THE ACOUSTICS AND PROBABILISTIC PHONOLOGY OF SHORT STOPPED-SYLLABLE TONES IN HONG KONG CANTONESE

Phil Rose

Phetics Laboratory, Linguistics Program, (Arts) Australian National University
philip.rose@anu.edu.au

Abstract

Mean fundamental frequency and duration data are presented for citation tones of Hong Kong Cantonese on short stopped syllables for five male and five female young native speakers. The relationship between the acoustics of the speakers' stopped and unstopped tones is examined and it is shown that the low-stopped tone F0 is for most speakers acoustically more similar to the low falling unstopped-tone than to the low level unstopped-tone. It is furthermore shown with likelihood ratios that, contrary to current assumptions concerning the phonology of Cantonese, and the notion of phonological categoricality in general, it is not possible to state for Cantonese that, on the basis of its F0, the low stopped-tone is related to any single unstopped tone.

1. Introduction

Many tone languages of East and South East Asia – this includes most varieties of Tai, Vietnamese and Sinitic – show apparent allotony conditioned by the phonological structure of the Rhyme. Specifically, allotones in these languages are typically conditioned by the absence or presence of a stop in the Coda. Such allotones are often called *checked*, or *dead*, or, as in this paper, *stopped*. In addition, many varieties, especially of Tai and Cantonese, show conditioning by the length of the nuclear vowel, and these allotones can be called "long- and short-stopped" tones. The form of allotones which occurs on stopped syllables has not received much attention, because the realisation of the usually more complex unstopped tone system is considered more interesting and basic. Their relationship has, however, been studied in detail for the Southern Thai dialects of Pakphanang and Hua Sai (Rose 1996a) and Roiphibun (Thompson 1998), and the NW Thai dialect of Phake (Rose 1990).

The aim of this paper is to give a quantified description of the acoustic properties (fundamental frequency and duration) of short stopped citation-form tones in Hong Kong Cantonese – tones which occur on stopped syllables with short nuclear vowels. The linguistic-tonetic properties of the Cantonese unstopped six-way system were quantified by Rose (2000) from the acoustics of 10 speakers. This paper quantifies these speakers' short stopped tones, and then compares the stopped and unstopped forms. There are many auditory descriptions of Cantonese tones in the literature, expressed in various representations. The are also many descriptions of the acoustics of individual speakers, e.g. Hashimoto (1972:122-126); Vance (1976). As far as I know, no multispeaker studies exist for the stopped tones, and, apart from Rose (2000), none exist for the unstopped tones either.

2. Phonology

As with most other stopped-tone languages, the number of surface tonal contrasts on citation monosyllables in Cantonese depends on the structure of the Rhyme - specifically the absence or presence of a syllable-final stop (p, t, k) in the Coda. On unstopped syllables, with phonologically at least two sonorants in the Rhyme, conservative Cantonese shows a six-way contrast (between tones with high, mid and lower-mid level pitches, two rising pitches, and a low falling pitch). On Cantonese stopped syllables, on the other hand, four different tones are usually recognised. (I use the term *tone* here analogously to *phone*, in its proper sense of audibly different pitch shapes that one might want to consider as potential allotones.) Examples with transcription, from Yuan (1983: 181), with glosses added, are: [sɪk 5] 識 *to know*; [siːt 33] 泄 *to leak*; [sɪk 2] 食 *to eat*; and [fat 5] 忽 *unexpectedly*; [faːt 33] 法 *way*; and [faːt 22] 乏 *tired out*. Note that there is some complementary distribution between vowel length and pitch height: long vowels followed by a stop cannot occur with a high pitch; and short vowels followed by a stop cannot occur with a mid pitch. This paper will describe the acoustics of the two short-stopped tones, since for the speakers used in this study they were relatively unproblematic. In contrast, the realisation of their long-stopped tones involved between-speaker systemic and incidential differences (for these terms see Rose (2002: 186-188). For example, some speakers did not contrast mid and low long-stopped tones, and for those that did, some had different tones for the same morpheme. Because of these complicating factors a description of the long-stopped tones is best postponed for another paper.

As far as the tonological integration of the stopped tones is concerned, the consensus, represented for example by Mathews and Yip (1984: 22) is that they are allotonically related, by virtue of phonetic equivalence, to the three unstopped level tones. Thus the high short-stopped tone is explicitly related to the unstopped high level tone (described as [55] in most sources); the mid long-stopped tone is related to the unstopped mid level [33] tone, and the short- and long-stopped low tones are related to the unstopped lower-mid [22] tone. This view looks back to observations on auditory similarity made since at least the 40's, in particular Chao (1947).

