HONG KONG CANTONESE CITATION TONE ACOUSTICS: A LINGUISTIC TONETIC STUDY
Phil Rose
Phonetics Laboratory, Linguistics Program, A.N.U.

ABSTRACT Mean fundamental frequency and duration data for the six citation tones of Hong Kong Cantonese on unstopped syllables are presented for five male and five female young native speakers. The phonetic-phonetic properties of the tones are specified from mean standard deviation normalised F0 and duration data. The effectiveness of the normalisation is shown to be much better than for some other Chinese dialects, and it is hypothesised that this is a function of the minimal nature of some of the Cantonese tone contrasts.

INTRODUCTION
The aim of this paper is to give a quantified description of the linguistic-phonetic acoustic properties (fundamental frequency and duration) of Hong Kong Cantonese citation tones. There are many auditory descriptions of Cantonese tones in the literature, expressed in various sentences. Descriptions can be found using the Chao tone letters, e.g. Yuan (1983), or musical notation, e.g. Jones and Woo (1912-36), or prose, e.g. Mathews and Yip (1984). The are also many descriptions of the acoustic properties of individual speakers, e.g. Hashimoto (1972:122-126; Vance (1976). However, there has to my knowledge been no study which specifies the linguistic-phonetic acoustic characteristics of Cantonese by normalising acoustic data from a number of speakers. This study remedies this situation, and also briefly examines how well tones normalise across dialects.

As with all such phonetic studies, phonological descriptions are logically prior (Ladefoged 1997: 136-139). In Cantonese, as in many varieties of Asian tone language, the number of surface tonal contrasts depends on the structure of the Rhyme - specifically on the absence or presence of a syllable-final stop (p, t, k) in the Coda. In syllables without a stop in the coda ('unstopped syllables'), conservative varieties of Cantonese have a six-way tonal contrast, and it is the acoustic allotones of these six tones that are described here. (The number of surface tonal contrasts on unstopped syllables is either two or three, depending on the phonological interpretation of vowel length.) On unstopped syllables, the Cantonese tone system contrasts one falling, three level and two rising pitches. Of the three level pitched tones, one (Tone1) sounds to be at the top of the speaker's normal pitch range; one (Tone3) is in the middle of the pitch range, or just below, and one (Tone6) lies just below T5.

For example, with transcription, from Yuan (1983: 185,186), are: (T1) [f u 55] husband; (T3) [1 f u 44] father; (T4) [1 f 22] father; (T6) [f 21] support. Of the two rising pitched tones, one (T3) has a pitch which onsets in the lower third of the speaker's pitch range and rises into the upper third, e.g. [f 34] bitter; and one (T4) has a pitch contour similar to T3, but which only rises into the third of the speaker's range, e.g. [f u 13] woman. The only falling pitched tone (T2) falls through the lowest third of the speaker's pitch range, e.g. [f u 21] to support. Often T2 falls below the speaker's normal pitch range, and its phonation type can become creaky or breathy towards the end, especially on open vowels. From this description, it can be seen that the Cantonese tone system does not utilise maximum pitch contrasts, instead. Two of the tones have mid-level T5 and lower-mid level T6, and between low-to-high-rising T3 and low-to-mid-rising T4 -- appear to rest on rather small differences in pitch, and indeed from the point of view of pitch realisation Hong Kong Cantonese is one of the world's nastier tone languages.

PROCEDURE
Corpus and elicitaton The corpus represented a compromise, given the phonactics of Cantonese and the availability of common possible morphemes, between having a balanced vowel height effect (to counteract intrinsic vowel F0) and a uniform consonantal effect (for comparison with existing tonal data from Chinese dialects). The items analysed are given in table 1. As can be seen, in four of the six tones the initial consonants are, with the one exception ([w] in T6), voiceless unaspirated stops [p, t, k].

