© *Australian Journal of Linguistics* **13** (1993), 185-220. Printed in Australia.

A LINGUISTIC-PHONETIC ACOUSTIC ANALYSIS OF SHANGHAI TONES[1]

Phil Rose

ABSTRACT

The linguistic phonetic properties of Shanghai tones are specified from normalised mean fundamental frequency and duration data of four male and three female speakers. Corroborative normalised F0 shapes are derived for an additional nine Shanghai speakers, and the Shanghai data compared with another Wu dialect. A linguistic phonetic contrast is demonstrated between the two varieties in their falling tones, and in consonantally induced F0 onset perturbations. The importance of retaining durational relationships in normalisation is demonstrated.

1. INTRODUCTION

Several auditory and acoustic descriptions of Shanghai tones exist (e.g. Sokolov 1965; Sherard 1972; Zee and Maddieson 1979), but to date knowledge is still lacking of their linguistic phonetic properties. Linguistic phonetic properties have two aspects. Within a language, they characterise 'any and all distinctions which may ever be manipulated systematically in a particular language...'; cross-linguistically, they constitute 'Any phonetic property...in terms of which utterances in Language X are systematically different from utterances in Language Y...' (Anderson 1978:133,134). The specification of language-internal phonetic properties must logically precede that of cross-linguistic properites. The first aim of this paper is to specify the phonetic properties of Shanghai tones in such a way as to permit comparison with other varieties (cf. Nolan 1982:3). The second aim is to demonstrate how such data can be used to investigate the nature of linguistic tonetic variation. The paper is structured as follows. In section 2 previous auditory descriptions of the Shanghai tones are discussed, and it is argued that these

1 This work, which is a revised and expanded version of Rose (1990c), is part of a larger project on linguistic tonetic variation supported by an Australian Research Council grant which I gratefully acknowledge. I am obliged to David Bradley for his many very helpful reviewer's comments. My thanks also to Mark Donohue for processing much of the Shanghai data, and especially to Zhu Xiaonong for supplying the raw mean values for the second Shanghai data set from his own current Ph.D. research on Shanghai phonetics.

are inadequate for linguistic tonetic purposes because they do not provide enough phonetic detail, or indicate between- or within-speaker variation. In section 3 linguistic tonetic representations of the Shanghai tones are derived from normalised acoustic representations from the tones of seven Shanghai speakers. Section 4 compares the Shanghai data, augmented with data from an additional nine speakers, with the tones of another Wu dialect (Zhenhai). Results are summarised and implications discussed in section 5.

2. AUDITORY TRANSCRIPTIONS OF SHANGHAI TONES

Shanghai is traditionally described as contrasting five tones on citation monosyllables or monosyllabic words. In Table 1 are given auditory descriptions of the five Shanghai tones from several major sources. The sources are given below the table. With the exception of the musical notation used in Chao (1928), all the transcriptions are conventional and readily understood. Chao's (1928) transcription is not that of his later (1930) five point 'tone letters'. Rather, it notates musically the pitch of a speaker's tone as imitated by a sliding pitch pipe. The integers refer to the notes of the major scale, with 1 equal to *do*, and the speaker's mid pitch range equal to 3^{b} (i.e. *mi* ^{b}/ *re* $^{\#}$). Accidentals are used to give a pitch scale in semitone units. For most sites, but not for Shanghai, the musical pitch value for 1 is given as a reference point, e.g. 1 = A^{b}. If the reference value of 1 in Chao's Shanghai transcriptions is assigned the plausible value of G above middle C, as is, for example, the case for the male Pudong speaker he transcribed, then the Shanghai tone 1 – transcribed 41 – would represent a glide from midddle C to G. Transitions between notes are also notated. The underscoring indicates that the individual notes are not prolonged, and that the pitch moves directly from 4 to 1: "41" would indicate a prolonged initial pitch component: [441]. (Underscoring in the other sources has its conventional interpretation of shortness.) It is worth noting that the musical basis of Chao's notation permits easy transformation into fundamental frequency (F0) values on the simplifying assumption that tonal pitch is a primary function of F0 – see Rose (1989) for a more detailed account of this relationship.

A further point to be noted is that the Chao (1928) transcriptions for tone 2 in Table 1 represent two apparently tonemically different pitch shapes. Chao (1928) gives $3^{b}2^{b}2$ and $^{2}2^{b}3$ as the reflexes of the historical *Yinshang* and *Yinqu* tonal categories respectively. These reflexes are not distinguished in modern Shanghai, and their transcribed pitch shapes are so similar that I have lumped them together under the same tone, the implication being that they represent free variants of tone 2. It should be noted, however, that Chao's transcriptions are reminiscent of some more conservative age- and locality-related Shanghai varieties reported in Xu and Tang (1988:57,74), e.g. 44 (*shang*) and 34 (*qu*) in Shanghai City Old Speech, or 44 (*shang*) and 335

(*qu*) in Pudong[2]. This is especially so if "$\underline{3}^{b}\underline{2}^{b}2$" is considered to indicate an effectively level, and "$^{2}\underline{2}^{b}\underline{3}$" an effectively rising pitch. The two transcriptions may therefore be tonemically distinct after all. If this is the case, then the pitch shape in Chao (1928) comparable to the tone 2 transcriptions in the other sources is presumably the rising, *Yinqu* reflex.

With the exception of source 5, all integers imply use of the Chao (1930) five point scale. The glottal stop before the integers in the sources under 3 indicates syllable-*final* occurrence.

Several generalisations can be extracted from Table 1 as to the pitch characteristics of the five tones, viz.:

Tone 1 has a falling pitch contour starting high in the pitch range. According to the transcriptions, it has a variable offset, ranging from mid ("[5]3") to low ("[4]1").

Tone 2 seems to incorporate a component which rises from somewhere in the mid pitch range. Again, the offset is apparently variable, ranging from high ("[3]5") to just below mid ("[$\underline{3}^{b}\underline{2}^{b}$]^{2}"). There may also be a level or falling component preceding the rise.

Tone 3 is described as having a rising pitch component which starts either at the lowest point in the range or just above, and, like tone 2, has an offset which ranges between high ("[1]4") and mid ("[11]3"). About half the transcriptions show tones 2 and 3 to have the same offset value.

Tone 4 is clearly shown to have a level or slightly falling pitch at the highest point in the pitch range. It may or may not have the same pitch onset height as tone 1. Some transcriptions also notate a shortness of pitch length and the syllable-final glottal stop as additional characteristics of tone 4.

Tone 5 has a pitch starting, like tone 3, at or just above the lowest point in the pitch range, and rising to a value in the mid point of the pitch range. As with tone 4, some transcriptions show short length and syllable-final glottal stop.

The Wu dialects are well known for the fact that differences between tones often involve several pervasive co-occurring and recurrent phonetic features in addition to pitch. Of these, those most usually mentioned for Shanghai are length, voice quality *qua* phonation type, and voicing offset. Tones 3 and 5 are described as having breathy or whispery voiced vowels. This is usually analysed, e.g. by Cao and Maddieson (1992), as a phonological function of an initial consonant. Tones 4 and 5 are described as short, with glottal stop offset; and tones 1, 2 and 3 are described as long. Vowels in the short tones 4

2 I thank Zhu Xiaonong for pointing this out to me.

TABLE 1: AUDITORY DESCRIPTIONS OF SHANGHAI CITATION TONES

	SOURCE				
TONE	1	2	3	4	5
1	"...sharply falling..."	42	53; 53 ~ 52	[HL]; High-falling	41
2	"Moderately high level...No discernable peak in pitch height in normal speech...in careful, overprecise speech – or dictation – there is a short audible rise in pitch toward the end of phonation. ...the informant cannot prolong the vowel indefinitely...: the obligatory rise in pitch at the end forces him to the upper limit of his normal pitch range eventually, causing a rapid – though not abrupt – cessation of phonation."	35	34; 334 ~ 434	[MM↑]; Mid-rising	$3^{b}2^{b}$ 2 ~ 2 $2^{b}3$
3	"Entire syllable...is in low register. Clearly audible tone contour with end point higher than onset..."	24	13~14; 113 ~ 223	[LM↑]; Low-rising	13^{b}
4	"Moderately high pitch with no discernable change of pitch."	55	5; ʔ55 ~ ʔ54	[H]; Short-high	$4^{\#}$
5	"Low register throughout with a short rise in pitch, ending with a glottal stop as in type [i.e. tone] 4."	23	2; ʔ12 ~ ʔ23	[LM↑]; Short-low	23

Index of sources:

1: Sherard (1972:59-63)
2: Norman (1988:202)
3: JSS (1960:116); Xu and Tang (1988:8)
4: Zee and Maddieson (1979:71); Jin (1986:2)
5: Chao (1928:76,77)

and 5 also tend to be auditorily less peripheral, although this is not usually mentioned. It can be seen from Table 1 that some of these features are included in some of the tonal transcriptions, and some are not.

The transcriptions in Table 1 differ in what they claim to be the phonetic pitch characteristics of the tones. The main differences involve relative values for lowest points in tones 1 and 3; relative peak values of the rising tones 2, 3, and 5; and relative onset values for the high onset tones 1 and 4. Some transcriptions imply that the falling tone 1 offsets at the same value as the lowest point of the low rising tone 3. Others imply that it does not fall as far. All sources agree that the three rising tones do not all have the same peak value. Apart from this, there is very little agreement between them. Tone 3 is shown to peak either at the same value or lower than tone 2; and the short tone 5 is shown mainly to peak at the same value or lower than tone 3. As far as the high onset tones are concerned, the transcriptions differ as to whether they have the same onset, or whether the short tone 5 is higher. In addition to these, differences are also noted for long rising tones 2 and 3 according to whether the rise is delayed or not.

The transcriptions represent an expected heterogeneity in the broad-narrow phonetic dimension, with the broader transcriptions bordering on the phonological. Chao's (1928) transcriptions are clearly the narrowest. They are derived from imitation with a musical instrument; use a pitch scale divided into semitones; and indicate the nature of pitch transition and duration. Those of Zee and Maddieson (1979), on the other hand, are much broader. They show pitch using symbols ('H' 'M' 'L') which are usually found representing phonological tones. Together with the vertical arrow diacritic, these tones appear to provide for a five point pitch scale (L, L↑, M, M↑, H). Duration and transitions are not indicated. The distinction between broad and narrow on the one hand, and phonetic and phonological on the other, together with the possibility that some transcriptions are based on utterances of one speaker, probably accounts for some of the transcriptional differences documented in Table 1. Other differences may reflect the different varieties of Shanghai already noted, and there is also of course the possibility that some of the differences are ascribable to between-transcriber differences, or, as Yip (1980:224) observes '...one man's 22 is another man's 32...'.

