EFFECT OF TEMPO AND TONAL CONTEXT ON FUNDAMENTAL FREQUENCY CONTOURS IN JAPANESE

Yasuko Nagano-Madsen

Abstract

Effect of tempo and tonal context on fundamental frequency contours was studied by using Japanese reiterant speech "nanana". Eight logically possible tonal patterns spoken with statement intonation were tested at three tempos. The Fo values of a number of reference points in the middle of an utterance varied systematically as a function of tempo and tonal context while those of other points close to the beginning and end of an utterance were not much affected by the two variables. When the notion of "grid" is applied to represent the overall shape made by the eight Fo contours, the relationship between all the reference points appeared basically constant across tempo variation while the slope of grid varied systematically as a function of tempo. Increase of tempo gradually raised the Fo values of Lows in the middle of utterance so that the Fo contour takes a short-cut between the beginning, peak, and end reference points. The phrase final L was found to be least variable across tempo and contextual variations. The relevance of these phonetic findings to the phonological analysis of tone is discussed briefly.

Introduction

Many studies have been devoted to "reduction" and "coarticulation" phenomena primarily in relation to the invariance issue. These studies have typically dealt with segmentals and less with suprasegmentals. Very little is known, for example, as to how Fo patterns and the relation between Fo and time dimension are affected by stylistic and tempo variations. Likewise, not much is known as to how the phonetic mechanisms of tonal coarticulation are affected by tempo change. One study of tempo effects on tonal and temporal patterns in Swedish dialects was made by Gårding (1975).

As for the tonal(1) patterns of Japanese, it has been pointed out that "casual speech" and "fast speech" should not be equated (Nagano-Madsen, 1987). In casual speech, the tonal contrast is often reduced in pitch range and sometimes even neutralized phonetically. On the other hand, in formal speech such as laboratory recording, a speaker tends to speak clearly. He becomes more aware of the tonal patterns and tries to keep them as much as possible even when directed to speak very fast. Such fast speech sounds like a "shout" presumably because of increased vocal effort. Thus, reduction and neutralization phenomena often observable in casual speech tend not to appear as a mere function of fast tempo in the laboratory.

In the present paper, the results of an exploratory study of tempo and contextual effect on the acoustic manifestations of various tonal patterns in the Kochi dialect of Japanese are reported. It controlled two variables (tempo and context) while other aspects of speech such as formality of style, type of utterance (statement), vocal effort, and segmental composition were kept constant.

The main questions are: what are the acoustic manifestations of the tone represented as H and L in this dialect of Japanese; which features vary and which features remain constant across tempo variation; is the tempo variation systematically represented in the Fo contour; does the degree of tonal coarticulation increase as tempo increases; if so, in what direction are the phonetic mechanisms of coarticulation directed?

Experiment

The test material consisted of trisyllabic reiterant Japanese speech /nanana/ with eight logically possible tonal patterns using binary (H and L) analysis. A real word or a phrase(2) corresponding to a particular tonal pattern was written in brackets as a reference. The test material is shown below:

Tonal pattern	reference phrase	lexicon
1. HHH	/hasiga/	"the edge is"
2. HHL	/hasida/	"it is an edge"
3. HLL	/hasida/	"it is a bridge"
4. LHL	/hasida/	"it is a chopstick"
5. LHH	/hasiga/	"a chopstick is"
6. LLH	/hadasi/	"bare feet"
7. LLL	/(hi)basiga/	"a fire-chopstick is"
8. HLH	/hasida(me)/	"a bridge is no good"

Note that no. 7 and 8 do not occur in the same environment as the other test items. All these phrases were meant as possible answers to the question "what did you say?"

A female speaker (present author) of the Kochi dialect, subdivision of the Kyoto(3) dialect, read the list very clearly with statement intonation at four different rates of utterance. The four speeds were: very fast, normal, slow (approximately adagio) and very slow (approximately largo). For the two degrees of slow, a metronome was placed outside the studio window to prevent great deviation in tempo. A minimum of five tokens were obtained for each tonal pattern at each rate. An informal listening test of the recorded material confirmed that all the words were perceived as having the tonal pattern intended.

A second female informant of the Kyoto dialect found the reiterant version difficult. Consequently, only (near) minimal sets of HLL and LLH patterns were recorded at three different rates for comparison.

Mingograms showing intensity curves, fundamental frequency curves, and duplex oscillograms were made for the entire material with a paper speed of 50 mm/s. Syllable duration and word duration were measured to the nearest 5 ms.

