Ann. Bull RILP (1982) No. 16, 11-19

INTERACTION BETWEEN ARTICULATORY MOVEMENTS AND VOCAL PITCH CONTROL IN JAPANESE WORD ACCENT*

Masayuki Sawashima, Hajime Hirose, Hirohide Yoshioka and Shigeru Kiritani

Introduction

Japanese word accent is characterized by changes in the vocal pitch, which are closely linked with the time segment of the syllable or so-called "mora". Thus, laryngeal control for vocal pitch should be expected to have an appropriate timing relative to supraglottic articulatory controls in the realization of Japanese word accent. It has been reported that there is a time lag between the articulatory and vocal pitch controls in two-mora words in the Kinki dialect accent type, where the articulatory control represented by the formant transition takes place earlier than the vocal pitch change (Fujisaki, et al. 1976; Fujisaki 1977). Recent observations on two-mora words with a Kinki accent (Sugito 1981) and with a Tokyo accent (Sawashima, et al. 1980) have revealed that the relative timing of the articulatory and vocal pitch controls varies with the difference in the order of the sequence of open and close vowels, and also that the relative timing in the rising accent type differs from that in the falling accent type. These recent findings are suggestive of an involvement of articulatory movements in the laryngeal mechanism of vocal pitch control. A possible explanation for this interaction between articulatory and laryngeal controls is that the sternohyoid muscle which is responsible for the opening of the jaw may have some effect on vocal pitch control, although other mechanisms such as the "tongue-pull" hypothesis (Ladefoged 1964; Lehiste 1970; Ohala and Eukel 1978) should also be taken into account.

In our present study, we used two-mora words, including both open and closed vowels uttered with both the rising and falling accent of the Tokyo dialect. The study was mainly focused on the interaction between jaw movements and vocal pitch control as observed in the activity patterns of the sternohyoid and cricothyroid muscles.

Experimental Procedures

Two adult male speakers of the Tokyo dialect, MS and HH, served as subjects. Test words were two-mora nonsense words with rising and falling accents. The two-mora words were /aa/, /ii/, /ai/, /ia/, /ama/, /imi/, /ami/ and /ima/. The test words were embedded in the following frame sentences.

* Paper to be submitted to "Pitch Analysis" special issue of phonetica. The work was in part supported by Grant-in-Aid for Scientific Research (No. 57115005), Ministry of Education, Science and Culture.

"sorewa ___ iroda" (It is a color called ___) for /ai/ and /ami/
"ii ___ ageru" (I'll give you a good ___) for /ia/ and /ima/
"sorewa ___ madeda" (It is up to ___) for /aa/ and /ama/
"ii ___ iroda" (It is a beautiful ___ color) for /ii/ and /imi/

In each of these frame sentences, the syllable immediately before the test word had the same vowel as the first mora of the test word in order to render the steady vowel state observable even in the running speech.

The subjects uttered the test sentences at their habitual speaking rate, intensity and vocal pitch. The vocal pitch was kept as constant as possible from 2 morae before the first mora of the test word until it changed according to the word accent. Each of the test sentences with a different test word was repeated ten times.

For each of the subjects, EMG of the cricothyroid (CT) and sternohyoid (SH) muscles were recorded, using hooked wire electrodes (Hirose, 1977), on a multichannel FM tape recorder simultaneously with the recording of the speech waves. The speech waves were also recorded on an audiotape for the acoustic analysis.

The EMG signals, as well as the speech waves on the FM Tape, were integrated with a time window of 10 msec, and the value of the successive windows was plotted in order to display the time curves of the EMG patterns, with the speech envelope on the same time axis for each utterance sample. No averaging or smoothing of the EMG data was made in this experiment. The onset of the changes in the EMG levels relevant to the articulatory and vocal pitch controls was determined by a visual examination of the time curves. This onset timing was measured with reference to the initiation of voicing determined for the speech envelope at the beginning of the frame sentence.