Now, whilst a broad type of phonetic representation might be suitable as observation data for tonological analyses (e.g. Yip 1980), it is clear that it is inadequate for specifying the tonetic properties of a variety – and *a fortiore* for comparing varieties with respect to tonetic properties – since a broad transcription characteristically ignores phonetic detail. In his original proposal for a method of tonal transcription Chao in fact explicitly disallowed several combinations in order to avoid making 'too fine distinctions' (1930:25). But even a transcription as narrow as that in Chao (1928), which looks as if it might be detailed enough to show tonetic differences, is inadequate for a

second, equally important reason, namely that it fails to show either between- or within-speaker variation. Knowledge of such variation is crucial in any linguistic phonetic comparison. It would clearly be nonsense to claim that two varieties differed with respect to a tonal feature – say [44] vs. [33] – if there was between-speaker variation of this type within a variety. Notice also that the five point, or any discrete scale, imposes restrictions on the nature of between-speaker variation observable: speakers may be either [33] or [44] or both, but not an intermediate value. In order to make linguistic tonetic comparison, therefore, representations are needed of individual varieties that are detailed enough and that provide an minimally distorted picture of between-speaker variation, as well as controlling for between-transcriber variation. Finally they must be quantifiable to permit statistical analysis. Opinions currently differ considerably as to the nature of a phonetic representation (Nolan 1990). One obvious candidate for linguistic tonetic work, however – one which allows more objective measures of similarity and difference between speakers and between dialects – is an acoustical representation. In sections 3 and 4 below an acoustic analysis of the tones of Shanghai will be presented based on data from sixteen speakers, and it will be shown how a representation of the variety can be derived from measurements of individual speakers.

3. ACOUSTIC ANALYSIS

3.1 *Procedure*

Four males and three females, aged between 30 and 50 and considered by their peers typical speakers of modern metropolitan Shanghai dialect, were recorded on professional equipment (Nagra 4.2 reel-to-reel tape recorder, Nakamichi CM300 cardioid mike) in a sound-proofed room. The corpus was compiled from data in a large scale dialect survey of Jiangsu and Shanghai (JSS 1960), and designed to control as strictly as the language would allow for intrinsic vocalic amplitude and fundamental frequency (F0), and consonantal perturbatory effect on F0. The corpus is given in Table 2 in the transcription used in JSS (1960). It consisted of Chinese characters exemplifying the five Shanghai tones on four CV monosyllables (where C was an unaspirated bilabial/dentalveolar plosive, and V was either an open or close vowel). The voiced stop symbols in Table 2 represent a morphophonemically separate series of syllable-initial consonants which are modally voiced word-internally, but are voiceless, with coincident VOT, word-initially. This series co-occurs with the low pitch onset tones 3 and 5. Two other morphophonemic series – voiceless unaspirated and voiceless aspirated – co-occur with the non-high pitch onset tones 1, 2, and 4. The former series is represented by the voiceless stop symbols in Table 2.

TABLE 2: CITATION TONE CORPUS

Tone 1	*ti* 'low'	*tu* 'capital'	*to* 'knife'	*pɑ* 'dad'
Tone 2	*ti* 'choose'	*tu* 'gamble'	*to* 'arrive'	*pɑ* 'worship'
Tone 3	*di* 'lift'	*du* 'stomach'	*do* 'flee'	*bɑ* 'arrange'
Tone 4	*tiə?* 'target'	*to?* 'earnest'	*pɑ?* 'eight'	*po?* 'north'
Tone 5	*diə?* 'enemy'	*do?* 'read'	*bɑ?* 'white'	*bo?* 'thin'

Prompt cards were prepared, each with the four different tokens of the same tone, e.g. [ti tu to pɑ]. All four 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 to read. The informants were instructed to pause 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.

The informants' productions deviated both from existing auditory descriptions and the transcritions in Table 2. There were clear differences in segmental quality. Some vowels were auditorily different from the implied transcription: "ɑ" was a more central [a>] for example, and "iə" was a monophthong [ɪ]. The initial stops in tones 3 and 5 were, as expected, voiceless. More importantly, the long tones 2 and 3 were produced with a clear syllable-final glottal stop in all cases. As pointed out above, it is normally assumed that only the short tones 4 and 5 end thus. This is significant because putative restricted occurrence of the glottal stop to tones 4 and 5 has been exploited phonologically (e.g. Jin 1986:2; Yip 1980:195) to allow tones 3 and 5 (and sometimes tones 2 and 4) to have the same underlying tone. Obviously the results of this study show this analysis can no longer stand as it is, although it still needs to be investigated to what extent the final glottal stop is a characteristic of citation form as opposed to unedited speech. A second noteworthy point concerned tone 3, which showed clearly audible between-speaker differences in pitch contour: low dipping, with concomitant longer duration (speakers M1 and M4) vs. low rising (the rest). This may be the difference reflected in the Xu and Tang (1988) transcriptions for tone 3 (113~223 vs. 13~14) shown in Table 1 above .

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 per cent points of effective vocalic duration (i.e from F0 onset to F0 minimum /inflection point in tones 1 and 4, and to F0 peak in tones 2, 3 and 5). Means and standard deviation values over all 16 tokens of each tone were calculated, and are given for the 7 speakers as functions of percentage points of duration in Table 3. (So for example the mean F0 value for speaker M1's 16 tokens of

tone 1 at 30 per cent of duration is 193 Hz, with a standard deviation of 13 Hz. The 30 per cent point occurs at csec. 5.3 (30 per cent of 17.5 csec.)) Mean values are shown graphically in Figure 1, which plots the F0 values for all seven speakers' tones as functions of absolute duration.

TABLE 3: Mean and standard deviation values (x, s) for fundamental frequency (Hz) and duration (csec.) of seven speakers' citation tones (M, F = male, female). n = 16.

Speaker	M1	M2	M3	M4	F1	F2	F3
TONE 1							
0%	215, 15	236, 16	240, 18	147, 15	278, 20	222, 9	238, 9
10%	211, 15	228, 15	234, 15	140, 10	274, 19	222, 8	237, 9
20%	203, 15	217, 14	227, 15	135, 10	268, 19	216, 8	234, 10
30%	193, 13	206, 13	213, 15	131, 9	260, 19	210, 7	230, 11
40%	185, 13	192, 11	194, 15	126, 8	248, 18	203, 6	224, 11
50%	175, 13	176, 9	172, 14	121, 7	235, 15	194, 7	216, 11
60%	162, 13	161, 8	151, 13	115, 8	221, 15	186, 6	207, 12
70%	148, 11	146, 8	133, 11	107, 8	203, 15	178, 5	196, 11
80%	136, 11	133, 9	121, 10	99, 6	185, 12	168, 6	181, 11
90%	124, 13	122, 9	111, 10	94, 5	170, 9	155, 9	161, 15
100%	112, 15	115, 9	104, 8	88, 7	158, 9	141, 11	136, 18
F0 offset	–	–	–	–	156, 17	–	–
Duration	17.5, 2.1	20.3, 1.9	19.3, 1.8	20.5, 2.2	21.5, 3.3	20.5, 1.8	15.8, 1.9
Duration to offset	–	–	–	–	28.2, 4.0	–	–
TONE 2							
0%	183, 11	184, 10	199, 13	125, 13	252, 15	212, 11	219, 14
10%	174, 14	172, 11	189, 12	113, 9	245, 17	203, 8	217, 13
20%	166, 14	170, 11	185, 12	106, 8	239, 15	197, 8	212, 13
30%	161, 14	170, 11	183, 12	102, 8	233, 16	192, 9	208, 13
40%	158, 14	171, 12	183, 13	100, 8	230, 18	190, 10	207, 13
50%	157, 14	173, 13	184, 12	99, 8	230, 18	188, 9	207, 13
60%	157, 15	174, 13	188, 13	99, 7	234, 20	189, 10	207, 12
70%	157, 15	176, 12	194, 13	99, 7	240, 21	189, 10	207, 14
80%	160, 15	177, 11	199, 14	99, 7	247, 22	191, 10	206, 14
90%	163, 16	175, 10	206, 13	101, 8	252, 21	193, 11	204, 16
100%	166, 15	171, 11	212, 11	104, 10	256, 22	196, 13	186, 28
F0 offset	149, 17	166, 12	165, 10	100, 14	243, 23	169, 15	161, 27
Duration	22.2, 2.3	29.0, 2.4	22.1, 1.4	28.9, 2.7	26.6, 4.2	23.7, 1.9	21.9, 1.3
Duration to offset	25.9, 1.2	30.2, 1.5	27.2, 1.2	31.2, 2.3	29.3, 3.8	27.4, 1.5	24.7, 1.2

Speaker	M1	M2	M3	M4	F1	F2	F3
TONE 3							
0%	146, 10	146, 8	129, 15	105, 7	208, 13	195, 12	201, 12
10%	135, 7	143, 6	123, 10	91, 6	199, 12	182, 9	194, 12
20%	130, 9	141, 5	125, 9	90, 6	194, 14	175, 7	187, 13
30%	131, 9	143, 7	130, 9	89, 7	195, 15	172, 8	181, 14
40%	131, 10	147, 8	135, 11	88, 6	201, 17	172, 9	179, 15
50%	132, 12	152, 8	141, 11	90, 6	212, 21	176, 10	179, 15
60%	135, 13	158, 9	150, 12	92, 6	222, 23	180, 11	181, 16
70%	140, 14	168, 11	162, 14	94, 6	232, 24	185, 12	185, 16
80%	148, 15	177, 12	174, 16	97, 8	241, 25	189, 13	188, 17
90%	160, 15	183, 13	184, 17	103, 11	250, 26	193, 13	192, 16
100%	169, 14	185, 12	192, 18	112, 12	252, 27	198, 13	194, 17
F0 offset	152, 17	161, 15	154, 17	100, 30	239, 25	172, 15	163, 18
Duration	24.4, 2.0	24.8, 1.7	21.3, 2.9	29.7, 2.5	26.0, 3.6	23.4, 1.8	18.3, 1.9
Duration to F0 offset	27.1, 2.1	28.7, 2.3	26.3, 3.1	31.1, 2.6	28.1, 3.0	26.8, 1.5	21.7, 1.7
TONE 4							
0%	207, 15	207, 13	233, 10	136, 11	275, 11	216, 12	243, 18
20%	203, 16	200, 13	230, 10	132, 11	274, 13	213, 9	240, 12
40%	198, 18	193, 12	225, 10	127, 12	269, 13	205, 9	236, 10
60%	192, 19	192, 13	218, 12	123, 14	259, 12	193, 12	229, 11
80%	179, 18	191, 13	199, 18	116, 14	245, 11	176, 17	212, 16
100%	163, 19	183, 15	175, 26	109, 15	226, 21	160, 22	181, 16
Duration	4.5, 1.3	9.0, 1.4	8.6, 1.2	5.4, 0.7	7.5, 1.2	8.8, 0.8	7.4, 1.3
TONE 5							
0%	145, 10	150, 10	147, 13	106, 8	215, 12	195, 7	200, 14
20%	140, 9	149, 9	151, 13	96, 6	212, 12	183, 7	191, 10
40%	139, 9	154, 9	160, 14	94, 5	208, 10	179, 8	185, 9
60%	142, 10	162, 11	177, 13	97, 6	215, 13	182, 7	185, 10
80%	148, 12	173, 9	202, 12	104, 8	228, 17	188, 6	191, 10
100%	153, 13	184, 10	222, 12	112, 9	241, 18	195, 5	195, 11
F0 offset	138, 11	167, 18	190, 15	107, 11	229, 18	173, 11	164, 15
Duration	8.7,0.4	13.5, 1.8	10.9, 0.6	9.7, 1.0	10.3, 1.8	12.7, 1.1	12.0, 1.2
Duration to F0 offset	10.8, 1.3	14.3, 2.0	13.7, 0.6	11.1, 1.0	12.9, 1.7	15.7, 1.0	14.4, 1.3

