Ann. Bull. RILP, No. 14, 53-63 (1980)

VOCAL PITCH CONTROL IN WORD ACCENT AND* SENTENCE INTONATION IN JAPANESE

Masayuki Sawashima

The contour of the vocal pitch, the voice fundamental frequency (Fo), plays an important role in the transmission of linguistic information concerning word meaning and sentence structure in many languages. It is known that in Japanese the Fo contour is the most important feature of word accent. Thus there have been fairly extensive experimental phonetic studies on vocal pitch control in the realization of Japanese word accent. In this paper, I will summarize recent works on this topic and also present some preliminary results of acoustic studies on sentence intonation in Japanese.

Physiological mechanism of vocal pitch control

The voice fundamental frequency is mainly controlled by the larynx. There have been numerous works on the laryngeal mechanism of Fo control during speech. It is generally agreed that among the laryngeal muscles the cricothryoid is the only muscle which is uniquely related to Fo control. The contraction of this muscle elongates the vocal folds, which results in an increase in the longitudinal tension and a decrease in the effective mass of the vocal fold tissue involved in vibration during phonation. Thus the contraction and relaxation of this muscle cause the Fo to rise and fall, respectively, especially in the higher pitch range.

Up and down movements of the larynx associated with Fo rise and fall have also been observed by many researchers. But the question of how these movements can affect the vocal pitch is still to be answered.

The Fo fall as well as the Fo rise are known to be important acoustic features in speech, especially in tone languages, and in Japanese word accent. This led us to raise the question whether or not there exists an "active" pitch lowering mechanism in addition to the "passive" control via the relaxation of the cricothyroid muscle. Experimental observations have revealed the activation of some of the extrinsic laryngeal muscles in correspondence to Fo lowering[1)-4)] or low Fo[5)]. However, these results are not so consistent as those obtained for the cricothyroid activity in relation to Fo changes. Furthermore, the timing of the activation of the extrinsic laryngeal muscles relative to the fall in Fo suggests that these muscles may not be causal mechanisms, but rather supplementary or maintaining mechanisms for Fo lowering following the relaxation of the cricothyroid.

In other words, these extrinsic muscles may actively contribute to the Fo lowering in the lower pitch range, while Fo fall in middle and higher pitch ranges is predominantly controlled by decrease in the cricothyroid

* This paper was presented at the symposium "The Roles of Prosody and Syntax in Language" sponsored by Grant-in-Aid for Special Project Research, Japanese Ministry of Education, Science and Culture, held in Tokyo, May 20, 1980. The author acknowledges the support of the grant to the works reported here.

activity. There seem to be some individual variations in the use of the extrinsic muscles for pitch lowering. Also, it should be noted that these extrinsic muscles belong to the infrahyoid muscles, which contribute to the downward movement of the larynx framework and also to some articulatory movements such as jaw opening.

At present, we may say that the activity pattern of the cricothyroid muscle provides the most straightforward physiological correlate in Fo control, while activation of the infrahyoid muscles may reflect the active pitch lowering gesture in the low pitch range. This figure of Fo control is considered to be a universal language mechanism, although the linguistic function of Fo changes may be different among different languages.

Examples of laryngeal muscle activities for Fo control in speech are shown in Figures 1 and 2. In Fig. 1, two different Fo contours for the sentences "i⌝zininaru" and "izi⌝ninaru" are illustrated. Lines 1 through 3 show the EMG patterns of the cricothyroid, the sternohyoid and the sternothyroid muscles for the two sentences. The activity pattern of the cricothyroid is seen to be uniquely related to the Fo changes. In this particular case, activity of the sternothyroid muscle, which is one of the infrahyoid muscles, also appears to contribute to the Fo lowering.

Figure 2 shows the Fo contours and EMG activities of the cricothyroid (CT) and sternohyoid (SH) muscles for the word /imi/ in the four different accent types of the Osaka (Kinki) dialect. Here also, the activity pattern of CT well represents the Fo contour. The activity of SH in this case appears to contribute to the lowering of Fo.

It is interesting to note that in Japanese word accent, and probably in some tonal patterns in other languages, a rise in Fo or a high Fo state precedes the rapid Fo fall.