The speech waves on the audiotape were sampled at 10 kHz. The formant frequencies were determined every 10 msec by linear prediction analysis with a time window of 20 msec. The fundamental frequency of voice was determined every 10 msec by detecting the peak of the auto-correlation function of the residual wave. The time curves of F_0, F_1, F_2 and F_3 thus obtained were displayed on the same time axis with the speech envelope for each utterance sample. The onset of the F_0 shift corresponding to the word accent was defined as the time point where the curvature of the F_0 curve showed a local peak. The procedures used for determing the local peak of the curvature were as follows:

1) the time curves demarcated by a time interval of 50 msec preceding and following a given time point were approximated by straight lines respectively; 2) the angle between the two lines was estimated as the measure of the curvature at the time point; 3) the estimation was repeated every 10 msec on the time curve.

These procedures were also applied for determining the onset of the formant transition. It was observed that the onset of the F_2 transition occurred earlier than those for F_1 and F_3 in the /ai/ samples while the onset of the F_3 transition was earliest in the /ia/ samples. Thus, the timing of F_2 and F_3 transitions were taken as the timing of the vowel transitions for /ai/ and /ia/, respectively. The onset of /m/ was determined by an abrupt decrease in the speech envelope. The onset timing for the F_0 shift, formant transition and /m/ was measured with reference to the initiation of voicing determined on the speech envelope at the beginning of the frame sentence.

Results

Acoustic data

The timing of the onset of the formant transition relative to the F_0 shift for /ai/ and /ia/ with rising and falling accent is displayed in Fig. 1. In the figure, aī and iā indicate the test words with a rising accent, and āi and īa those with a falling accent. The filled circles indicate sample values and the upward arrows indicate mean values.

It can be noted that the formant transition for /ai/ took place earlier than that for the F_0 shift, irrespective of accent type in both subjects. It is apparent that the formant transition for /ia/ took place later than for /ai/ with reference to the onset of the F_0 shift. In other words, the relative timing of the articulatory and vocal pitch controls appears to be affected by the direction of the vowel sequence, at least at the level of its acoustic manifestation. The mean values of the onset timings for the formant transition in the /ia/ samples were very close to the onset of the F_0 shift, irrespective of accent type in both subjects. There was little difference in the relative timing of the formant transition between the rising and falling accents, in both the /ai/ and /ia/ samples.

The timing of the onset of /m/ relative to the onset of F_0 shift in /ami/, /ima/, /ama/ and /imi/ in rising and falling accents is shown in the lower part of Fig. 1. It should be noted that there was little difference in the timing of the onset of /m/ among the different combinations of the vowels in both of the subjects. The mean values of the onset timings for /m/ in all the test words were fairly close to the onset of the F_0 shift. There was no difference in the relative timing for /m/ between the rising and falling accent types.