TABLE 3 *(continued)*

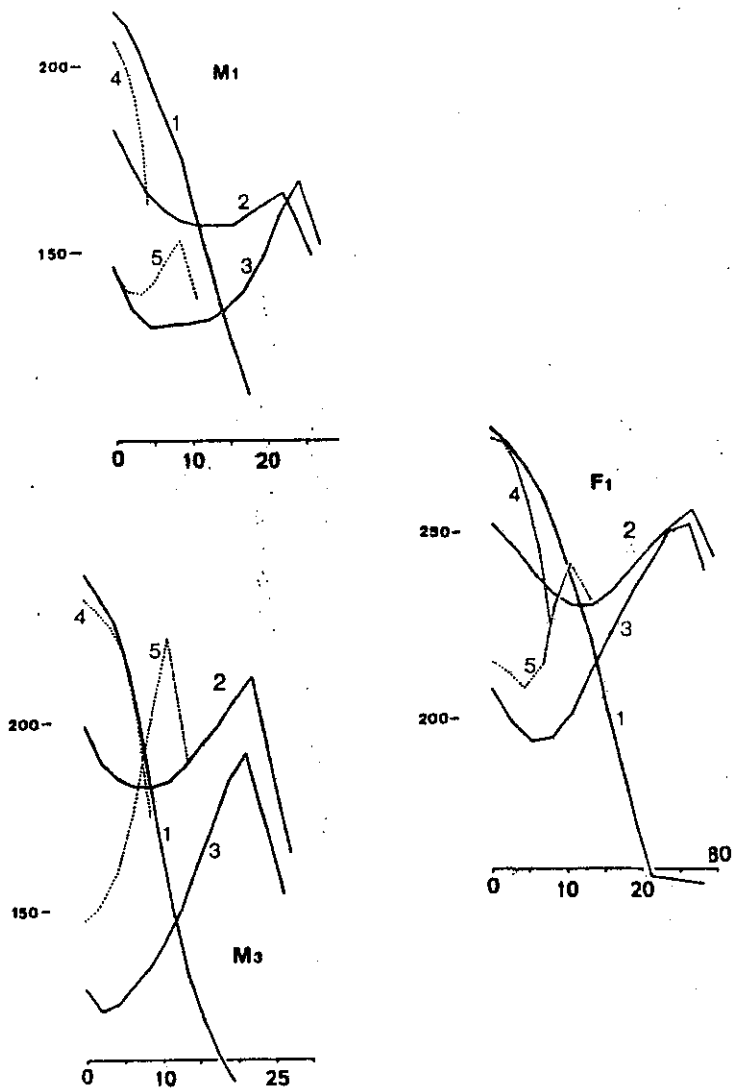

FIGURE 1: Mean fundamental frequency shapes for the citation tones of seven Shanghai speakers. Solid lines show long tones; interrupted lines show short tones. Vertical scale is mean fundamental frequency (Hz). Horizontal scale is duration (csec.).

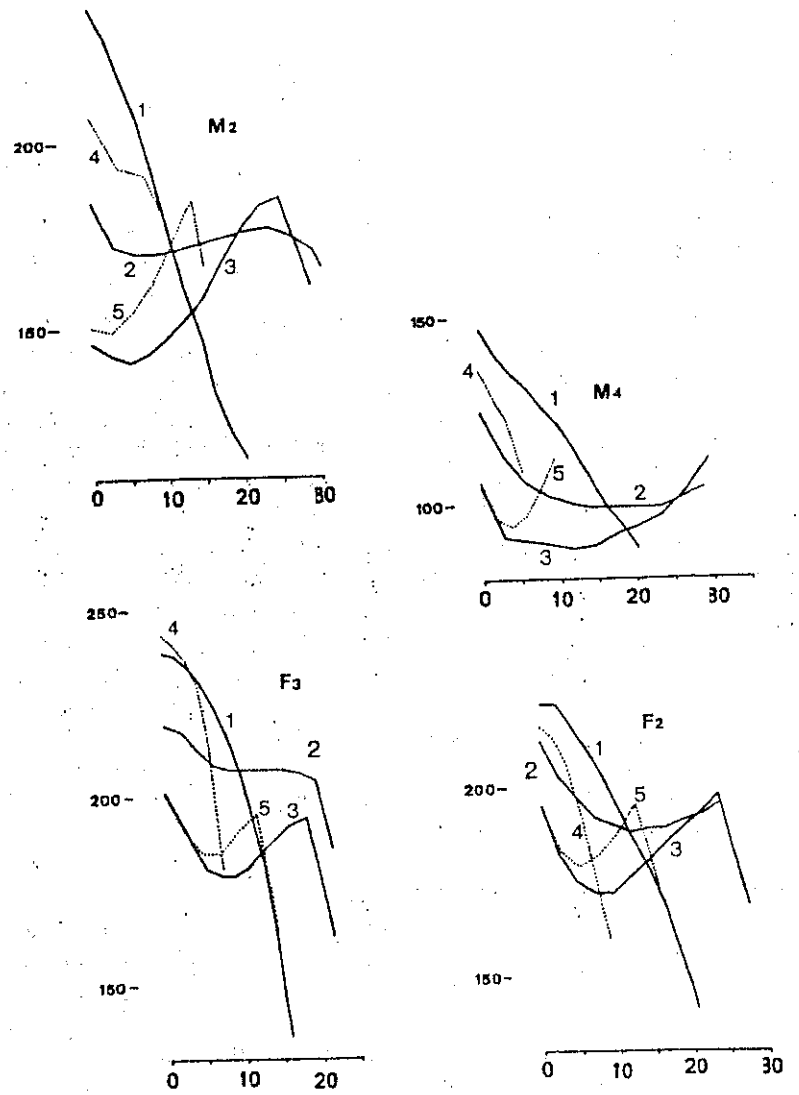

FIGURE 1 *(continued)*

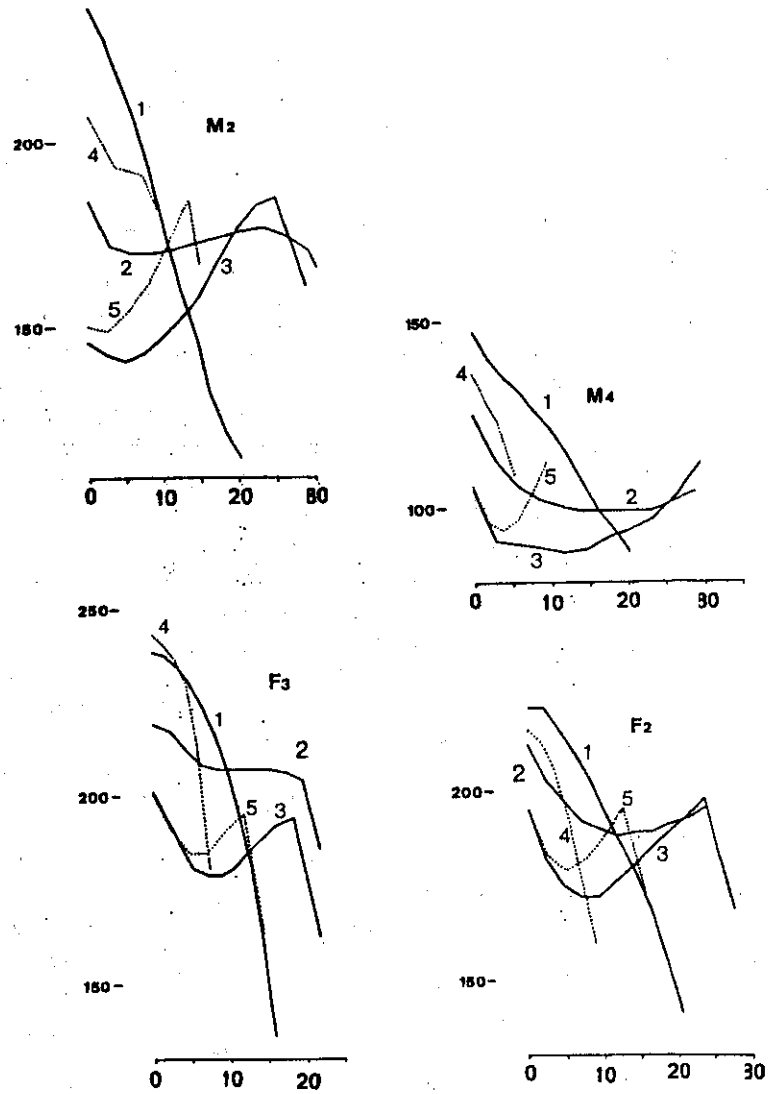

FIGURE 1 *(continued)*

3.3 *Results*

Figure 1 shows that all speakers share a configuration characterised by an abruptly falling F0 for tone 1, which traverses the whole of the speakers' range, intersecting the converging shapes of the two other long tones (tone 2, tone 3). The F0 shapes of the two short tones (tone 4, tone 5) lie within the range bounded by the onset of tone 1 and the lowest point of tone 3. The abrupt fall in F0 after peak in all except tone 1 (which also occurs but is not shown for M4) appears to be an acoustical reflex of the syllable-final glottal stop which, as mentioned above, contrary to existing descriptions clearly characterises the long tones 2 and 3 as well as the short tones. Two sets of onset points can be distinguished. The long tones show three roughly equidistant high, mid and low onset points. The short tones onset either at (i.e. at a point not significantly different from) or between the high and mid points (tone 4), or at or between the low and mid points (tone 5). With two exceptions (tone 5, M3; tone 1, F2) all speakers have similar abruptly falling F0 values over the first 10-15 csec. after onset, with approximately equal rates of fall for each tone. This, together with the shape of tone 1 tends to give the whole configuration an overall falling appearance.

Duration relationships are the same for all speakers. Tones 2 and 3 do not differ systematically in duration, but are always the longest. Expressing tonal duration as a percentage of the mean of these two longest tones gives the following mean values for the seven speakers: tone 2, tone 3 (100%) > tone 1 (80%) > tone 5 (46%) > tone 4 (30%). Thus, of the three long tones, the high falling tone 1 is on average 20% shorter than tones 2 and 3. The short tones 4 and 5 are always shorter than tone 1, and tone 4 is shorter than tone 5. These relationships presumably reflect two kinds of differences: extrinsic differences between long and short tones – the short tones have somewhat less than half the duration of the long tones – and intrinsic differences within the two groups correlating with F0 contour and height (Rose 1982:43-45).

Several between-speaker differences can be noted. Tone 3 and especially tone 2 show differences in F0 contour. For tone 3, speakers differ most importantly with respect to the prolongation of low F0 values, and this correlates with the already noted 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. Tone 2 shows the greatest amount of between-speaker variation, the main difference being in the maximum F0 attained after inflection point. For speaker M3, this value lies considerably above the tone's F0 onset value; for F1 it is about the same. M1, M2, M4, and 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. The between-speaker variance in tone 2 notwithstanding, for no speaker does it reach the same

height at peak as the onset of tone 1, as implied in some of the auditory transcriptions.