3. Procedure

3.1 Speakers

The central idea in the description of tonal acoustics is to obtain data that are representative of the variety in question,

Proceedings of the 10th Australian International Conference on Speech Science & Technology
Macquarie University, Sydney, December 8 to 10, 2004.

Accepted after full paper review

here Hong Kong Cantonese. Obtaining representative data means, firstly, using several speakers, since with a single speaker there is no way to know which of the acoustic characteristics, even after normalisation, reflect the variety and which are characteristic of the individual (Rose 1987). Accordingly, the corpus was recorded by ten young native speakers of Hong Kong Cantonese, five male and five female, all students at the A.N.U. The speakers are referred to below as CM(ale) / CF(emale) 1-5. This is actually still four speakers short of the preferred minimum for phonetic descriptive work suggested by Ladefoged (1997: 140). Ladefoged does not justify this particular minimum number, and perhaps a better indication of how many speakers are necessary is given statistically, by when the standard deviation around the speakers' mean normalised curves becomes asymptotic to a given value.

Table 1: Corpus. (YR = Yale romanisation)

High stopped tones				Low stopped tones			
YR	gloss	character	typical realisation	YR	gloss	character	typical realisation
bat	*Un-, non-*	不	[pat	baht	*uproot*	拔	[pat
dak	*virtue*	德	tak	duhk	*study*	讀	tok tɔk tɔʊk tɒk
dik	*target*	的	tɪk tek teɪk	dihk	*enemy*	敵	tek tɛk
guk	*valley*	谷	kʊk kok]	guhk	*office*	局	kok kɔk kɒk kɔʊk]

3.2 Corpus

The stopped corpus is shown in table 1. Forms are given in the Yale romanisation, together with their Chinese character, and typical segmental realisations are also shown. Morphemes with high stopped tones are on the left; low are on the right. It can be seen that the high and low short-stopped tones were represented by four CVC morphemes each. The discrepancies in vowel quality between the actual realisation and the romanisation should be noted. The Yale romanisation implies a phonemisation whereby the highest vowels in short-stopped syllables are grouped with high vowels in unstopped syllables, whence the transcribed "i" and "u". Several sources have pointed out that the highest vowels in the Cantonese short-stopped syllables are in fact more open, as noted in table 1 for these speakers, and that they are consequently to be identified as allophones of phonemes other than /ii/ and /uu/ (Li 1985: 28-30, Pulleyblank 1997: esp. 200-202).

Since one of the aims of this paper is a comparison between stopped and unstopped tones, the choice of stopped morphemes was basically determined by considerations of comparability with the unstopped corpus. The unstopped corpus had CV structure, with V balanced for vowel height, within the constraints of Cantonese phonological structure, with each tone containing equal numbers of high / mid-high [i/ɛɪ, u/ɔʊ] and low/mid-low [a/ɔ] vowels. However, it is not possible to select a short-stopped corpus that is totally comparable with respect to phonetic height in high vowels. This is because, as already mentioned, in Cantonese, the highest vowels in the short-stopped tones are supposed to be lax or half-close [ɪ/e] and [ʊ/o], whereas the highest vowels in the unstopped corpus were higher: [i/ɛɪ, u/ɔʊ]. Nevertheless, it can be seen that, within these constraints, the stopped corpus is well balanced for vowel height, with low / mid vowels and (relatively) high vowels.

As can be seen in table 1, all stopped-tone morphemes were chosen with unaspirated initial stops. This means that it is highly comparable to the unstopped corpus, where all morphemes began with voiceless unaspirated initial stops, except those with the low-falling tone (because unaspirated stops are disallowed phonotactically in this tone, tokens in the unstopped corpus had aspirated stops).

3.3 Elicitation

Since this was part of a larger experiment to elicit data for HK Cantonese tones on both stopped and unstopped syllables, the eight morphemes in table 1 were randomly combined with additional long-stopped forms, and the unstopped forms already described, 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 reading each character. Subjects read the set of characters at least four times, and were recorded on professional equipment in the studio of the Faculties' (then) Department of Linguistics at the A.N.U. Tokens from the first four repeats were analysed. There were thus in all (10 speakers * 8 morphemes * 4 repeats = 320 syllables measured.

