Tone 4 and Tone 3 discrimination in Modern Standard Chinese

Eva Gårding, Paul Kratochvil, Jan-Olof Svantesson and Jialu Zhang

INTRODUCTION

Unless affected by tone sandhi (see Cheng 1973:46-53 and Kratochvil 1985), the pitch contour of Tone 3 in the mid position in Modern Standard Chinese declarative sentences is generally falling, the same as that of Tone 4, but the pitch of the two tones differs in such respects as the relative level and the shape of the contour, as seen in the existing descriptions of Standard Chinese tones, such as Howie 1976, and in our own material (see Gårding, Zhāng and Svantesson 1983 and Gårding 1985). These two tones may also differ as regards the presence or absence of creak; see Hockett 1947:256, where glottalization (i.e. creak) is referred to as one of the properties of Tone 3.

It has, however, neither been known which, or the combination of which of these differences were distinctive, nor have the differences been understood clearly in quantitative terms. The aim of the experiment described in this paper was to examine the pitch characteristics distinguishing from each other Tone 4 and Tone 3 in such circumstances. It was hoped that the experiment would also contribute to the understanding of tone discrimination in continuous Standard Chinese speech in general.

Figure 1 shows the F0 contours of these two tones on

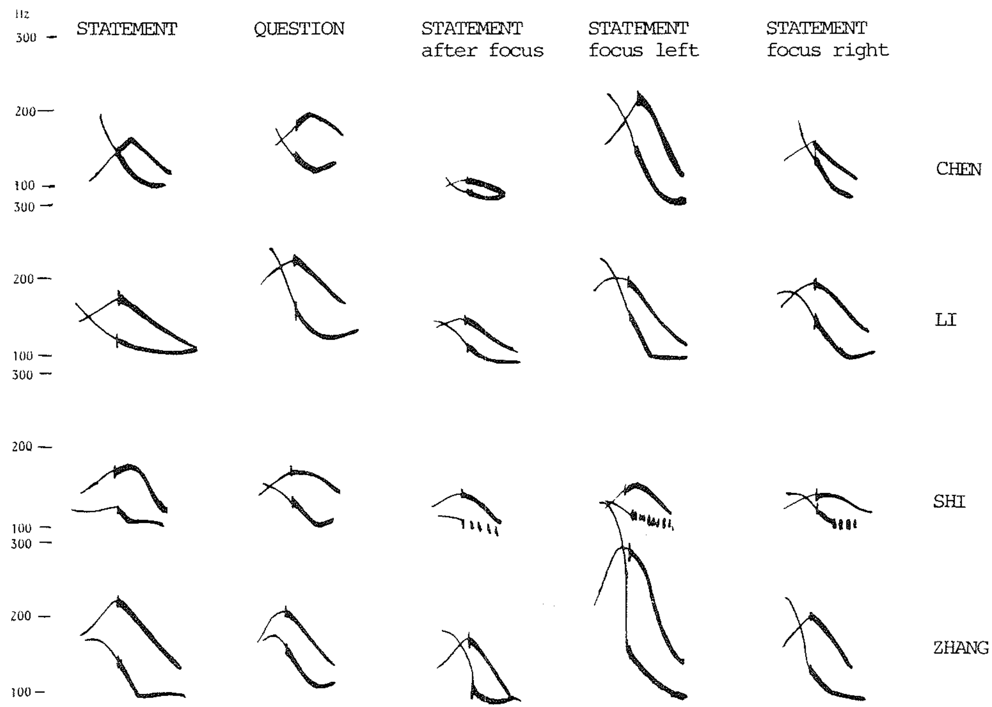


Fig. 1 Superimposed sequences of mai derived from 4 speakers' productions of five prosodic patterns. The falling tone 4 is above and the low tone 3 below. The C/V boundary is the time reference for each pair. The thick line marks the vocalic segment.

segmentally identical syllables (mài "sell" and mǎi "buy"), as said by four male speakers of Standard Chinese in five prosodic patterns. These syllables were said in comparable statements with stress equally distributed over the phrase.

Comparison of Tone 3 and Tone 4 in our material and in other studies shows that although the two tones are realized in different ways by different speakers, certain features seem to be invariant. Tone 4 generally starts with or quickly reaches a peak, from which it falls gradually towards a minimum which is often not reached until in the beginning of the next syllable. The third tone generally has a relatively low pitch level throughout the second half of the vowel. This can be achieved either by starting lower than for Tone 4, or by making the pitch fall quicker. These observations suggested various manipulations for changing Tone 4 into Tone 3 in our experiment.