The beginning (BEG) and the syllable boundary (i.e. at the end of the vowel [a]) were chosen as reference points. A syllable boundary was chosen as a reference because when there is a clear turning point(4), it typically occurs there. The syllable boundaries were designated L1, H2 etc. depending on the position in the phrase and on whether they correspond to a High tone or a Low tone. H3 and L3 generally correspond to the end of the utterance, except for some cases at slow tempo where there is a Fo peak in the last syllable followed by a slight fall. For example, the HLL pattern has the following reference points: BEG, H1, L2, L3 (=END).

It was found that the slowest version (largo) had frequent short silent intervals within the utterance. Consequently, only one slow tempo (adagio) was measured and used for the discussion.

Results and Discussion

1. Acoustic manifestation of H and L

What are the acoustic manifestations of the H and L tones in

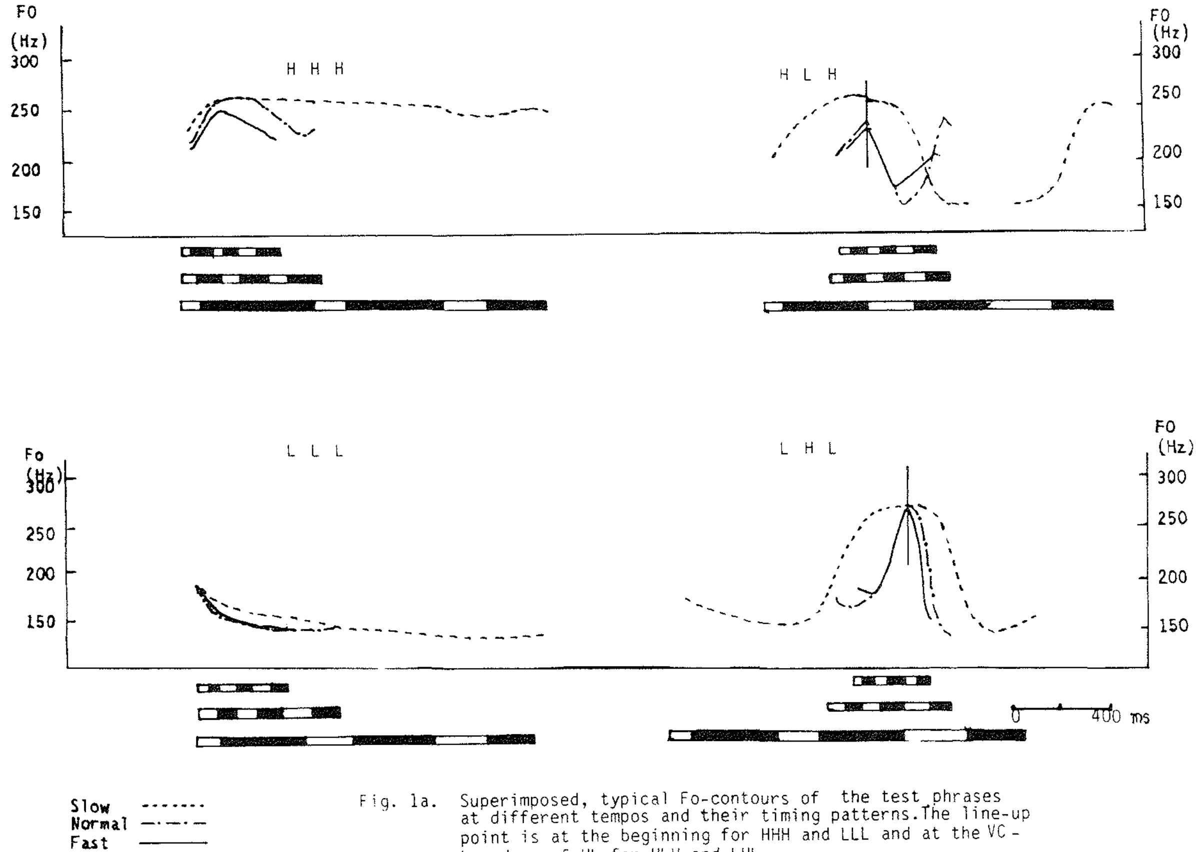
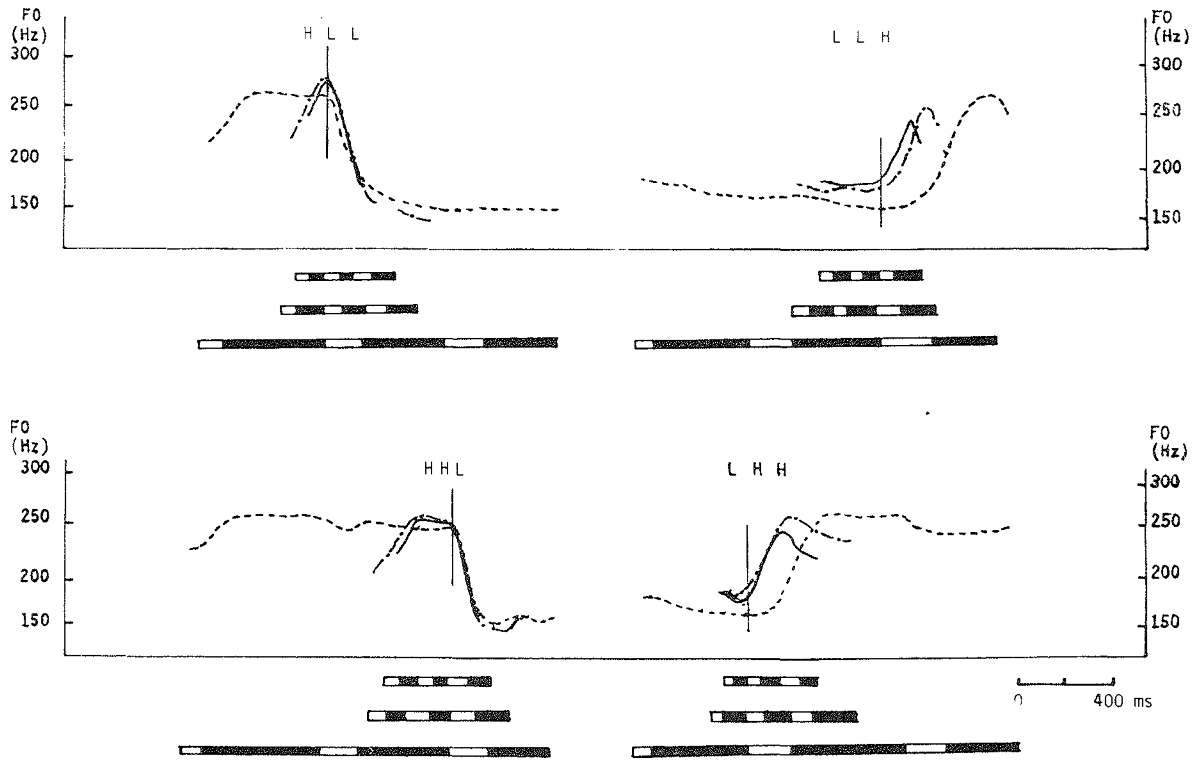


