LINGUISTIC PHONETIC ASPECTS OF SHANGHAI TONAL ACOUSTICS

Phil Rose

Department of Linguistics (Faculties) A.N.U.

ABSTRACT Mean fundamental frequency and duration data are presented for the tones of 3 female and 4 male speakers of Shanghai dialect. Normalised F0 shapes for Shanghai and Zhenhai tones are compared, and a linguistic tonetic contrast demonstrated between both the falling and rising tones. The importance of retaining durational relationships in normalisation is demonstrated.

INTRODUCTION

Several auditory and acoustic descriptions of Shanghai tones exist (eg. Sherard 1972; Zee & Maddieson 1979), but up to now no-one has attempted to quantify the linguistic phonetic properties of Shanghai tones by deriving normalised acoustic representations from the tones of several speakers. That is the first aim of this paper. The second aim is to demonstrate the indispensibility of such data for investigating the nature of linguistic tonetic variation.

Shanghai contrasts 5 tones on citation monosyllables or monosyllabic words. The following, from, respectively, Norman (1988:202); JSS (1960:116); Zee & Maddieson (1979:71); and Sherard (1972:59-63) is a representative group of pitch descriptors for the 5 tones. **Tone 1**: 42; 53; [HL]; "...sharply falling...".**Tone 2**: 35; 34; [MM$^{\uparrow}$]; "Moderately high level.... ...in careful...speech...[has a] short audible rise...toward the end of phonation." **Tone 3**: 24; 13~14; [LM$^{\uparrow}$]; "[a]...low register. ...tone...with end point higher than onset..." **Tone 4**: 55; 5; [H]; "Moderately high pitch with no discernable change..." **Tone 5**: 23; 2; [LM$^{\uparrow}$]; "Low register... with a short rise in pitch..."

As in many Wu dialects, differences between Shanghai citation tones involve several pervasive co-occurring and recurrent phonetic features in addition to pitch. Of these the most salient are voice quality (i.e. phonation type), length, and voicing offset. Tones 3 & 5 are described as having breathy voiced vowels; tones 4 & 5 are described as short, with glottal stop offset; and tones 1, 2 & 3 are described as long. Vowels in tones 4 & 5 also tend to be less peripheral, although this is not usually mentioned.

PROCEDURE

4 males and 3 females, aged between 30 and 50 and considered by their peers typical speakers of metropolitan Shanghai dialect, were recorded on professional equipment in the A.N.U. Linguistics Department phonetics laboratory. The corpus, listed below, was designed to control strictly for intrinsic vocalic amplitude and fundamental frequency (F0), and consonantal perturbatory effect on F0. It consisted of Chinese characters exemplifying the 5 tones on four CV monosyllables (where C was a voiceless unaspirated bilabial/dentalveolar plosive, and V was either an open [a ɔ ɒ] or close [i ɩ u] vowel).

Corpus

Tone 1: ti 'low', tu 'capital', tɔ 'knife', pa 'dad'; **Tone 2:** ti 'choose', tu 'gamble', tɔ 'arrive', pa 'worship', **Tone 3:** ti 'lift', tu 'stomach', tɔ 'flee', pa 'arrange'; **Tone 4:** tɩʔ 'target', tɒʔ 'earnest', paʔ 'eight', pɒʔ 'north'; **Tone 5:** tɩʔ 'enemy', tɒʔ 'read', paʔ 'white', pɒʔ 'thin'.

Prompt cards were prepared, each with the 4 different tokens of the same tone. All 4 possible morphemic sequences were used for each tone. This gave a total of 5 tones x 4 morphemes x 4 permutations = 80 tokens. The cards were randomised, and placed one at a time in front of the informant for him to read. The informants paused between reading each token on the card, but typically a pitch difference was audible between the first and last tokens on each card, as if the speaker were treating each card intonationally as a discourse unit with its own declination.

F0 and duration were measured from narrow-band expanded scale (ca. 1 KHz: 3 cms), and wide-band spectrograms respectively. F0 was sampled at a rate high enough to resolve details of its time course, namely 10% points of effective vocalic duration (i.e from F0 onset to F0 minimum /inflection point in tones 1 & 4, and to F0 peak in tones 2, 3 & 5). Mean values for all 16 tokens of each tone were calculated, and are shown for the 7 speakers in fig. 1.

