Tone Production by the Speakers of Different Age-and-Gender Groups

Wai-Sum Lee

Department of Chinese, Translation and Linguistics, City University of Hong Kong, Hong Kong
w.s.lee@cityu.edu.hk

Abstract

This paper is an acoustic analysis of the pitch/F0 of the six long tones [55 33 22 21 25 23] in Cantonese produced by the male and female adult speakers and male and female child speakers. Results show that (i) the F0 patterns of the Cantonese tones for the speakers of different age-and-gender groups are similar, but the absolute F0 values differ. (ii) The difference in F0 between the adult and child speakers is large, but less between the child and female adult speakers. (iii) The difference in F0 is noticeable between the adult speakers of different genders, but not between the male and female child speakers. (iv) The difference in F0 across the speaker groups is uniformly scaled for different tone types. And (v) for all the six tones, the F0 value for the child speakers is approximately an octave higher than that for the male adult speakers and 1.2 to 1.3 times higher than that of the female adult speakers, and the F0 for the female adult speakers is slightly over a half octave higher than the F0 for the male adult speakers.

Index Terms: F0/pitch, tones, age, gender, acoustic analysis

1. Introduction

It has been known that the age and gender have an effect on the F0/pitch value of speech. In general, the pitch is lower for males than for females, and it is lower for adults than for children. However, it has also been reported that there is no difference in pitch between the preadolescent children of different genders ([1, 2, 3, 4, 5]) and the difference in pitch is not pronounced between the female adults and children ([1, 6]). All of these past studies of pitch production are mainly based on the speech data from English-speaking subjects. As English is a non-tone language, the present study aims at obtaining the F0 of the speech of a tone language by investigating the Cantonese tones produced by the speakers of different ages and genders.

In Cantonese, there are a total of nine citation tones on monosyllables ([7]). Six of them are long tones associated with the non-checked CV (C = syllable-initial consonant; V = vowel or diphthong) or CVN (N = syllable-final nasal) syllables. The other three are short tones associated with the checked CVS (S = syllable-final stop) syllables. The long tones include three level tones [55 33 22] ([1 4]), one falling tone [21] ([1]), and two rising tones [25 23] ([1 4]), and the three short tones are [5 3 2] ([1 4]). Phonologically, [5 3 2] are considered as the short variants of the three long level tones [55 33 22], respectively. In this study, only the six long tones are investigated.

2. Method

2.1. Test material

Table 1 shows the test syllables in Cantonese associated with the six long tones [55 33 22 21 25 23]. They are all meaningful monosyllabic words. All of the test syllables contain a low vowel [a]. Four of the test syllables have a zero-initial consonant, whereas the other two have an optional syllable-initial velar nasal [ŋ]. In Cantonese, the occurrence of the syllable-initial [ŋ] is optional ([8]). The test words [(ŋ)əa] with two different tones used in this study, i.e., [(ŋ)əa 21] (‘tooth’) and [(ŋ)əa 23] (‘tile’), were pronounced as [a 21] and [a 23], respectively, by all the speakers participated in the study. Thus, all of the test words were produced with a single low vowel [a], associated with different tones.

<table>
<thead>
<tr>
<th>Tones</th>
<th>Test syllables</th>
</tr>
</thead>
<tbody>
<tr>
<td>[55]</td>
<td>[a 55] (‘twig’)</td>
</tr>
<tr>
<td>[33]</td>
<td>[a 33] (‘Asia’)</td>
</tr>
<tr>
<td>[22]</td>
<td>[a 22] (‘taking up space’)</td>
</tr>
<tr>
<td>[21]</td>
<td>[(ŋ)əa 21] (‘tooth’)</td>
</tr>
<tr>
<td>[25]</td>
<td>[a 25] (‘dumb’)</td>
</tr>
<tr>
<td>[23]</td>
<td>[(ŋ)əa 23] (‘tile’)</td>
</tr>
</tbody>
</table>

The six test monosyllables were randomized in a word list and each was placed in a carrier sentence [23 jiu 33 tuk 2 __ pei 25 lei 23 t³ŋ 55] (‘I want (to) read __ for you (to) listen’).

2.2. Subjects

The speech data were provided by 8 native Cantonese speakers, including 2 male and 2 female university students who were all in their early twenties and 2 male and 2 female preadolescent children of 9-10 years of age. All the speakers were born and grew up in Hong Kong and did not have a history of speech and hearing problems. The subjects were paid for their participation.

2.3. Procedure and analysis

Audio recording was performed for each of the speakers in a sound-proof booth in the Phonetics Laboratory at the City University of Hong Kong. The subjects were instructed to read the word list that consisting of five repetitions of the six test syllables in the carrier sentence at a normal rate of speech. There were in total 240 test tokens (6 test words x 5 repetitions x 8 speakers).