3.4 Measurement

Not all the F0 on a syllable is tonally relevant. F0 on a syllable-initial sonorant, for example, is best discounted, whereas a sonorant Coda in the Rhyme carries tonally relevant F0 (Rose 1992). Details of F0 at offset are also not necessarily always tonally relevant, and can reflect the way in which a speaker typically ceases phonation (Rose 1996b). Thus, as with the sampling of tonal F0 in unstopped tones, some consideration has to be given to determining an appropriate sampling base for tonal F0 in stopped tones. The point at which the sampling base can be adjudged to start is not problematic, and can be taken to be the first strong glottal pulse after the release of the Onset consonant (here a voiceless unaspirated stop). However, the offset point of the sampling base is not so clear due to one aspect of stopped tones that is only infrequently mentioned in the descriptive and phonological literature, namely their source coarticulation. Fibre-optic studies have shown that syllable-final stops in Cantonese (and in other Sinitic varieties) are coarticulated with ventricular adduction, which is assumed to be part of the articulation of a glottal stop. The adduction begins typically before the supraglottal closure, becoming more rapid afterwards (Iwata et al 1981: 47). Acoustically, laryngealisation was observed in the data typically manifesting as jitter and/or shimmer, and, consistent with the fibre-optic descriptions, beginning somewhat before the onset of the supralaryngeal closure of the Coda stop as indicated by F-pattern transitions. Figure 1, which shows the morpheme baht拔 *to uproot* from two different male speakers, illustrates this. It can be seen for both tokens that the onset of jitter and/or shimmer occurs a couple of centiseconds before the

Accepted after full paper review

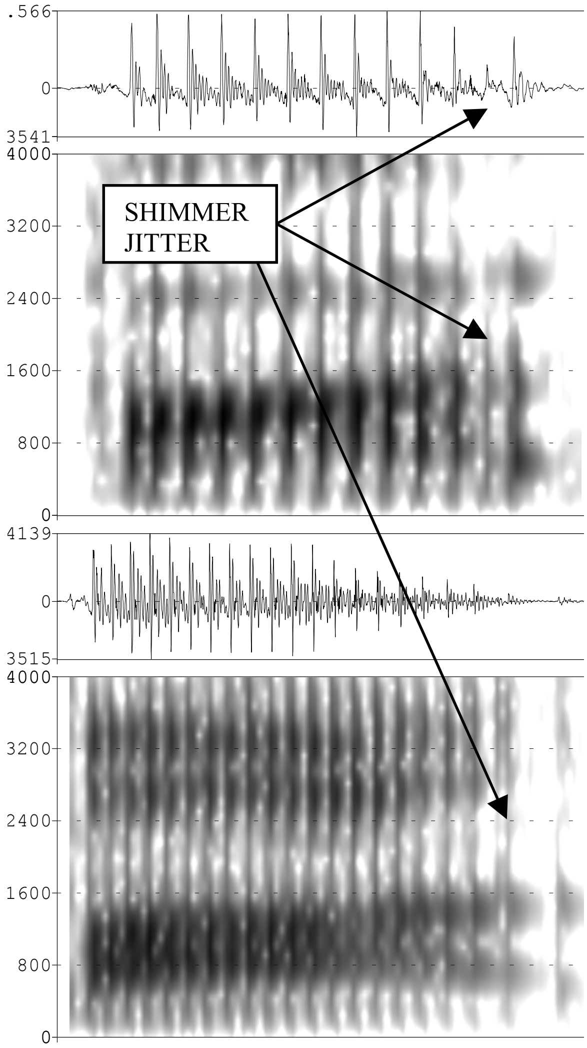

Figure 1: Wide-band spectrograms and waveforms of the low stopped-tone word baht 拔 *to uproot* in two different male speakers, showing typical jitter and shimmer associated with laryngealisation of syllable-final stops. Note also change in F-pattern indicating presence of an oral stop (in this case t).

onset of the Coda stop – whose existence can be inferred by the clear drop in F1- can be assumed. (Actually, it is not clear at what point the onset of hold occurs, since there is no source energy after the last strong glottal pulse). The laryngealisation is thus similar to place-relevant formant transitions in being properties of the consonant occurring before its actual hold. Since both laryngealisation and supralaryngeal closure are properties of the final consonant, there are two possibilities for the choice of offset to F0 sampling base: point of supralaryngeal closure, or onset of laryngealisation. Because there is only a very small difference in time between the onset of laryngealisation and the presumed onset of the hold phase (it was found to be of the order of ca. 1-2 csecs.), there will be little difference in duration between these two sampling bases. The former was chosen because it involved less direct measurement of F0 when the automatic extraction failed as a result of the laryngealisation.

