for each syllable were averaged for each speaker and submitted to a test of variance. The realizations of the target tones were examined by the following three parameters: The F0 Height; The F0 Dispersion; and The F0 Contour. The F0 height was the mean of the F0 in T-value for each target tone. The F0 dispersion was a measure of the distance of the two level tones by computing the difference of the F0 heights of T3 and T6 in T-value. The F0 Contour was the graphical representation of the trajectories of the target tones. The values for the first two numerical parameters were averaged for each participant and then for each age group. In order to capture closely the difference between the tones produced by the non-mergers and mergers of each age group, the F0 Height of the initial, medial and final portion of the vocalic segment was computed for each tone, which referred to the mean F0 in the

T-value for the first two (0% and 10%), the middle two (40% and 50%) and the last two (90% and 100%) time points respectively.

3.3 Results

First off, the productions of the non-mergers will be presented. The production of T3 and T6 was compared within each age group, then across the three age groups using T-tests. Second, the productions of the mergers will be shown. Again, a comparison was made of the production of T3 and T6 in each age group, then across age groups. Finally, comparisons of the productions between the non-mergers and the mergers in each age group will be presented.

3.3.1 The profile of the canonical tones produced by the non-mergers

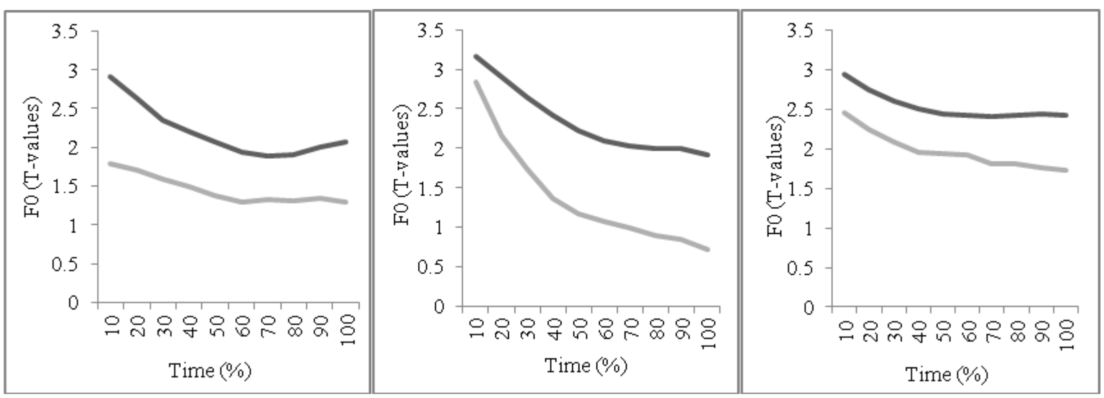
Table 3.2 displays the F0 Height and F0 Dispersion for the non-mergers. Within each age group, the F0 Height of T3 was significantly different from that of T6 ($p < .05$ for all). That is to say the contrast of T3 and T6 was well maintained among this group of participants, which was also confirmed by the F0 Contours as shown in Figure 3.1(a) – (c).

Across the three age groups, no significant difference was found for the two numerical parameters for the two tones. However, it should be noted that the dispersion between T3 and T6 was decreasing, implying that the two tones were becoming more and more similar even among the participants who still maintained their contrast.

Table 3.2

The F0 Height and F0 Dispersion for the non-mergers of each age group. Asterisks indicate significance

Parameter / Age group	F0 Height (T-value)		F0 Dispersion (T-value)
	T3	T6	
Senior	2.7*	1.6	1.1
Middle-aged	2.4*	1.4	1
Young adult	2.4*	1.9	0.5



(a) (b) (c)

Figure 3.1 (a) – (c). The F0 contours of T3 (black line) and T6 (grey line) for the non-mergers of the senior (a), the middle-aged (b) and the young adult group (c).

3.3.2 The profile of the new level tone produced by the mergers

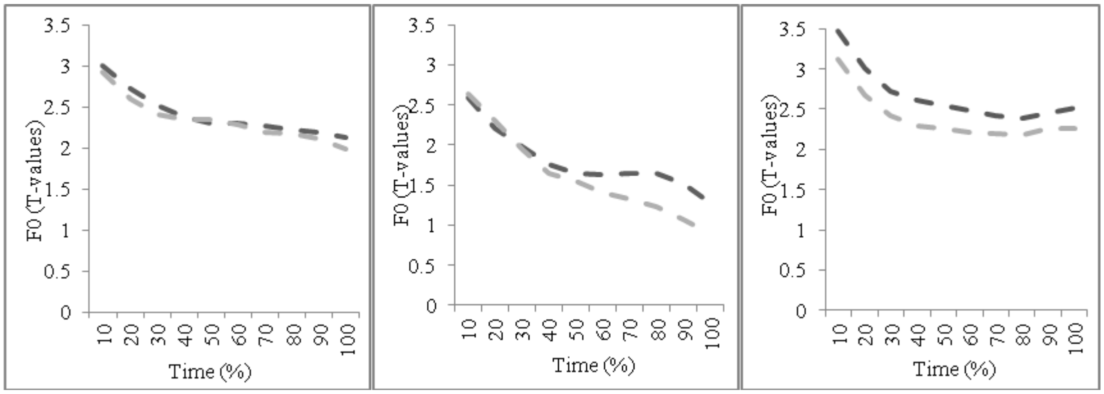
Table 3.3 displays the F0 height and F0 dispersion for the mergers. For each age group, no significant difference was found between the F0 height for T3 and T6. No significant difference of F0 Dispersion across groups was found either. In other words, the new level tone was realized as one phonetic contour among the mergers of all age groups. This is also confirmed by the F0 Contours shown in Figure 3.2 (a) – (c).

The F0 Height of the merged tone produced by each age group was computed, as shown in Table 3.3. Note that the F0 height of the new level tone produced by the middle-age mergers was significantly lower than those of the other two groups (p < .017 for both, significance threshold corrected for multiple comparisons equals .017).

