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 linguistic-tonetic properties of the tones are specified from mean and 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 tonal contrasts.

INTRODUCTION

The aim of this paper is to give a quantified description of the linguistic-tonetic 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 representations. 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 (1994). The are also many descriptions of the acoustics of individual speakers, e.g. Hashimoto (1972:122-126); Vance (1976). However, there has to my knowledge been no study which specifies the linguistic-tonetic 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: 138,139). 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 tonemes that are described here. (The number of surface tonal contrasts on stopped 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 (T(one)1) sounds to be at the top of the speaker's normal pitch range; one (T5) lies in the middle of the pitch range, or just below, and one (T6) lies just below T5. Examples with transcription, from Yuan (1983: 185,186), are: (T1) [fu 55] *husband*; T5 [fu 33]; *trousers*; (T6) [fu 22] *father*. Of the two rising pitched tones, one (T3) has a pitch which onsets in the lowest third of the speaker's pitch range and rises into the upper third, e.g [fu 35] *bitter*; and one (T4) has a pitch contour similar to T3, but which only rises into the mid third of the speaker's range, e.g. [fu 13] *woman*. The only falling pitched tone (T2) falls through the lowest third of the speaker's pitch range e.g [fu 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 contrastivity. Two of the contrasts -- between 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 elicitation The corpus represented a compromise, given the phonotactics 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). 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 selected 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 1996). 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 phonotactic restriction in Cantonese on the combination of some stops with high vowel monophthongs e.g. *pi, *ti, *ki, *pu, *tu, diphthongs with high vowel offglides [ɛi, ɔu] 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.

T1	kɛi	*basis*	ku	*father's sister*	kɔ	*song*	ka	*to add*
T2	p^hɛi	*skin*	fu	*support*	p^hɔ	*woman*	p^ha	*scramble*
T3	tɔu	*gamble*	ku	*ancient*	tɔ	*hide*	ta	*hit*
T4	$tɕ^h$i	*similar*	fu	*woman*	p^hɔu	*hug*	sɛ	*society*
T5	pɛi	*back*	ku	*cause*	kɔ	*classifier*	pa	*tyrant*
T6	tɛi	*earth*	pɔu	*part*	wa	*speak*	pa	*cease*

Table 1. Corpus

Speakers The corpus was recorded by ten young native speakers of Hong Kong Cantonese, five male and five female, all students at the Australian National University. 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 when the standard deviation around the mean normalised curves becomes asymptotic to a given value. The speakers are referred to below as M(ale)/F(emale) 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 four repeats were analysed. There sounded to be some confusion between the morphemes pa5 *tyrant* and pa6 *to cease*. It sounded as if M4, M5 and possibly F5 read the character for *tyrant* (pa5) with tone 6, and as if F2 read the character for *to cease* (pa6) as pa5. Comparison using the tonal acoustics (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 at 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 =) 8640 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 pa5 and pa6 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 160 Hz, compared to an average female range extending from ca 165 Hz to 260 Hz. Although these mean ranges do not overlap, quite a few individual ranges do. M2's highest value, for example, is higher than four females' lowest values, and F2's lowest value is lower than four males' highest