BASIC DATA

The basis for the experiment was one of the isolated sentences Sòng Yán mài niúròu "Sòng Yán sells beef", mentioned above. Since the sentence is voiced throughout after the initial voiceless fricative [s] in sòng, its alternating Tones 4 and 2 gave its pitch contour the shape of a regular sequence of peaks and troughs. The key syllable mài "sells" coincided with the top of the middle peak (see Figure 2).

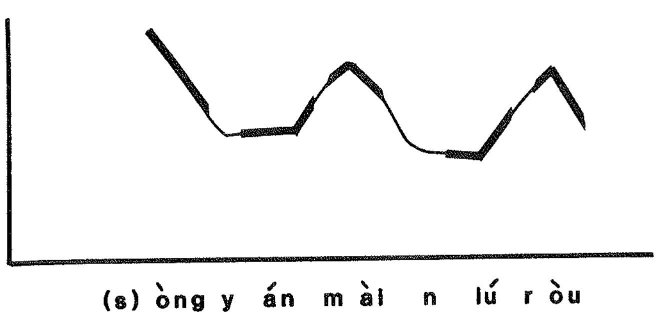


Figure 2. Stimulus 1, segmented.

This and the immediately adjoining parts of the sentence pitch curve were modified in various controlled ways, in order to find out what modification or modifications made the syllable become perceived as carrying Tone 3, i.e. as <u>mǎi</u> "buys".

The processing of the original sentence incorporating the modifications was carried out by using the programs of the ILS package. Although only the F0 in the relevant part of the sentence was manipulated, this resulted in some changes in the corresponding amplitude values due to the automatic adjustments carried out by the given program.

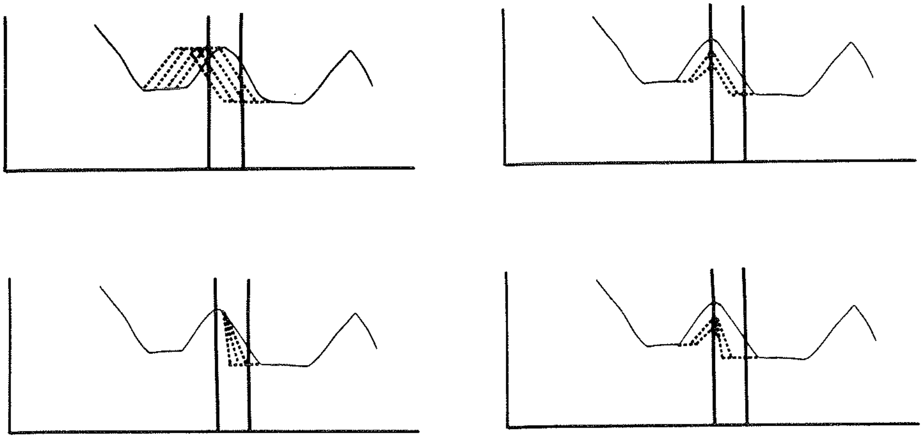
The F0 curve of the whole voiced part of the sentence was first extracted and segmented by referring to the waveform. In particular, the borderlines of the vocalic and nasal components in the area of the transition from the syllable <u>yán</u> to the syllable <u>mài</u> were established as precisely as possible. F0 has a maximum reached about 30 ms after the nasal-vowel borderline in <u>mài</u>, and then it falls to a minimum reached in the vocalic part <u>iu</u> of the following syllable <u>niú</u>.

The periodicity of the waveform was undisturbed throughout the whole syllable, and there was no sign of the kind of irregularities in it which signal creak.

MODIFICATION OF THE DATA

The F0 curve of the original sentence was manipulated in various ways, and each time the respective variant of the sentence was recorded on a magnetic tape, so that it could serve as a stimulus in the subsequent perceptual test. In all, 18 different stimuli were prepared.

The manipulation followed three main strategies as suggested from inspection of the realizations of the tones in our material. Figure 3 summarizes these modifications, which are labelled as follows:



Stimulus	Peak shifted	Peak lowered	Steepness increased	Creak added
1				
2	32 ms			
3	64 ms			
4	96 ms			
5	128 ms			
6				+
7	32 ms			+
8	64 ms			+
9	96 ms			+
10	128 ms			+
11		20 Hz		
12		40 Hz		
13			32 ms	
14			64 ms	
15			96 ms	
16		20 Hz	32 ms	
17		20 Hz	64 ms	
18		40 Hz	32 ms	

Figure 3. Design of the stimuli.