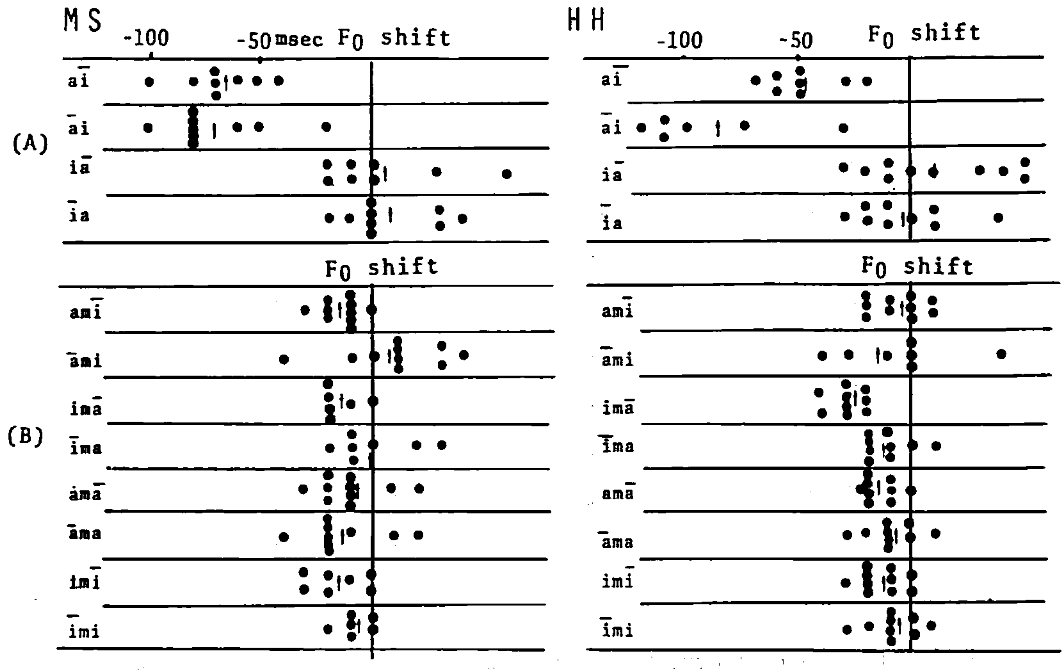


Fig. 1 *Timing of the onset of formant transition (A) and /m/ (B) relative to that of the F_0 shift. XX̄: rising accent; X̄X: falling accent; ↑: mean value.*

EMG data

As expected, in both subjects, there was a discernible change in the EMG level of the CT muscle corresponding to the F_0 shift for all the utterance samples; that is, an increase in the EMG level for the F_0 rise and a decrease for the F_0 fall. The occurrence of an increase in SH activity showed a quite different patterns between the rising and falling accent types. For the subject MS, an increase in the SH activity was observable in /ia/ and /ima/ with a rising accent, and in all the test words with a falling accent. For the subject HH, activation of SH muscle was discernible only in /ia/ with a rising accent, and in /aa/, /ama/, /ami/, /ia/ and /ima/ with a falling accent. The results are summariged in Table I. It can be concluded that SH activity is increased for pitch lowering as well as for jaw opening, although there is some individual variation in the occurrence of the activation. The occurrence of the SH activation appears to be consistent with both F_0 lowering and jaw opening.

Subject	MS		HH	
Muscle	SH	CT	SH	CT
aā	-	+	-	+
amā	-	+	-	+
iī	-	+	-	+
imī	-	+	-	+
aī	-	+	-	+
amī	-	+	-	+
iā	+	+	+	+
imā	+	+	-	+
āa	+	+	+	+
āma	+	+	+	+
īi	+	+	-	+
īmi	+	+	-	+
āi	+	+	-	+
āmi	+	+	+	+
īa	+	+	+	+
īma	+	+	+	+

Table I *Changes in CT activity and an increase in SH activity observable for test words with both rising and falling accent.* $X\bar{X}$*: rising accent;* $\bar{X}X$*: falling accent.*

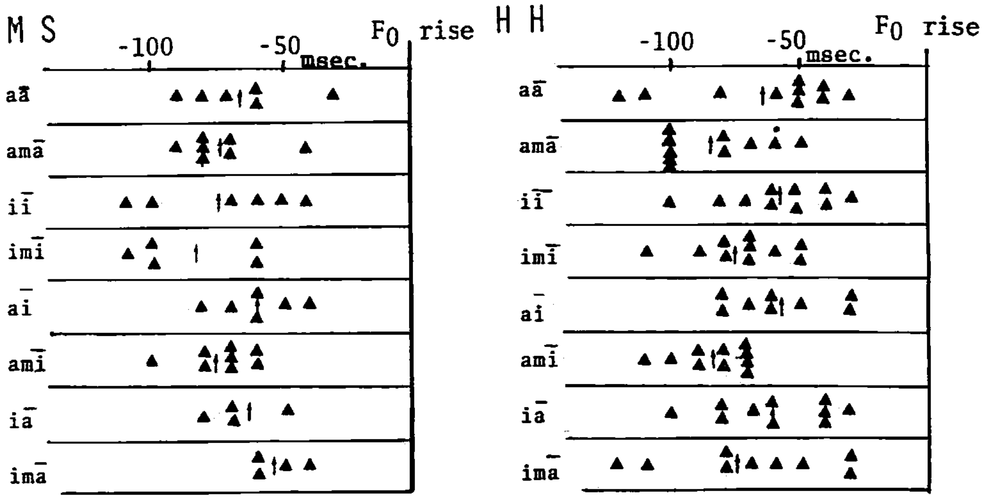


Fig. 2 *Timing of the onset of the increase in CT activity relative to that of the F_0 rise with a rising accent.*

It was observed that the change in CT activity took place earlier than F_0 shift. The timing of the onset of the increase in CT activity relative to the onset of the F_0 rise in the test words with rising accent is shown in Fig. 2. In the figure, the filled triangles indicate sample values and the upward arrows indicate mean values. There was little difference between the /ia/ or /ima/ samples and the samples for the other test words. This indicates that the activation of the SH muscle does not affect the F_0 raising mechanism of the CT muscles. In all the test words, the mean values were within the range of 100 msec to 50 msec before the onset of the F_0 rise, in both of the two subjects.