Tokens were digitised at 10K, and *Praat*'s 'Edit' command was used to generate aligned wave-form, spectrogram and superimposed F0 trace. The F0 sampling base was determined by inspection of the spectrogram and wave-form. The adjudged onset was the first strong glottal pulse of the Rhyme; adjudged offset was the point in the wave-form where the laryngealisation was first observed as a clear discontinuity, either in peak to peak amplitude or period. F0 was then sampled automatically at 20% points of the sampling base, a sampling frequency assumed high enough to resolve the details of the F0 time course. Pitch-synchronous measurements were made by hand in cases where the automatic F0 extraction was obviously incorrect, or failed as a result of the laryngealisation.

4. Results

All speakers produced a clear contrast between high and low short-stopped morphemes. The high morphemes were generally said with high level pitch; the low with a low level, or low falling pitch (although it is difficult to be certain about pitch contour when the stimulus lasts for such a short time). As shown in table 1, there was a certain amount of between-speaker variation in vowel quality for the non-low vowels, although again it is difficult to be sure about small differences in vowel height because of the short stimulus.

For each of the ten speakers, mean F0 and duration values for the high and low stopped-tones were calculated from the means of their four separate morphemes. Figure 2 shows the mean tonal F0 of each of the ten speakers' high and low short-stopped tones as a function of mean absolute duration. To show the relationship between stopped and unstopped tones, mean F0 from four of the unstopped tones (taken from Rose 2000) is also shown. The four unstopped tones plotted are the high, mid and lower-mid level tones, and the low falling tone (it is obvious which is which). F0 is plotted as a function of absolute duration in order to show between-tone durational differences, but the duration scales have been equalised between-speakers. The vertical axes have been given differing ranges and scales in order to make the speakers' F0 ranges visually comparable.

Figure 2 shows that the F0 of the upper stopped-tone is either relatively flat or slightly falling, and in all except one speaker (CF4) has a short offset perturbation with a greater or lesser increase in F0 derivative. Its Rhyme duration is rather short ($\bar{x}$ = 10.6 csec), but there are big differences in the duration relative to that of the high unstopped tone. On average the unstopped-tone Rhyme is about 3.5 times longer than the short. But the range is big: for example, CM1's stopped-tone Rhyme duration is only a fifth of his unstopped; whereas CF3's is just under half.

The low stopped-tone has a falling F0 shape, with a slight offset perturbation visible for some speakers. Its mean Rhyme duration is 11.4 csec, and for a given speaker it is about the same as, or slightly longer than, the duration of the speaker's high stopped-tone.