1. Peak shifted.

The middle peak was moved to the left (that is, in the direction of the beginning of the sentence) in four steps of 32 ms each (each step corresponding to 5 frames in the ILS program terms), without changing its general shape and the shape of the terminal peaks. The movement was thus achieved only at the cost of shortening the trough preceding the middle peak and lengthening the trough following it. See Figures 3 and 4 for the four steps of the movement to the left (stimuli 1-5).

2. Peak lowered.

The peak was lowered by 20 Hz (stimulus 11) and by 40 Hz (stimulus 12) without changing its shape and steepness, and without proceeding below the level of the surrounding troughs. The movement thus prolonged the troughs. Stimulus 2, i.e. step 1 of the leftward movement by which the middle peak was made to coincide with the nasal-vowel borderline in màì, was used as the starting point. See Figures 3 and 5 for the first step of the leftward movement and the two steps of lowering the peak (stimuli 2, 11, 12).

3. Steepness increased.

The original drop of the curve to the level of the following trough was made steeper in three steps (stimuli 13-15), by aiming the curve to points which were 32 ms, 64 ms and 96 ms closer to the middle peak (again made to coincide with the nasal-vowel borderline of màì first). The movement prolonged the trough following the middle peak. See Figures 3 and 6 for the first step of the leftward movement and the three steps of making the fall of the F0 curve steeper (stimuli 2, 13-15).

2+3. Combination of peak lowered and steepness increased.

In addition, three stimuli were prepared which combined the last two strategies. The first and the second of these stimuli (16-17) combined lowering of the peak by 20 Hz with making the following curve drop to points 32 ms and 64 ms closer to the peak respectively, and the third (stimulus 18) combined

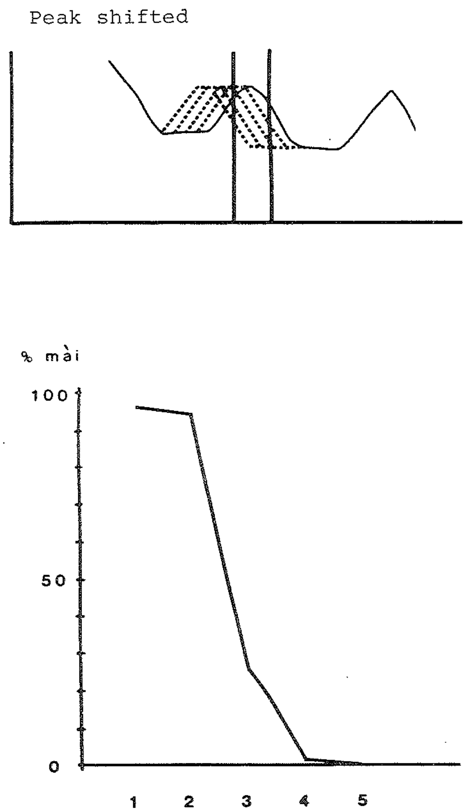


Figure 4. F0 curves and results for stimuli 1-5.

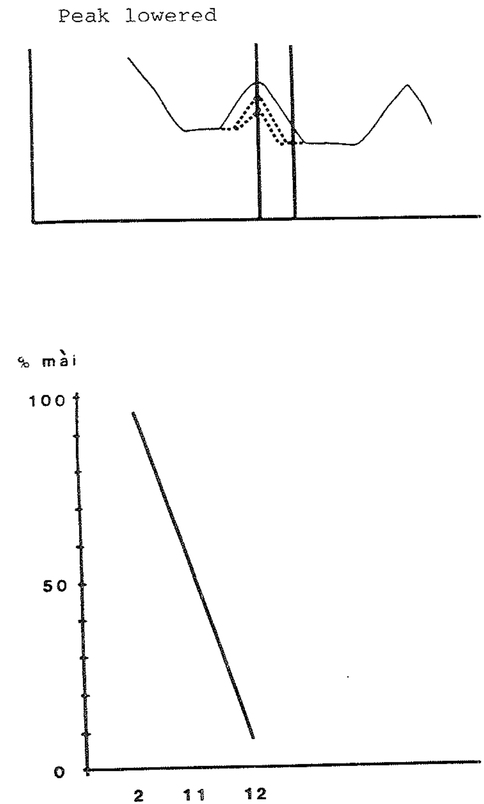


Figure 5. F0 curves and results for stimuli 2, 11 and 12.