The speech data were analyzed for F0/pitch curves of the tones on the test syllables, using the pitch synchronous F0 tracing program available in Kay’s CSL (Computerized Speech Lab) 4400 speech analysis software. For each F0 contour, 11 data points were sampled. The data points consist of the onset of the F0 contour as the initial data point and the points at every 10% of the overall duration of the F0 contour of the tone.

For each tone, the mean F0 value for each of the 11 data points was obtained by averaging across the values for the corresponding data points of the F0 contours of the five
repetitions of the tone. The mean F0 values of the 11 data points for the tone were then plotted on the y-axis against the mean duration of the five repetitions of the tone on the x-axis.

3. Results and discussion

3.1. The F0 pattern of the six long tones

Figures 1-8 show the mean F0 contours of the six long tones, \([55]\) (thick dark line), \([33]\) (thick grey line), \([22]\) (thin dark line), \([21]\) (thin grey line), \([25]\) (dark broken line), and \([23]\) (grey broken line), for each of the eight Cantonese speakers. The figures have the same time scale on the x-axis, but differ in the frequency scale on the y-axis which varies according to the frequency range of the tones for any individual speakers.

As shown in the figures, while the absolute F0 values for the speakers differ, the F0 patterns of the six tones for the speakers are similar. The F0 contours of \([55 33 22]\) are basically level, though \([33]\) and \([22]\) have the tendency to fall slightly. Among the three level tones, the F0 level of the contour of \([55]\) is the highest of all three, to be followed by the F0 contours of \([33]\) and \([22]\), as can be seen the F0 level for \([33]\) is only slightly higher than that for \([22]\).

Figures 1-8: The mean F0 contours of \([55 33 22 21 25 23]\) for each of the eight Cantonese speakers.
As for the F0 contours of the tones [21 25 23], they overlap quite extensively in the first half of the F0 contour and differ mainly in the second half. The F0 contour of [21] falls continuously toward the end of the tone, subsequently reaching the F0 level for the lower end of the pitch range for any speakers. The F0 contours of [25] and [23] are rising, and a slight dip is observed in particular in the F0 contour of [25]. The main difference between [25] and [23] is in the height that the offset of the F0 contour reaches.

3.2. The F0 differences of the tones between different age-and-gender groups

To show the differences in F0 contours of the tones across the speaker groups, the mean F0 contours of each of the six tones [55 33 22 21 25 23] produced by the eight Cantonese speakers are plotted in the same figure for comparison. Figures 9-14 show the mean F0 contours of the individual tones for the child speakers, male (thick broken lines) and female (thin broken lines), female adult speakers (thin solid lines), and male adult speakers (thick solid lines).

As shown in the figures, the difference in the F0 contours of the tones between any groups of speakers is in the F0 level rather than shape. For any tone, the F0 level is the highest for the child speakers, male and female, to be followed by the female adult speakers and male adult speakers in descending order. There is no overlap between the F0 contours of any tone by the different groups of speakers, except the male child and female child speakers. The F0 level of each tone for the male adult speakers is markedly lower than that for the female adult speakers. Such F0 difference is not found in the speech of the male and female child speakers. As shown in Figures 9-14, the F0 levels of any tone for the male and female child speakers are similar. This is especially so for the tones [33] (Figure 10), [22] (Figure 11), [25] (Figure 13), and [23] (Figure 14), where the F0 contours of any tone for the male children (thick broken lines) and female children (thin broken lines) overlap extensively. This shows that not only in English ([1, 2, 3, 4, 5]), a non-tone language, but also in
Cantonese, a tone language, the difference in \( F_0 \) between the male and female child speakers is insignificant.

As for the space between the \( F_0 \) contours of any tone produced by the child and adult speakers, it is smaller between the \( F_0 \) levels for the child speakers (both male and female) and female adult speakers than between the \( F_0 \) levels for the child and male adult speakers. Similar results in the non-tone English language have also been reported ([1, 6]).

Table 2 shows the mean \( F_0 \) values for the six tones for the eight Cantonese speakers. The \( F_0 \) value for the level tone [55], [33], or [22] is the value for the mid or the sixth \( F_0 \) data point of the mean \( F_0 \) contour of the tone. As for the contour tones [21 25 23], they are represented by two \( F_0 \) values. The first \( F_0 \) value is the value for the third data point of the mean \( F_0 \) contour of the tone. As for the second \( F_0 \) value, for [21] it is the \( F_0 \) value for the offset of the mean \( F_0 \) tone contour, but for [25] and [23] it is the \( F_0 \) value for the highest point of the rising portion of the mean \( F_0 \) tone contour.

