

ANALYSIS, SYNTHESIS, AND PERCEPTION OF WORD ACCENT TYPES
IN JAPANESE*

Hiroya Fujisaki, Hajime Hirose and Miyoko Sugito**

Fundamental frequency of the glottal source (F_0) is the primary acoustic correlate of prosodic features such as word accent and intonation in spoken Japanese. In particular, in the Tokyo dialect as well as in many other dialects, it is used to draw the binary distinction of subjective pitch associated with each mora of a word and to discriminate between a number of word pairs which are phonemically identical, such as [āme] (rain) and [amē] (candy). In these dialects, each word (except auxiliaries) is associated with a definite type of pitch accent, i. e., a rise-fall pattern of subjective pitch. These binary patterns, however, never manifest as such in the F_0 -contour. Due to the smoothing characteristics of various neural, muscular, and pneumatic components in the control mechanism of glottal oscillations, the discrete linguistic information gives rise to a continuous F_0 -contour which smoothly rises and falls at the accented morae, superposed on a base line that initially rises and then gradually decays toward the end of an utterance.

In order to separate linguistically relevant information from characteristics of the glottal mechanism, one of the present authors has proposed a functional model for the control mechanism of glottal oscillations that generates the F_0 -contour from binary commands of voicing and accent.¹ Analysis-by-Synthesis of F_0 -contours of various accent types in the Tokyo dialect has revealed that the onset and the offset of the extracted accent command relative to the segmental features of each mora are most essential in characterizing various accent types, while parameters such as magnitude and rate of rise and fall of F_0 -contours characterize the glottal mechanism and its manner of control by individual speakers. The model has also been extended to F_0 -contours of sentences.²

While subjective patterns of pitch variation are always synchronized with successive morae in the Tokyo dialect, this is not the case for Kinki dialects such as the Osaka and Kyoto dialects. Namely, these dialects possess accent types which are marked with pitch transitions within a mora, such as in [è] (food) and [é] (picture). In the present study, F_0 -contours of these peculiar accent types as well as those of other types are analyzed to see if the above-mentioned model is also applicable in these cases. The perceptual relevance of the extracted parameters is also examined by identification tests using synthetic words generated by a digital computer.





* Abstract of paper presented at the Eighth International Congress of Phonetic Sciences, Leeds, August 17-23, 1975.

** Department of Japanese Literature, Faculty of Education and Liberal Arts, Osaka Shoin Women's College.

Analysis of F_0 -contours

Since the peculiar patterns are found only in one- and two-mora words uttered in isolation, analysis of F_0 -contours was conducted first on those two-mora words for which all the accent types exist as meaningful words with the identical phoneme sequences, as shown by the examples of Table 1. While Types A, C, and D are common to both Tokyo and Osaka dialects, Type B with a marked downward pitch transition in the second mora is peculiar to the Osaka dialect. These words were pronounced in

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

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)

random order by one male and two female speakers of the Osaka dialect. The male speaker also had a good command of the Tokyo dialect. In the case of the two female speakers, simultaneous recordings of EMG from lateral crico-arytenoid (LCA), crico-thyroid (CT), and sterno-hyoid (SH) muscles were also conducted. The speech material was sampled at 10 kHz with an accuracy of 10 bits, and the fundamental frequency was extracted at intervals of 12.8 msec. Parameters characterizing each F_0 -contour were then extracted by Analysis-by-Synthesis, based on the above-mentioned model. They are: onset (T_0) and offset (T_3) of voicing command, onset (T_1) and offset (T_2) of accent command, amplitude (A_v) and rate (α) of response to voicing command, amplitude (A_a) and rate (β) of response to accent command, and lower bound (F_{min}) of fundamental frequency.

Figure 1 shows the measured F_0 -contour, its best approximation obtained by Analysis-by-Synthesis, and the timing of the accent command extracted from each of the four accent types of [ame] uttered by the male speaker. It indicates that the model is valid for Type B as well as for the other types. Figure 2 shows results of an analysis of six utterances each of the four types by the same speaker, in terms of the onset (T_1) and the offset (T_2) of the accent command relative to the onset of the utterance, together with various segment boundaries. The analysis also indicates that Types A and B are characterized by somewhat larger values of α as compared to Types C and D, suggesting that rapid downward pitch transitions in Types A and B may require some additional adjustment of the glottal mechanism. Similar analysis was also made of the three types (Type A, C, and D) of the one-mora word [e]. It was found that F_0 -contours of Types A and D were quite similar to those of two-mora words, and Types A and C clearly indicated the existence of pitch control within a syllabic unit.

Comparison of F_0 -contours and the extracted commands with the EMG activity in the two female subjects indicated that CT activity corresponded well to the accent command in both subjects, while some individual differences were observed in the activity of SH. Lateral crico-arytenoid activity obtained from only one of the subjects appeared to correspond to the voicing command.