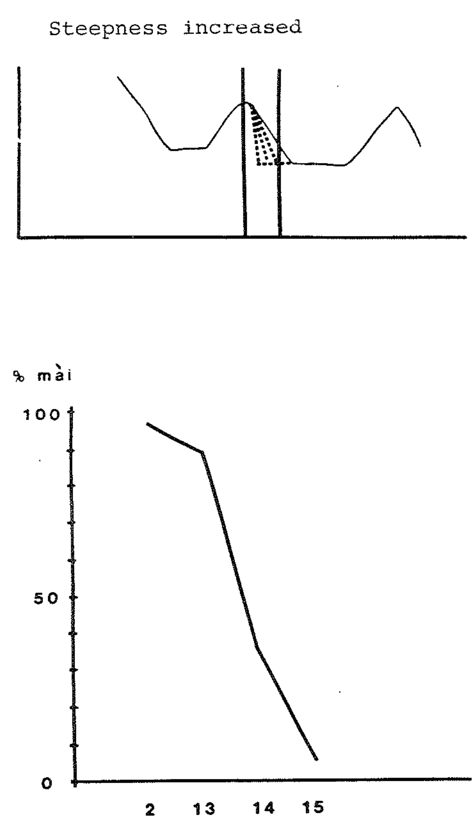


Figure 6. F0 curves and results for stimuli 2 and 13-15.

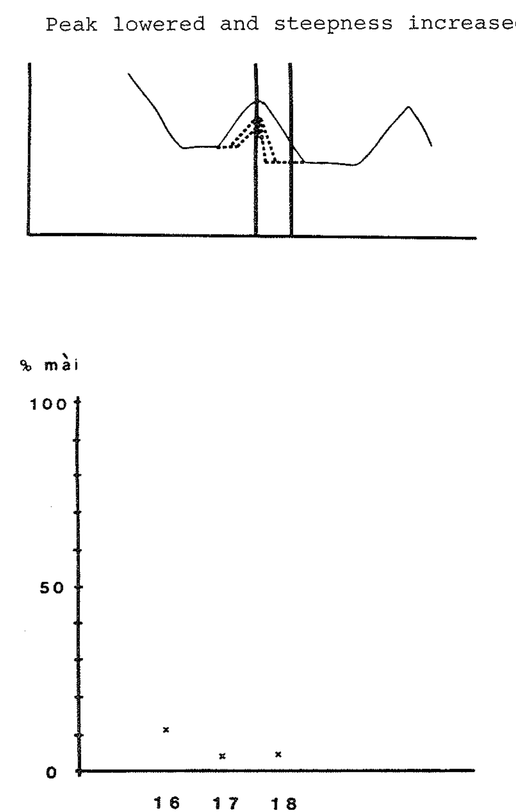


Figure 7. F0 curves and results for stimuli 16-18.

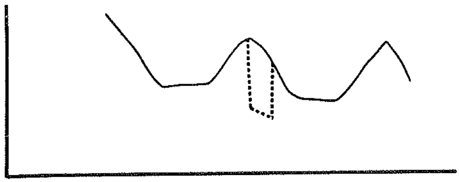


Figure 8. Stimulus 6. Creak added.

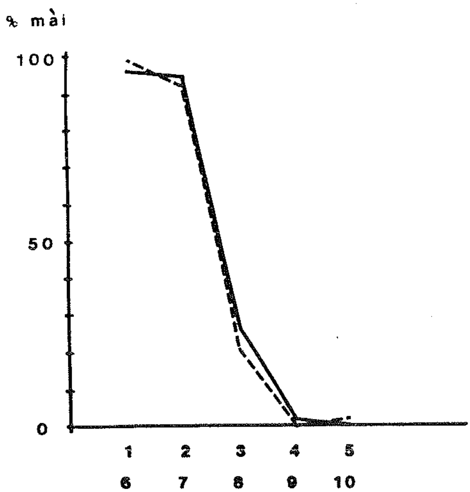


Figure 9. Result of added creak (stimuli 6-10, dashed line) as compared to the same stimuli without creak (stimuli 1-5, solid line).

lowering of the peak by 40 Hz with dropping the curve to the point 32 ms closer to the peak. (See Figures 3 and 7; stimuli 2, 16-18).