Table 2. The \( F_0 \) values (in Hz) of the six tones for the eight Cantonese speakers; one \( F_0 \) value for [55 33 22] and two \( F_0 \) values for [21 25 23].

<table>
<thead>
<tr>
<th>Speakers</th>
<th>[55]</th>
<th>[33]</th>
<th>[22]</th>
<th>[21]</th>
<th>[25]</th>
<th>[23]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male Adult 1</td>
<td>128.4</td>
<td>104.4</td>
<td>95.2</td>
<td>94.8</td>
<td>95.6</td>
<td>93.8</td>
</tr>
<tr>
<td>Male Adult 2</td>
<td>167.2</td>
<td>130.4</td>
<td>120.2</td>
<td>114.0</td>
<td>114.2</td>
<td>115.4</td>
</tr>
<tr>
<td>Female Adult 1</td>
<td>235.0</td>
<td>197.6</td>
<td>181.8</td>
<td>180.6</td>
<td>178.2</td>
<td>179.6</td>
</tr>
<tr>
<td>Female Adult 2</td>
<td>248.0</td>
<td>207.8</td>
<td>190.6</td>
<td>182.0</td>
<td>178.2</td>
<td>184.0</td>
</tr>
<tr>
<td>Male Child 1</td>
<td>297.4</td>
<td>258.4</td>
<td>233.6</td>
<td>223.6</td>
<td>219.6</td>
<td>221.6</td>
</tr>
<tr>
<td>Male Child 2</td>
<td>347.2</td>
<td>262.6</td>
<td>237.0</td>
<td>250.8</td>
<td>231.2</td>
<td>229.4</td>
</tr>
<tr>
<td>Female Child 1</td>
<td>338.8</td>
<td>262.6</td>
<td>235.8</td>
<td>223.0</td>
<td>222.0</td>
<td>232.2</td>
</tr>
<tr>
<td>Female Child 2</td>
<td>308.8</td>
<td>258.2</td>
<td>232.0</td>
<td>243.8</td>
<td>178.0</td>
<td>228.6</td>
</tr>
</tbody>
</table>

As shown in Table 2, the absolute \( F_0 \) values of any tone for the speakers of different age-and-gender groups differ and the \( F_0 \) values are in general higher for the child speakers (both male and female) than the male and female adult speakers. The \( F_0 \) range for the level tones [55 33 22] is larger for the female child speakers (323.8 Hz - 233.9 Hz = 89.9 Hz) than for the male child speakers (307.3 Hz - 235.3 Hz = 72.0 Hz), the female adult speakers (241.5 Hz - 186.2 Hz = 55.3 Hz), and the male adult speakers (147.8 Hz - 107.7 Hz = 40.1 Hz). However, the ratios of the \( F_0 \) values of [55 33 22] show that the difference in \( F_0 \) across the speaker groups is not pronounced. For instance, the ratios of the \( F_0 \) values of [55]/[33] and [33]/[22] are 1.24 and 1.11 for the female child speakers, 1.19 and 1.10 for the male child speakers, 1.19 and 1.09 for the female adult speakers, and 1.26 and 1.09 for the male adult speakers.

Table 3 shows the ratios of the \( F_0 \) values for the six tones for the child speakers and the male and female adult speakers. Due to the similarity in \( F_0 \) data for the male and female child speakers, they are lumped together. As shown in the table, the ratios of the \( F_0 \) values between any two speaker groups are similar across the tonal categories. The \( F_0 \) ratios are 2.0 or slightly over for all the six tones for the child speakers and male adult speakers, which means the \( F_0 \) for the child speakers is approximately an octave higher than the \( F_0 \) for the male adult speakers. The ratios for the child speakers and female adult speakers range from 1.17 to 1.31, and the ratios for the female and male adult speakers range from 1.60 to 1.74. The data also show that the \( F_0 \) for the female adult speakers is slightly over a half octave higher than the \( F_0 \) for the male adult speakers. Thus, the difference in \( F_0 \) between the male and female adult speakers is larger than that between the female adult and child speakers.

4. Conclusions

This paper has presented the \( F_0 \) contours of the six long tones in Cantonese. The \( F_0 \) patterns of the six tones for the speakers of different age-and-gender groups are similar, but the absolute \( F_0 \) values differ. The difference in \( F_0 \) is large between the child speakers and male adult speakers, but less between the child speakers and female adult speakers. There is a difference in \( F_0 \) between the male and female adults, but not between the child speakers of different genders. The results in this study also indicate that the difference in \( F_0 \) across the speaker groups is uniformly scaled for the different tone types.

5. References