The timing of the onset of the decrease in CT activity relative to the onset of the F_0 fall in the test words with a falling accent is displayed in Fig. 3. There were considerable sample variations. The mean values are distributed in the range of 120 msec to 200 msec before the onset of the F_0 fall in the subject MS, the mean values for the subject HH being in the range of 80 msec to 150 msec. It can be noted that the values for /ai/ and /ami/ are considerably greater than those for the other test words in MS. This appears to have been responsible for the greater range of the distribution of the mean values in this particular subject. In HH, the values for /ai/ and /ami/ are not very different from those for the other test words. With a falling accent, the activity of the SH muscle increased for all of the test words

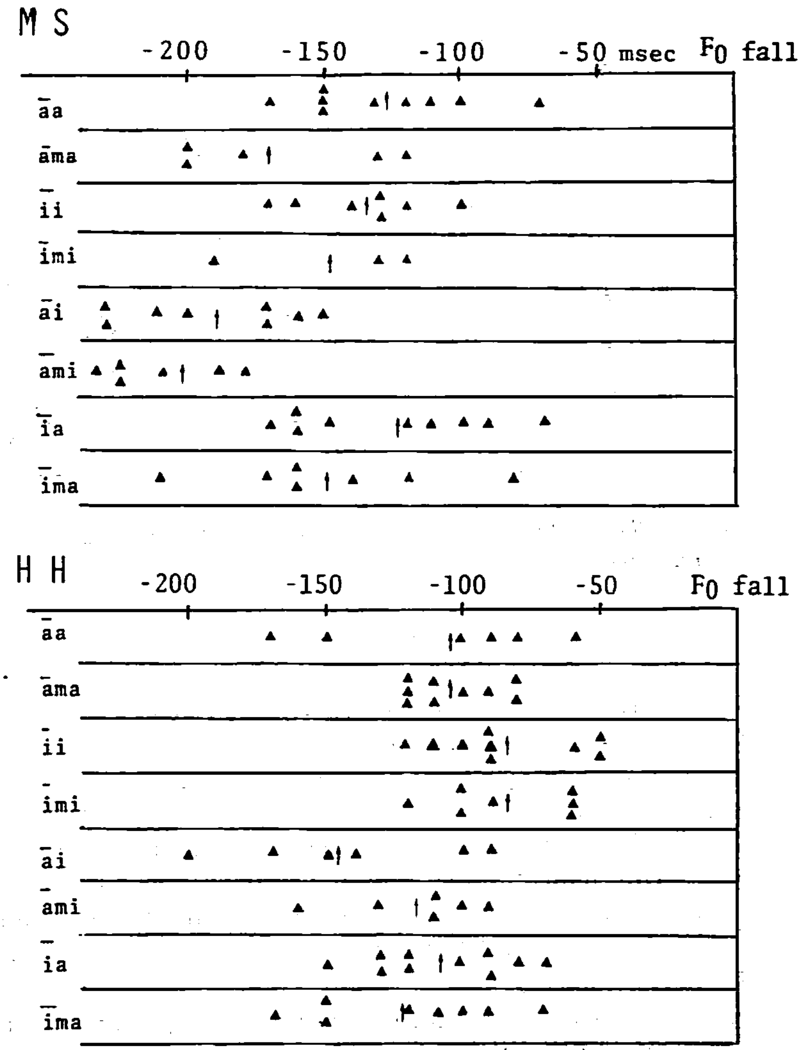
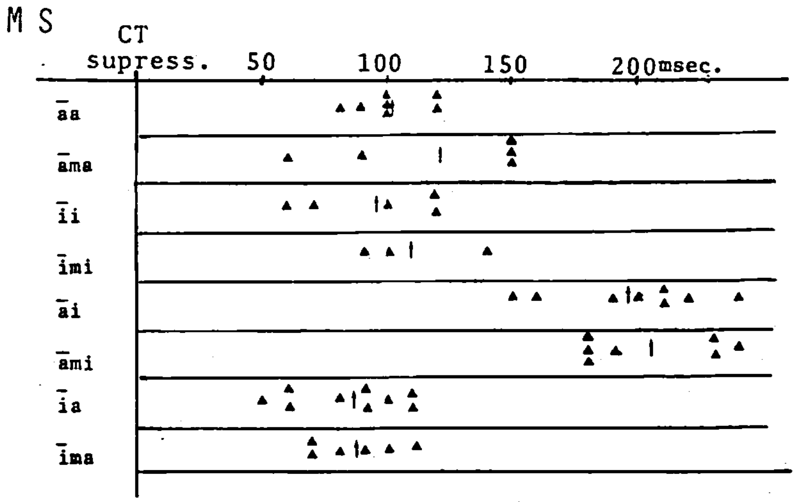