4. Creak added.

Finally, five stimuli (6-10) were prepared, which were identical to stimuli 1-5, except that simulated creak was added. Creak was simulated by abrupt halving of F0 over the middle part of the vowel in mài. (Halving of F0 brought about by biphasic phonation is one of the most common features of creak in Standard Chinese; for biphasic phonation see Lehiste 1970:59.) See Figure 8 for the F0 curve of stimulus 6.

PERCEPTUAL TEST

The perceptual test in which the subjects were exposed to the individual stimuli and simply asked whether they heard Sòng Yán mài niúròu "Sòng Yán sells beef" or Sòng Yán mǎi niúròu "Sòng Yán buys beef" in each case, was carried out in the Institute of Acoustics, Academia Sinica, in Běijīng. There were 18 subjects, all native speakers of Běijīng dialect. Altogether 60 stimuli recorded on magnetic tape were presented to the subjects. Of these stimuli, the first six were used with the intention of getting the subjects accustomed to the data, and responses to them were subsequently discarded. The remaining 54 stimuli were the 18 stimuli described above, each repeated three times in the corpus, and randomly ordered.

RESULTS AND DISCUSSION

The results of the perceptual test are found in Table 1 and shown in the diagrams in Figures 4-7 and 9.

The results indicate that all the three strategies according to which the F0 curve was manipulated (leftward shifting of the turning-point in the syllable mài, decreasing of the F0 value at the turning-point, and increasing of the steepness of the fall after the turning-point) proved successful in changing Tone 4 to Tone 3.

A common denominator of these changes is a decrease of the F0 average over the segments [ai] of the syllable mài. Figure 10 shows the mài percentage plotted against the F0 average for the stimuli which did not have simulated creak. It shows that listeners separated the stimuli into three distinct groups:

1. Stimuli which were perceived as carrying Tone 3 (stimuli 4, 5, 12, 15-18).

2. Stimuli which were perceived as carrying Tone 4 (stimuli 1, 2, 13).

Stimulus	Responses			F0 average
	mài	mǎi	% mài	(Hz)
1	52	2	96	184
2	51	3	94	170
3	14	40	26	148
4	1	53	2	126
5	0	54	0	112
6	53	1	98	-
7	49	5	91	-
8	11	43	20	-
9	0	54	0	-
10	1	53	2	-
11	28	26	52	146
12	4	50	7	129
13	47	7	87	161
14*	18	35	34	147
15	2	52	4	133
16	6	47	11	133
17	2	52	4	122
18	2	52	4	120

*One response is missing for this stimulus

Table 1. Responses to the different stimuli and F0 average over [ai].

3. Stimuli for which the results did not indicate any clear preference for one of the tones (stimuli 3, 11, 14).

The stimuli of group 1 have a low, rather flat interval at the end of the vocalic segments.

Group 2 have a high, rather flat or rising interval at the beginning of the corresponding segments.

With this description in mind, it is interesting to inspect the dubious cases of group 3, which have percentages between 20 and 50. Number 3 (26%) is falling throughout the pertinent segments, number 11 (52%) is also falling with a very short

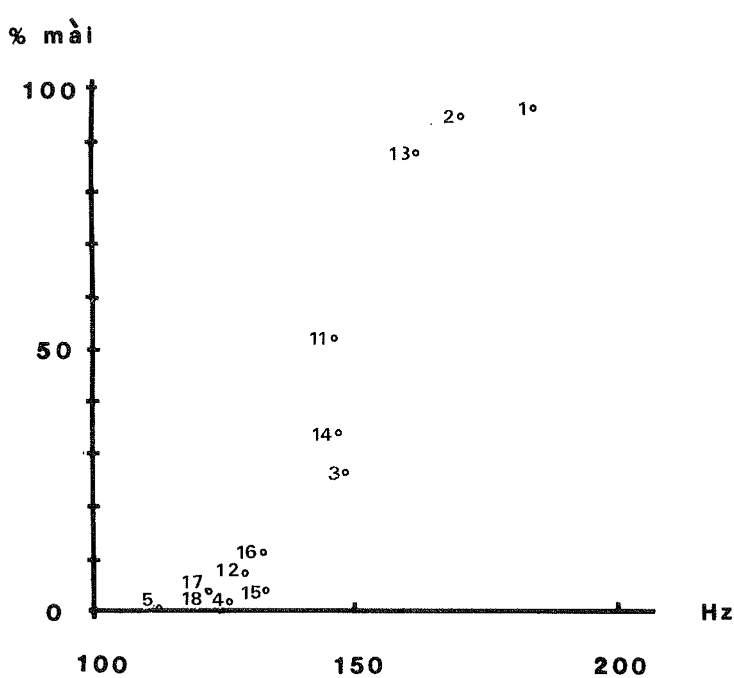


Figure 10. Tone 4 percentage plotted against F0 average.