Accepted after full paper review

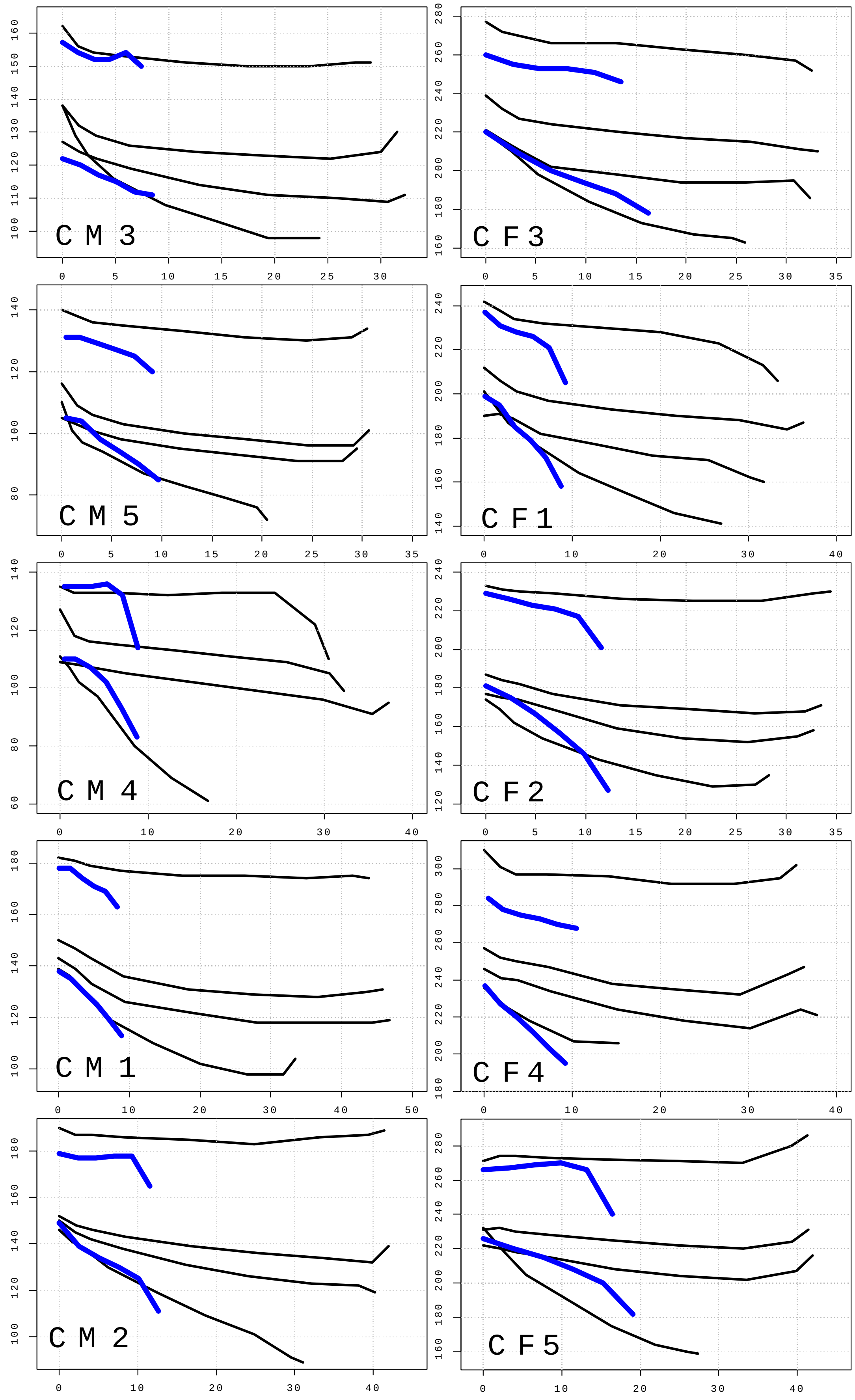

Figure 2: Individual Cantonese speakers' mean F0 values (Hz) for their high and low short stopped tones (thick lines), and high, mid, lower-mid and low falling unstopped tones (thin lines). Horizontal axis = duration (csec.)

4.1 Comparison with unstopped tones

4.1.1 High stopped-tone

The F0 of the high stopped-tone is most similar to that of the high-unstopped tone, but the relationship differs depending on the speaker. For three speakers (CM3, CM4, CF5) it appears to be at the same height as the high unstopped-tone; for the rest it lies somewhere between the high unstopped-tone and half-way between the high and mid unstopped-tones (CF4). I checked to see if these differences related to any between-speaker differences in intrinsic vocalic F0 height, but they did not appear to do so (i.e. speakers with [ʊ] and [ɪ] did not necessarily have F0 in their upper stopped-tone more similar to their high unstopped-tone).

Appealing to the time-hallowed principles of phonetic similarity and complementary distribution, the high stopped-tone can be considered as allotonically related to the high unstopped tone, (although it would be nice to know what, if anything, is causing the between-speaker differences in its realisation, and thus the variability in the phonetic similarity to the high unstopped-tone).

4.1.2 Low stopped-tone

With the exception of CF5, figure 2 shows that the F0 of the low stopped-tone is clearly most similar to that of the unstopped low-falling tone, especially towards the former's end. As long as the falling F0 is extrinsic, and not in some way an intrinsic function of the final stop, this finding appears to be at odds with the generally accepted claim, mentioned above, that the low stopped-tone is allotonically related to the low-mid level tone. However, in the spirit of Probabilistic Linguistics and Phonology (e.g. Pierrehumbert 2003: 177-228), and partly in the spirit of the Reverend Bayes, it is still necessary to evaluate the evidence in support of this hypothesis. The proper way of doing this is with a likelihood ratio (Robertson & Vignaux 1995). The strength of evidence in favour of one hypothesis over another (for hypotheses cannot be evaluated on their own) is given by the ratio of the probabilities of getting the evidence under the competing hypotheses. This is shown at (1), where E = evidence and H_1, H_2 = competing hypotheses.