Table 3.3

The F0 Height and F0 Dispersion for the mergers of the three age groups, as well as the F0 Height for the merged tone. Asterisks indicate significance

Parameter / Age group	F0 Height (T-value)			F0 Dispersion (T-value)
	T3	T6	The merged tone	
Senior	2.5	2.3	2.4	0.2
Middle-aged	1.8	1.6	1.7*	0.2
Young adult	2.6	2.3	2.5	0.3



(a) (b) (c)

Figure 3.2 (a) – (c). The F0 contours of T3 (black line) and T6 (grey line) for the mergers of the senior (a), the middle-aged (b) and the young adult group (c).

3.3.3 Comparing the new level tone with the canonical tones

3.3.3.1 The senior group

The F0 Height of the initial, medial and final portions of the tones produced by the mergers and the non-mergers are illustrated in Table 3.4. The F0 Height of all the portions of T3 produced by the non-mergers and mergers were comparable ($p > .05$ for all). On the contrary, the F0 Height of all three portions of T6 produced by the mergers were significantly higher than those of the non-mergers (for the initial and medial portion, $p < .01$; for the final portion, $p < .05$). The findings suggested that among the senior mergers, the trajectories of T3 remained relatively intact, whereas those of T6 had undergone an upward movement and approximated those of T3. This was also confirmed by the result that there was no significant difference found for the T6 produced by the mergers between the T3 produced by the non-mergers ($p > .05$ for all three portions).

Table 3.4

The F0 Height of the initial, medial and final portion of T3 and T6 for the non-mergers and mergers of the senior group. Asterisks indicate significance

Parameter / Group	F0 Height of T3			F0 Height of T6		
	initial	medial	final	initial	medial	final
Non-merger	2.8	2.0	2.0	1.7	1.3	1.3
Merger	2.6	2.2	2.2	2.5*	2.2*	2.0*

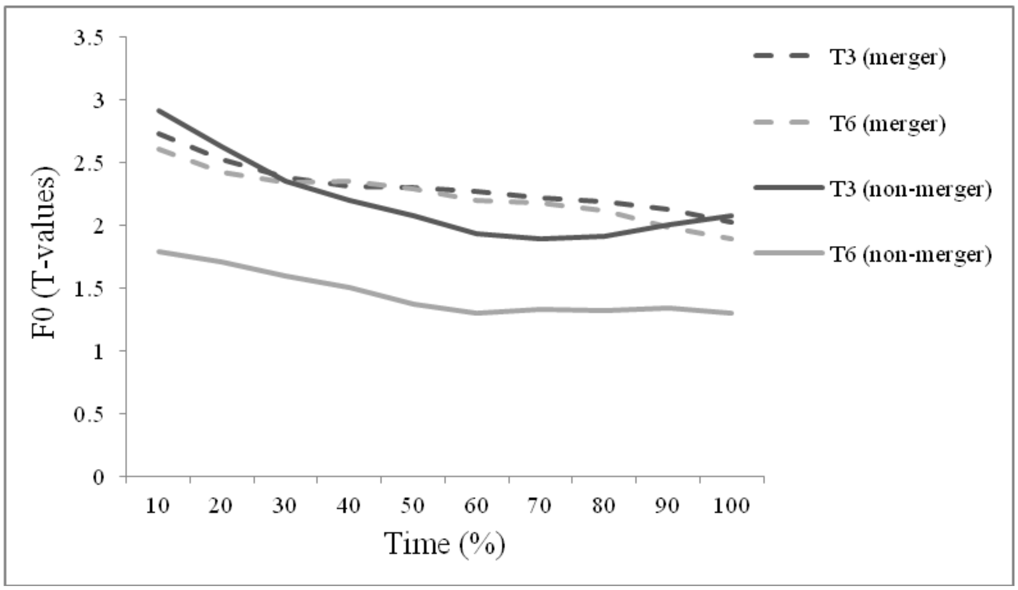


Figure 3.3. F0 contours of the T3 and T6 produced by the non-mergers and mergers of the senior group.

3.3.3.2 The middle-aged group

Table 3.5 shows the F0 Height of the initial, medial and final portions of the tones produced by the mergers and non-mergers in the middle-aged group. Differing from the results for the senior group, the F0 Height of the three portions of T3 produced by the mergers were significantly lower than those of the non-mergers ($p < .05$ for all). Although no significant difference was found for the productions of T6 between non-mergers and mergers ($p > .05$ for all), the F0 Height of the medial and final portions were higher than those of the non-merger counterparts, as illustrated in Figure 3.4. The final portion of T6 produced by non-mergers was significantly lower than that of the T3 produced by the mergers ($p < .05$), while marginal significance was found for the medial portion ($p = .054$). This finding suggested an intermediate form for the new level tone with T3 moving down and T6 moving up.

Table 3.5

The F0 Height of the initial, medial and final portion of T3 and T6 for the non-mergers and mergers of the middle-aged group. Asterisks indicate significance

Parameter / Group	F0 Height of T3			F0 Height of T6		
	initial	medial	final	initial	medial	final
Non-merger	3.0	2.2	2.0	2.4	1.1	0.7*
Merger	2.4*	1.6*	1.4*	2.5	1.4	1.0

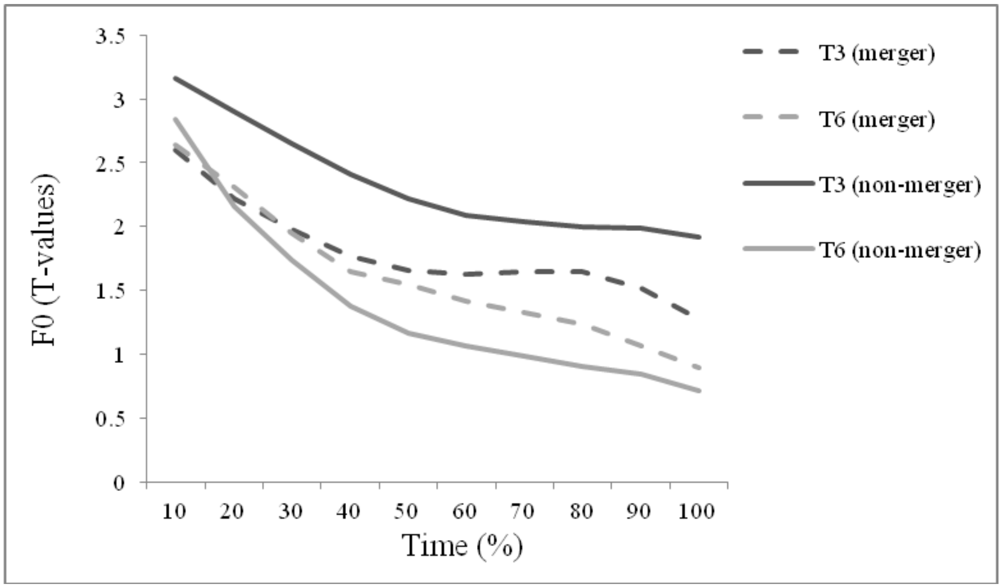


Figure 3.4. F0 contours of the T3 and T6 produced by the non-mergers and mergers of the middle-aged group.

