Perceptual Analysis of the Six Contrastive Tones in Cantonese

LI, Po-Yee (李葆怡)

Phonetics Lab
Department of Chinese, Translation and Linguistics
City University of Hong Kong
eleanor.li@student.cityu.edu.hk

Abstract

The experiment here is a perceptual study of the six contrastive tones in Cantonese. The monosyllable [si] was synthesized, and a large number of closely spaced F_0 contours were applied to it. The identification tests were separated into three parts. In Part I of the tests, subjects were asked to identify each synthetic stimulus as a Cantonese word with a high-high, mid-high, or mid-low tone. In Part II of the tests, subjects were asked to identify each synthetic stimulus as a Cantonese word with a mid-low, low-rising or high-rising tone. In Part III of the tests, subjects were asked to identify each synthetic stimulus as a Cantonese word with a mid-low or low-low tone. The influence of F_0 on the perception of Cantonese words was investigated and the categorical boundaries for the six contrastive tones were established. The identification results may prove Chao's and Vance's descriptions of the level tones are more accurate than Hashimoto's description.

1. Introduction

There are six contrastive tones in Cantonese. The pitch values are transcribed according to the convention of using ascending tone numbers, "1" to "5", with "1" representing the lowest pitch and "5" the highest pitch. The descriptions of the six tones by Chao (1947), Hashimoto (1972) and Vance (1976) are as follows:

Chao:	53	21	33	22
Hashimoto:	53	21	44	33
Vance:	55	11	33	22
	(high-high)	(low-low)	(mid-high)	(mid-low)

35	23
35	24
35	13
(high-rising)	(low-rising)

Some of the descriptions of the six contrastive tones by Chao, Hashimoto and Vance are different. The tone mark of the mid-high tone and mid-low tone (i.e., 33 and 22) given by Chao and Vance was lower than that given by Hashimoto (i.e., 44 and 33). However, Hashimoto gave a higher rising point (i.e., 24) for the low-rising tone than Chao (i.e., 23) and Vance (i.e., 13)

To summarize, the high-high, mid-high and the mid-low tones are regarded as level tones; the high-rising and the low-rising tone are regarded as rising tones and the low-low tone is regarded as falling tone.

2. Method

2.1 Recording of the natural speech sounds to collect F_0 acoustic data

Six Cantonese words corresponding to the six contrastive tones were read by 3 native male Cantonese speakers. The speech sounds were recorded for four times. The words they read were *詩* [si] (high-high) ("poet"), *嗜* [si] (mid-high) ("hobby"), *事* [si] (mid-low) ("matter"), *市* [si] (low-rising) ("city"), *史* [si] (high-rising) ("history") and *時* [si] (low-low) ("time").

2.2 Instruments used

All the stimuli were created by using the High-Level Parameter Speech Synthesis System (Version 2.2).

2.3 Design of the stimuli

2.3.1 F_0 of the level tones

The average F_0 of the 10 evenly distributed points for each token each tone was obtained. Finally, the average F_0 values of the 10 points of each tone were calculated. According to the average F_0 of the highest level tone (i.e., high-high) and the lowest level tone (i.e., mid-low), the F_0 range for the stimuli with level tones was between 102 Hz and 138 Hz. The F_0 from the starting and the end points were the same. The increment in F_0 between two stimuli was in 3 Hz steps.

2.3.2 F_0 of the rising and the falling tones

For the low-rising tone and the low-low tone, the tone marks transcribed by Chao (1947), Hashimoto (1977) and Vance (1976) are different. As a result, the acoustic findings of the rising tones and the falling tone collected here were used as reference for synthesizing the rising and the falling tones. According to the method used by Vance (1977), the F_0 of the six contrastive tones from the acoustic data (in Hz) were converted to a five-point scale as shown in Table 1.