Functional model of Fo control in Japanese word accent

A functional model for the generation of Fo contours of Japanese word accent has been proposed by Fujisaki[6,7]. The model is based on acoustic analysis of speech waves by the method of Analysis-by-Synthesis. The basic concept of the model is illustrated in Fig. 3. In this model, linguistic factors of voicing and accent are both assumed to take the form of stepwise binary commands to the Fo control mechanism. Commands for voicing and accent are smoothed separately by the low-pass characteristics of their respective control mechanisms. This smoothing effect represents the response characteristics of the biomechanical system of the human vocal organs to the input of discrete linguistic units. The smoothing process is approximated by a critically damped second-order linear system, and outputs of both the voicing and accent control mechanisms are combined to control the fundamental frequency of glottal oscillations through a nonlinear mechanism.

Using this model, it is possible to extract, from the speech waves, parameters that characterize the timing of voicing and accent commands as well as those of various properties of the control mechanism as defined in the model. The timing of the accentual command has been found to be crucial in characterizing various word accent types.[8,9]

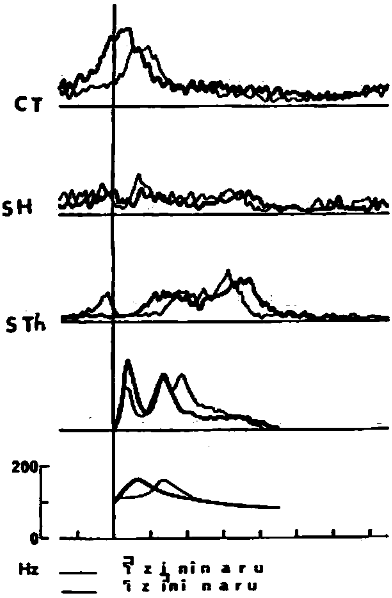

Fig. 1 Fo and EMG patterns for "i zininaru" and "izi ninaru". Simada and Hirose. (1971)[1]

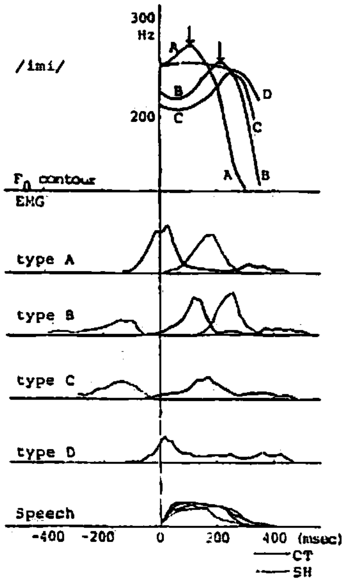

Fig. 2 Fo and EMG patterns for four accent types found in the Osaka (Kinki) dialect. Sugito and Hirose (1978)[4]

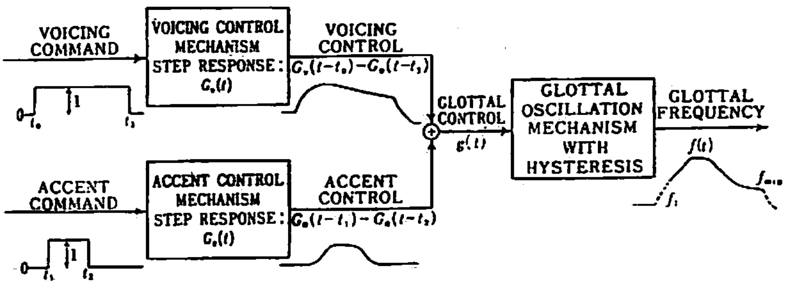

Fig. 3 Model of control mechanism of glottal frequency. Fujisaki and Sudo (1971)[6]

Examples of Analysis-by-Synthesis of Fo contours of four accent types of two-mora words are presented in Fig. 4. Accent types A through D are listed in Table 1. Types A and C are common to both the Tokyo and Osaka dialects, while type B, with a marked downward pitch transition in the second mora, as well as Type D are peculiar to the Osaka dialect.