3.3.3.3 The young adult group

The F0 Height of the initial, medial and final portions of the tones produced by the mergers and non-mergers are illustrated in Table 3.6. F0 Height of all three portions of T3 produced by the non-mergers and mergers were comparable ($p > .05$). As for T6, the F0 Height of the initial and final portions produced by the mergers were significantly higher than those of the non-mergers ($p < .05$ for both). The findings suggested that, among young adult mergers, the trajectories of T3 remained relatively intact, whereas those of T6 had moved upward to those of T3. This is also corroborated by the result that no significant difference was found between the productions of T6 by the mergers and the productions of T3 by the non-mergers ($p > .05$). Note that the F0 Height of the three portions of T3 produced by the mergers in this group were slightly higher than those of T3 by the non-mergers. This suggested that, while T6 moved up, T3 also moved up its trajectory.

Table 3.6

The F0 Height of the initial, medial and final portion of T3 and T6 for the non-mergers and mergers of the young adult group. Asterisks indicate significance

Parameter / Group	F0 Height of T3			F0 Height of T6		
	initial	medial	final	initial	medial	final
Non-merger	2.8	2.4	2.4	2.4	1.9	1.7
Merger	3.2	2.5	2.5	2.9*	2.2	2.2*

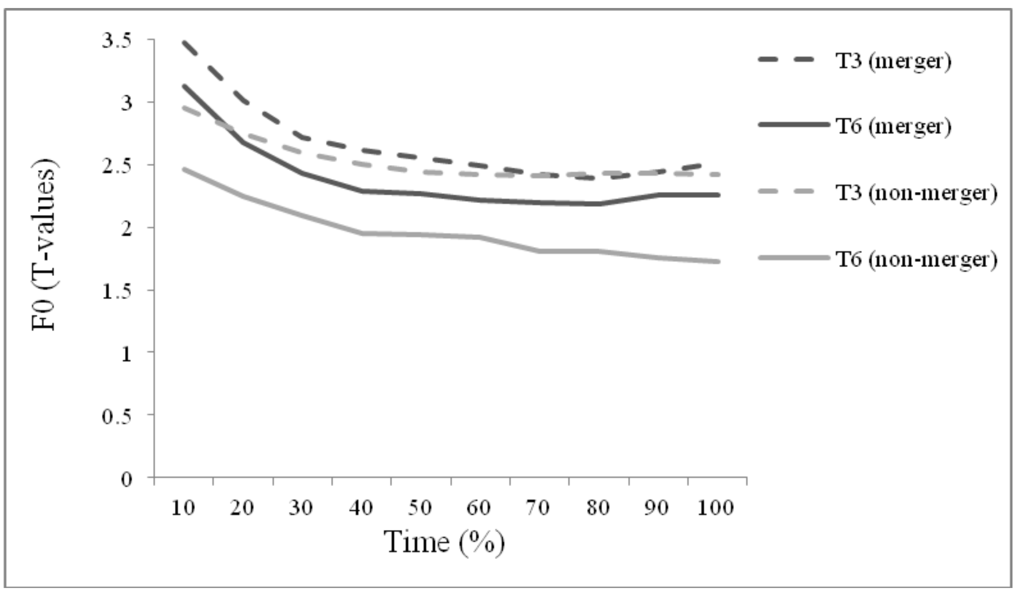


Figure 3.5. F0 contours of the T3 and T6 produced by the non-mergers and mergers of the young adult group.

3.4 Summary and discussion

This acoustic study examined the phonetic realizations of T3 and T6 produced by the mergers and the non-mergers of the three age groups. The results showed that the newly evolved level tone was realized as one phonetic contour among the mergers. As for the phonetic value of the realization, this differed across age groups. In the senior group, the new level tone took up the pitch height of the canonical T3. In the middle-aged group, it was an intermediate of the canonical T3 and T6. For the young adults, the F0 height of the new level tone was even higher than the canonical T3. In short, the new level tone of each age group was realized as a mid-level tone, with pitch value similar to that of the canonical T3.

The difference in the movements of tonal trajectories implies different mechanisms of tone merging adopted by each age group. In both the senior and young adult group, trajectories of T6 transferred upward to those of T3, suggesting that both groups adopt the merger-by-transfer mechanism in production. Nonetheless, the way in which this mechanism was implemented by these two age groups may have been somewhat different. In the senior group T3 remained relatively intact with T6 transferring up to T3. In the young adult group, T3 moved up, with T6 also shifting up to approximate T3. Differing from the above two age groups, the results from the middle-aged mergers indicated that the trajectory of T3 moved down while that of T6 moved up and they approximated each other. T3 retained its initial point while reducing the later vocalic segment, whereas T6 retained the later vocalic segment while increasing the initial point, leading to an innovative category that showed a mean F0 value in between those of the two canonical tones. A different mechanism, merger-by-approximation, is proposed to be adopted by the middle-aged mergers. The merged tone results from the gradual approximation of the two level tones until they are non-distinct. Implications drawn from the different tone merger mechanisms will be discussed further in Chapter 4.

Chapter 4

General Discussion and Conclusion

In this chapter, the main findings of the three experimental tasks and the acoustic study will be summarized to reveal the whole picture of the tone merger phenomenon in contemporary GZC. Implications drawn from the tone mergers identified in GZC will then be discussed. Furthermore, the tone merger in HKC will be presented, followed by a comparison between the two varieties of Cantonese. In the last section, the concluding remarks for this study will be provided.