Fig. 3 *Timing of the onset of the decrease in CT activity relative to that of the F_0 fall with a falling accent.*

in MS, and for some of the test words in HH. The timing of the onset of the increase in SH activity relative to the onset of CT suppression is shown in Fig. 4. There was some time lag between CT suppression and SH activation. It is apparent that in MS the time lag was greater in /ai/ and /ami/ than in the other test words. It can be concluded that the onset of SH activation was suppressed for these two test words which involve a closing action of the jaw. This could have caused the greater time lag between the CT suppression and the F_0 fall in these particular test words.



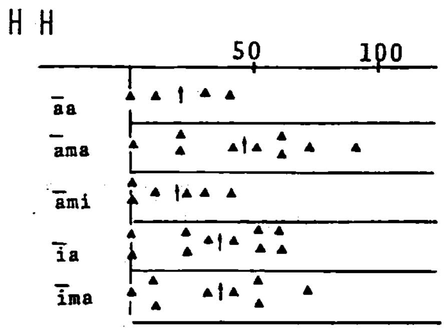


Fig. 4 *Timing of the onset of the increase in SH activity relative to that of the decrease in CT activity with a falling accent.*

Discussion

The acoustic analysis in the present study reveals that the articulatory lip closure for /m/ takes place nearly at the same time as the onset of the shift in vocal pitch in the realization of Japanese word accent. The synchronization of lip closure and vocal pitch shift is quite stable irrespective of difference in vowel sequences and accent types. This would suggest that the timing for consonant articulation provides a good reference in determining mora-boundaries. The timing of vowel transition relative to the shift in vocal pitch for /ai/ is considerably different from that for /ia/. This difference is consistent for both accent types. If we carefully examine the formant transition in the /ami/ and /ima/ samples, we can observe the formant transition occurs in advance of the /m/ closure in the /ami/ samples, almost at the same time for the /ai/ samples with reference to the F_0 shift, while it is not so in the /ima/ samples. This finding is supported by our previous observation of tongue movements using the X-ray microbeam system (Kiritani et al. 1977). The relative timing of the tongue and lip movements was observed in $V_1 m V_2$ sequences with different vowel combinations and with a flat accent type. It was noted that the tongue movement from the first vowel to the following vowel took place much earlier, with reference to /m/ closure, when the first vowel was /a/ as compared to when the first vowel was /i/. Thus, the asymmetry in the relative timing of vowel transition for /ai/ and /ia/ appears to be a matter of the inherent nature of the vowel articulation, rather than the interaction between the articulatory and phonatory mechanisms.

As already described above, the SH muscle is observed to be active for pitch lowering as well as jaw opening. The role of the extrinsic and intrinsic muscles in vocal pitch control has been extensively reviewed by Ohala (1978). Experimental observations have revealed an activation of some of the extrinsic muscles in corresponding to a lowering of the vocal pitch (Simada and Hirose 1971, Sawashima, et al. 1973; Erickson 1976; Sugito and Hirose 1978) or a low vocal pitch (Atkinson 1978). However, these results are not as consistent as those obtained for the cricothyroid activity in relation to changes in vocal pitch. It can be suggested that these muscles may not be causal mechanisms, but rather supplementary or maintaining mechanisms for vocal pitch lowering following the relaxation of the cricothyroid. In other words, these extrinsic muscles may actively contribute to the lowering of vocal pitch in the lower pitch range, while the F_0 fall in the middle and higher pitch range is predominantly controlled by a decrease in the cricothyroid activity.