Figure 4 shows the measured Fo contour, its best approximation obtained by Analysis-by-Synthesis, and the timing of the accent command extracted from the speech samples of [ame] with four different accent types. The results indicate that the model is valid for all of the accent types. It was also revealed that the activity of the cricothyroid muscle showed a good correspondence to the timing of accent command estimated by the model.[9] Perceptual experiments using synthetic speech of [ame] with varying timing of the segment boundaries relative to the accent command have revealed that the identification of an accent type is based predominantly on the relative timing of the accent command and a specific segment boundary within a word.[10].

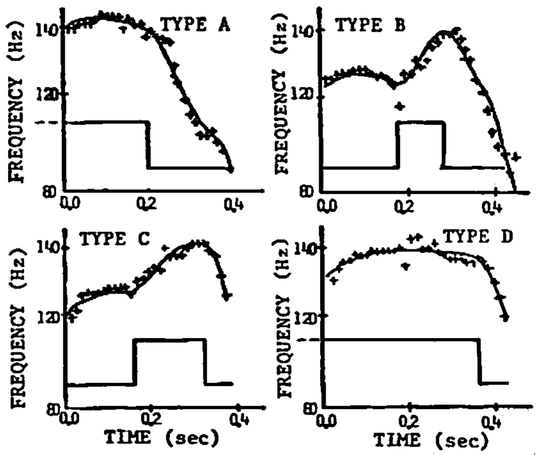

Fig. 4 Analysis-by-Synthesis of pitch contours of four accent types of two-mora words [ame] together with extracted accent command. Fujisaki et al. (1976)[9]

Accent type	A	B	C	D
Subjective pitch				
[iki]	abandon (Osaka)	chic (Osaka)	breath (Osaka)	going (Osaka)
[ame]	rain (Tokyo)	rain (Osaka)	candy (Tokyo)	candy (Osaka)

Table 1 Examples of accent types of two-mora words in the Osaka and Tokyo dialects

Temporal organization of phonatory and articulatory controls in the realization of word accent in Japanese

In Japanese, the accent pattern as represented by the Fo contour is closely linked with a time segment of the syllable or so-called "mora". Thus laryngeal Fo control should be expected to have some appropriate timing relative to supraglottic articulatory control. For analyzing dynamic properties of articulatory control, Fujisaki[7] proposed a functional model which is based on the same principle as his model of phonatory control. In this model, formant trajectories are analyzed on speech waves of a series of connected vowels. The timing of initiation of the formant transition for each of the successive vowel phonemes as well as their target formant frequencies and other dynamic parameters are estimated.

Using these functional models of phonatory and articulatory controls, two-mora words with Osaka accent types were analyzed. The results revealed that there was a time lag between the onset of articulatory and accent controls. The articulatory control takes place earlier than accent control.[11] Our preliminary experiments with the Tokyo accent[12] have also revealed the same tendency as shown in Figures 5 and 6. Figure 5 shows the acoustic data of speech samples of two-mora words with a rising accent pattern /ai⌉/ and /ia⌉/. The Fo contour and formant shift data obtained by Fujisaki's method are displayed. The vertical line Tp indicates the estimated timing of accent command, and Tf the onset of formant transition. It is apparent that the onset of formant transition takes place earlier than that of Fo transition.