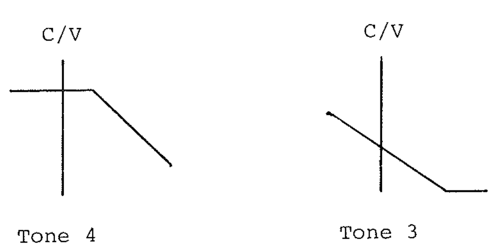
	First half	Second half
Tone 4	m=188 s=1.5	m=155 s=21.6
Tone 3	m=143 s=14.1	m=108 s=2.6

Figure 11. Mean values and standard deviations (in Hz) for F0 over the first and second half of the segments ai for stimuli judged as carrying Tone 4 or Tone 3.

flat part at the end, and number 14 (34%) is falling with flat parts at both ends.

If we assume the hypothesis that a flat or rising part at the beginning of a fall is characteristic of Tone 4 and a flat or rising part at the end of a fall is characteristic of Tone 3, one case above, number 14, has conflicting cues, and the others (3 and 11) have no distinctive tonal cues at all. Hence our hypothesis fits the three dubious cases. Also our production data give support to this hypothesis. Figure 2, for instance, shows how the falling tones are preceded by rising intervals (except the first, beginning with voiceless [s]) and the rising tones are preceded by flat ones.

Another feature which strengthens the hypothesis is indicated by Figure 11, which shows the mean fundamental frequency values and Standard deviations for the first and the second halves of the vocalic segments of the synthetic stimuli. For those judged as Tone 4, the values of the first half have very little variability compared to those of the second half. For Tone 3 the situation is reversed. Here the second half has little variability compared to the first half. One may perhaps conclude that the <u>change to</u> a fall from a certain level is important for the identification of Tone 4, and a <u>change from</u> a fall to a certain low level is important for Tone 3. If we use the notion of turning points[1] in our interpretation of the results, we might say that Tone 4 has an early high turning point introducing a fall in the vocalic segments [ai], whereas Tone 3 has a late low one ending a fall in the same segments[2]. The following schematic figure illustrates the situation:



Our hypothesis may not be without general perceptual significance. A turning point marks the change from one clearly perceived mode (falling, rising, level) into another clearly perceived mode and should therefore be a salient perceptual event.

The fact that manipulation of F0 was sufficient to change Tone 4 into Tone 3 shows that creak is not a necessary perceptual cue for Tone 3. It also turned out that the addition of simulated creak had very little or no effect (see Figure 9). This indicates that the presence of creak is not an important cue for the identification of Tone 3. However, it is also possible that creak simulated by greater disturbance of the periodicity of the waveform would have a greater effect.

NOTES

1. Turning points are basic concepts in the generative model of intonation developed at Lund.

2. As the F0 properties of the prevocalic part in the syllable mài had no effect on the interpretation of the tone the syllable carried, the results indicated support to the hypothesis about the domain of tone in Standard Chinese formulated by Howie 1974:129-48. According to this hypothesis, the tone-carrying part of the syllable in Standard Chinese is its voiced part with the exclusion of the voiced non-vocalic component which may occur at the beginning of the syllable.

REFERENCES

Cheng, C.C. 1973. A synchronic phonology of Mandarin Chinese. The Hague.

Gårding, E., Zhāng Jiālù and J-O. Svantesson 1983. A generative model for tone and intonation in Standard Chinese. Working Papers 25, 53-65. Department of Linguistics, Lund University.

Gårding, E. 1985. Constancy and variation in Standard Chinese tonal patterns. Working Papers 28, Department of Linguistics, Lund University.

Hockett, C.F. 1947. Peiping phonology. Journal of the American Oriental Society 67.

Howie, J.M. 1974. On the domain of tone in Mandarin. Phonetica 30.

Howie, J.M. 1976. Acoustical studies of Mandarin vowels and tones. Cambridge.

Kratochvil, P. 1985. The case of the third tone. In Papers presented to Wang Li on his eightieth birthday. Hong Kong.

Lehiste, I. 1970. Suprasegmentals. Cambridge, Mass.