Large (apparently sex-related) between-speaker variation also obtains in the relative values of the offset in tone 1 and the inflection point in tone 3. In males the difference is one quarter or less than the range between tone 1 high onset point and tone 3 inflection point; in females it is one half or greater. Since the speakers' tone 3 all sound as if they occupy the same position in their pitch range, it seems best to consider the tone 1 offset value as the speaker-dependent variable.

3.4 *Normalisation*

3.4.1 *Procedure*

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). This provides a way of abstracting the Individual content of the speech signal from the Linguistic and Accentual content, and quantifies the between-speaker variation involved. (For Linguistic, Accentual and Individual content, see Ladefoged (1967:104)). In order to normalise F0 in this way, normalisation parameters of a speaker's overall mean and standard deviation F0 are first calculated. Each individual F0 value for that speaker is then normalised by subtracting it from their overall mean, and dividing the result by the standard deviation. This transform thus expresses each F0 observation as so many standard deviations above or below a speaker's overall mean. For example speaker M1's tone 1 10 per cent point F0 value of 211 Hz is z-score normalised thus: 211−158.3/24.2 = 2.18. This indicates that it lies just over two standard deviations above his overall mean F0 value. Table 4 gives the values for the normalisation parameters for all speakers. The normalisation parameters were derived from F0 values at all sampling points on all five tones except at onset, and offset in tones 2–5. These latter are values which tend to show between-speaker differential effects in consonantally induced F0 perturbations (Rose 1987:349), and are therefore best omitted. Linear regression of standard deviation on mean gives no significant linear relationship.

To give a visual impression of how well the normalised F0 shapes of the different speakers cluster, they are shown in Figure 2.

TABLE 4: SHANGHAI CITATION TONE Z SCORE NORMALISATION PARAMETERS (HZ). N=39.

Speaker	M1	M2	M3	M4	F1	F2	F3
Overall mean	158.3	169.8	176.5	105.8	230.2	188.1	200.2
Standard deviation	24.2	24.0	35.9	14.4	27.9	15.6	21.8

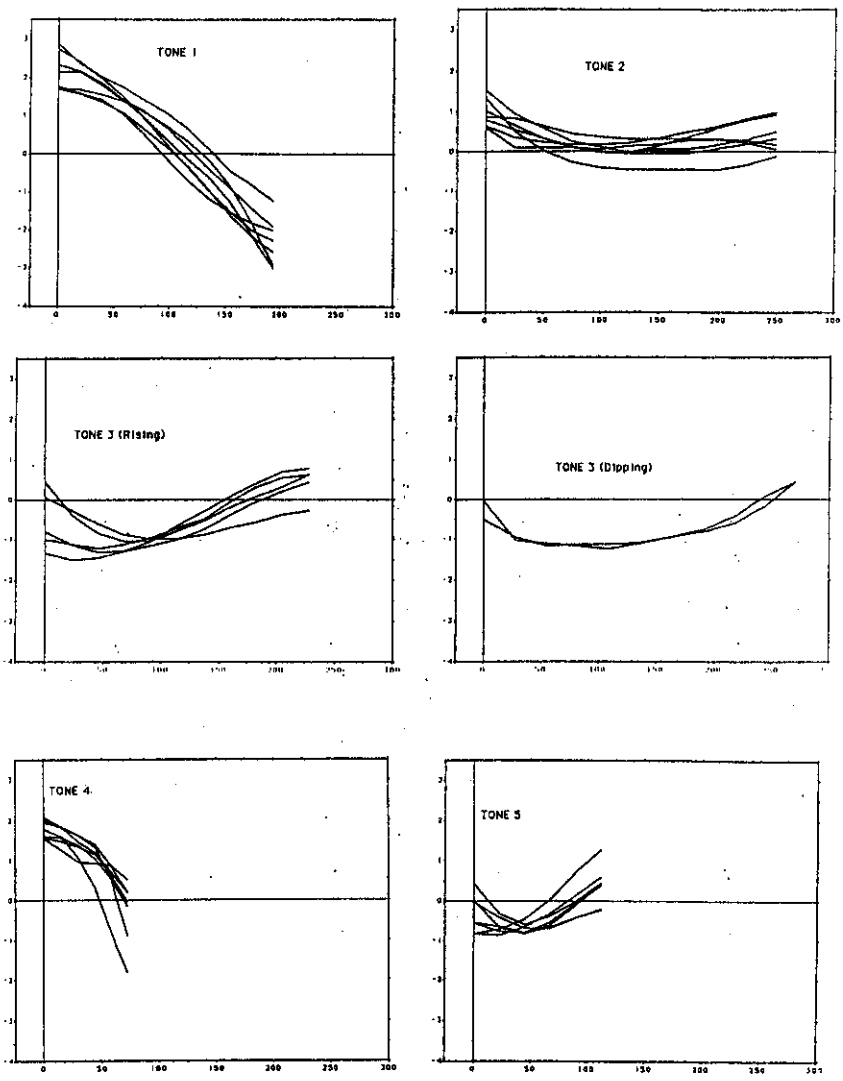

FIGURE 2: Z-score normalised values for the seven speakers' tones in Figure 1. Vertical scale shows normalised FO in units of standard deviation away from overall mean. Horizontal scale shows mean absolute duration in centiseconds.

It should be noted that one of the problems of displaying the kind of data shown in Figure 2 lies in the selection of an appropriate duration base. The problem of the normalisation of tonal duration has not yet been addressed, but two procedural points seem clear from previous research. Firstly, it has been shown that, within a given tone, normalisation is more effective – leads to a significantly greater reduction in between-speaker variance, and in agreement with auditory impression – if the duration of normalised F0 shapes is taken into account. In Zhenhai dialect, for example, the F0 of one speaker's falling tone was found to fall relatively much lower, but over a relatively greater duration, than with other speakers (Rose 1987:345). In this case, when durational differences were taken into account, speakers' tones were shown to have essentially the same contour over a substantial part of their duration. If durational differences were equalised, however, the tones were resolved with different contours. Secondly, it has been shown that, in the comparison of different tones, if durational differences are ignored, distortions arise which have undesirable consequences for linguistic tonetic comparison. For example Rose (1990c:392) shows that a clearly audible linguistic tonetic difference between the low rising tones of two different varieties is obliterated if duration is not taken into account. These results indicate that in the normalisation of tone, F0 should not be expressed as a function of *equalised* duration, and should ideally be expressed as a function of *normalised* duration.

Figure 2 incorporates these considerations insofar as the preservation of duration differences between different tones is concerned. Thus the normalised F0 values of each tone have been plotted as functions of the mean duration of that tone averaged across all speakers. (The mean duration values are given below in Table 5.) Note that the 'dipping' versions of tone 3 of M1 and M4 have been plotted separately from the 'rising' versions. This is an additional clear example of the importance of preserving duration differences: if all versions of the F0 of tone 3 were plotted against the same duration base, no acoustical correlate of the audible difference between the dipping and rising versions could be demonstrated. Within a tone, however – apart from the two versions of tone 3 – normalised F0 values of different speakers have been plotted as functions of equalised duration.

3.4.2 *Discussion*

It can be appreciated from a visual comparison between the raw and normalised data in Figures 1 and 2 that the normalisation has resulted in considerable clustering of F0 shapes. The clustering is such that no overlap occurs between the normalised F0 of tones of different phonological categories unless it also characterises an individual's raw F0 values. Thus the linguistic resolution of the normalisation is satisfactory.

The nature of the clustering is not uniform for all tones. In tones 1 2 3 (dipping) and 4, the normalised F0 values show about the same spread – about 1 standard deviation – throughout their duration. The low tones 3 (rising) and 5 do show a point of maximum clustering, which occurs near the inflection point of the low tones 3 and 5 (at about 40 per cent 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 and 5 cluster better at onset than the comparable long tones 1 and 3, but the long rising tones 2 and 3 show considerably less variance at peak 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.

The effectiveness of the normalisation can be quantified in terms of the amount it reduces the between-speaker variance in the raw data. In this case the reduction is by a factor of just under seven. This value – the normalisation index – is the ratio of the dispersion coefficients of the raw and normalised data. The dispersion coefficient is the ratio of the mean between-speaker variance to overall sample variance, and is a measure of the degree to which individual speakers' values cluster (Earle 1975:133ff.). The raw dispersion coefficient for the seven Shanghai speakers' five tones was (1615/1851) = 87 per cent. The corresponding normalised dispersion coefficient was (0.125/ 0.977) = 12.8 per cent. The figure of 12.8 per cent constitutes just under a seven-fold reduction in variance from the raw dispersion coefficient of 87 per cent.

Figure 3 gives a clearer picture of the results of the normalisation by showing the mean normalised F0 shapes for the five Shanghai tones. Curves are again plotted as functions of mean absolute duration. In addition to the mean normalised curves, Figure 3 also shows one standard deviation above and below the mean. Apart from the dipping version of tone 3, which has been omitted from calculations because of the low number (2) of speakers involved, there is no significant difference in the mean standard deviations of the individual tones, which range from 0.41 (tone 1) to 0.3 (tone 3, rise). (The mean standard deviation is the mean of the standard deviations around the mean normalised curve.) This can be taken to indicate that there is no overall difference between the tones in their degree of clustering (although there may be some differences in the nature of the clustering – see above). The overall mean standard deviation for all the normalised Shanghai data except dipping tone 3 is 0.33 (s=0.125 n=39). This constitutes about 7 per cent of the mean normalised range of 4.49 standard deviations.

The mean and standard deviation of the normalised curves as shown in Figure 3 indicate the magnitude of expected variation in the normalised F0 of Shanghai dialect. Assuming normally distributed normalised values, two

FIGURE 3: Mean z-score normalised fundamental frequency values for Long (left) and Short (right) Shanghai tones. Scales as for Figure 2. Vertical bars indicate one standard deviation above and below mean.

standard deviations around the mean will include about 68 per cent of all observations. These statistics predict, in other words, that for every further *n* speakers examined under the same conditions, n x .68 of these would be expected to have isolation tone F0 contours lying within the standard deviation corridors shown in Figure 3. (Four standard deviations around the mean would predict 95 per cent.) It can now be seen that normalisation is necessary, not to obtain mean curves, since these could be derived from a simple transform of the mean of the raw curves, but to gain an estimate of the between-speaker variance involved.