$$LR = p(E \mid H_1) \,/\, p(E \mid H_2) \qquad (1)$$

In this case, the evidence was taken as the mean F0 values of the low stopped-tone at two points - one early (20%) and one late (80%) - in its time course. The probability density was then calculated of these values assuming the distribution of (a) the low falling tone and (b) the low level tone at the corresponding time point. Normality was assumed for all distributions, their standard deviations being based on the number of items in the sample (usually four morphemes * four replicates = 16). The LR is then the ratio of the probability density assuming (a) to that assuming (b), values greater than unity giving support to the hypothesis that the F0 of the low stopped-tone is more similar to the falling tone, and values less than unity indicating the opposite. The strength of evidence is proportional to the magnitude of the LR. Unity or values close to unity indicate that the evidence would be equally likely under both hypotheses and that it is therefore of no use in evaluating them. (Forensically, equivocal evidence is a much underrated concept, since it is often assumed that if one hypothesis is not tenable, the alternative is. The possibility of equivocal evidence shows that not only is one hypothesis useless - both are.)

Table 2: Details of likelihood ratio evaluation of hypothesis that F0 of low stopped-tone is more similar to F0 of low level tone than to F0 of low falling tone. LS = low stopped-tone; LL = low level tone; LF = low falling tone; LR = likelihood ratio. est x̄ *sd* = estimated mean and standard deviation. F0 values in Hz. LR (LL) = likelihood ratio assuming low stopped-tone F0 is from the low level tone.

	Comparison at 20% of stopped tone					Comparison at 80% of stopped tone			
	LS x̄	LF est x̄ *sd*	LL est x̄ *sd*	LR		LS x̄	LF est x̄ *sd*	LL est x̄ *sd*	LR
Speaker									
M1	135	135 *6.2*	139 *5.3*	1.1		119	121 *4.3*	131 *4.1*	44.9
M2	139	139 *6.4*	145 *5.1*	1.5		125	122 *5.3*	135 *4.3*	7.7
M3	120	128 *4.7*	124 *4.5*	3.4 (LL)		112	111 *2.8*	118 *3.5*	5.7
M4	110	104 *8.7*	108 *3.7*	2.6 (LL)		93	86 *8.4*	105 *2.2*	4.5 E5
M5	104	100 *6.0*	103 *5.7*	1.3 (LL)		90	86 *3.4*	97 *2.9*	8.8
F1	195	193 *8.3*	189 *6.3*	1.1		171	173 *4.9*	183 *2.5*	2.1 E4
F2	175	165 *7.3*	175 9.5	2.0 (LL)		146	145 *5.9*	165 *6.8*	148.1
F3	209	207 *10.4*	211 *9.1*	1.1 (LL)		188	178 6.6	196 *7.4*	2.0 (LL)
F4	227	228 *13.0*	242 *7.4*	5.0		203	212 *12.4*	234 *7.2*	4.5E3
F5	220	218 *10.9*	220 *9.3*	1.2 (LL)		200	183 *6.6*	214 *7.7*	4.3 (LL)

Table 2 gives the results of the likelihood ratio testing. These are divided into two groups - the left for the comparison at the 20% of duration in the stopped-tone, and the right for comparison at 80%. In table 2 it can be seen for example that the LR of 1.1 for speaker M1's comparison at 20% means that one would be just about as likely to get his low stopped-tone 20% mean F0 value of 135 Hz, assuming that it had come from the distribution of his low falling tone (with its estimated mean and standard deviation at that point in duration of 135 Hz and 6.2 Hz), as assuming that it had come from the distribution of his low level tone (with its estimated mean and standard deviation at that point in duration of 139 Hz and 5.3 Hz.). The LR of 3.4 for M3's comparison at 20% shows that

Accepted after full paper review

one would be about three and a half times more likely to get his low stopped-tone F0 assuming it had come from the distribution of his low level tone rather than his low falling tone: this is weak evidence in support of the hypothesis that his stopped tone F0 is more similar to his low level tone at that point than to his low falling tone.