4.1 The overall picture in GZC

4.1.1 The mergers identified

In this study, the tonal variations and mergers in both perception and production were investigated using three experimental tasks and an acoustic study. The main findings of the discrimination task, identification task, production task and the acoustic study, are summarized in Table 4.1.

Based on these main findings, three types of merger have been identified: 1) T6/T3, a full-merger; 2) T6/T4, a near-merger; 3) T2/T5, a potential merger. The confusion between T3 and T5 is proposed as a case of alternative character readings.

Table 4.1

Summary of the main findings of the experimental and acoustic study

	Discrimination Task	Identification Task	Production Task	Acoustic Study
Senior group	T6/T4 T6/T3	T6 as T4 T6 as T3	T6 as T3 T5a as T3	T6 moves up to T3
Middle-aged group	T6/T3 T6/T4	T6 as T3 T6 as T4	T6 as T3 T5a as T3 T3 as T6	Intermediate between T3/T6
Young adult group	T6/T3	T6 as T3	T6 as T3 T5a as T3 T2 as T5 T3 as T6	T6 moves up to T3; T3 also moves up

4.1.2 The full-merger: T3/T6

4.1.2.1 The phonetic realization of the new level tone

The findings of the acoustic study demonstrated that the two level tones had been completely merged into one tonal category with one phonetic form among some GZC speakers from all three of the age groups. This is a mid-register tone with pitch value approximately equal to that of T3. In other words, the new level tone adopts the phonetic feature of the canonical T3.

4.1.2.2 The apparent time hypothesis

Adopting the apparent time hypothesis that the distribution of linguistic variables across age levels is indeed a reflection of change in progress, it could be inferred that the linguistic variations attested among the senior generation reflects the earlier stage of a sound change; that of the middle-aged speakers the intermediate stage, and that of the young adult speakers the current stage. Hence, the diachrony of the tone change can thus be constructed based on the synchronic tonal variations exhibited from the older to the younger generation.

4.1.2.3 The mechanism of the full-merger

As introduced in section 1.6, Labov (1994, p. 321) proposed the following three major mechanisms of sound merger: merger by transfer; merger by approximation; and merger by expansion. The findings of the acoustic study have suggested that this level tone merger was a unidirectional process among the seniors. It resembled the merger-by-transfer model in that the tokens of T6 took up the phonetic form of T3 and gradually transferred to that category whereas the phonetic realization of T3 remained intact in this process.

Among the middle-aged, the level tone merger was no longer a unidirectional process but a bidirectional one. It corresponded to the merger-by-approximation model in the sense that both the former level tones had moved away from their canonical phonetic forms and approximated each other's phonetic forms, resulting in one pitch contour with the mean value of the two canonical level tones.

As for the young-adults, the picture is a bit complicated. The production tasks revealed the co-existence of two merging routes: one group of speakers adopted the unidirectional mode in which T6 was transferred to T3; the other group of speakers adopted the reverse direction in which T3 was transferred to T6. Some speakers even adopted the bidirectional mode, in which some tokens of T3 were realized as T6 and vice versa. Thus, the merger was a bidirectional merger-by-expansion mode. However, the acoustic analysis failed to capture this phenomenon since the two subjects who scored the lowest in the production and discrimination tasks happened to have their T6 transferred to T3.

Different mechanisms of merger indeed operate at different rates. As proposed by Labov (1994, p. 323), merger by expansion is the fastest, followed by merger by approximation then meager by transfer. In GZC, the mechanism of T3/T6 merger has changed from merger by transfer, to merger by approximation, and then the merger by expansion. Obviously, the merger is accelerating in the community.

4.1.3 The near-merger: T4/T6

Since its introduction (Labov, Yaeger, & Steiner, 1972, Chapter 6), the notion ‘near-merger’ has been highly controversial. As Labov (1994, p. 368) remarked, ‘how does a speaker learn to articulate members of one phoneme in one way, and members of the other phoneme in another way, provided that he/she consistently fails to perceive the difference between these two phonemes?’ However, apart from the early studies of vowel near-mergers (e.g., Di Paolo, 1988; Trudgill, 1974; Milroy & Harris, 1980), the T4/T6 near-merger in Cantonese has provided another piece of empirical evidence to support this highly controversial phenomenon in sound change.

To answer why the near-merger is possible, Labov (1994, p. 391) has proposed “suspension of phonemic contrast” as a possible explanation. He suggested that a reduced role played by the semantic contrast of two sounds will result in the lack of attention or awareness to this distinction of the stimuli presented to those who even manage to differentiate clearly the two sounds in their own production. The reduced role of semantic contrast is gradually emerging due to the extensive contacts between speakers who make a phonemic distinction with those who do not. During this process, the two-phoneme speakers may often misunderstand the words by the one-phoneme speakers, since the allophonic variation is interpreted as a phonemic distinction by them. Gradually, the two-phoneme speakers may cease to give attention to this phonemic distinction themselves because of its unreliability, but they may continue to produce the distinction. This is probably where the near-merger will emerge. In view of the fact that some portions of speakers collapse this distinction but some others maintain it, near-merger – a counterintuitive merger though – is indeed a mere transitional state in the process of sound change. As this process continues, mergers exhibited in the speech of one-phoneme speakers may rise to the awareness of two-phoneme speakers, which eventually leads the two-phoneme speakers to abandon the distinction entirely. However, in the current study, the T4/T6 near-merger had ceased and the contrast became well maintained again in both production and perception among the young adults in GZC.

4.1.3.1 Two competitive mergers: T3/T6 vs. T4/T6

The cessation of the T4/T6 near-merger in the young adult group may be a result of the inception of another sound change – the T3/T6 merger. As remarked in section 2.3.1.2, the senior and middle-aged groups differed in the extent to which T6 was confused with T3 or T4. Specifically, T6 was confused more readily with T4 than with T3 by the older generation, whereas it was confused more readily with T3 than with T4 by the middle-aged generation. Taken together, it is highly likely that T6 started to merge with T4 in the earlier stage, but that some tokens of T6 started to merge with T3 in the later stage. Gradually the T4/T6 merger slowed down its process, further making way for the T6/T3 merger. The T3/T6 merger accelerated its way through by changing from a unidirectional to a bidirectional mechanism. As a result, the T4/T6 merger was suspended.