values. 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

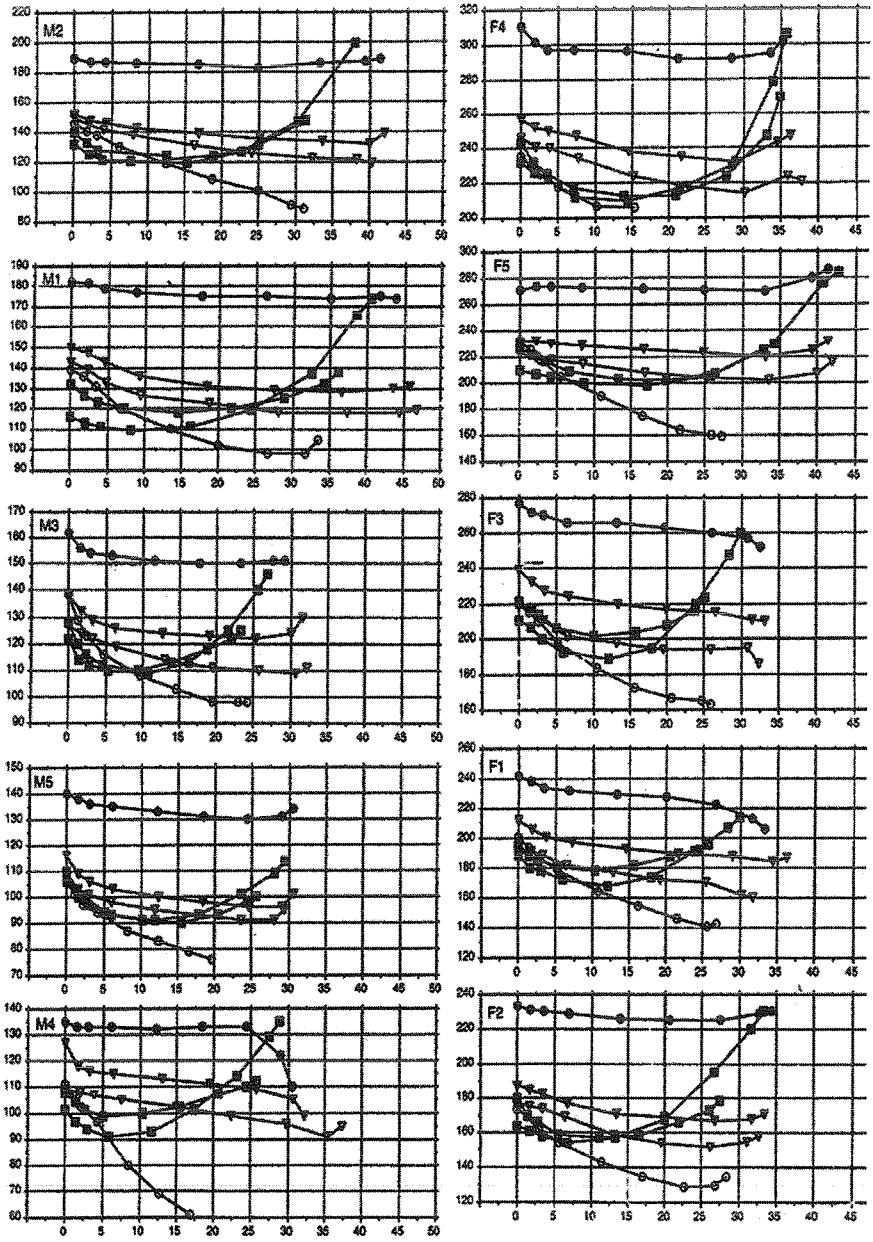

Figure 1. Individual Cantonese speakers' mean F0 values for their six tones on unstopped syllables.

are similar to M'2s 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 between speakers. The other, which is used in this paper, is acoustically based: to chose 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 NP's, as well as the values near the onset (0%, 5%, 10%) and offset (95%, 100%). The mean and standard deviation NP's 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.

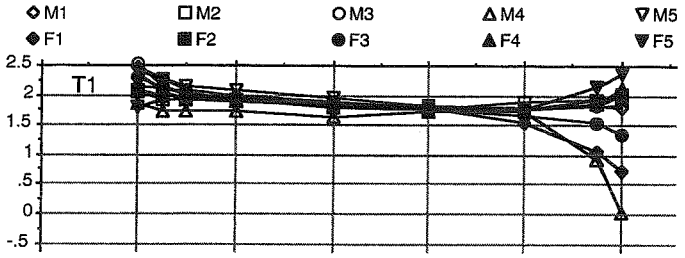

Figure 2. Normalised F0 values for ten Cantonese speakers' high level tone.

	M1	M2	M3	M4	M5	F1	F2	F3	F4	F5
$\bar{X}$	133.4	141.3	123.2	109.4	102.5	190.8	176.9	216.0	237.8	223.2
sd	22.6	23.7	15.3	13.7	15.7	20.9	27.1	26.1	30.6	26.5

Table 2. Values for normalisation parameters (Hz).

Thus the normalised value of speaker M1's mean F0 value of 182 Hz at 0% in T1 is ((182 -133.4) /22.6 =) 2.15, which means it lies just over 2 standard deviations above his mean F0. The results of the normalisation for T1 are shown in figure 2. 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 value 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) x 100 =) 107%.

	0%	5%	10%	20%	40%	60%	80%	95%	100%	D
T1	2.20	2.05	1.98	1.92	1.84	1.78	1.74	1.69	1.62	107
T2	0.28	-0.02	-0.27	-0.65	-1.21	-1.61	-1.85	-1.98	-1.81	81
T3	-0.31	-0.57	-0.72	-0.91	-0.95	-0.65	0.15	1.31	1.75	102
T4	0.14	-0.19	-0.37	-0.61	-0.73	-0.65	-0.39	-0.02	0.22	88
T5	0.75	0.49	0.34	0.19	0.01	-0.10	-0.19	-0.19	-0.07	111
T6	0.15	0.03	-0.09	-0.31	-0.56	-0.75	-0.84	-0.86	-0.81	111
T1	0.248	0.165	0.111	0.087	0.078	0.032	0.091	0.396	0.718	
T2	0.314	0.205	0.184	0.146	0.356	0.531	0.758	1.124	0.863	
T3	0.296	0.233	0.223	0.192	0.148	0.138	0.249	0.479	0.548	
T4	0.182	0.142	0.17	0.119	0.091	0.136	0.155	0.16	0.31	
T5	0.313	0.16	0.115	0.132	0.147	0.137	0.132	0.207	0.348	
T6	0.175	0.117	0.096	0.128	0.098	0.105	0.106	0.298	0.361	

Table 3. Means (above) and standard deviations (below) for z-score normalised F0 and duration for ten Cantonese speakers' citation tones. D = duration.