Fig. 1a. Superimposed, typical Fo-contours of the test phrases at different tempos and their timing patterns. The line-up point is at the beginning for HHH and LLL and at the VC-boundary of HL for HLH and LHL.



Slow - - - - - - -
Normal —·—·—
Fast ————

Fig. 1b. Superimposed, typical Fo-contours of the test phrases at different tempos and their timing patterns. The line-up point is at the VC-boundary of HL (HLL and HHL) or LH (LLH and LHH).

statement intonation in this dialect of Japanese? Figures 1 (a) and (b) show mingogram tracings of Fo and timing for a typical sample of each tonal pattern at different tempos. The line-up point is BEG for the level patterns (HHH and LLL) and the VC boundary where the first pitch change occurs for other patterns. It is only at slow tempo that Hs and Ls can be identified as a continuation of Fo at a certain pitch level, i.e. about 140 to 165 Hz for a L tone and about 240 to 270 Hz for a H tone. At normal and fast speeds, the Fo contour is continuously moving up and down unless there is a succession of the same tone elements.

The timing of Fo in relation to segments is a keynote in accent analyses in Japanese as well as in Swedish. The present data from the Kochi dialect of Japanese also show that the most constant acoustic feature across tempo is the relationship between rise and fall in connection with the segmental timing. Since the test material used only /nanana/ in order to avoid Fo variations caused by segments, these relationships appear regularly. Unless preceded by the same tone element, H is typically associated with Fo rise while L is associated with Fo fall, pitch change taking place in the vicinity of the syllable boundary. However, the association with Fo fall for a L is only clearly manifested when it is preceded by a H. In initial position, the manifestation of a L is basically level Fo but may decline or slightly move upward depending on tempo and tonal context.