Table 1: *Transcriptions of the six contrastive tones according to the three male native speakers of Cantonese.*

	first speaker		second speaker		third speaker	
詩 (high-high)	peak : 5.00	decay : 3.50	peak : 5.00	decay : 3.24	peak : 5.00	decay : 2.81
嗜 (mid-high)	peak : 4.04	decay : 2.57	peak : 3.34	decay : 1.46	peak : 3.69	decay : 1.96
事 (mid-low)	peak : 2.78	decay : 1.36	peak : 2.40	decay : 1.24	peak : 3.85	decay : 1.50
市 (low-rising)	dip : 1.89	peak : 2.90	dip : 1.61	peak : 2.30	dip : 1.97	peak : 3.21
史 (high-rising)	dip : 2.04	peak : 4.40	dip : 1.68	peak : 2.86	dip : 1.91	peak : 3.69
時 (low-low)	peak : 2.08	decay : 1.00	peak : 2.16	decay : 1.00	peak : 1.78	decay : 1.00

We can see that the F_0 of the dip of the two rising tones are very similar. They differ only in the rising extent. When we compare the dip of the rising tones with the mid-low tone, their values are more or less similar too. Therefore, in this study, the F_0 for the starting points of the stimuli with rising tones were the same as the mid-low tone. Also, we can see that the F_0 of the peak of the low-low tone is more or less similar to the mid-low tone too. Therefore, in this study, the F_0 for the starting points of the stimuli with a falling tone were the same as the mid-low tone as well.

To produce the rising and falling contour, according to the instrumental data provided by Hashimoto (1972) and Vance (1976), the Cantonese rising and falling tones can be approximated by linear ascending or descending contour. As a result, the F_0 for the end points of the two rising tones was linearly ascended and in 3 Hz steps, and thus, the range for the end points of the rising F_0 was between 105 Hz and 138 Hz. On the other hand, the F_0 for the end points of the falling tone was linearly descended and in 3 Hz steps and thus, the range for the end points of the falling F_0 was between 66 Hz and 99 Hz.

To summarize, the description of the six Cantonese words corresponding to the six contrastive tones used in this study was 詩 [si^{55}], 嗜 [si^{33}], 事 [si^{22}], 市[si^{23}] 史[si^{25}] and 時 [si^{21}].

2.4 Duration and the vowel quality of the stimuli

The duration of the stimuli were all kept constant with 330 ms (Zee, 1995) with the F_1, F_2 and F_3 of the vowel [ɪ] follow the data collected by Zee (2003) (i.e., F1=264.8 Hz; F2=2369 Hz and F3=3327 Hz). All the stimuli differ only in F_0 over the entire tested syllables.

2.5 Subjects and tasks

The subjects were 20 male and 20 female who were native speakers of Cantonese. They ranged in age from 20-25. The identification tests were divided into three parts, and subjects were instructed to pick the word that they heard. In Part I of the tests, the subjects were asked to identify the sound stimuli as 詩 [si^{55}], 嗜 [si^{33}] or 事 [si^{22}] every time they heard the stimulus. This part of the test is to concentrate on the identification of the level tones. In Part II of the tests, subjects were asked to identify the sound stimuli as 事 [si^{22}], 市 [si^{23}] or 史 [si^{25}] every time they heard the stimulus. This part of the test is to concentrate on the identification of the rising tones. In Part III of the tests, subjects were asked to identify the sound stimuli as 事 [si^{22}] or 時 [si^{21}] every time they heard the stimulus. This part of the test is to concentrate on the identification of the falling tone. There was a practice session in the beginning; subjects can hear the six different tones at least twice or more. It is to familiarize them with the format of the experiment and to determine a comfortable loudness level. The 13 stimuli in Part I of the tests yielded 520 responses (40 subjects × 13 stimuli), the 12 stimuli in Part II of the tests yielded 480 responses (40 subjects × 12 stimuli) and the 12 stimuli in Part III of the tests yielded 480 responses (40 subjects × 12 stimuli). A CD player played all the randomized stimuli once with a 3 second interval between two stimuli in all parts of the tests.