It is evident that the level-tone contrasts are not well distinguished by GZC listeners. Labov (1994, p. 328) proposed two potential explanations for sound merger. The first one was concerned with the discriminability of the phonetic feature. To be specific, the higher the degree of discriminability of the feature on which the contrast depends, the less likely it is that the contrast will diminish. The other explanation was concerned with the number of distinctions; that is, the larger the number of contrasts already made along one phonetic dimension, the more likely it is that some of them will tend to collapse.

In Cantonese, there are four level tones: T1, T3, T4 and T6. This rich level-tone contrast is, in fact, relatively rare among the tone languages in the world (Yip, 2002, p. 26). In the literature about Cantonese tone perception, the relative F0 level, the direction of F0 change and the magnitude of F0 change have been identified as the important perceptual correlates (e.g., Fok, 1974; Vance, 1977; Gandour, 1981; Khouw and Ciocca, 2007). Applying the three dimensions in analyzing the contrasts among the four level tones, it is evident that the four level tones differ from each other in the F0 level alone. Thus, widespread confusions between T3/T6 and between T4/T6 are understandable in view of the degree of F0 level difference between each tonal contrast.

The sequent question is: Why does the T4/T6 merger yield to the T3/T6 merger? The two competitive mergers may be viewed in the light of the history of Chinese language. To account for the evolution of tones, Lien (1986) identified two modes of tone mergers: horizontal tone merger and vertical tone merger in Chinese dialects. Horizontal merger refers to the coalescence among the four traditional tonal categories *Ping, Shang, Qu, Ru*. Vertical merger refers to the coalescence between the *Yin* and *Yang* register within the same traditional tonal category. As described in Section 1.5.1, after the splitting of the four tonal categories into eight in Middle Chinese following the devoicing of the consonant onsets, many dialects have merged some of the eight categories. Among the Chinese dialects, merging the *Yin Shang, Yang Shang* tone in the vertical dimension has been the prevalent pattern among dialects. In his survey of 480 northern Chinese dialects, Lien (1986) observed that the *Qu* tone was a target for tone merger, which has never been found to be merged with other tonal categories horizontally. Turning back to the tone merger in GZC, the potential merger of T6 (*Yang Qu* category) into T4 (*Yang Ping* category) occurring in the earlier stage would be highly uncommon in Chinese dialects. On the contrary, the merging of T6 (*Yin Qu*) and T3 (*Yang Qu*) vertically is commonly found in many dialects. The "win out" of the T3/T6 merger in this competition may probably be due to the influence of other dialects.

4.1.4 The alternative character readings: T3/T5

The contrast is well maintained in perception (both discrimination and identification), while shaky in production of the T5a group only. This finding corroborates our hypothesis concerning the influence of Mandarin on Cantonese. The four syllables in the T5a group (*Yang Shang* tone) are all pronounced as T4 (*Qu* tone) in

Mandarin, while those in the T5b group are pronounced as T3 (*Shang* tone). The Cantonese reading of a Chinese character by GZC speakers who have strong Mandarin tie may be adversely affected by the Mandarin reading of that character. The finding that the senior group in this study confused T5a words with T3 notably less than did their younger counterparts also lends support for our hypothesis. As mentioned in Section 1.2, although Cantonese is used extensively as the local dialect in GZ, it is not the medium of education in the city. Putonghua (Standard Mandarin) has been the only spoken language that is government-sanctioned for use in the school system since the late 1980s. As a result, the older GZC speakers would generally only have a passive knowledge of Putonghua (mainly gained through social contact or the mass media) and the cross-dialect tonal correspondence of Putonghua to Cantonese would be barely established in their linguistic system. Therefore, the older speakers would confuse T5a words as T3 less than do the younger speakers, as observed in the present study. However, there are two questions remaining unanswered if the influence of Mandarin is a cause of this phenomenon. The first one is concerned with the target of the merger: why T5a merges with T3 (*Yin Qu*), but not T6 (*Yang Qu*). The second question is concerned with the extent of the Mandarin influence: why the alterative readings of T5a syllables were only observed in the production task, but not in the identification task in which lexical information was also provided. Further investigation using larger data sets is needed in order for these queries to be answered.

4.2 Comparison between HKC and GZC

4.2.1 Tone merger in HKC

As described in Section 1.4.1, it has been found that the T3/T6, T2/T5 and T4/T6 pairs are in the process of merging in HKC. Furthermore, the alternative character readings of T5/T3 have also been reported. Apparently, the two varieties of Cantonese exhibit a similar pattern of tone merger. However, the rate of each merger is in a non-parallel manner.

4.2.1.1 T3/T6

According to the HKC findings reported by Fung and Wong (2010a), this tonal contrast is well maintained in perception among the HKC speakers, even though they fail to produce the differences between these two tonal categories. Therefore, T3/T6 is a possible merger in HKC but a full-merger in GZC. The GZC variety is more advanced in tone change in reference to this level tone merger.

4.2.1.2 T2/T5

The tone pair T2/T5 has been confirmed to be merged in a sub-community of HKC. However, it is just a possible merger in GZC. The HKC variety is more advanced in tone change with reference to this rising tone merger.

As discussed in Section 4.1.3, the merging of *Yin Shang* and *Yang Shang* (i.e., T2 and T5 in Cantonese) is the major trend in contemporary Chinese dialects. It can be hypothesized that the coalescence of the *Shang* category should precede that of the *Qu* category in Mandarin. The hypothesis is based on the observation that dialects which preserve the *Yin-Yang* contrast in the *Qu* category but collapse that of the *Shang* category are much more commonly found than the reverse. GZC seems to be quite uncommon in that the coalescence of the *Qu* tone precedes that of the *Shang* tone. On the contrary, the tone merger pattern in HKC conforms to the norm of tone evolution observed in other dialects. It has been widely assumed that the simplification of the tonal system in Cantonese has stemmed from the strong influence of Mandarin. However, our findings may refute such an assumption, since GZC is more heavily influenced by Mandarin as compared with HKC. In sum, the influence of Mandarin on tone merger in Cantonese may not be a possible motivation of the merger.