In the present data, the maxima before the fall were the most easily identifiable turning points and were more rigorously fixed to the syllable boundary than the other kind of turning point, i.e. minima before rise (cf. HLL and LLH in fig. 1b). Since the same situation has been observed in a number of studies of intonation from different languages, this may be an automatic consequence of the production mechanisms involved. In relation to the asymmetry between a fall and a rise, Hirose (1981) notes that "the activation of muscle is achieved by asynchronous excitation of many different motor units, whereas at the time of relaxation all the units can stop their activity almost synchronously".

2. Effect of tempo on Fo

How is the tempo variation manifested in the acoustic signals? Is it only the time dimension that is affected while the frequency dimension remains basically the same? Figures 1 (a) and (b) show that all the Fo curves do undergo change in pitch range as well as change in temporal pattern.

An increase of tempo most notably affects the level portion of the Fo contour making the curves look steeper while the rising and falling parts are left less affected.

In the frequency dimension, the degree and direction of change differs in a more complicated way. The Fo values for Ls and Hs as well as BEG are presented in Tables 1 and 2. An overall tendency is that of centralization, i.e. Fo values of Hs become lower and those of Ls become higher as tempo increases. This tendency appears quite regularly for all Hs and Ls except for L3 whereas BEG and L3 were less affected by tempo variation.

The foregoing observations make more sense when the same Fo contours in Figures (1) and (2) are rearranged according to each tempo (cf. Figure 2). It appears that the eight Fo contours together make an overall shape that becomes gradually steeper as tempo increases. Within this overall shape, the relationship between Hs and Ls is basically constant except for the HLH pattern for which L and Hs are much centralized.

In a model of intonation developed at Lund, such a global shape is interpreted as a manifestation of declarative sentence intonation and expressed by two (upper and lower) reference lines whose slope is supposed to be dependent on the length of the phrase and the initial and final frequencies (Bruce and Gårding, 1978). Later the term "grid" is given to those reference lines (Gårding, 1983). The tendency for the slope of the grid or top line to vary according to tempo is reported for Chinese (Gårding and Zhang, 1987), English (Cooper and Sorensen, 1981), and Swedish (Gårding, 1975). The role of grid (phrase component) for intonational phonology has been discussed at some length by Ladd (1984, 1985).

3. Contextual effect

How are the Fo values of Ls and Hs affected by their tonal context? Does L1 become gradually higher as it changes its environment from _LL, _LH, _HL, and _HH? Is H2 higher in the environment of H_H than in the environment of L_L? Does the influence of context increase as tempo increases? If so, in what way?

Contextual effects were examined from the figures presented in Table 1. It was seen that even at slow tempo, the effect of context was present. For example, the Fo value of L2 is lowest in LLL and highest in LLH at all tempos. The difference in Fo values between

Table 1. Mean Fo values in Hz ($\overline{x}$), standard deviations (s), and the number of tokens (n) for each Low and High in different tonal context.

Tone	Context	Slow			Normal			Fast		
		$\overline{x}$	s	n	$\overline{x}$	s	n	$\overline{x}$	s	n
L1	L H H	165	5.5	7	191	8.2	5	190	5.7	5
	L H L	165	5.5	7	191	8.2	5	186	5.1	5
	L L H	172	5.5	6	183	5.1	5	181	5.3	5
	L L L	150	0	6	157	3.7	5	161	2.0	5
L2	L L H	161	5.3	6	184	3.7	5	189	4.1	5
	H L L	150	0.7	7	173	3.4	5	184	10.1	5
	H L H	153	3.8	6	163	6.0	5	181	1.9	6
	L L L	142	1.8	6	150	7.1	5	153	3.1	5
L3	H H L	160	0.8	5	156	4.9	5	laryngealized		
	H L L	148	2.0	7	142	1.9	5			
	L H L	143	2.1	5	148	3.2	5			
	L L L	141	1.0	5	141	0.8	5			
H1	H L L	262	2.1	7	273	2.9	5	260	6.9	5
	H H L	260	2.7	7	262	6.2	5	255	1.5	6
	H H H	263	8.0	7	252	6.0	5	243	2.0	6
	H L H	264	3.4	6	232	8.0	5	241	5.2	6
H2	L H L	271	1.2	6	268	4.1	5	262	3.0	5
	L H H	260	3.5	7	257	4.3	5	255	4.8	5
	H H L	252	5.2	8	257	2.9	5	247	3.6	6
	H H H	257	7.5	7	251	6.6	5	244	5.6	6
H3	L L H	261	5.8	6	252	7.9	5	242	1.8	5
	L H H	242	6.4	7	230	6.9	5	234	4.0	5
	H H H	250	5.0	7	217	5.1	5	226	4.5	6
	H L H	240	8.7	6	237	3.3	5	209	6.7	6

Table 2. Mean Fo values in Hz ($\bar{x}$), standard deviations (s), and the number of tokens (n) for BEG in different tonal context and tempo.