RESULTS

Fig. 1 shows that all speakers share a configuration characterised by an abruptly falling F0 for T1, which traverses the whole of the speakers' range, intersecting the converging shapes of the 2 other long tones. The F0 shapes of the 2 short tones lie within the range bounded by the onset of T1 and the lowest point of T3. The abrupt fall in F0 after peak in all except T1 (which also occurs but is not shown for M4) is an acoustical reflex of a syllable-final [ʔ] which, contrary to existing descriptions, clearly characterises the long tones 2 & 3 as well as the short tones. Two sets of onset points can be distinguished. The long tones show 3 roughly equidistant high, mid and low onset points, and the short tones onset either at (i.e. at a point not significantly different from) or between the high and mid points (T4), or at or between the low and mid points (T5). With 2 exceptions (T5, M3; T1,F2) all speakers have similar abruptly falling F0 onsets over the first 10-15 csec, with approximately equal rates of fall for each tone. This, together with the shape of T1 tends to give the whole configuration an overall falling appearance. Duration relationships are the same for all speakers: T2, T3 > T1 > T5 > T4.

Several between-speaker (BS) differences can be noted. T3 and especially T2 show differences in contour. For T3, speakers differ most importantly with respect to the prolongation of low F0 values, and this correlates with a pitch difference between (low) dipping (M1, M4) and (low) rising (others). The latter group also differ with respect to the timing of an inflection point, i.e. point of most abrupt change in F0 derivative: relatively early in M2, M3, F1 (and M1, M4); later in F2, F3. T2 shows the greatest amount of BS variation, the main difference being in the maximum F0 attained after inflection point. In M3, this value lies considerably above T2 onset value, in F1 it is about the same. M1, M2, M4, & F2 are characterised by F0 peaks located between values at inflection point and onset, and F3's value does not differ significantly from her value at inflection point. Large, apparently sex related BS variation also obtains in the relative values of T1 offset and T3 inflection point. In males the difference is 1/4 or less than the range between T1 high onset point and T3 inflection point; in females it is 1/2 or greater. Since the speakers' T3 all sound to occupy the same position in their pitch range, it seems best to consider the T1 offset value as the speaker-dependent variable.

NORMALISATION

In order to obtain a quantified representation of the F0 characteristic of Shanghai tones the speakers' mean F0 values were z-score normalised (Rose 1987). The normalisation parameters (mean, sd) were derived from values at all sampling points on all 5 tones except at onset, and offset in tones 2 - 5). Fig. 2 shows the results for this normalisation, which reduces the amount of variance in the raw data by a factor of 6.9. (This value - the normalisation index - is the ratio of the dispersion coefficients of the raw and normalised data. The dispersion coefficient (DC) is the ratio of the mean BS variance to overall sample variance, and is a measure of the degree to which individual speakers' values cluster. The raw DC for the 7 Shanghai speakers' 5 tones was [2297.9/ 1851] = 124% ; the corresponding normalised DC was [0.175/ 0.974] = 17.96%.) Fig. 2 shows that the linguistic resolution of the normalisation is satisfactory, in that no overlap occurs between the normalised F0 of tones of different phonological categories unless it also characterises an individual's raw F0 values. The onsets of tones 1, 4, and 2 differ significantly from each other and from those of tones 3 & 5, which share the same onset. Peak values for the rising tones 2, 3 & 5 do not differ significantly. The point of maximum clustering occurs near the inflection point of the low tones 3 & 5 (at 40% of duration), where speakers' values fall within the comparatively narrow range of between 0.3 and 0.4 standard deviations. Differences in clustering are also apparently related to the length distinction. Short tones 4 & 5 cluster better at onset than the comparable long tones 1 & 3, but the long rising tones 2 & 3 show considerably less variance at peak