4.2.1.3 T4/T6

T4/T6 is an on-going near-merger in HKC but has ceased as a near-merger in GZC. As discussed in Section 4.1.3.2, the T6/T4 near-merger may be intercepted by the T6/T3 merger. However, the T3/T6 is an emergent sound change in HKC and has not exerted strong interference on the T6/T4 merger. Therefore, they may explain why this near-merger is still on-going in the HKC variety.

4.2.1.4 T3/T5

Alternative character readings between T3/T5 have been observed in both varieties. However they differed in the lexical items involved as well as the direction of the tone change. In GZC, the lexical items of T5a in the present study: ‘奮 /fɐn/’, ‘誘 /jɐu/’, ‘婦 /fu/’, and ‘市 /si/’ were pronounced as T3. Note that, ‘婦 /fu/’, and ‘市 /si/’ were confused readily as T2 in HKC (Fung and Wong, 2010a). According to the work on HKC (e.g., Cheng, 2002, 2003), lexical items such as ‘臼 /k^hɐu/’, ‘舅 /k^hɐu/’, ‘似

/ts^{h}i/' and '鱔 /sin/'of T5, were pronounced readily as T3. Therefore, it is highly possible that this tonal confusion has carried out and diffused through different lexical items in the two Cantonese varieties. As for the direction of this change, only T5 to T3 was attested in GZC in the present study. However, the change was bi-directional in HKC, T5 to T3 and vice versa. According to Cheng (2003), the direction T3 to T5 is more commonly found than the reverse in the present stage of HKC. In GZC, this tonal confusion is probably linked to speakers' knowledge of the Putonghua pronunciation, although the possible cause of this tonal confusion in HKC or Cantonese as a whole still remains unclear. This issue warrants further investigation.

In both of the Cantonese varieties, there seems to be a bipartition of the T5 syllables during this tone change, with part of the T5 syllables confused with T3 and the other T5 syllables confused with T2. For instance, the '婦 /fu/', '市 /si/', '社 /se/', '耳 /ji/', '也 /ja/' are confused with T2 in HKC (e.g., Yiu, 2009; Fung and Wong, 2010a). In GZC, there is a tendency for syllables in T5b: '引 /jɐn/', '耳 /ji/', '軟 /jyn/', and '偉 /wɐi/' to be pronounced as T2. This bipartition may eventually lead to the collapse of T5, as is observed during the historical tone development among many other Chinese dialects.

4.3 Conclusion

Cantonese is known for the richness of its tonal inventory. However, the tones in this complex system are in the process of merging. Most of the previous studies focused on HKC, whereas few studies concentrated on GZC, which is also a major variety of Cantonese. The current study was the first comprehensive investigation of the tone merger phenomenon in GZC. The findings of the three experimental tasks and an acoustic study provide us supporting evidence that the process of tone merging is underway in GZC. Drawing from the present findings, it is proposed that the two level tones: T3 and T6, as well as the two rising tones: T2 and T5 may eventually merge into one level and one rising tone respectively, as illustrated in Table 4.2. It follows that after the completion of the tone mergers, the tonal inventory of GZC would bear strong resemblance to that of Mandarin, as schematized in Table 4.3.

Table 4.2

Schematization of the future development of the tonal system in GZC

Upper Register	T1	T2 ⇕	T3 ⇕
Lower Register	T4	T5	T6

Table 4.3

Schematization of the GZC tonal system after completion of the tone merger

<table>
<tr><td>Upper Register</td><td>T1</td><td rowspan="2">T2/T5</td><td rowspan="2">T3/T6</td></tr>
<tr><td>Lower Register</td><td>T4</td></tr>
</table>

Appendix I Questionnaire for personal and family linguistic backgrounds

Name：_______________ Gender: ☐ Male ☐ Female

Year of Birth：_________

Education Level: ☐ Primary ☐ Secondary ☐ Junior High ☐ Senior High

☐ Diploma ☐ Bachelor ☐ Master ☐ PhD

1. Your current residential district: ☐ Dongshan ☐ Yuexiu ☐ Haizhu ☐ Liwan

Others: __________

2. District where you grew up: ☐ Dongshan ☐ Yuexiu ☐ Haizhu ☐ Liwan

Others: __________

3. Other than Guangzhou, have you lived in other places for more than one year?

☐ No ☐ Yes, __________

4. Other than Cantonese, do you speak other dialects?

☐ No ☐ Yes, __________ (such as Hakka, Chaozhou)