In the sense that they estimate parameters (here, central tendency and dispersion) for a population, the third order normalised data displayed in Figure 3 constitute an acoustical representation of Shanghai tones. This statement must be qualified in several important ways. Firstly, Figure 3 shows only the acoustic parameters of normalised F0, and duration. It might be the case that Shanghai tones are differentially characterised *vis a vis* other varieties by other acoustical parameters like amplitude, or spectrum *qua* correlate of phonation type, or F-pattern. If it is clear that these characteristics cannot be ascribed to other aspects of Shanghai phonology, they also belong in an acoustical characterisation of the tones. Secondly, effects due to concomitant segmental articulation and aerodynamics must be incorporated. It has long been known, for example, that the manner of a syllable-initial consonant induces F0 onset perturbations. Rose (1982:33) shows the differences in F0 which are caused by the presence of sonorant vs. obstruent at F0 onset. Segmental features of the Rhyme will also cause differences in F0. For example, Zhu (1992) gives details of the effect of vowel height on F0 in Shanghai; Rose (1992) shows that significant differences in tonal F0 are related to the presence of a syllable-final nasal coda. The corpus upon which the present analysis is based contained tokens with a well-defined, highly restricted syllable-structure. It is probably the case, therefore, that Figure 3 shows the effect of perturbations due to syllable-initial voiceless unaspirated stops and monophthongal Rhymes. One likely candidate for the effect of the syllable-initial consonant, for example, is the initial falling portion evident in all tones. It would be necessary to find out how much of the F0 shapes can be attributed to these effects, and a more comprehensive representation of tones would have to show F0 targets free of such perturbations. Thirdly, although very little is known about this factor at present, a possible effect of differences in elictation must be considered. The tokens in this study were arranged, and read out, in groups of four of the same tone, and it will be recalled that some kind of discourse intonation appeared to be being used. It may be that another approach to elicitation, for example with randomly occurring tones, would yield slightly different results. Finally, and clearly related to the problem of elicitation, is the important question of the relationship between citation forms vs. forms in

unedited speech. Again, with the notable exception of work by Kratochvil on Peking dialect (e.g. Kratochvil 1985), very little has been done on how citation tones relate to tones in unedited speech. It might turn out that the difference between the two types of tone 3 is clarified by this kind of investigation. With these caveats in mind, the normalised F0 of Shanghai isolation tones as shown in Figure 3 can now be described.

3.4.3 *Results*

When the normalisation is based on normalisation parameters from all tones, the tonal F0 is located within a normalised range of about two standard deviations above and below the mean. This agrees well with Jassem's (1971:64) proposed specification of two standard deviations above and below the mean as a speaker's normal pitch (sic) range. The curve for tone 1 traverses the whole of this normalised range, falling more or less linearly from about two standard deviations above to two standard deviations below the middle of the range.

Tone 2 has a mildly concave contour within one standard deviation above the mid range point.

Figure 3 shows that both dipping and rising versions of tone 3 fall to about one standard deviation below the mid range point, and rise to peak just above the mid range point. They differ in that the dipping version simply has a prolonged lower portion, and is significantly longer overall. This suggests that both versions have the same F0 contour, and differ primarily in overall duration. It was pointed out that some of the pitch transcriptions in Table 1 imply that tone 3 dips as low as the offset of tone 1. This is clearly not the case with the normalised F0, where the lowest point of tone 3 is still about one standard deviation significantly higher than the offset of tone 1.

Tone 4 onsets just below two standard deviations above the mid range point, and falls, probably as an intrinsic function of its syllable final glottal stop, to offset at the mid range point. According to the pitch transcriptions in Table 1 above, tone 4 has either the same or higher onset than tone 1. From the point of view of the normalised F0, however, *t* tests show the short tone 4 may have a significantly *lower* onset ($p=0.076$) than tone 1.

Tone 5 dips from just below the mid range point, and then rises to peak just above the mid range point. It was noted above that the pitch transcriptions of Shanghai tones differed with respect to the values assigned to peak in tones 2, 3 and 5. As far as the normalised F0 of these tones is concerned, *t* tests show there is no significant difference between their peak values: they all offset at about half a standard deviation above the mid range point.

Several phonological facts, both specific to Shanghai, and of general application, point to the need to distinguish high and low tonal registers.

Tones 1, 2 and 4 belong to the upper (or *Yin*, or +Upper) register; tones 3 and 5 belong to the natural class of lower (or *Yang*, or -Upper) register. It is apparent that the results of the z score normalisation provide a nice phonetic correlate of this phonological feature. The phonological difference between high and low register in Shanghai (or +/- Upper in Yip's (1980) terms) is reflected in location of post onset tonal values with respect to the overall mean. Tones with values at about 5 centiseconds after onset that are higher than the overall mean value – that is above 0 standard deviations – are upper register, and tones with post onset values below the overall mean are lower register. It is necessary to specify post onset values, because the onset value of an F0 shape may reflect the perturbatory effect of an initial consonant (Hombert 1978.)

It is not clear, for the purposes of this paper, that a linguistic tonetic acoustic representation should consist of any more analysis than already demonstrated, that is, indicators of normalised central tendency and dispersion. Ideally, both of these would be expressed non-discretely in terms of polynomials which relate values to a duration base[3] (cf. Kratochvil 1985). Should a discrete representation be required, as a convenient descriptive tool perhaps, the data in Figure 3 lend themselves to such a quantification in the following way. A scale can be used corresponding to standard deviation units, with a lower bound value of 1 set at -2 standard deviations. The arrow diacritic (↑) of Zee and Maddieson (1979) can be used to signify a shift of (about) half a standard deviation in the following integer value, as well as Chao's (1930) use of underlining to notate prolonged vs. abrupt pitches. With this method, the Shanghai tones are represented as follows: tone 1: **51**; tone 2: **43↑3**; tone 3: **↑22(2)↑3**; tone 4: ↑**44** or **54**; tone 5: ↑**22↑3**. This representation is subject to the same caveats as already pointed out above. Thus, for example, it is not clear whether the initial drop in tone 2 (**43**...) is to be attributed intrinsically to the initial consonant, or is an extrinsic part of the tone.

The apparent similarity of this standard deviation scale to the Chao five point pitch scale should not obscure the fact that more than five values in the standard deviation range – eight in fact – are needed to capture the distinct points in the Shanghai normalised F0 configuration: 5, ↑4, 4, ↑3, 3, ↑2, 2, 1. This is an indication that for linguistic tonetic purposes a five point scale (of whatever nature) is inadequate.

---

3 It will be shown below that it may not be necessary to include the mean standard deviation, since the available evidence suggests that it is not a linguistic tonetic variable. Evidence on this question can be expected to emerge from comparison with normalised mean and standard deviation data from other dialects, but at present there is little available.

4. LINGUISTIC TONETIC COMPARISON

In the sections above, a lingusitic tonetic representation was derived by normalisation in order to characterise the F0 and duration of the isolation tones of Shanghai. The other use of this kind of 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 and duration are concerned. This section demonstrates this use by comparing normalised Shanghai data from two sources with those from another Wu dialect for which normalised data are available.

The county of Zhenhai lies about 100 miles south of Shanghai. The varieties spoken there belong to the same (Wu) group, and, provided the lexicon used is not too local, are mutually intelligible with Shanghai, although their tonologies differ. Zhenhai has far more complicated tone sandhi than Shanghai, and an additional low convex toneme (Rose 1990a). Phonetically, however, the pitch of the other Zhenhai citation tones sounds very similar to those of Shanghai, with two exceptions. The Zhenhai falling and low rising tones (i.e. tones corresponding to Shanghai tones 1 and 3) both typically have short initial level components. This means that the Zhenhai falling tone sounds in some sense less abrupt than Shanghai tone 1 – I have transcribed it [441] elsewhere, e.g. Rose (1990a) to reflect this – and the Zhenhai long low rising tone sounds like the dipping version of Shanghai tone 3. 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, with essentially the same phonology) was documented 60 years ago. Chao (1928:chart 4/2) transcribes the Ningpo falling tone in his musical notation as 3$\underline{1}$ (cf. $\underline{41}$ for Shanghai). Differences between Ningpo and Shanghai are thus indicated both in pitch onset height: Shanghai is higher, and pitch transition: the initial pitch in Ningpo is prolonged, in Shanghai the fall is immediate. It should also be noted that, despite the phonetic similarity, the cognate relationships between the Shanghai and Zhenhai tones are not totally straightforward because of differential historical merger. The Zhenhai falling tone is a reflex of the historical *Yinping* and *Yinqu* categories (the latter having merged with the former), whereas the Shanghai falling tone represents *Yinping* only. The Shanghai mid dipping tone (tone 2) is a reflex of *Yinshang* and *Yinqu*, the corresponding Zhenhai tone is a reflex of *Yinshang* only. The Shanghai low rise/dipping tone (tone 3) represents a merger between all the non-*ru yang* tones; the corresponding Zhenhai low dipping tone represents a merger, in citation form only, between *Yangshang* and *Yangqu*.

Figure 4 shows the results of a comparison between the mean z-score normalised F0 shapes of Zhenhai tones from Rose (1987:350) and Shanghai tones from two sources. Like the Shanghai data presented above, the Zhenhai data were based on seven speakers, four male and three female, but the

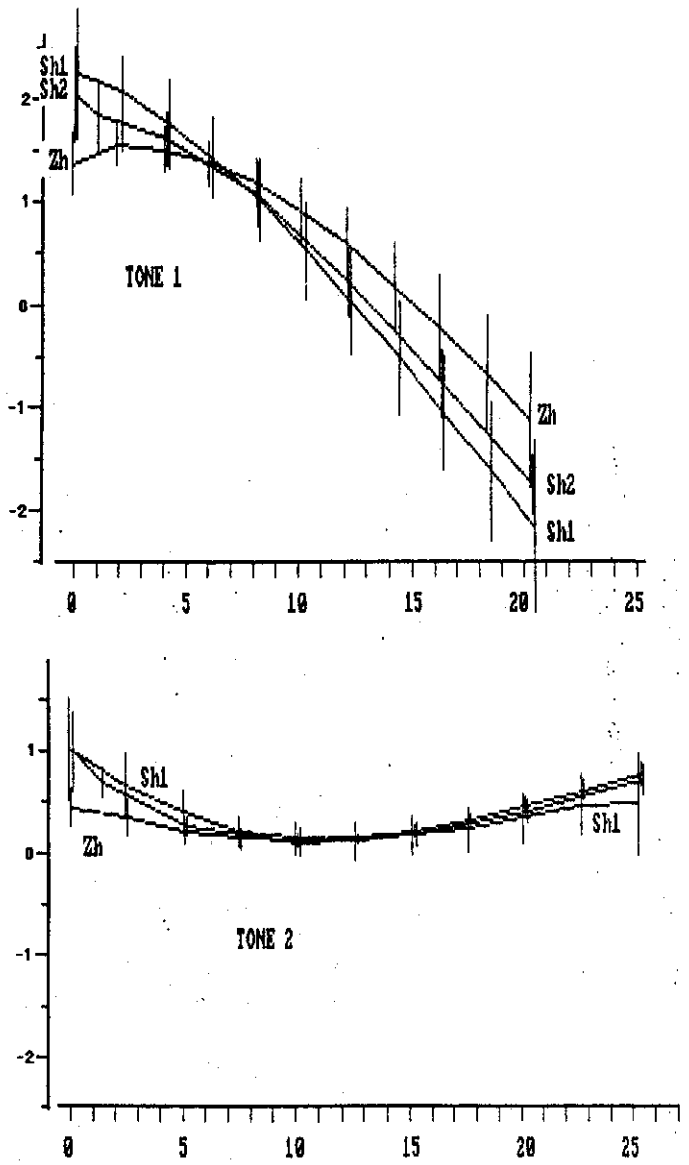

FIGURE 4: Mean z-score normalised FO values compared for seven Zhenhai speakers (Zh), seven Shanghai speakers (Sh 1), and nine Shanghai speakers (Sh2). Scales as for Figure 2. Vertical bars indicate one standard deviation above and below mean.