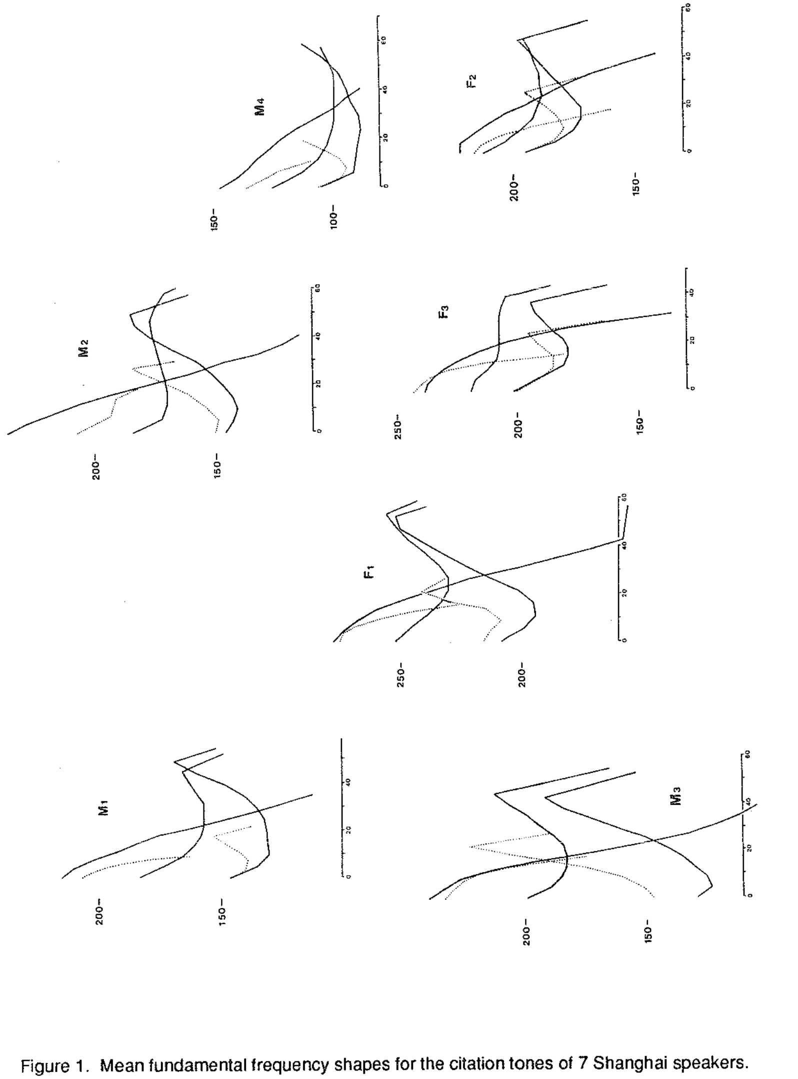
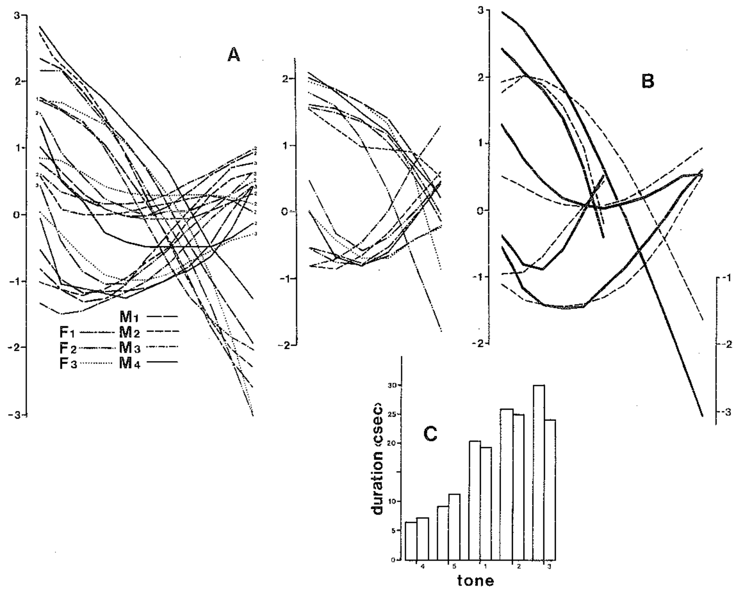


Figure 1. Mean fundamental frequency shapes for the citation tones of 7 Shanghai speakers.

Figure 2. Z-score normalised F0 values for Shanghai tones (A); comparison of mean normalised F0 shapes for Shanghai and Zhenhai tones (B); comparison of mean duration values for Shanghai (on right) and Zhenhai tones (C).



than the short tone 5. Greater variance is also associated with the offset value in the short tone 4 than with the long tone 1.

LINGUISTIC PHONETIC COMPARISON

The mean and standard deviation of the normalised curves indicate the magnitude of expected variation in the normalised F0 of Shanghai dialect (4 standard deviations around the mean will include about 95% of all normally distributed observations). The other main use of this kind of linguistic acoustic-phonetic representation is in comparisons with normalised representations of other varieties, in order to ascertain the nature of linguistic tonetic variation - at least as far as tonal F0 goes. A normalised representation of the tonal F0 of Zhenhai dialect was given in Rose (1987:350). Zhenhai lies about 100 miles south of Shanghai, belongs to the same (Wu) group, and is mutually intelligible with it. Apart from an additional low convex tone, Zhenhai citation tones sound very similar to those of Shanghai. However, two aspects of their pitch contour keep them apart. The Zhenhai falling and low rising tones both typically have short initial level components, (this gives the latter a dipping pitch), whereas the Shanghai falling tone (T1) sounds more abrupt than Zhenhai, and the long low rising tone (T3) usually rises without a dip. (Recall, however, that both M1 & M4 have dipping versions of T3.) It is interesting to note that a similar contrast in the pitch contour of the falling tones in Shanghai and Ningpo (a town about 20 km from Zhenhai) was documented 60 years ago (Chao 1928:chart 4/2).