3. Result

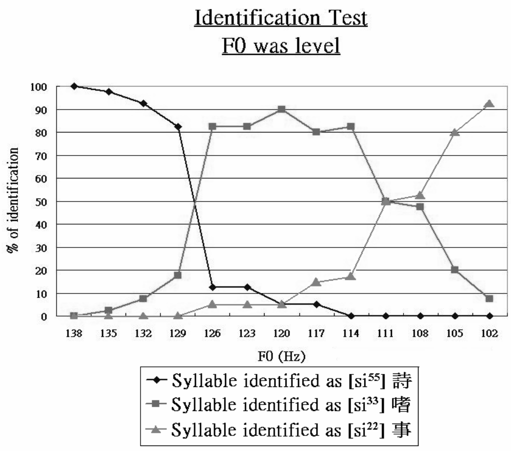

Figure 1: *Identification functions of the synthetic 詩 [si^{55}], 嗜 [si^{33}] and 事 [si^{22}] in which the F_0 was level.*

3.1 Level tones

Figure 1 shows the identification of the stimuli as 詩 [si^{55}] starts from 100% (F_0 = 138 Hz) and continues to drop. It drops dramatically from 82.5% to 12.5% when the F_0 of the stimuli changed from 129 Hz to 126 Hz. The percentage of identification continues to drop with the F_0 of the stimulus lowered continually. It reaches 0% when the F_0 is 114 Hz. The identification of the stimuli as 嗜 [si^{33}] starts from 0% (F_0 = 138 Hz) and continues to rise. It rises dramatically from 17.5% to 82.5% when the F_0 of the stimuli changed from 129 Hz to 126 Hz. The percentage of identification continues to rise until the F_0 is 120 Hz. After that, the percentage drops slowly with the F_0 lowered continually. Finally, it reaches 7.5% when the F_0 of the stimuli is 102 Hz. The identification of the stimuli as 事 [si^{22}] starts from 0% from the F_0 is 129 Hz to 138 Hz. The percentage rises slowly with the F_0 lowered continually. Finally, it reaches 92.5% when the F_0 of the stimuli is 102 Hz.

The categorical boundary between 詩 [si^{55}] and 嗜 [si^{33}] occurs when the F_0 is in the range of 126 Hz to 129 Hz (50% of identification) and the categorical boundary between 嗜 [si^{33}] and 事 [si ^{22}] occurs when the F_0 is in the range of 108 Hz to 111 Hz (50% of identification).

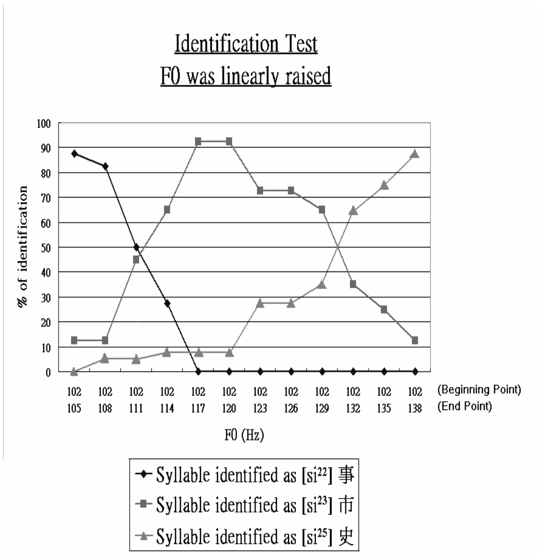

Figure 2 : *Identification functions of the synthetic* 事 *[si^{22}],* 市*[si^{23}] and* 史 *[si^{25}] in which the F_0 was linearly raised.*

3.2 Rising tones

Figure 2 shows the identification of the stimuli as 事 [si^{22}] starts from 87.5% (F_0 of the end point = 105 Hz) and continues to drop with the F_0 of the end point raised continually. The percentage of identification reaches 0% when the F_0 of the end point is 117 Hz. The identification of the stimuli as 市 [si^{23}] starts from 12.5% (F_0 of the end point = 105 Hz) and continues to rise until the F_0 of the end point of the stimuli is 120 Hz (92.5%). After that, the percentage drops with the F_0 raised continually. Finally, it reaches 12.5 % when the F_0 of the end point is 138 Hz. The identification of the stimuli as 史 [si^{25}] starts from 0% (F_0 = 138 Hz) and continues to rise with the F_0 of the end point raised continually. Finally, the percentage of identification reaches 92.5% when the F_0 of the end point is 138 Hz.