From table 2 it can be seen that early on in its time course there is no real evidence for the stopped-tone F0 being more similar to either the level or the falling tone: all LRs are below 5 and most are very near to unity. Late in its time course, however, there is strong evidence for four out of the ten speakers (M4, F1, F2, F4), and moderate evidence for one more (M1), that their low stopped-tone F0 is more similar to their low falling tone. For the remaining five, with LRs below 10, again there is no real evidence either way.

The main point here, perhaps, is that it is not that for five out of the ten Cantonese speakers the low stopped-tone can be considered as being allotonically related to the low falling tone. Rather it is that are no grounds for categorically saying that the Cantonese low stopped-tone F0 is related to either one or the other unstopped tones. We are accustomed to the notion of phonology being categorical, but in the case of where the low stopped-tone belongs, the phonology appears not to be thus. For some speakers it is clearly more related to the falling tone; for others it is clearly not related to either.

5. Acknowledgments

I thank my informants and my referees for their (different kinds of) patience.

6. References

Chao Y.R. (1947) *A Cantonese Primer*. Harvard University Press: Cambridge, Mass.

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

Iwata, Ray Sawashima, Masayuki & Hirose, Hajime (1981) “Laryngeal Adjustments of Syllable-Final Stopes in Cantonese”, *Ann. Bull. Research Institute of Logopedics and Phoniatrics* 15: 45-54.

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

Li Xingde 李行德 (1985) 廣州話元音的音質及長短對立 [Vowel Quality and Length Contrast in Cantonese] 方言 *Fang Yan* 1: 28 – 38.

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

Vance, Timothy J. (1976) "An experimental investigation of Tone and Intonation in Cantonese". *Phonetica* 33, 368-392.

Pierrehumbert, Janet B. (2003) 'Probablilistic Phonolgy: Discrimination and Robustness'. In Bod, Rens Hay , Jennifer and Jannedy, Stefanie *Probabilistic Linguistics*. MIT Press: Cambridge,Mass.: 177-228.

Pulleyblank, Edwin G. (1997) “The Cantonese vowel system in historical perspective”, in Wang Jialing & Norval Smith (eds.) *Studies in Chinese Phonology*. Linguistic Models 20. Mouton de Gruyter: Berlin, New York: 184 – 217.

Robertson, B. and Vignaux, G.A. (1995) *Interpreting Evidence*. Wiley: Chichester.

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

Rose, P. (1992 "Bidirectional Interaction between Tone and Syllable-Coda: Acoustic Evidence from Chinese". In J.Pittam (ed.) *Proc. 4th Australian Intl. Conf. on Speech Science and Technology*, Australian Speech Science and Technology Association: 292-297.

Rose, P. (1990) "Thai-Phake Tones: Acoustic, Aerodynamic and Perceptual Data on a Tai Dialect with Contrastive Creak". In R. Seidl (ed.) *Proc. 3rd Australian Intl. Conf. on Speech Science and Technology*, Australian Speech Science and Technology Association: Canberra: 394-300.

Rose, P. (1996a) “The Realisation of Stopped-Syllable Tones in Hua Sai and Pakphanang.” In McCormack & Russel (eds.) *Proc. 6th Australian Intl. Conf. on Speech Science and Technology*, Australian Speech Science and Technology Association: 605-610.

Rose, P. (1996b) “Aerodynamic Involvement in Intrinsic F0 perturbations – Evidence from Thai Phake.” In McCormack & Russel (eds.) *Proc. 6th Australian Intl. Conf. on Speech Science and Technology*, Australian Speech Science and Technology Association: 593-598.

Rose, P. (2000) “Hong Kong Cantonese Citation Tone Acoustics – A Linguistic-Tonetic Study.” In S. Barlow (ed.) *Proc. 8th Australian Intl. Conf. on Speech Science and Technology*, Australian Speech Science and Technology Association: 198-203.

Rose, P. (2002) *Forensic Speaker Identification*. Taylor and Francis: London & New York.

Thompson, N. (1998) “Tone Sandhi Between Complex Tones in a Seven Tone Southern Dialect“. In Robert Mannell and Jordi Robert-Ribes (eds.) *Proc. 5th Intl. Conf. on Spoken Language Processing*. The Australian Speech Science and Technology Association. Vol 2: 53-56.

Yuan Jiahua et al. 袁家華等著 (1983) 漢語方言概要 Hanyu Fangyan Gaiyao [A survey of Chinese Dialects], 2nd ed, Wenzigaige chubanshe: Peking.

Accepted after full paper review