In the present study, we examined the influence of the SH activity associated with jaw movement upon the pitch control mechanism of the CT activity. It is reasonable to assume that the effect of SH activity on the laryngeal pitch control mechanisms, if any, may well be manifested as a variation in the time lag between the change in CT activity and the shift in vocal pitch. With rising accent type, there was little influence of SH activity on the time lag for both of the two subjects. This would indicate a strong predominance of CT activation in the pitch raising mechanism. With falling accent type, the SH activity appears to have changed the time lag in one subject, but not so in the other. In MS, the time lag tended to be elongated by the suppression of the onset of the SH activity in the /ai/ and /ami/ samples, involving jaw closing action. The finding may be interpreted as indirect evidence of the positive contribution of SH activity to the pitch lowering in the other test words where the time lag is shorter with earlier onset of SH activity.

There is also subject-to-subject variation in the contribution of SH activity to the pitch lowering. It is assumed that there are several mechanisms which may contribute to pitch lowering, and the use of the SH muscle is optional for speakers especially in the modest pitch fall of the Tokyo accent. There is another condition to be taken into account in the present experiment. If MS was using the vocal pitch at the lower part of his vocal range as compared to HH, then it was natural that subject MS relied more on SH muscles for pitch lowering than subject HH. At present, we have no data to answer this question.

It was noted that the time delay from the change in CT activity to the pitch change was considerably different between the rising and falling accent types. It should also be pointed out that the relative timing of the articulatory event and the vocal pitch change at the acoustic level showed no difference between the two accent types. These findings indicate that the relative timing of the articulatory and vocal pitch controls is organized so as to compensate for timing variations in the internal mechanisms and maintain constant time relation in the final acoustic output.

References

Atkinson, J.E. (1978); Correlation analysis of the physiological factors concerning fundamental voice frequency. JASA 63: 211-233.

Erickson, D.M. (1976); A physiological analysis of the tones of Thai. Univ. Conneticut Ph. D. Thesis.

Fujisaki, H. (1977); Functional models of articulatory and phonatory dynamics. Temporal organization of articulatory and phonatory controls in realization of Japanese word accent. In Dynamic Aspects of Speech Production, edited by Sawashima, M. and Cooper, F.S. 346-366 (Univ. Tokyo Press).

Fujisaki, H., H. Morikawa and M. Sugito (1976); Temporal organization of articulatory and phonatory controls in realization of Japanese word accent. Ann. Bull. RILP, 10: 177-190.

Hirose, H. (1977); Electromyography of the larynx and other speech organs. In Dynamic Aspects of Speech Production, edited by Sawashima, M. and Cooper, F.S. 46-67 (Univ. Tokyo Press).

Kiritani, S., K. Itoh, H. Hirose and M. Sawashima (1977); Coordination of the consonant and vowel articulations. Proc. 9th Internat. Congr. Acoust. Madrid 1-12.

Ladefoged, P. (1964); Sited from Ohala J.J. and Eukel B.W. (1978).

Lehiste, I. (1970); Suprasegmentals (MIT Press).

Ohala, J.J. (1978); The production of tone. Report of the Phonalogy Lab. Berbeley 2: 63-117.

Ohala, J.J. and B.W. Eukel (1978); Explaining the intrinsic pitch of vowels. Report of the Phonology Lab. Berbeley 2: 118-125.

Sawashima, M., Y. Kakita and S. Hiki (1973); Activity of the extrinsic laryngeal muscles in relation to Japanese word accent. Ann. Bull. RILP, 7: 19-25.

Sawashima, M., H. Hirose, K. Honda and M. Sugito (1980); Relative timing of articulatory and phonatory controls in Japanese word accent: a study on nonsense two-mora words. Ann. Bull. RILP, 14: 139-147.

Simada, Z. and H. Hirose (1971); Physiological correlates of Japanese accent patterns. Ann. Bull. RILP, 5: 41-49.

Sugito, M. (1981); The relation between voice lowering for accent and articulation of vowels. Gengo Kenkyu, 79: 1-32.