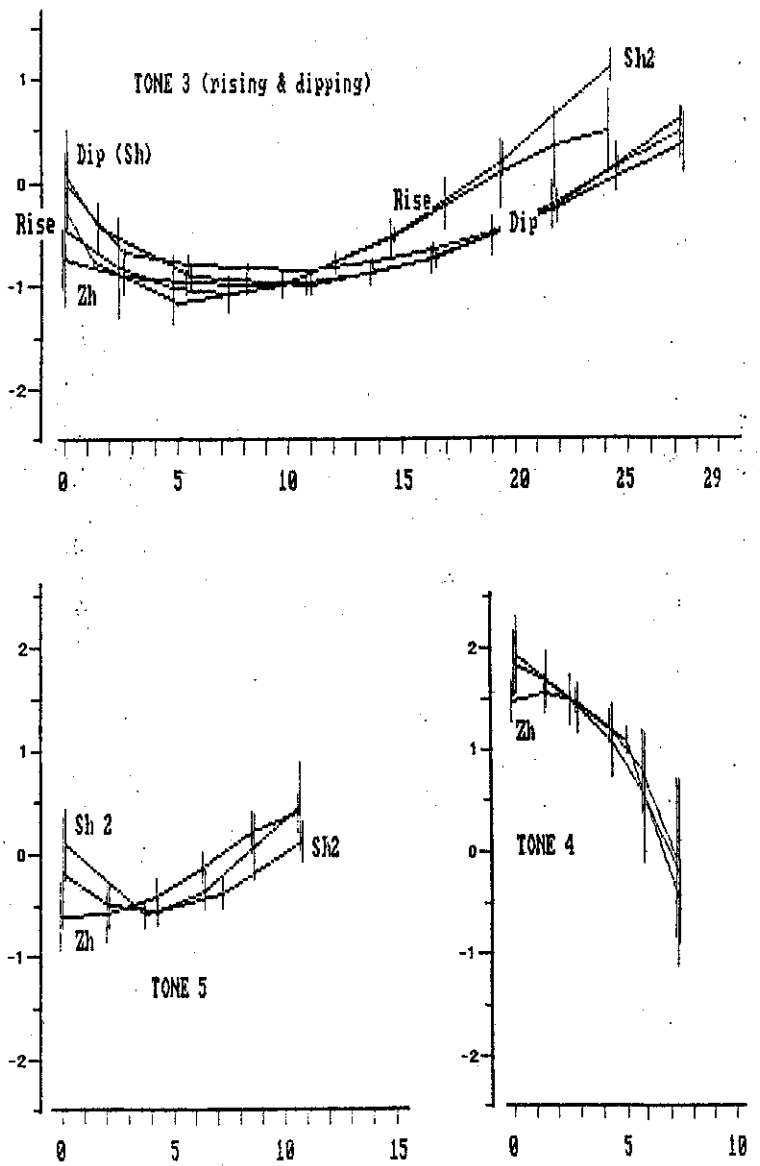

FIGURE 4 *(continued)*

elicitation differed in that citation tones were recorded singly, in random order. The two sets of Shanghai data are those presented above ("Sh 1"), and a second set ("Sh 2"). The second set is based on the eleven speaker corpus, six male and five female, reported in Zhu (1992), who generously supplied the raw mean data for normalisation by the author. Two of the male speakers in this second set were the same as in the Sh 1 corpus, and, because it was important to investigate the degree of agreement between two sets from different speakers, their data was excluded. The Sh 2 corpus therefore shows data from nine more speakers. (A comparison of the Sh 2 sets with and without the two speakers showed in fact that their exclusion had insignificant effect on the normalised values obtained.) The elicitation for the second Shanghai set was the same as for the Zhenhai.

In order to achieve comparability between the three sets of tones (Sh 1, Sh2 and Zh), it was necessary to renormalise using paramenters based on comparable tones. Thus the long falling tones in both varieties were excluded from calculation, as well as the extra Zhenhai convex tone (which is not shown in Figure 4). Excluding the long falling tones results in new normalisation parameters of mean and standard deviation from those quoted in Table 4 above, and different normalised values for the tones. For convenience in comparability, these new normalised values have been converted to the scale for the original normalisation based on parameters from all tones. The selection of an appropriate duration base for the curves was, fortunately, made easy by the high comparability between the raw mean durations of the three varieties, as shown in Table 5 below. As can be seen, corresponding pairs of tones differed by the very small amount of between 0.8 and 2.8 csec. A two-factor ANOVA on the duration data showed, of course, a highly significant ($p = 0.0001$) between-tone effect, but no significance for either the between-variety effect ($p = 0.272$), or an interaction effect ($p = 0.812$)[4]. This is an important result, apart from the fact that it avoids the need of addressing the problem of normalisation of duration, since it means that observed differences in normalised F0 shape cannot be due to differences in duration base. In view of the general lack of significant differences in the duration figures, the normalised F0 shapes have been plotted, with one standard deviation above and below the mean, as functions of mean absolute duration calculated from the three varieties. For example, normalised shapes for tone 1 are plotted as a function of a duration base of (19.3 + 21.7 + 20.3/3 =) 20.4 csec. Each F0 curve has been slightly offset to show the standard deviation bars more clearly.

4 The only individually significant differences concern the large value of 7.1 csec. between the long low Zhenhai tone and the rising version of Shanghai tone 3, and the small value of 2.1 csec. between the two low short tones in Zhenhai and Sh 1.

TABLE 5: Mean and standard deviation values (x,s) for duration (csec.) in Shanghai and Zhenhai citation tones. Except where indicated, n = 7 (Shanghai 1, Zhenhai); 9 (Shanghai 2).

Tone	1	2	3 (dipping)	3 (rising)	4	5
Shanghai 1	19.3, 2.0	24.9, 3.2	27.1, 3.8 n=2	22.8, 3.1 n=5	7.3, 1.7	11.3, 1.5
Shanghai 2	21.7, 4.0	25.9, 5.6	24.9, 5.9 n=7	25.5, 4.5 n=2	8.2, 1.5	11.6, 3.4
Zhenhai	20.3, 4.4	25.8, 3.7	29.9, 3.9	–	6.5, 1.7	9.2, 1.8

The first thing to note from Figure 4 is that the curves look very similar for all three varieties, with all except tone 1 showing a high degree of congruence. The two Shanghai data sets in particular appear excedingly similar, showing very little overall difference in both mean normalised value and mean standard deviation. There are no significant differences at the 0.5 level between the normalised mean values of the two Shanghai sets, and the magnitude of the difference between the mean values is also very small: the mean modulus difference between the two sets' means is 0.2 sd, or about 3.5 per cent of the maximum range. Differences between the two sets in standard deviation are also very small: the difference between their mean standard deviations (0.34 for Sh 1 (sd=0.13, n=23); 0.2 for Sh 2 (sd=0.09, n=23)) is about 3.0 per cent of maximum range[5]. Paired *t* tests show that, despite the magnitude of the difference, the two Shanghai sets do in fact differ very highly significantly in mean standard deviation ($p = 0.0001$). This may reflect the different number of speakers in the two sets and/or the difference in elicitation method. If the former is the case – if the additional two speakers in the second data have significantly reduced the mean standard deviation – this would indicate that more than seven speakers are necessary to obtain a more accurate estimation of the between-speaker variation[6]. However, it would

5 Because standard deviations from normalisations using parameters based on different subsets of tones are not comparable, these figures are from a normalisation based on parameters derived from all tones, as in section 3 above. The value of 0.34 for the mean standard deviation of the first Shanghai data set differs from the 0.33 quoted in section 3 above because it is based on a slightly different sample number: 23 instead of 33. This was originally done to ensure comparability with the second Shanghai data set, which was sampled at the smaller number of points. However, the minimal difference between the two standard deviation values thus obtained indicates that such a difference in sample number is of no consequence.

6 Including the two oringinally excluded speakers in the Sh 2 corpus non-significantly increased the mean standard deviation from 0.2 to 0.22. This suggests that the mean standard deviation curve has become asymptotic by eleven speakers.

seem that an overall difference of 0.14 of a standard deviation is too small a difference, however significant, to be of any practical consequence.

The near-identity of the two sets of normalised Shanghai F0 shapes is especially striking when it is recalled that they reflect two different groups of speakers, each speaker with their own different raw F0 values as input. This can be taken as an indirect indication of the reliability of the normalisation method in extracting the phonetic properties of the tones of the variety.

In both tone 2 and the rising version of tone 3 there is a small difference between the two Shanghai varieties with the first Shanghai data set tones showing a flatter profile at peak. As pointed out above, however, the difference does not reach the 0.5 level of significance. Again, this must be considered a possible result of the difference in elicitation. It is possible that this is the same effect as that already observed in the falling tone of the two Shanghai sets.

The introduction of the second Shanghai data set throws additional light on the relationship between the rising and dipping versions of tone 3: specifically, on what underlies the linguistic tonetic contrast between these two types. This relationship is somewhat obscured in the tone 3 curves in Figure 4, and therefore needs to be clarified. The first Shanghai data set (Figure 3 above) showed the rising and dipping versions of tone 3 to differ primarily in overall duration. Now, *t* tests on the data in Table 5 show that for the second Shanghai data set the rising and dipping versions of tone 3 (which have different F0 shapes, and sound different) do not differ significantly in duration. Moreover, the duration of both these versions lies intermediate between, and not significantly different from, the rising and dipping versions of the first data set. Note that Figure 4 shows the rising and dipping versions of tone 3 in the second Shanghai data set still have different F0 shapes. Furthermore, these shapes are almost identical with the corresponding rising and dipping allotones of the first Shanghai set when plotted against equalised duration. These additional data suggest, then, that the linguistic phonetic contrast between the dipping and rising versions is implemented primarily not in overall duration, as was initially assumed, but in F0 shape, with a possible durational reinforcement. Figure 4 shows that in both Shanghai data sets, the rising version is distinguished from the dipping version by a earlier rise of the F0 and concomitant lower F0 onset. The additional difference in overall duration observed in the first set may be another function of the difference in elicitation. Whereas the first Shanghai data set showed a possibly significant difference between the high falling tone 1 and high short tone 4 at onset, no such difference is observable for the second set. This effect may also be referable to elicitation.

As far as the cross-linguistic comparison is concerned, there are, despite overall similarity, clearly several differences between Shanghai and Zhenhai in their mean normalised curves. Firstly, they differ with respect to onset

perturbations. The Shanghai curves all onset higher than in Zhenhai, and these differences persist up to about 5 centiseconds after phonation onset. The magnitude of the difference is slightly under one standard deviation (the mean difference in onset between Sh 2 and Zhenhai for example is 0.88 sd). *T* tests show the vast majority of the differences to be significant ($p > .05$) at the first two sampling points. The perturbations also differ in sign and rate: Zhenhai has either mildly positive (tones 1, 4, 5) or negative (tones 2, 3) perturbations, compared to the strongly negative ones of Shanghai. There is also a possible differential for the long/short tones. The mean difference in onset between Shanghai and Zhenhai is 0.97 sd for long tones, compared with 0.65 sd for short.