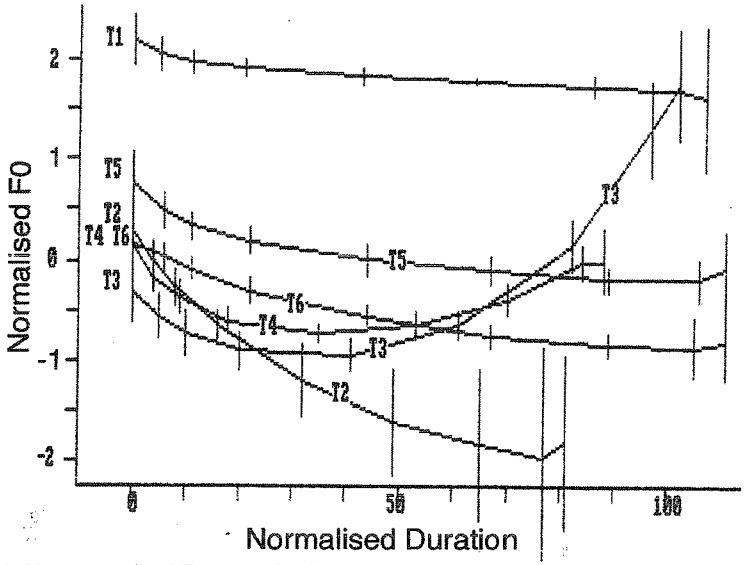

Figure 3. Mean normalised F0 curves for the unstopped tones of Cantonese citation monosyllables. Vertical bars show one standard deviation above and below mean.

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 tonetic 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 unstopped 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 unstopped tones lie between +/- 2 standard deviations (sds) above and below the mean. Since the low falling T2 usually involves changes in

below the mean, but at about 1, and that the normal F0 range therefore lies between 2 sds and –1sds. The normalised F0 of the three level pitch tones 1, 5 and 6 lies parallel, but not equidistant, with tones five and six separated by much less than either is from T1. T5 and T6 are separated by about 0.5 sds, whereas T5 lies between 1.5 and 2 sds lower than T1. The normalised F0 shapes of high-rising T3 and low rising T4 are similar, but T4 is truncated at about the same height as the offset of T5. Offset of T6 is at about the same height as the lowest point of T3 and T4. Four onset points can be distinguished: the highest for T1, the next highest for T5, then an onset for tones 2, 4 and 6, and the lowest for T3. It is unlikely that the relative position of all these onset points is tonally relevant, however. This is because they probably show differential conditioning by laryngeal features of their onset consonant. The onset value of T4 and T2, for example probably strongly reflects the [+spread glottis] feature, and it is likely that T4 is actually more congruent with T3 than their contours suggest.

Evaluation The effectiveness of the normalisation can be assessed by quantifying the amount by which the between-speaker variance in the raw (i.e. unnormalised) data is reduced in the normalised data (Earle 1975: 133ff). This is done by finding the ratio of the dispersion coefficients (DC) of the raw and normalised data, (this ratio is called the normalisation index (NI). The DC is 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%, 80%, 95%,100%). As far as the tonally relevant data were concerned, the DCs for the raw and normalised data were 92.5% and 4.1% respectively. This means that in the raw data the amount of between-speaker variance was almost as much as the overall variance, but that it was reduced to about 4% of the overall variance by normalisation. This normalisation 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. (dc_{raw} = 90.7%, dc_{norm} = 8.8%). It is interesting to note that the value of 22.3 far exceeds the NIs 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

Earle, M.A. (1975) An acoustic phonetic study of North Vietnamese tones. Monograph 11, Santa Barbara: Speech Communication Research Laboratories Inc..

Jones, D. & Woo (1912) A Cantonese Phonetic Reader. London: University of London Press.

Hashimoto, Oi-kan Yue (1972) Studies in Yue Dialects 1. Phonology of Cantonese. Cambridge Cambridge Universtiy Press.

Ladefoged, P. (1997) "Instrumental techniques for fieldwork." In Hardcastle & Laver (eds.) Handbook of Phonetic Sciences. London: Blackwell, 137-166.

Mathews, Stephen and and Yip, Virginia (1994) Cantonese A Comprehensive Grammar. London and New York: Routledge.

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

Rose, P. (1987) "Some considerations in the normalisation of the fundamental frequency of linguistic tone." Speech Communication 6/4: 343-352.

Rose, P. (1993) "A Linguistic-Phonetic Acoustic Analysis of Shanghai Tones." Australian Journal of Linguistics 13: 185-220.

Rose, P. (1996) "Aerodynamic Involvement in Intrinsic F0 perturbations – Evidence from Thai Phake." In McCormack & Russel (eds.) Proceedings of the Sixth International Conference on Speech Science and Technology. Canberra: ASSTA, 593-598.

Yuan Jiahua et al. (1983) Hanyu Fangyan Gaiyao [A Survey of Chinese Dialects]. 2nd ed., Peking: Wenzigaige Chubanshe.