The categorical boundary between 事 [si^{22}] and 市 [si^{23}] occurs when the F_0 of the end point is near 111 Hz (50% of identification) and the categorical boundary between 市 [si^{23}] and 史 [si^{25}] occurs when the F_0 of the end point is in the range of 129 Hz to132 Hz (50% of identification).

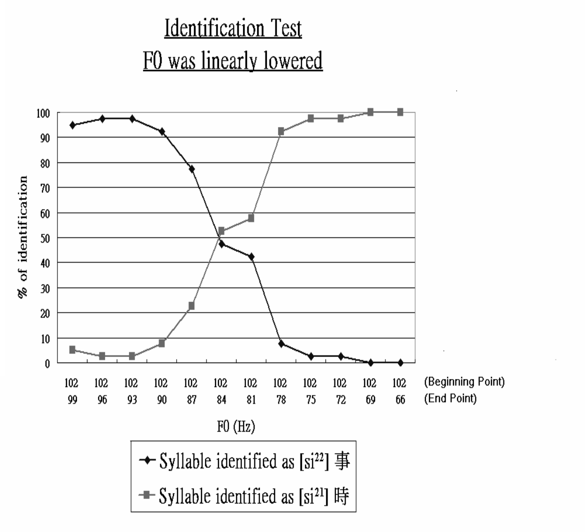

Figure 3 : *Identification functions of the synthetic* 事 *[si^{22}] and* 時 *[si^{21}] in which the F_0 was linearly lowered.*

3.3 Falling tone

Figure 3 shows the identification of the stimuli as 事 [si^{22}] starts from 95% (F_0 of the end point = 99 Hz). It starts to drop when the F_0 of the end point is 90 Hz (92.5%). The percentage of identification continues to drop with the F_0 of the end point lowered continually. It reaches 0% when the F_0 of the end point is 69 Hz. The identification of the stimuli as 時 [si^{21}] starts from 5% (F_0 of the end point = 99 Hz). It starts to rise when the F_0 of the end point is 90 Hz (7.5%). The percentage of identification continues to rise with the F_0 of the end point lowered continually. It reaches 100% when the F_0 of the end point is 69 Hz.

The categorical boundary between 事 [si^{22}] and 時 [si^{21}] occurs when the F_0 of the end point is in the range of 81 Hz to 84 Hz (50% of identification).

4. Discussion

4.1 The categorical boundaries of F_0 of levels tones are consistent with those of the end point of F_0 of the rising tones

The categorical boundaries of F_0 of the levels tones 詩 [si^{55}] (high-high tone) and 嗜 [si^{33}] (mid-high tone) is in the range of 126 Hz to 129 Hz and between 嗜 [si^{33}] and 事 [si^{22}] (mid-low tone) is in the range of 108 Hz to 111 Hz match with the categorical boundaries of the end point of F_0 of the rising tones. The categorical boundary of the end point of F_0 between 史 [si^{25}] (high-rising tone) and 市 [si^{23}] (low-rising tone) is in the range of 129 Hz to 132 Hz while between 市 [si^{23}] and 事 [si^{22}] is near 111 Hz. The categorical boundaries of F_0 of the level tones and the end point of F_0 of the rising tones are consistent with each other.

4.2 Subjects had less confusion to identify between high-high tone and mid-high tone than to identify between mid-high tone and mid-low tone

Figure 1 shows that the identification curve (rising part) of *嗜* [si^{33}] (mid-high tone) is steeper than that of *事* [si^{22}] (mid-low tone). It appears that subjects had less confusion to identify between *詩* [si^{55}] (high-high tone) and *嗜* [si^{33}] (mid-high tone) than to identify between *嗜*[si^{33}] (mid-high) and *事* [si^{22}] (mid-low tone). If it is true, it seem that *詩* [si^{55}] and *嗜* [si^{33}] are easier to be identified than between *嗜* [si^{33}] and *事* [si^{22}]. At the same time, it may prove Chao's and Vance's description of the level tones (i.e., 53, 33 and 22) is more accurate than Hashimoto's description (i.e., 53, 44 and 33). It is because if the difference of F_0 between *詩* [si^{55}] and *嗜* [si^{33}] is bigger than between *嗜* [si^{33}] and *事* [si^{22}], it is likely that subjects will be more easily to identify between *詩* [si^{55}] and *嗜* [si^{33}] than between *嗜* [si^{33}] and *事* [si^{22}].