5. Mother tongues of your parents

Father: ☐ Cantonese ☐ Mandarin ☐ Hakka ☐ Chaozhou other _________

Mother: ☐ Cantonese ☐ Mandarin ☐ Hakka ☐ Chaozhou other _________

Appendix II Demographic information of the participants

Subject No.	Age	Gender	Education Level	Current residential district	District where the participant grew up	Other Yue dialects spoken	Mother tongue of mothers	Mother tongue of fathers
1	35	M	6	YX	DS	N/A	Cantonese	Cantonese
2	36	F	7	Others	DS	N/A	Cantonese	Cantonese
3	35	M	7	YX	YX	N/A	Cantonese	Cantonese
4	23	F	6	DS	DS	N/A	Cantonese	Cantonese
5	35	M	6	LW	YX	N/A	Cantonese	Cantonese
6	40	M	6	HZ	YX	N/A	Cantonese	Cantonese
7	44	F	6	YX	YX	N/A	Cantonese	Cantonese
8	24	F	6	LW	LW	N/A	Cantonese	Cantonese
9	41	M	2	DS	DS	N/A	Cantonese	Cantonese
10	25	M	2	Others	DS	N/A	Cantonese	Cantonese
11	43	M	7	LW	DS	N/A	Cantonese	Cantonese
12	22	F	6	LW	LW	N/A	Cantonese	Cantonese
13	22	F	6	YX	YX	N/A	Cantonese	Cantonese
14	22	M	6	LW	LW	N/A	Cantonese	Cantonese
15	21	M	6	YX	YX	N/A	Cantonese	Cantonese
16	21	F	6	DS	DS	N/A	Cantonese	Cantonese
17	21	F	6	Others	DS	N/A	Cantonese	Cantonese
18	20	F	6	Others	DS	N/A	Cantonese	Cantonese
19	21	M	6	HZ	HZ	N/A	Cantonese	Cantonese
20	20	M	6	LW	LW	Hakka	Hakka	Cantonese
21	22	F	6	Others	LW	N/A	Cantonese	Cantonese
22	20	F	6	Others	LW	N/A	Cantonese	Cantonese
23	22	F	6	Others	LW	N/A	Cantonese	Cantonese
24	24	M	7	HZ	DS	N/A	Cantonese	Cantonese
25	57	M	3	HZ	HZ	N/A	Cantonese	Cantonese
26	55	F	4	HZ	HZ	N/A	Cantonese	Cantonese
27	42	M	4	HZ	HZ	N/A	Cantonese	Cantonese
28	56	F	1	HZ	HZ	N/A	Cantonese	Cantonese
29	43	F	3	HZ	HZ	N/A	Cantonese	Cantonese
30	44	F	4	HZ	HZ	N/A	Cantonese	Cantonese
31	42	F	4	HZ	DS	N/A	Cantonese	Cantonese
32	24	M	7	HZ	HZ	N/A	Cantonese	Cantonese
33	36	F	6	HZ	HZ	N/A	Cantonese	Cantonese
34	58	F	3	HZ	LW	N/A	Cantonese	Cantonese
35	37	F	6	HZ	LW	N/A	Cantonese	Cantonese
36	58	F	4	HZ	LW	N/A	Cantonese	Cantonese
37	56	F	4	HZ	LW	N/A	Cantonese	Cantonese
38	23	M	6	HZ	HZ	N/A	Cantonese	Cantonese
39	21	M	5	HZ	HZ	N/A	Cantonese	Cantonese
40	38	M	6	LW	LW	N/A	Cantonese	Cantonese
41	25	M	6	HZ	HZ	N/A	Cantonese	Cantonese

42	36	F	5	HZ	HZ	N/A	Cantonese	Cantonese
43	23	M	5	HZ	HZ	N/A	Cantonese	Cantonese
44	22	F	5	LW	LW	N/A	Cantonese	Cantonese
45	24	F	6	LW	LW	N/A	Cantonese	Cantonese
46	56	F	4	HZ	LW	N/A	Cantonese	Cantonese
47	36	F	6	HZ	HZ	N/A	Cantonese	Cantonese
48	36	F	5	DS	DS	N/A	Cantonese	Cantonese
49	55	F	4	LW	LW	N/A	Cantonese	Cantonese
50	58	M	4	LW	LW	N/A	Cantonese	Cantonese
51	40	M	6	LW	YX	N/A	Cantonese	Cantonese
52	36	F	6	LW	HZ	N/A	Cantonese	Cantonese
53	40	M	4	LW	DS	N/A	Cantonese	Cantonese
54	67	M	3	YX	YX	N/A	Cantonese	Cantonese
55	61	F	1	LW	HZ	N/A	Cantonese	Cantonese
56	41	M	4	LW	YX	N/A	Cantonese	Cantonese
57	54	F	4	LW	YX	N/A	Cantonese	Cantonese
58	54	M	3	LW	LW	N/A	Cantonese	Cantonese
59	24	F	6	LW	LW	N/A	Cantonese	Cantonese
60	24	F	6	LW	LW	N/A	Cantonese	Cantonese
61	24	F	5	LW	LW	N/A	Cantonese	Cantonese
62	37	F	5	DS	DS	N/A	Cantonese	Cantonese
63	25	F	6	Others	DS	Hakka	Cantonese	Hakka
64	39	M	7	Others	YX	N/A	Cantonese	Cantonese
65	24	M	6	Others	YX	N/A	Cantonese	Cantonese
66	36	F	2	LW	LW	N/A	Cantonese	Cantonese
67	52	F	5	HZ	HZ	N/A	Cantonese	Cantonese
68	56	M	6	HZ	LW	N/A	Cantonese	Cantonese
69	24	M	7	HZ	DS	N/A	Cantonese	Cantonese
70	35	M	7	Others	HZ	N/A	Cantonese	Cantonese
71	23	F	6	HZ	HZ	N/A	Cantonese	Cantonese
72	44	M	3	HZ	HZ	N/A	Cantonese	Cantonese
73	43	M	3	DS	DS	N/A	Cantonese	Cantonese
74	45	M	4	YX	YX	N/A	Cantonese	Cantonese
75	43	M	4	LW	LW	N/A	Cantonese	Cantonese

Note. M = Male; F = Female; Education level: 1= Primary school; 2 = Secondary school; 3 = Junior High; 4 = Senior High; 5 = Diploma; 6 = Bachelor; 7 = Master; District: DS = *Dongshan*; LW = *Liwan*; HZ = *Haizhu*; YX = *Yuexiu*.

References

Bauer, R. S., & Benedict, P. K. (1997). *Modern Cantonese phonology*. Berlin: Mouton De Gruyter.

Bauer, R.S., Cheung, K. H., & Cheung, P. M. (2003). Variation and merger of the rising tones in Hong Kong Cantonese. *Language Variation and Change,* 15, 211-225.

Chao, Y. R. (1947). *Cantonese primer*. New York: Greenwood Press.

Chen, M., & Newman, J. (1984). From Middle Chinese to modern Cantonese (Part 1). *Journal of Chinese Linguistics*, 12(1), 148-198.

Cheng, S. K. (2002). *Yangshang – Yinqu alternation in modern Cantonese: irregularity in pronunciation or diffusion through lexicon?* Paper presented in the Annual Research Forum of LSHK, Hong Kong University of Science and Technology.

Cheng, S. K. (2003). *3a-2b tonal alternations: an on-going sound change in Hong Kong Cantonese*. Paper presented to Lelab members, City University of Hong Kong.

Cheung, K. H. (1986). *The phonology of present day Cantonese* (Doctoral dissertation). University of London.

Di Paolo, M. (1988). Pronunciation and categorization in sound change. In K. Ferrara, B. Brown, K. Walters., & J. Baugh (Eds.), *Linguistic change and contact: Proceedings of the 16th Annual Conference on New Ways of Analyzing Variation in Language* (pp. 84-92). Austin: University of Texas.