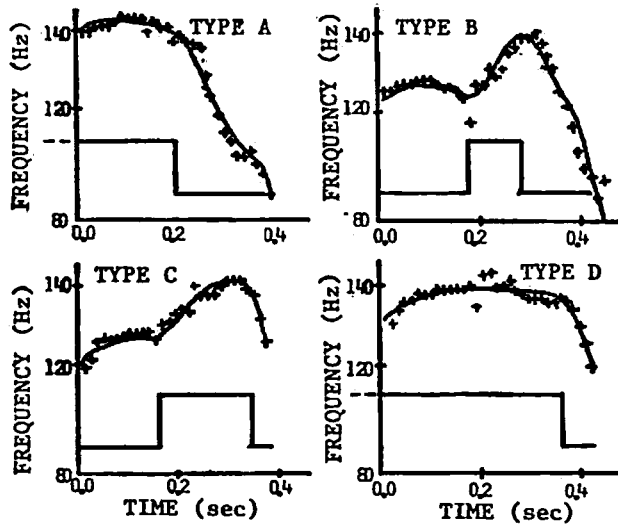


Fig. 1. Analysis-by-Synthesis of pitch contours of four accent types of two-mora words [ame] together with extracted accent command.

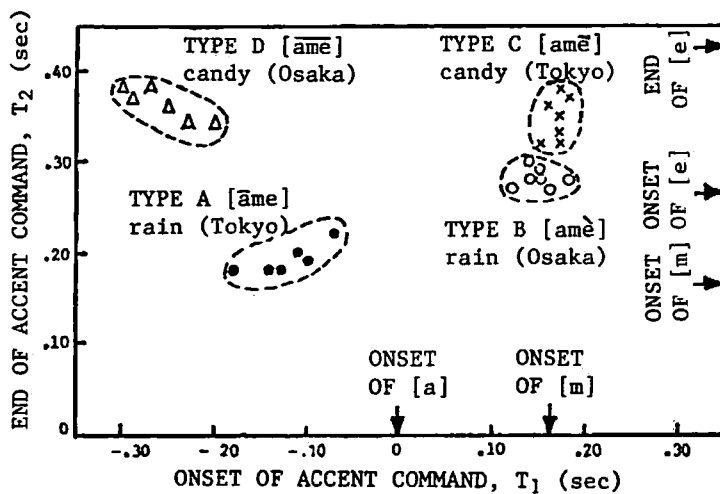


Fig. 2. Representation of four word accent types of [ame] in terms of onset (T_1) and end (T_2) of the accent command.

Synthesis and Perception of Word Accent Types

Although the analysis of F_0 -contours revealed the importance of the timing of the accent command relative to segmental features of an utterance, its perceptual relevance has yet to be confirmed by listening experiments. Furthermore, it is not clear from the analysis of natural utterances which segment boundary serves as the crucial reference for the timing of onset and offset of the accent command in each accent type. In order to clarify these points, synthetic words [ame] with the identical segmental features but with a variety of F_0 -contours were generated by digital computer simulation of a series-type terminal-analog speech synthesizer with five formants.

First, forty stimuli were selected on the T_1 - T_2 quadrangle whose vertices were the four points representing typical utterances of the four accent types. These stimuli were then randomized and used for identification tests of accent types. The subjects consisted of five female adults with normal hearing; all were speakers of the Osaka dialect, though they were also familiar with the Tokyo dialect. Results of the identification tests were analyzed individually to give estimates of category boundaries between accent types on the stimulus continuum; the individual differences in boundaries were found to be quite small, however.

Having thus determined the boundaries between accent types on the T_1 - T_2 continuum, we then measured the influence of a shift in various segment boundaries on the perceptual boundaries between accent types. The main findings of these experiments are:

- (1) A shift in the position of the intervocalic [m] causes a significant shift in the boundary between Type A and Type B, while it scarcely affects other boundaries.
- (2) A shift in the duration of the final vowel [e] causes almost the same amount of shift in the boundary between Type B and Type C in terms of T_2 , indicating that a certain length of the final vowel is required to realize the downward pitch transition characteristic of Type B.

These results confirmed that perception of a particular accent type is closely tied to a specific temporal relationship between a segment boundary and the onset or the offset of the accent command.

References

1. Fujisaki, H. and S. Nagashima (1969), "A Model for the Synthesis of Pitch Contours of Connected Speech," Annual Report of the Engineering Research Institute, Faculty of Engineering, University of Tokyo, 28, 53-60.
2. Fujisaki, H. and H. Sudo (1971), "A Model for the Synthesis of Prosodic Pitch Contours of Connected Japanese," J. Acoust. Soc. Japan, 27, 396-397.