The higher and more abruptly falling onset characterises both sets of Shanghai curves, and therefore it cannot be a function of their different elicitation. This appears therefore to constitute a *bona-fide* linguistic phonetic difference between Shanghai and Zhenhai. However, since the difference occurs in all the tones, it seems better not to consider it as lingusitic *tonetic*, but rather related to some other, non tonal feature of the syllable. Their occurence over the first few centiseconds of the F0 seems to indicate perturbatory influence of the syllable-initial consonant. Additional, indirect evidence for this lies in the abovementioned possibly smaller differences in the short tones: smaller differences in consonantally-induced F0 onset perturbations as a function of the long/short tonal distiction have also been documented in Thai Phake (Rose 1990b:397). Also, they are a little smaller on the average than the smallest onset difference between two tones of the same variety[7]. It should be recalled that the nature of the syllable-initial stop was maximally tightly controlled: it was a voiceless unaspirated stop in all three sets of data. Therefore the systematic difference in onset perturbation observed cannot be referred to extrinsic voice onset time differences. Perhaps Shanghai is characterised against Zhenhai in another parameter, related to vocal cord tension during stop articulation. Whatever may underlie this difference, however, identifying the onset differences as non-tonal is important, since it allows us to say that tones 2, 3, 4, and 5 all constitute occurrences of the same linguistic tonetic units – the same tonal target – in the two different dialects.

The most conspicuous difference between the Shanghai and Zhenhai curves in Figure 4 is in the long falling tone 1 which starts higher, falls more linearly, and offsets lower in Shanghai. *T* tests show the Zhenhai tone to be significantly different in its latter half from both Shanghai tone 1 curves at at least the 0.05 level. The difference between the two Shanghai varieties in the

7 The smallest onset difference between tones of the same variety is 1.23 sd, for tones 1 and 2 in Sh 2; this is effectively the same as the largest onset diference in the data of 1.2 sd between Zhenhai and Sh 1 tone 1 – cf. the mean difference of 0.88 sd noted above.

rate of fall may correlate with the difference in elicitation. There are several ways this difference can be interpreted: that Shanghai differs from Zhenhai in having a lower offset tonal target than Zhenhai, and the Zhenhai shouldered shape is intrinsically determined from the combination of non-tonal positive onset perturbation plus offset tonal target. Or, the Zhenhai shoulder is an extrinsic tonal target, to which the offset is assimilated in height. Or both shoulder and offset are independant targets. Irrespective of interpretation, this is clearly a *bona fide* linguistic tonetic difference between the two varieties. In terms of the descriptive standard deviation scale suggested above, the contrast could be represented discretely as 51 (Shanghai) vs. ↑4↑42 (Zhenhai).

It is worth pointing out at this point how the descriptive approach used in this paper – deriving a representation from the normalised physical phonetic measurements of a representative set of individual speakers – allows us to see the nature of individual variation which underlies this linguistic tonetic contrast. It is not the case that all individual Shanghai speakers have "51", and all Zhenhai speakers have "↑4↑42" for their falling tone. Rather, they are distributed around these normalised values, such that there is a small overlap between the two varieties. For example, there is one Shanghai speaker in the first set (M1) whose offset value is high enough to overlap with the Zhenhai offset range; and a few speakers in the second Shanghai set have a prolonged shouldered onset to their F0. In the light of this result, it would be interesting to conduct a perceptual experiment to ascertain if the abruptly falling pitch contour with the lower offset is better recognised as the Shanghai version by native speakers. The nature of individual variation revealed by the approach highlights the necessity of using many speakers to estimate between-speaker variation. Suppose Shanghai and Zhenhai were being compared with respect to offset in tone 1 on the basis of only two Shanghai speakers, one of whom, like Shanghai M1, had a relatively higher offset. The erroneous conclusion would be reached that the two varieties did not differ in offset. More interesting is the possibility that the observed variation reflects the continuous nature of a dialect continuum.

The Zhenhai data are also comparable to Shanghai in the amount of variance around the mean normalised curves. The mean standard deviation for all six of the Zhenhai tones is 0.21 ($s = 0.09$, $n = 49$), or 6.3 per cent of maximum normalised range (cf. 0.34 and 0.2 for the two Shanghai data sets.) Moreover, these figures also agree well with those from a z score normalisation of six tones of eleven North Vietnamese speakers, where the mean standard deviation was found to be 0.25 ($s = 0.11$, $n = 11$) or 6.8 per cent of maximum normalised range (Rose 1987:351). It is possible, therefore, that the mean normalised standard deviation, or the ratio of mean normalised standard deviation to maximum normalised range is a effective constant, and because of that cannot be a lingusitic tonetic parameter.

5. SUMMARY

This paper has fulfilled a descriptive function in specifying the acoustics of Shanghai citation tones. This was done in the most precise way presently available for characterising the tones of a variety, namely by deriving a representation from the normalised physical phonetic measurements of a group of representative speakers. This statistically based multi-speaker approach was shown, by good agreement in F0 and duration with a second normalised set from eleven additional speakers, to provide an accurate specification of the tonal population of the variety. The main features of Shanghai tonal acoustics to emerge from this study are: (1) the identity of peak for the three rising tones; (2) the probable identity of tones 1 and 4 at onset (and hence equivalence of onset for corresponding pairs of long and short tones 1 and 4, and 3 and 5); (3) the relatively lower offset of tone 1 compared with the lowest point of tone 3; and (4) the two low dipping and low rising allotones of tone 3. It was also pointed out how the phonetic feature of overall mean provided a definition of the phonological Register feature.

Comparison of the two Shanghai sets revealed small differences that may be referable to differences in elicitation. Producing the same tone in groups of four, as was the case in the first Shanghai data set, resulted for example in a slight but uniform lowering effect on the offset value of the F0 in all long tones except the dipping tone 3. The onset of tone 4 might also have been lowered by the same effect. Although the magnitude of these effects is small, it is best to be aware of them, in order to ensure maximal comparability in future data acquisition.

Good agreement was found between the two Shanghai sets, not only in mean normalised F0, but also in mean standard deviation. Additional data from Zhenhai and North Vietnamese also showed good agreement in mean standard deviation, and this was taken to reflect a constant. There is reason to suppose that, if z-score normalisation parameters are derived from all tones in a system, the magnitude of the mean standard deviation around the mean normalised F0 curve will be between 0.2 and 0.5, and can be estimated at 5 to 7 per cent of the maximum normalised range. Apart from indicating that the normalised standard deviation is probably not a linguistic tonetic parameter, this finding has a practical value in that it obviates the need to do separate z score normalisations on all speakers when specifying the tonal acoustics of a variety. It was pointed out in section four above that the main reason for normalising speakers separately is to obtain an estimate of the variance around the normalised mean to enable comparison with other varieties. If this variance is a constant, it is only necessary to calculate the mean F0 values from all speakers, convert to z score, and assume a standard deviation value of 7 per cent of the maximum z score range.

A major methodological point to arise from this study is that success of linguistic tonetic work depends on incorporation of a suitable time base into the normalisation of F0 and, presumably, other time-varying parameters like amplitude and F-pattern. If F0 shapes had been expressed as a function of equalised, as opposed to normalised duration, for example, the difference in F0 shape between the two types of dipping and rising tone in Shanghai would have been obscured. In this study, it was possible to avoid the problem because, fortunately, the absolute mean durations were so comparable, and it was therefore possible to compare F0 shapes as functions of absolute mean duration. This enabled us to say, among other things, that the difference between the F0 shapes of the falling tones was not due to different durations. It seems clear, however, that in future work, where mean absolute durations might not be so comparable, the time-varying acoustic correlates of tone will have to be expressed and compared as functions of normalised, as opposed to equalised, duration. Attention will therefore have to be given to the normalisation of duration.

In cross-dialectal comparison, Shanghai was shown to differ systematically from Zhenhai in F0 onset perturbation. This effect was assumed to be due to a difference in the production of the syllable-initial consonant, even though the consonant was auditorily the same (voiceless unaspirated obstruent) for both varieties. Apparently, controlling for syllable-initial consonant in terms of auditory identity will not necessarily permit the assumption that remaining differences are tonal. More importantly, of course, it may be necessary to distinguish between-variety differences in perturbatory effect of syllable-initial consonant on F0 (and probably amplitude) before tonetic comparison can procede.

This study has demonstrated that the multi-speaker normalised acoustic approach is superior, for linguistic tonetics, to representations based on pitch descriptors as they are currently used. This is because it quantifies the between-speaker variance within a variety and quantifies the detail of the tonal F0 time course in an effectively continuous scale[8]. These factors enable a statistically based specification of the tonal contours. It was this that made possible the resolution of some of the disagreements that characterise the various existing pitch descriptions of Shanghai such as the identity of the peak values in the rising tones.

It can be noted here that it *is* possible for some features of tonal pitch to be transcribed in such a way that shows considerable agreement with the acoustics. This can be seen in Figure 5, which compares normalised F0 values derived from Chao's (1928) musical transcription of one speaker's

8 Since it was shown that there are more than five significantly different points in the mean normalised F0 range of Shanghai tones, the current five point scale will inevitably be procrustean.

Shanghai tones with the mean normalised F0 values from the first data set[9]. Given the complex relationship between tonal F0 and perceived musical pitch (Rose 1989) it is not of course to be expected that the original F0 of Chao's informant can be recovered. A direct comparison is unjustified, even if we could be sure that Shanghai tones had not changed in the interim. Nevertheless, Figure 5 shows that the two sets of curves agree fairly well in range and shape, considering their different (acoustic vs. auditory) bases. The long tones are separated by very similar intervals at onset, for example, and the offset values of the long rising tones are almost identical, as are also the onset values of the short tones. The main difference between the two sets is that the Chao curves for the long tones appear displaced downwards over the first half of their duration, relative to the first Shanghai data set, and the curve for the short rising tone 5 lies much higher. The reason for the downward displacement is probably a normalisation artifact, resulting from an overall higher mean for the Chao set, itself a function of the higher tone 5, the not so low offset to tone 1, and the lack of fall in tone 4. These features in turn are probably to be referred to factors in (Chao's) F0 perception.

Apart from being yet another example of Chao Yuen Ren's musical and phonetic virtuosity, this comparison shows that it is in practice possible to get some indication of the important points in the tonal pitch profile using auditory methods, although the use of an instrument – the sliding pitch pipe – might well disqualify this approach from consideration as exclusively auditory. It is therefore conceivable that a linguistic phonetic description of the important points in the tonal pitch of a variety is possible based on normalised musical transcriptions of several different speakers. However, the essential role of the duration base in contributing to the goodness of fit between the two configurations in Figure 5 must not be overlooked. The duration base for Chao's transcriptions had, after all, to be supplied from the duration data of the first Shanghai set. To the extent that it has been shown in this paper that a duration base is essential for linguistic phonetic work, the auditory representation must be said to be inadequate, and it must be

9 This was done by first converting Chao's musical values to F0, assuming a reference value of 1 = G. For example, the pitch value of 41 for Shanghai tone 1 would indicate a direct drop in F0 from ca. 131 to 98 Hz (C = 130.8 Hz). F0 values for the two putative *Yinshang* and *Yinqu* allotones (see section 2) were lumped to obtain a mean value for tone 2. F0 values were then interpolated by linear regression to get a set of values comparable to those used in the first Shanghai data set. These F0 values were then normalised by z score transform in the same way as for the first Shanghai data set, and plotted as functions of the mean tonal durations of the first Shanghai data set. The inflection points in tones 2 and 3 were fixed to occur at 50% of duration.