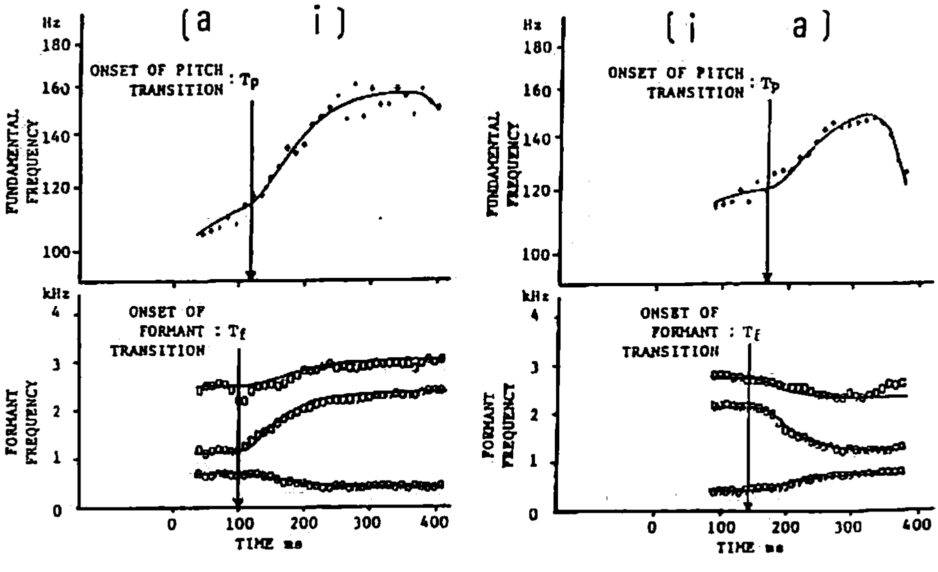

Fig. 5 Comparison of Fo and formant transitions for /ai⌉/ and /ia⌉/. Sawashima (1979)[12]

Figure 6 shows integrated and smoothed EMG curves for each of the tokens used for the acoustic analysis. CT indicates the cricothyroid muscle for Fo control and GG the genioglossus muscle for tongue movement, which corresponds to the formant shift. The thick vertical bar indicates the onset of the vowel of the first mora. Time points of Tp and Tf obtained by acoustic analysis are also indicated by thin vertical bars on the EMG curves. The open circle on the CT curve indicates the onset of the increase in CT activity for the pitch rise. The solid circle on the GG curve of /ai / indicates the onset of the increase in GG activity corresponding to the tongue fronting from /a/ to /i/ The solid circle on the GG curve for /ia / indicates the onset of the decrease in GG activity, which presumably corresponds to the tongue backing from /i/ to /a/. For these samples, the relative timing between phonatory and articulatory controls observed from EMG data is in close agreement with that observed from the acoustic data.

We next undertook another experiment.[13] The experiment was designed to answer the following two questions: (1) whether or not the relative timing of articulatory and phonatory controls in rising accent type may differ from that in falling accent type, and (2) whether or not this relative timing may differ with different combinations of the open and close vowels.

In the experiment a subject pronounced the two-mora words /ai/, /ia/, /ami/, /ima/, /ama/ and /imi/ in isolation ten times each, in random order with both rising and falling accent types. EMG of laryngeal muscles and articulatory muscles were recorded simultaneously with sound recording in order to observe relative timing in the action of different speech muscles. Acoustic analysis of speech was then made using a sound spectrograph, and the onset of articulatory events and Fo changes was determined by visual examination of sound spectrograms. The EMG data are not completely prepared yet, and here I will present some acoustic data.

Figure 7 summarizes the timing of the onset of formant transition relative to Fo change for /ai/ and /ia/ in rising and falling accent types. The time point of 0 on the time axis indicates the onset of Fo change for each sample. Filled circles indicate onset of formant transition for the samples with falling accent type, and open circles indicate that for the samples with rising accent type. Comparing /ai/ and /ia/, it is apparent that the formant transition for /ai/ takes place earlier than for /ia/ with reference to the onset of Fo change. As far as the present data are concerned, there appears to be some interaction between the vowel articulation and the Fo control. In other words, the data suggest that the temporal organization of articulatory and phonatory controls may be affected by the direction of vowel transition.

It is also noted that the formant transition tends to occur earlier in the rising accent type than in the falling accent type for each of the /ai/ and /ia/ sequences. This suggests that the temporal organization of articulatory and phonatory controls may also be affected by the direction of Fo change.