Tones 2 and 4 do not commonly occur in Cantonese with voiceless unaspirated stops, so [+spread glottis] segments (i.e. voiceless aspirated stops or voiceless fricatives) were specified instead. This systematic difference between tones 2 and 4 on the one hand and tones 1, 3, 5 and 6 on the other can be expected to be reflected in F0 differences at Rhyme onset, since [+ spread glottis] segments can be expected to have an intrinsically higher F0 at Rhyme onset (Rose 1998). As far as the vowels are concerned, it can be seen that each tone has two Rhymes containing a high, and two with a non-high vocalic segment. Due to the phonactctic restriction in Cantonese on the combination of some stops with high voiced monophthongs e.g. [pi], [ti], [ki], [pu], [tu], diphthongs with high vowel offglides ([i], [u]) were substituted. As can be seen from their transcription, the first vowel target in these diphthongs is actually lower-mid in height, so the diphthongs [+ high] status is not totally clear. The T4 forms are slightly unbalanced for vowel height, and lack a second low vowel rhyme.

Since this was part of a larger experiment to elicit data for Cantonese tones on both stopped and unstopped syllables, the 24 morphemes in table 1 were randomly combined with additional stopped syllable forms, written with Chinese characters, and presented on four lists to subjects to read out. In order to avoid list-initial and list-final intonation, dummy characters were inserted at the beginning and end of each list, and to avoid listing intonation, subjects were instructed to pause between each character.

<table>
<thead>
<tr>
<th>T1</th>
<th>kei</th>
<th>bas</th>
<th>ku</th>
<th>father’s sister</th>
<th>ko</th>
<th>song</th>
<th>ka</th>
<th>to add</th>
</tr>
</thead>
<tbody>
<tr>
<td>T2</td>
<td>p’ci</td>
<td>skin</td>
<td>tu</td>
<td>support</td>
<td>p’</td>
<td>woman</td>
<td>p’</td>
<td>scramble</td>
</tr>
<tr>
<td>T3</td>
<td>t’u</td>
<td>gamble</td>
<td>ku</td>
<td>ambient</td>
<td>f</td>
<td>kids</td>
<td>tu</td>
<td>air</td>
</tr>
<tr>
<td>T4</td>
<td>t’es</td>
<td>similar</td>
<td>tu</td>
<td>woman</td>
<td>p’</td>
<td>hug</td>
<td>tu</td>
<td>society</td>
</tr>
<tr>
<td>T5</td>
<td>peti</td>
<td>back</td>
<td>ku</td>
<td>cause</td>
<td>ko</td>
<td>classifier</td>
<td>p’</td>
<td>tyrant</td>
</tr>
<tr>
<td>T6</td>
<td>tel</td>
<td>earth</td>
<td>p’u</td>
<td>part</td>
<td>wa</td>
<td>speak</td>
<td>p’</td>
<td>cease</td>
</tr>
</tbody>
</table>

Table 1. Corpus

This composition is four better than the minimum for quantified phonetic work mentioned by Ladefoged (1997: 140), but still short of his preferred minimum of six male and six female informants. It might also be argued that the required number of subjects is indicated statistically, by what the standard deviation around the mean normalised curves becomes asymptotic to a given value. The speakers are referred to below as M(ale)/F(female) 1-5. The set of characters was read at least four times, and recorded on professional equipment in the A.N.U. phonetics lab. Studio. Tokens from the first and fourth reading were analysed. There sound to be some confusion between the diphthongs / tyrant and p’ai, and if F2 read the character for /cease (p’ai) as p’ai. Comparison using the tonal and F0 and duration) supported this hypothesis, and these speakers’ forms were excluded from the analysis.

Measurement. Tokens were digitised at 10K and analysed with the CSL pitch (sic) extraction routine (the frame length was set to 20 ms and the frame advancement at 10 ms.) F0 was sampled at a frequency adjudged high enough to resolve the details of its time course. This was at the following nine percentage points of the tone’s sampling base: 0%, 5%, 10%, 20%, 40%, 60%, 80% 95%, 100%. The sampling base was from adjudged phonation onset to F0 peak in the rising pitch tones T3 and T4; and to adjudged phonation offset in the other tones. In all (6 tones x 4 morphemes x 9 sampling points x 4 repeats x 10 speakers = ) 6540 individual F0 measurements were made.