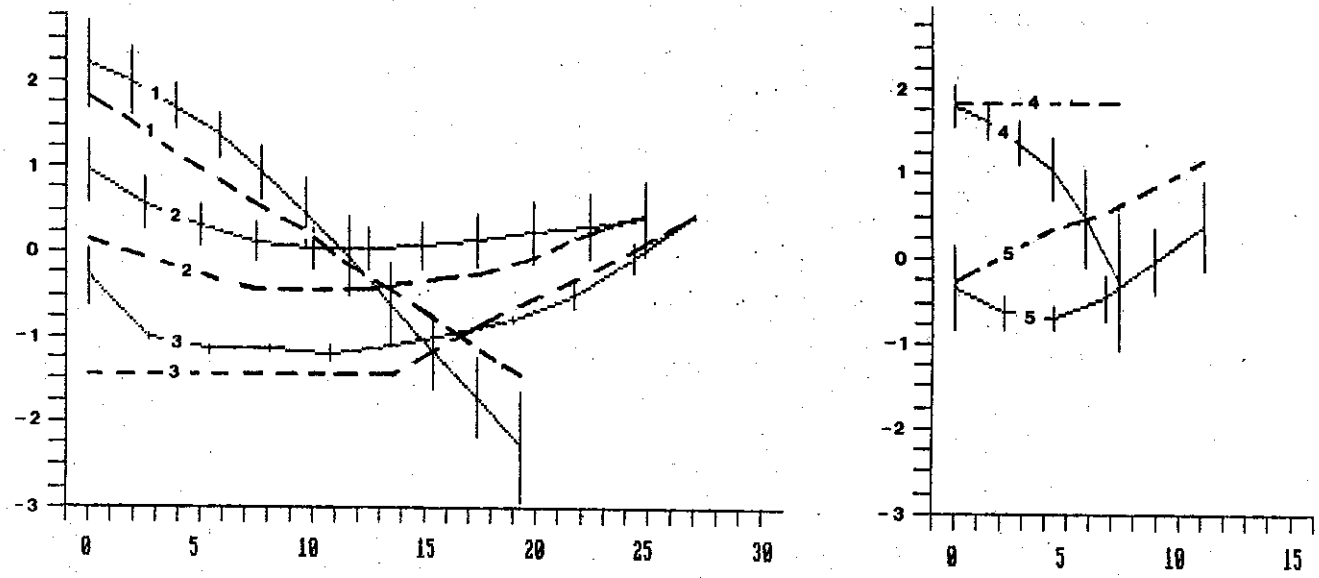

FIGURE 5: Mean normalised fundamental frequency shapes for Shanghai tones (solid lines) compared with derived fundamental frequency values for Chao's (1928) transcriptions (interrupted lines). Scales as for Figure 3.

concluded that it is in the specification of the F0 contour through a time base that the instrumental approach is probably ultimately superior to any auditory method.

The first part of this study used normalised data from seven speakers to demonstrate six linguistic phonetic contrasts in the tones of Shanghai. These results were virtually replicated in the second part of the study, with a second set of Shanghai data from nine other speakers, elicited in a slightly different way. The second part of the study also used the same method of comparison with multi-speaker normalised data to investigate how tones can differ between two related dialects. Of the five pairs of comparable tones in the two dialects (falling, mid concave, low dipping, high short, low short), the falling pair was found to differ in offset and contour. The other pairs were shown, once the onset differences were discounted, to constitute examples of the same tonal targets.

Chao (1928:73-79) presents a very large number of different transcriptions for the tones of the Wu dialects. But, as noted above, the amount of between-speaker variation for a given dialect is not shown. As mentioned above, the present study has shown that out of five tones, four are clearly shared between the two varieties of Shanghai and Zhenhai: the two short tones, and the two long rising tones. This points to the possibility that the degree of linguistic-tonetic variation in this area is not so great as implied by the Chao transcriptions. Of course, there is always the possibility that the situation has changed over the intervening 65 years. Shanghai is the local prestige variety and neighbouring dialects (e.g. Pudong) are heteronymous with respect to it (Xu 1989). It is possible that the linguistic tonetic identity demonstrated has resulted from phonetic convergence of Zhenhai on Shanghai[10].

It is not surprising that dialects can differ in the phonetic nature of their tones. It is, perhaps, surprising to find cross-linguistic tonetic *identities* between two phonologically different dialects separated by 100 miles in an area of rather large dialectal diversity. These identities highlight in particular several points in the relationship between the phonetics and phonology of tone. Firstly, it might be thought that in a variety with fewer tones, speakers would show greater variation in the realisation of those tones. This is not the case with the data presented. The phonetic aspect of the degree to which speakers' normalised values cluster (reflected in the mean normalised standard deviation) is independent of the number of tones in the system. This is shown by the fact that Shanghai, with five tones, has the same degree of clustering as Zhenhai, or North Vietnamese, with six. Secondly, it might be thought that lack of contrastivity in tone features might result in greater between-speaker latitude in tone production. From a structural point of view,

10 I have noted personally, however, that, although convergence on Shanghai readily occurs with Zhenhai segmentals, it is rare with suprasegmentals.

Shanghai only needs one distinctive tone feature, to handle the contrast between tones 1 and 2: [+/- Fall]. All other features of the five tones are predictable. For both Shanghai and Zhenhai the mean standard deviation around the mean normalised curves was found to be less than 7 per cent of the maximum normalised range. Whilst there is still no objective measure of what constitutes a small and what a large degree of spread, the extent to which the tones cluster nevertheless appears rather large. Speakers are producing their tones with considerable precision, irrepective of the number of tones, or distinctive features, in the system.

A third reason why the cross-linguistic identity of some of the tones is interesting comes from the different relationship between these citation forms and their sandhi shapes in the two dialects. The mid concave tone 2, for example, is related in a phonetically transparent way to its sandhi shapes in both Shanghai and Zhenhai, but the relationship is different in both varieties. In Shanghai, disyllabic lexical expressions with tone 2 on the first syllable have a pitch and stress shape ['33 4] (e.g. [tsəŋ dɤ] 'pillow'). The corresponding shape in Zhenhai ([tɕɪŋ dœy]) is ['334 52]. The Shanghai shape plausibly results from a spreading of the mid convex tone over both syllables. This is often represented (e.g. Zee and Maddieson 1979) as the rightwards spread of the second H tone of the LH sequence underlying tone 2. In the Zhenhai case, however, there is no spreading. The rising contour remains on the first syllable, and the second syllable pitch represents a default falling shape the higher onset of which is determined by inertial effects from the first syllable (Rose 1990a:31,32). These examples provide us with a clear case of phonetically the same entity being treated in phonologically different ways.

REFERENCES

Anderson, Stephen R. 1978. Tone features. In Fromkin, V. (ed.) 1978. 133-175.

Cao J. and Maddieson, I. 1992. An exploration of phonation types in Wu dialects of Chinese. *Journal of Phonetics* **20**. 77-92.

Chao Yuen Ren. 1928. *Studies in the modern Wu dialects*. Monograph No.4. Peking: Tsinghua College Research Institute.

Chao Yuen Ren. 1930. ə sɪstəm əv 'toun letəz'. *Le Maître Phonétique* **45**. 24-27.

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

Fromkin, Victoria. 1978. *Tone: a linguistic survey*. New York: Academic Press.

Hombert, Jean-Marie. 1978. Consonant types, vowel quality, and tone. In Fromkin, V. (ed.) 1978. 77-111.

Jassem, Wiktor. 1971. Pitch and compass of the speaking voice. *Journal of the International Phonetics Association* **1, 2.** 59-68.

Jin Shunde. 1986. *Shanghai morphotonemics.* Indiana University Linguistics Club.

JSS 1960. *Jiangsusheng he Shanghaishi fangyan gaikuang* [A survey of the dialects of Jiangsu and Shanghai]. Nanjing: Jiangsu Renmin Chubanshe.

Kratochvil, Paul. 1985. Variable norms of tones in Beijing prosody. *Cahiers de Linguistique, Asie Orientale* **14, 2.** 153-174.

Ladefoged, P. 1967. *Three areas of experimental phonetics.* London: Oxford University Press.

Nolan, F. 1982. The nature of phonetic representations. *Cambridge papers in Phonetics and Experimental Linguistics* **1.**

Nolan, F. 1990. Who do phoneticians represent? *Journal of Phonetics* **18.** 453-464.

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

Pittam, J. (ed.) 1992. *Proc. 4th Australian intl. conf. on speech science and technology.* Brisbane: University of Queensland.

Rose, P. 1982. Acoustic characteristics of the Shanghai-Zhenhai syllable types. In Bradley, D. (ed.) *Tonation. Pacific Linguistics* A-62. 1-53.

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

Rose, P. 1989. On the non-equivalence of fundamental frequency and linguistic tone. In Bradley, D. et al. (eds) *Prosodic analysis and Asian linguistics to honour R.K. Sprigg. Pacific Linguistics* C-104. 55-82.

Rose, P. 1990a. Acoustics and phonology of complex tone sandhi. *Phonetica* **47.** 1-35.

Rose, P. 1990b. Thai-Phake tones: acoustic, aerodynamic and perceptual data on a Tai dialect with contrastive creak. In Seidl, R. (ed.) 1990. 394-398.

Rose, P. 1990c. Linguistic phonetic aspects of Shanghai tonal acoustics. In Seidl, R. (ed.) 1990. 388-393.

Rose, P. 1992. Bidirectional interaction between tone and syllable-coda: acoustic evidence from Chinese. In Pittam, J. (ed.) 1992. 292-297.

Seidl, R. (ed.) 1990. *Proc. 3rd Australian intl. conf. on speech science and technology.* Canberra: Australian National University.

Sherard, M. 1972. Shanghai phonology. Ph.D. Thesis, Cornell University.

Sokolov, M.V. 1965. Eksperimental'noe issledovanie tonov Shankhajskogo dialekta [An experimental investigation of Shanghai dialect tones]. *Phonetica* **12.** 197-200.

Xu Baohua and Tang Zhenzhu (eds) 1988. *Shanghai shiqu fangyanzhi* [A description of urban Shanghai dialects]. Shanghai: Shanghai Jiaoyu Chubanshe.

Xu Weiyuan. 1989. The sociolinguistic patterns of Pudonghua in Duhang. Unpublished M.A. Thesis, Australian National University.

Yip, Moira. 1980. *The tonal phonology of Chinese.* Indiana University Linguistics Club.

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

Zhu Xiaonong. 1992. Intrinsic vowel F0 in a contour tone language. In Pittam, J. (ed.) 1992. 501-506.

Phil Rose
Linguistics Department
Arts Faculty
Australian National University
Canberra ACT 0200