4.3 The two rising tones are less easy to be distinguished than between low-low tone and low-rising tone

Figure 2 shows that the identification curve (falling part) of *市* [si^{23}] is less steep than that of *事* [si^{22}]. It shows that the two rising tones *市* [si^{23}] and *史* [si^{25}] are less easy to be distinguished than between a level tone and a rising tone. Probably, it is because the two rising tones have the same contour shape in which they differ only in the increment of F_0 at the end point, with high-rising tone rises to a larger extent.

4.4 The perception of the six contrastive tones in Cantonese is continuous

Eimas (1963) compared the difference of the identification functions of synthetic stop consonants /b, d, g/ in his study with the vowels /i, ε, æ/ in the study of Fry et al. The perception of the stop consonants /b, d, g/ is categorical as they have steep and abrupt boundaries while the perception of vowels /i, ε, æ/ are continuous as they do not have steep and abrupt boundaries. We can see that the curve pattern of the Figure 1, 2 and 3 in this study are similar to the kind of graph representing the identification functions of the vowels /i, ε, æ/ in Figure 4. The identification functions of the Cantonese tones are similar to that of vowel. Probably, it is because the identification of the contrastive tones in Cantonese is influenced by linguistic context like vowels. Chan (1974) also noticed the relative F_0, in relation to the neighboring tones is responsible for the identification of a Cantonese tone.

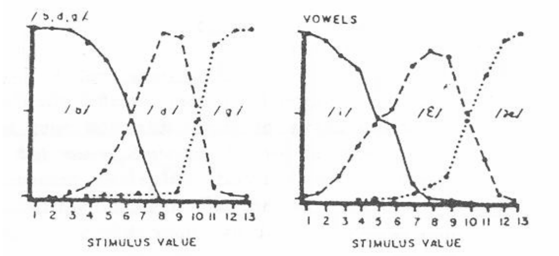

Figure 4 : *Identification functions of the synthetic stop consonants /b, d, g/ from Eimas (1963) and the vowels /i, ε, æ/ from Fry et al..*

5. Conclusion

The influence of the F_0 on the perception of Cantonese words has been found. Subjects perceived the synthetic speech sounds as different words when different F_0 was applied to the stimuli. As a result, the categorical boundaries of each tone have been established. From the result, we can see that the categorical boundaries of F_0 of levels tones are consistent with those of the end point of F_0 of the rising tones. Secondly, subjects had less confusion to identify between high-high tone and mid-high tone than to identify between mid-high tone and mid-low tone. Thirdly, distinguishing two rising tones is less easy then distinguishing between a level tone and a rising tone as the contour shape is the same in the two rising tones. Finally, the perception of the six contrastive tones in Cantonese is continuous like the perception of vowels.

6. Reference

[1] Chao, Yuen Ren, 1947. *Cantonese primer*. Greenwood Press: New York.

[2] Eimas, Peter D, 1963. *The relation between identification and discrimination along speech and non-speech continua*. Language and Speech.

[3] Fok Chan, Yuen Yuen, 1974. A perceptual study of tones in Cantonese. University of Hong Kong.

[4] Hashimoto, O.- K. Y, 1972. *Studies in Yüe dialects. 1. Phonology of Cantonese.* Cambridge University Pres: Cambridge.

[5] Vance, T. J., 1976. *An experimental investigation of tone and intonation in Cantonese.* Phonetica.

[6] Vance, T. J., 1977. Tonal Distinctions in Cantonese. Phonetica *34*: 93-107.

[7] Zee, Eric, 1995. *Temporal Organization of Syllable Production in Cantonese Chinese*. ICPHS 95 Stockholm, vol. 3, p. 250-254.

[8] Zee, Eric, 2003. *Frequency Analysis of the Vowels in Cantonese from 50 Male and 50 Female Speakers*. ICPhS Barcelona, vol. 2, p. 1117-1120.