RESULTS

Figure 1 shows the mean F0 of each of the ten speakers’ tones. Males are on the left; speakers are in descending order of overall F0 values. Except where p’ai and p’ai tokens were excluded, each data point is the mean of 16 observations (4 morphemes x 4 repeats). F0 is plotted as a function of absolute duration in order to show between-tone durational differences, and in order to show between-speaker durational differences, the duration axis scale and range have been fixed. Because of space considerations, however, between-speaker differences in F0 have not been visually preserved, and it will be noted that, on the contrary, the F0 axes have differing ranges and scales in order to make the speakers’ F0 ranges visually comparable.

Figure 1 shows between-speaker differences in F0 values correlating with sex. The mean male F0 range, at about 75 Hz, is about three-quarters that of the mean female range of about 100 Hz. There is a slight overlap in F0 range, with M2’s range of 100 Hz being the same as F2 and F5’s range, and slightly greater than the 90 Hz ranges of F4 and F1. The male average F0 range extends from ca. 85 Hz to 165 Hz, compared to an average female range extending from 245 Hz to 263 Hz. Although these mean ranges do not overlap, quite a few individual ranges do. M2’s highest value, for example, is higher than females’ lowest values, and F2’s lowest value is lower than four males’ highest
This overlap means that examples can be found of different tones having similar F0 values in different speakers. F1's mid level tone 5, for example, lies between 200 Hz and 190 Hz, which values are similar to M2's high level T1, at ca. 190 Hz. There are of course within-sex differences in F0 range which also create overlap in the F0 values of different tones. Figure 1 also shows between-speaker differences in overall tonal duration, with values ranging between ca. 30 csec (M5, M3, M4) and ca. 45 csec (M1, F5). There is no between-sex difference in overall duration, with both sexes having a mean of ca 33 csec, but males, with a standard deviation of 6.1 csec, show greater variability than females (sd = 3.3 csec).

Despite these differences in raw mean F0 and duration values, it is easy to see that all ten speakers show very similar F0 configurations. The F0 shapes of the three level pitched tones (T1, T5, T6) lie unevenly spaced with T1 at the top and T5 and T6 in the middle of the speaker's F0 range. The two rising pitched tones (T3, T4) have dipping F0 shapes that onset in the middle of the speaker's F0 range and remain congruent for much of their time course, with the high rising tone T3 rising to about the height of the high level tone, and low rising T4 offsetting at the about the level of the mid level T5. The F0 of the low falling T2 falls from the speakers' mid F0 range to the bottom, and has the shortest duration of the tones.

NORMALISATION

The purpose of normalisation is to extract the linguistic and accentual content of the speech signal by getting rid of as much as possible of the individual content. The speakers' mean F0 values were z-score normalised, which involves subtracting a speaker's F0 value from their mean F0 and dividing by their standard deviation F0, so that the F0 values are expressed as multiples of so many standard deviations above or below their mean (Rose 1987). There are two approaches to selecting the F0 values from which to calculate the normalisation parameters (NPs) of mean and standard deviation. One is auditorily based: to use F0 values of tones which sound to have the same pitch shape and dividing their mean and standard deviation. The other, which is used in this paper, is acoustically based: to choose those F0 values which appear to show the least amount of between-speaker acoustic variation. Since many of the T2 F0 values seemed to show a large amount of between-speaker variation, these were excluded from the NPs, as well as the values near the onset (0%, 5%, 10%) and offset (85%, 100%). The mean and standard deviation NPs were thus calculated from 20 F0 observations (at 20%, 40%, 60%, and 80% in all tones except T2). The values of the speakers' normalisation parameters are given in table 2.