Context	Slow			Normal			Fast		
	$\bar{x}$	s	n	$\bar{x}$	s	n	$\bar{x}$	s	n
L L H	183	3.4	6	183	6.0	5	178	5.7	5
L H H	185	4.4	7	189	9.8	5	174	4.9	5
L H L	176	3.4	6	172	4.3	5	173	3.4	5
L L L	170	3.9	6	172	6.4	5	168	4.1	5
H H L	238	9.6	7	224	5.5	5	221	2.7	6
H H H	226	3.3	6	218	5.1	5	222	3.7	6
H L L	223	5.2	7	222	2.7	5	226	8.6	5
H L H	214	4.5	6	209	4.9	5	223	4.7	6

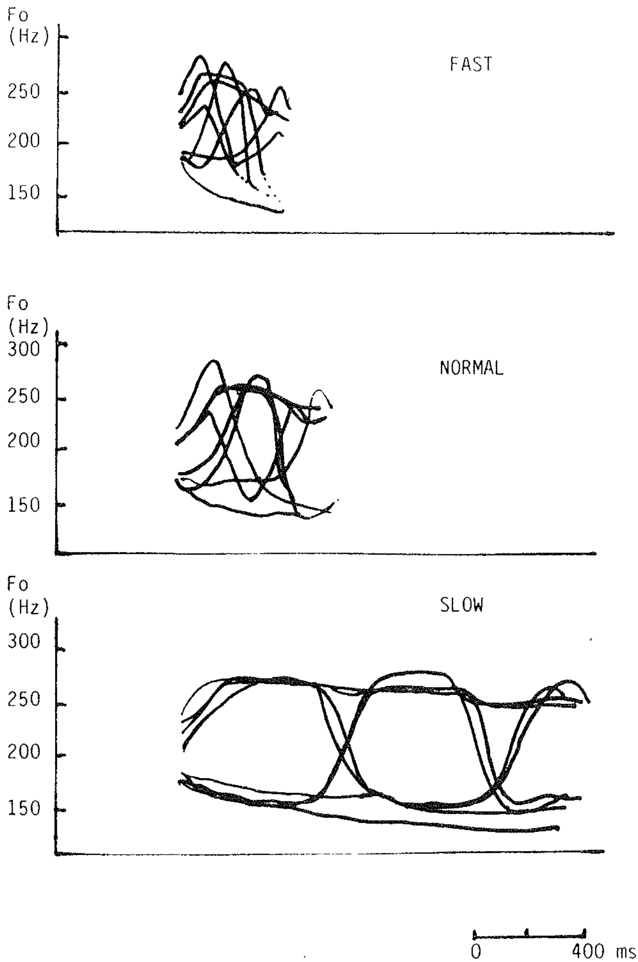


Fig. 2. Superimposed, typical Fo-contours of the test phrases grouped according to the tempo.

the two groups increases with tempo. At fast tempo, this difference comes close to 40 Hz. L3 appeared to be the least variable tone across tempo and contextual variations.

Contextual effects examined within the same tempo showed that all Low tones were affected in their Fo values not only by an adjacent but also by a non-adjacent High tone in the same word. On the other hand, High tones were not affected by any Low tones except for H1 at normal and fast speed and H3 at fast speed in the HLH pattern. Note that the Fo values for L1 and L2 increase systematically with tempo. Thus, the Fo contour corresponding to the first two syllables in the LLH pattern moves downwards before a rise at slow tempo but becomes level at normal tempo and then becomes a gradual rise at fast tempo. Since the Fo values of BEG and END points were not much affected by tempo, it looks like that the Fo contour takes a gradual short-cut from BEG to H and to the END.

Contrary to our preliminary assumption, the Fo value of H2 was highest in the environment of L_L while it was lowest in the environment of H_H at all tempos. Clearly, the tonal coarticulation mechanism can not be captured, at least for this language, by a simple HL analysis.