Figure 8 shows the timing of segment /m/ relative to the onset of Fo change in /ami/ and /ima/ in rising and falling accent types. It is noted that in both /ami/ and /ima/ series the articulatory closure for /m/ tends to occur earlier in the rising accent type than in the falling accent type. The tendency is consistent with the /ai/ and /ia/ series. On the other hand, there is little difference in the timing of the articulatory closure between the /ami/ and /ima/ series. The data suggest that the onset timing of the

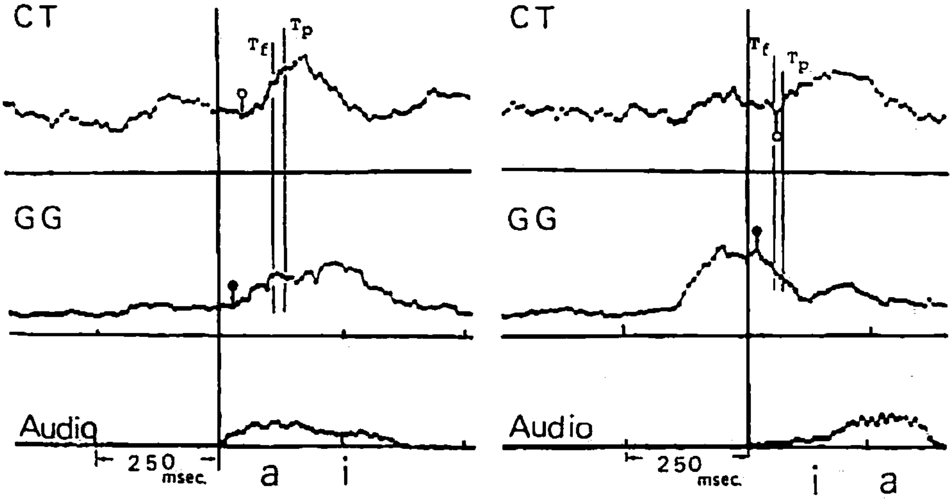

Fig. 6 Comparison of EMG curves of cricothyroid (CT) and genioglossus (GG) muscles for the same samples as Fig. 5. Sawashima (1979)[12]

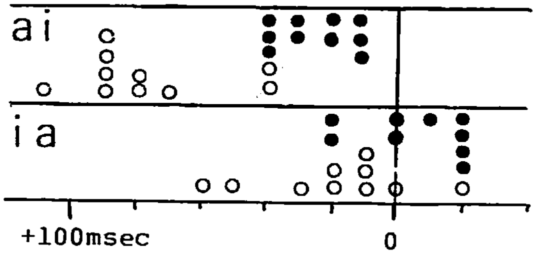

Fig. 7 Timing of the onset of formant transition relative to that of Fo change.
o: rising accent, ●: falling accent.
Sawashima et al.(1980)[13]

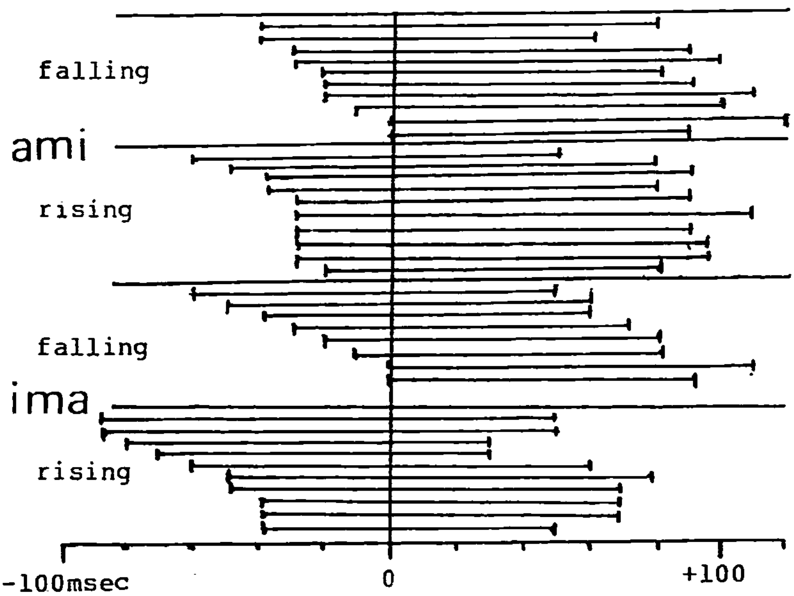

Fig. 8 Timing of the segment [m] relative to the onset of Fo change. Sawashima et al. (1980)[13]

consonant articulation relative to Fo change is fairly stable within the same accent type. It is quite probable, on the other hand, that the timing of vowel transition relative to /m/ in /ami/ is considerably different from that in /ima/.