Fok-Chan, Y. Y. (1974). *A perceptual study of tones in Cantonese*. Hong Kong: University of Hong Kong, Centre of Asian Studies.

Fung, R., & Wong, C. (2010a). *Mergers and near-mergers in Hong Kong Cantonese tones*. Presented at Tone and Intonation 4, Stockholm, Sweden.

Fung, R., & Wong, C. (2010b). *The mechanisms of tone mergers in Hong Kong Cantonese*. Presented at 15th International Conference on Yue Dialects, Macau.

Gandour, J. (1983). Tone perception of far eastern languages. *Journal of Phonetics*, 11, 149-175.

Halle, M. (1962). Phonology in generative grammar. *Word*, 18, 54-72.

Hashimoto, A. O. (1972). *Phonology of Cantonese*. Cambridge: Cambridge University Press.

Howie, J. M. (1974). On the domain of tone in Mandarin. *Phonetica*, 30, 129-148.

Kei, J., Smyth, V., So, K. H., Lau, C. C., & Capell, K. (2002). Assessing the accuracy of production Cantonese lexical tones: a comparison between perceptual judgment and an instrumental measure. *Asia Pacific Journal of Speech, Language and Hearing*, 7, 25-38.

Khouw, E., & Ciocca, V. (2007). Perceptual correlates of Cantonese tones. *Journal of Phonetics*, 35, 104-117.

Killingley, S. Y. (1986). *A new look at Cantonese tones: Five or Six?* UK: Newcastle upon Tyne.

Labov, W., Yaeger, M., & Steiner, R. (1972). *A quantitative study of sound change in progress*. Philadelphia: U.S. Regional Survey.

Labov, W. (1994). *Principles of linguistic change. Vol. 1: Internal factors*. Oxford & Cambridge, Mass.: Blackwell.

Lam, H. W., & Yu, K. M. (2010). *The role of creaky voice quality in Cantonese tonal perception*. Presented at the Acoustic Society of America, Baltimore, US.

Lehiste, I. (1970). *Suprasegmentals*. Cambridge. MA: MIT Press.

Li, S. X. (2008). *Guan yu Guangzhou hua yin qu diao he yang qu diao de ting bian shi yan* [Auditory discrimination of *Yinqu* vs. *Yangqu* tone of Cantonese]. *Fangyan*, 1, 34-39.

Li, X. K. (1994). *Guangdong de fan yan* [Dialects in Guangdong]. Guangdong: Guangdong *ren min* Press.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of speech code. *Psychological Review*, 74, 431-461.

Lien, C. (1986). Tone merger in the dialects of northern Chinese. *Journal of Chinese Linguistics*, 14(3), 243-291.

Luo, W. H. (1997). *Zhong gu quan zhuo shang sheng yu xian jin guang zhou hua sheng diao* [The *Shang* tone in Middle Chinese and the tones in contemporary Guangzhou Cantonese]. In B. H. Zhan (Ed.), *Proceedings of the 5th International Conference on Yue dialects* (pp. 23-27). Guangzhou: *Jinan* University Press.

Milroy, J., & Harris, J. (1980). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology*, 115, 39-57.

Mok, P., & Wong, P. (2010a). *Perception of the merging tones in Hong Kong Cantonese: preliminary data on monosyllables*. Speech Prosody 2010.

Mok, P., & Wong, P. (2010b). *Production of the merging tones in Hong Kong Cantonese: preliminary data on monosyllables*. Speech Prosody 2010.

Ng, W. M. (2007). The phenomenon of tone changes in Hong Kong Cantonese. *Cantonese Studies*, 2, 48-53.

Norman, J. (1988). *Chinese*. New York: Cambridge University Press.

Ohala, J. (1981). The listener as a source of sound change. In C. S. Masek, R. A. Hendrick, & M. F. Miller (Eds.), *Papers from the Parasession on Language and Behavior* (pp. 178-203). Chicago: Chicago Linguistics Society.

Shi, F. (1986). *Tian jin fang yan shuang zi zu sheng diao fen xi* [Analysis of tones on disyllabic words in Tianjin dialect]. In F. Shi (Ed.), *Yu yin xue tan wei* [Exploration in Phonetics] (pp. 84-100). Beijing: Peking University Press.

Shi, Q. S. (2004). *Yi bai nian qian Guang zhou hua de yin ping diao* [Notes on *Yin Ping* Tone of Guangzhou Cantonese one hundred years ago]. *Fangyan*, 1, 34-46.

So, L. K-H. (1996). Tonal changes in Hong Kong Cantonese. *Current issues in Language & Society*, 3(2), 186-189.

Trudgill, P. (1974). *The social differentiation of English in Norwich*. Cambridge: Cambridge University Press.

Vance, J. T. (1976). An experimental investigation of tone and intonation in Cantonese. *Phonetica*, 33, 368-392.

Vance, J. T. (1977). Tonal distinction in Cantonese. *Phonetica*, 34, 93-107.

Wang, S-Y. W. (1967). Phonological features of tone. *International Journal of American Linguistics*, 33, 93-105.

Wang, S-Y. W. (1987). A note on Tone development. In the Chinese Language Society of Hong Kong (Ed.), *Wang Li Memorial Volumes* (pp. 435-443). Hong Kong: Joint Publishing Co.

Wong, T. S. (2008). The beginning of merging of the tonal categories B2 and C1 in Hong Kong Cantonese, *Journal of Chinese Linguistics*, 36(1), 155-174.

Yip, M. (2002). *Tone*. Cambridge: Cambridge University Press.

Yiu, Y. M. (2009). A preliminary study on the change of rising tones in Hong Kong Cantonese: An experimental study. *Language and Linguistics*, 10(2), 269-291.

Yuan, J. H. (2001). *Han yu fan yan gai yao* [Reports on the Chinese dialects]. Beijing: *Yu Wen* Press.

Zhan, B. H. (2002). *Guang dong fang yan gai yao* [An outline of the Yue dialects in Guangdong]. Guangzhou: *Jinan* University Press.

Zhu, X. N. (2010). *Yu yin xue* [Phonetics]. Beijing: The Commercial Press.

Zhu, X. N. (2012). Multitregisters and four levels: a new tonal model. *Journal of Chinese Linguistics*, 40(1), 1-17.