Fig. 2B shows the results of a comparison between the normalised F0 shapes of Shanghai (bold) and Zhenhai. (An attempt was made to achieve comparability between normalisation parameters (mean & standard deviation) by excluding the long falling tones in both varieties from their calculation, as well as the extra Zhenhai convex tone. However, it must be pointed out that such attempts still suffer from the circularity imposed by using normalisation parameters derived from the tones themselves (Rose 1988). Does, for example, the Shanghai falling tone really have a lower offset than in Zhenhai, or is it that the Shanghai T3 is located higher?) The most conspicuous difference between the two dialects in fig. 2B is in the long falling tone, which starts higher and falls much lower in Shanghai. There is also a contour difference in that the abruptly falling Shanghai tone lacks the Zhenhai shoulder. (This is part of a second, general difference between both sets, in that all the Shanghai tones have greater negative F0 onset perturbations than in Zhenhai.) It is of course tempting to see in this difference a reflection of the abovementioned contrast in pitch contour between the falling tones of both varieties.

Ignoring the onset perturbations, all the other tones are resolved in surprisingly similar fashion, and seem to constitute occurrences of the same linguistic tonetic units in the two different dialects (although there may be differences between the two varieties in the amount of variance around the mean). Thus the second putative auditory difference between Shanghai and Zhenhai in low dipping vs. low rising pitch for the long low tone does not appear to be reflected in the normalised data. However, the identity of a contour is assured also by its duration, and a comparison of the durations of the tonal F0 in Shanghai & Zhenhai reveals that a contour difference between the two varieties has in fact been obscured by equalising durations during normalisation. Fig. 2C shows the mean durations of the two sets of tones for the 7 Shanghai and 7 Zhenhai speakers. It can be seen that all durations are very similar, except for the long low tone which is about 6 csec significantly longer in Zhenhai. (The difference extends to 7 csec if the auditorily dipping tones of M1 & M4 are excluded). Now, if allowance is made in normalisation for this durational difference, the Shanghai tone can be seen to rise earlier, thereby reflecting the second difference in pitch noted above for the two varieties. The obvious implication of this finding is that success of linguistic tonetic comparison may depend on incorporation of a time base into the normalisation of F0 (and amplitude) shapes. It was fortunate with these data that the absolute mean durations were so comparable (this enabled us to say, too, that the difference between the F0 shapes of the two falling tones was not due to different durations). It is likely, however, that in some cases the time-varying acoustic correlates of tone will have to be compared as functions of normalised - as opposed to equalised - duration.

One final observation must be made on the apparent nature of linguistic tonetic variation as revealed by this study. Of the 7 Shanghai speakers, 2 had Zhenhai-like low dipping T3. Among the 7 Zhenhai speakers, 1 (ZSC) had a Shanghai-like low rising tone and also a falling tone with a much lower offset. This suggests a model of discrete linguistic phonetic tones that are differentially distributed within a dialect continuum. Obviously, more comparison has to be carried out between normalised values before it can be demonstrated whether tones are in fact discrete at this level.

ACKNOWLEDGEMENTS This work is part of a larger project on linguistic tonetic variation supported by an ARC grant which I gratefully acknowledge. My thanks also to Mark Donohue for processing much of the Shanghai data.

REFERENCES

Chao Yuen Ren. (1928) *Studies in the Modern Wu Dialects*, Tsinghua College Research Institute Monograph No.4.

Norman J. (1988) *Chinese*, (Cambridge University Press: Cambridge).

JSS (1960) *Jiangsusheng he Shanghaishi Fangyan Gaikuang*, (Jiangsu Renmin Chubanshe: Nanjing).

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

Rose, P. (1988) "Normalisation of tonal F0 from long term F0 distributions", in M. Wagner ed. Proc. 2nd Aus. Intl. Conf. on Speech Science & Technology (Australian National University: Canberra) 68-73.

Sherard, M. (1972) Shanghai Phonology, Ph.D. Thesis, Cornell University.

Zee, E. & Maddieson, I. (1979) "Tones and tone sandhi in Shanghai: phonetic evidence and phonological analysis", *Glossa* 14 45-88.