<table>
<thead>
<tr>
<th>T</th>
<th>M1</th>
<th>M2</th>
<th>M3</th>
<th>M4</th>
<th>M5</th>
<th>F1</th>
<th>F2</th>
<th>F3</th>
<th>F4</th>
<th>F5</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1</td>
<td>226</td>
<td>232</td>
<td>163</td>
<td>137</td>
<td>157</td>
<td>20.9</td>
<td>27.1</td>
<td>26.1</td>
<td>30.6</td>
<td>26.6</td>
</tr>
<tr>
<td>T2</td>
<td>133.4</td>
<td>141.3</td>
<td>123.2</td>
<td>109.4</td>
<td>102.5</td>
<td>190.8</td>
<td>178.9</td>
<td>216.0</td>
<td>237.8</td>
<td>223.2</td>
</tr>
<tr>
<td>T3</td>
<td>226</td>
<td>232</td>
<td>163</td>
<td>137</td>
<td>157</td>
<td>20.9</td>
<td>27.1</td>
<td>26.1</td>
<td>30.6</td>
<td>26.6</td>
</tr>
<tr>
<td>T4</td>
<td>133.4</td>
<td>141.3</td>
<td>123.2</td>
<td>109.4</td>
<td>102.5</td>
<td>190.8</td>
<td>178.9</td>
<td>216.0</td>
<td>237.8</td>
<td>223.2</td>
</tr>
</tbody>
</table>

Table 2. Values for normalisation parameters (Hz).

Figure 2. Normalised F0 values for ten Cantonese speakers' high level tone.

This figure shows very similar normalised values for most of the tone's duration (from ca 10% to 80%), but small between-speaker differences at the onset and very large between-speaker differences at the end. Speakers thus obviously differ considerably in how they end citation tone phonation.

It has been shown (Rose 1993) that it is important for linguistic tonetic comparison to retain information on relative tonal duration, so the duration values were normalised as a percentage of a speaker's mean duration calculated from all tones. Thus, since M1's mean duration was 41.07 csec, the normalised value for his T1 duration of 43.8 csec was \((43.8/41.07) \times 100 = 107\%\).
Deviations

Deviations

Expected

Figure

Unstopped Citation Tones

For these are shown in Table 3, together with mean normalised durations for each tone. The values that

Mean

standard deviation normalised

Vertical

Figure

Example


normalised index (Ni). The DC is then the ratio of overall variance to mean between-speaker variance. Two Ni's were calculated: one for all the data; and one for tonally relevant data. Tonally relevant data are values which were taken to represent the tonal rather than the individual, or segmentally determined features in the normalised curves. The tonally non-relevant data to be excluded were taken to be values at onset (i.e. 0%, 5%) in all tones; values at offset (i.e. 95%, 100%) in level tones (T1, 5, 6); and values in the second part of T2 (60%, 90%, 95%, 100%). As far as the tonally relevant data were concerned, the DCs for the raw and normalised data, this ratio is called the normalisation index (Ni). This normalisation index has therefore resulted in about a twenty-fold reduction (Ni = 22.3) in between-speaker variance. The Ni for all the data was slightly less than half this at 10.3. (\( \Delta_{c_{3/4}} = 90.7\%; \Delta_{c_{67}} = 6.8\% \).) It is interesting to note that the value of 22.3 far exceeds the Ni's of 12.9 and 7.0 for two other Chinese dialects of Zhenhai (Rose 1987: 350) and Shanghai (Rose 1993: 200). These dialects have six and five tones respectively (of which two are on stopped syllables), so the difference in Ni from Cantonese is unlikely to be related to the number of tones per se. Perhaps it relates to the nature of the contrasts: as has been shown, Cantonese tones 3 and 4 are very close, as are tones 5 and 6.

Acknowledgements

I should like to thank my two referees for their helpful comments.

References


Figure 3. Mean normalised F0 curves for the unstoppable tones of Cantonese citation monosyllables. Vertical bars show one standard deviation above and below mean.

Mean and standard deviation normalised F0 values were then calculated for the ten speakers, and these are shown in Table 3, together with mean normalised durations for each tone. The values in this table give linguistic tonic information about Cantonese tonal acoustics. From Table 3 it can be seen for example that the F0 onset value of Cantonese citation T1 is a little more than two standard deviations above the mean F0, and that about 66 out of a hundred individual speakers' values can be expected to lie between 2.45 and 1.95 standard deviations above the mean. The data in Table 3 are shown graphically in Figure 3. This figure plots the mean normalised F0 curves of the Cantonese unstoppable citation tones as a function of their normalised duration. Vertical bars show one standard deviation above and below the mean normalised F0.

Figure 3 shows that the mean normalised curves for the unstoppable tones lie between +/- 2 standard deviations (sds) above and below the mean. Since the low falling T2 usually involves changes in