Data for the /ama/ and /imi/ series also indicate the earlier onset of /m/ in the rising accent type than in the falling accent type.

Vocal pitch pattern in the sentence

The last topic is the control of Fo contour in the sentence. It is known that the basic feature of Fo contour in speech is characterized by a gradual declination from the onset toward the end of utterance. This basic form is modified by the local prominence corresponding to word accent in the simple declarative sentence.

Analysis-by-Synthesis using Fujisaki's functional model described above works well also for analyzing the underlying voicing component and superposed accent component in the spoken sentence.[14] Here I will simply present some results in Fig. 9. In the figure, Fo contours extracted from speech waves for various sentences are presented by successive "o"'s, and the best approximations of Fo contours by the model are indicated by solid lines. Broken lines indicate the underlying voicing component. As seen in (a) and (b) of the figure, Fo contours of short sentences are well approximated by one stroke of voicing component with superposed accent component. When we assume one stroke of voicing component for the sentence in (c),

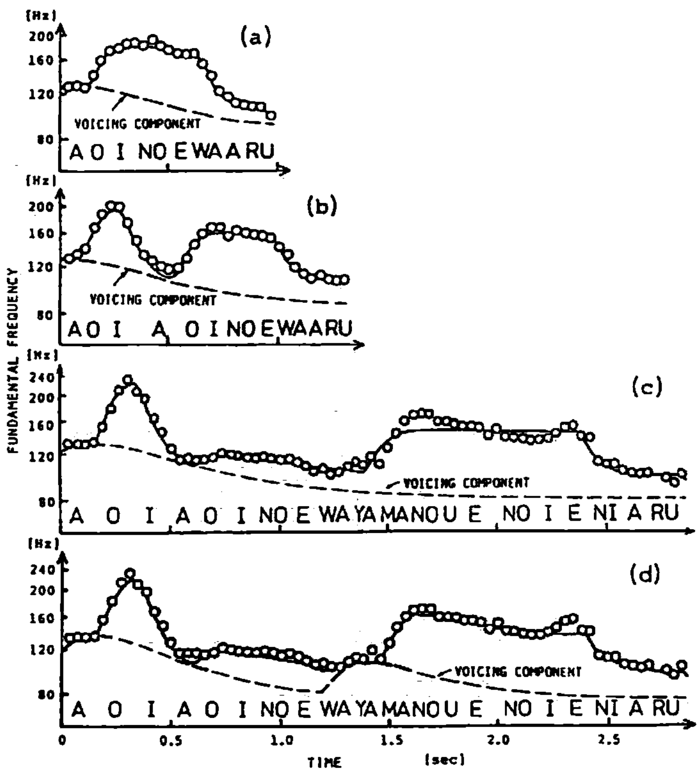

Fig. 9 Analysis-by-Synthesis of sentence Fo contours.
Fujisaki et al. (1979)[14]

there appears some discrepancy between the observed Fo and the model approximation especially in the latter half of the sentence. The approximation becomes much more accurate by using two strokes of voicing component as seen in (d) of the figure. The result indicates that there is a resetting of voicing in the speech sample for this sentence, although the sentence was read in one breath. It is also noted that the revoicing takes place in close correlation with the syntactic structure of the sentence.

Another work to be briefly mentioned here is the relationship between Fo contours and complex sentence constructions as studied by Uyeno and others.[15] In this study, three groups of test sentences were used: relative clause constructions (RC), a pair of simple sentences uttered in succession (SS), and a coordinate conjunction constructions (CC). Pitch patterns for SS and CC were found to be almost identical. In SS, there was a Fo rise at the initiation of the succeeding sentence. In CC, there was a Fo rise at the initiation of the succeeding clause.

RC construction was divided into two types: left-branching (RCL) and center-embedding (RCC). Examples of RCL, RCC and SS with the same number of mora and the same word accent are as follows:

RCL: [[ototoi naita] $_{S_1}$ otona-ga waratta] $_S$

day before yesterday / cried / adult / laughed

The adult who cried the day before yesterday laughed.