Why is the assimilation rule asymetrical, i.e. why can HLH become HHH but LHL not become LLL? Why does word final L often resist assimilation rules? These questions can be answered, at least in part, when phonetic mechanisms involved in the production of tone and intonation are understood. In the foregoing section, it has been shown that the recognition of the grid makes the change in Fo values for Hs and Ls consistent, which conforms well with the perceptual impression. Without this notion, the Fo changes are difficult to explain. The L3 tone was found to be least affected by tempo and context effect.

4 Comparison with the second informant

The main difference between the two informants was that of pitch range. For the first informant, the total pitch range varied from 140 Hz to 270 Hz at slow tempo. At slow tempo this informant used basically the same pitch range for Hs and Ls regardless of the tonal pattern in which they occurred. The second informant used a range extending from 160 Hz to 240 Hz at slow tempo. She has two sets of pitch ranges for Hs and Ls depending on whether they are part of a rising movement (e.g. LLH) or part of a falling movement (eg. HLL). The pitch range used for the rising movement was 20-30 Hz more reduced (centralized) than that of falling one. This difference may be due to stylistic variation. The first informant read the list very clearly (hyper speech) while the second informant read it more naturally. It should be also noted that the two dialects (Kochi and Kyoto) have considerable differences impressionistically.

Conclusion

The purpose of this paper has been to show how Fo changes with tempo and to show how tempo affects tonal coarticulation. The major points made were: (1) Fo change as a function of tempo manifests itself as a slope of the grid lines. Seen within the grid, the change in Fo due to tempo did not alter the relationship between Ls and Hs. (2) The mechanisms of tonal coarticulation can not be captured in terms of such linguistic categories as H and L alone. Instead the steady reference points, grid, and the production mechanisms involved for rise and fall should be considered.

The present study has dealt with only limited material. Some of the points made in this paper, therefore, may not be characteristic of more natural and longer utterances from real life. Since word citation forms are known to be inadequate for many other features in prosody (Bruce, 1977), another study is in preparation in which each tonal pattern appears in a different position and context in relation to a sentence.

Acknowledgments

I would like to thank Gösta Bruce and Eva Gårding for their very helpful comments and advice on earlier versions of this paper.

Footnotes

(1) Tokyo Japanese is considered to be a typical example of a pitch-accent system while some dialects of Japanese spoken in the Kansai area have more complicated systems. In this paper, the term "tonal" was used throughout without making distinction between "tonal" and "accentual".

(2) In the original list, this was written in Chinese characters.

(3) It has also been called "Osaka" or "Kansai" dialect.

(4) The turning points (local Fo maxima and minima) were found to be useful for the analysis of intonation in many languages (Gårding, 1983).

Bibliography

Bruce, Gösta. 1977. Swedish word accents in sentence perspective. Travaux de l'Institut de Linguistique de Lund XII. Gleerups, Lund.

Bruce, Gösta and Gårding, Eva. 1978. A prosodic typology for Swedish dialects. In Nordic Prosody, Department of Linguistics, Lund University:219-228.

Cooper, William and Sorensen, John. 1981. Fundamental frequency in sentence production. Heidelberg, Springer.

Gårding, Eva. 1975. The influence of tempo on rhythmic and tonal patterns in three Swedish dialects. Working Papers 12, Department of Linguistics, Lund University:71-84.

Gårding, Eva. 1983. A generative model of intonation. In Cutler, A. & Ladd, D.R. (eds.) Prosody:Models and Measurements. Heidelberg, Springer.

Gårding, Eva and Zhang, Jialu. 1987. Tempo and prosodic patterns in Chinese and Swedish. Working papers 31, Department of Linguistics, Lund University.

Hirose, Hajime. 1981. Comments in the discussion after Fujisaki et al. 's paper. In Stevens, K. & Hirano, M. (eds.) Vocal Fold Physiology. University of Tokyo Press:361.

Ladd, D. Robert. 1984. Declination: a review and some hypotheses. Phonology yearbook, 1:53-74.

Ladd, D. Robert. 1985. Intonational phonology is not a single autosegmental tier. Work in progress 11, Department of Linguistics, Edinburgh University:47-59.

Nagano-Madsen, Yasuko. 1987. Stylistic variations in the production of Japanese accent. Paper presented at the Mini Conference on Stylistics, Stockholm, May 4-5.