RCC: [ototoi [naita] $_{S_1}$ otona-ga waratta] $_S$

The adult who cried laughed the day before yesterday.

SS: [otona-ga naita] $_S$ [otona-ga waratta] $_S$

adult / cried / adult / laughed

An adult cried. An adult laughed.

An example of mean Fo contours from these types of sentences is shown in Fig. 10. It is noted that RCL and SS show Fo contours which are similar to each other. In RCC Fo contour shows a prominent pitch rise corresponding to the center-embedded relative clause, the Fo value being almost as high as that at the onset of utterance.

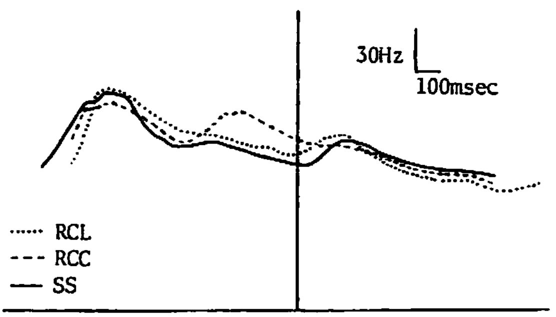

Fig. 10 Mean Fo contours of RCL, RCC and SS. Uyeno et al. (1979)[15]

The results may be summarized as follows:

i. The pitch contour of a declarative sentence rises at the onset of an utterance and gradually declines toward the end.

ii. Pitch curves rise at the clause initial. This is interpreted as a function of Fo contour which identifies a clause boundary.

iii. Variation in Fo contour may contribute to clarify certain complex sentences which would otherwise be structurally ambiguous.

The study was further extended to a perceptual experiment. The results revealed that a remarkably higher pitch assignment to the portion of a relative clause is necessary in order to obtain the clear interpretation of the listener for a center-embedded construction.[16]

References

1) Simada, Z. and H. Hirose (1971). "Physiological correlates of Japanese accent patterns." Ann. Bull. RILP 5, 41-49.
2) 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.
3) Erickson, D.M. (1976). "A physiological analysis of the tones of Thai." Univ. Connection Ph.D. Dissertation.
4) Sugito, M. and H. Hirose (1978). "An electromyographic study of the Kinki accent." Ann. Bull. RILP 12, 35-51.
5) Atkinson, J.E. (1978). "Correlation analysis of the physiological factors concerning fundamental voice frequency." J. Acoust. Soc. Am. 63, 211-222.
6) Fujisaki, H. and H. Sudo (1971). "A model for the generation of the fundamental frequency contours of Japanese word accent." J. Acoust. Soc. Japan. 27, 445-453.
7) 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 M. Sawashima and F. S. Cooper, 346-366, Univ. Tokyo Press.
8) Fujisaki, H., Y. Mitsui and M. Sugito (1974). "Analysis, synthesis, and perception of accent types of two-mora words in Tokyo and Osaka dialects." Transaction of the Committee on Speech Research, Acoust. Soc. Japan, S73-51.
9) Fujisaki, H., H. Hirose and M. Sugito (1976). "Analysis, synthesis, and perception of word accent types in Japanese." Ann. Bull. RILP 10, 173-176.
10) Fujisaki, H. and M. Sugito (1976). "Acoustic and perceptual analysis of two-mora word accent types of the Osaka dialect." Ann. Bull. RILP 10, 157-171.
11) 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.

12) Sawashima, M. (1979). "Temporal patterns of articulatory and phonatory controls." Ann. Bull. RILP 13, 1-11.
13) 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, dysillabic words." Ann. Bull. RILP 14. 139-147
14) Fujisaki, H., K. Hirose and K. Ohta (1979). "Acoustic features of the fundamental frequency contours of declarative sentences in Japanese." Ann. Bull. RILP 13, 163-173.
15) Uyeno, T., H. Hayashibe and K. Imai (1979). "On pitch contours of declarative, complex sentences in Japanese." Ann. Bull. RILP 13, 157-187.
16) Uyeno, T., H. Hayashibe, K. Imai, H. Imagawa and S. Kiritani (1980). "Comprehension of relative clause construction and pitch contours in Japanese." Ann. Bull. RILP 14, 225-236.