Ann. Bull RILP. No. 10. 157-171 (1976)

ACOUSTIC AND PERCEPTUAL ANALYSIS OF TWO-MORA WORD ACCENT TYPES IN THE OSAKA DIALECT

Hiroya Fujisaki and Miyoko Sugito*

I. Introduction

It has been widely recognized that word accent, which is one of the major constituents of prosody in a number of dialects of spoken Japanese, manifests itself primarily as the pattern of temporal variations in the fundamental frequency of voiced segments (hereafter to be referred to simply as pitch contour). It has not been made clear, however, which of the features found in the pitch contour as a continuous function of time constitute acoustic correlates of the word accent type as a discrete unit of linguistic information intended by the speaker and thus perceived by the listener. This situation may have been ascribable firstly to the lack of means for quantitative measurement of pitch contours as a more concrete basis of investigation than is provided by mere introspection; secondly, to the lack of knowledge concerning those processes of generating pitch contours which will help us to infer and extract the essential factors characterizing a pitch contour; and thirdly, to the lack of experimental techniques for examining the roles of such factors in the perception of accent types.

Recent advances in experimental techniques for analysis and synthesis of speech, as well as for observation of processes of speech production, however, have greatly changed the situation by making possible the elucidation of the processes of production, transmission, and perception of the accentual information. A technique for automatic extraction of the fundamental frequency of speech by means of a digital computer has been introduced by one of the authors;[1] it was later improved for greater accuracy of pitch contour measurement.[2] Furthermore, a functional model has been presented for the processes of generating pitch contours from a set of linguistic commands;[3] it has proved capable of closely approximating observed pitch contours.

Analysis-by-Synthesis of pitch contours based on this model has made possible separation of linguistic factors from physical and physiological properties of various mechanisms involved in the control of the fundamental frequency of glottal vibration.[4] Results of such an analysis have indicated that the major acoustic correlate of the word accent type is the temporal relationship of accentual command relative to segmental features of speech, thus suggesting that perception of accent types may depend not on the value of fundamental frequency at any particular point on the pitch contour, but rather on the instants of onset or offset of accentual command initiating upward or downward pitch transition. The model has been applied, however, mostly to the analysis of word accent types of the Tokyo dialect, so that its generality has yet to be tested on other dialects such as the Kinki dialects, which possess a wider variety of word accent types than that found

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

in the Tokyo dialects. Furthermore, the perceptual significance of various parameters of the model needed to be confirmed by experiments using synthetic speech sounds with precise control of these parameters.

The present study was thus conducted as part of an investigation aimed at establishing a quantitative theory of production and perception of prosody in spoken Japanese. The validity of the above-mentioned model was tested on two-mora word accent types of the Osaka dialect, which possesses peculiar accent types not found in the Tokyo dialect. The perceptual significance of temporal relationships between the accentual command and the segmental features was further examined by identification tests of word accent types using synthetic speech stimuli.

II. Accent Systems of Kinki and Tokyo Dialects

A number of written accounts of word accent, especially of dialects of the Kinki district, have been found in classic literature and Buddhist texts over the past millenium. These have proved to be instrumental in revealing a certain regularity in diachronic changes of the word accent system of the Kinki dialects, which were accepted as normative dialects until the latter half of the 19th century.[5] For the most part, these records are based on binary notation of the subjective pitch associated with each mora, i.e., a unit of quantitative meter; there exist, however, certain instances where strict association of the binary pitch level with a mora does not seem to be appropriate.

Written records of the word accent of the Tokyo dialect, which has been adopted as standard Japanese since the Meiji reform, are rather recent: in 1892 M. Yamada adopted a binary system for the accentual description in his dictionary of Japanese.[6] A comparison of word accent systems of the Kinki and Tokyo dialects was made in 1928 by E. D. Polivanov,[7] also on the basis of binary notation. Limitations of the binary descriptions, however, have been pointed out by several researchers.

A comparison of word accent systems of the Tokyo and Osaka dialects is shown in Table 1 which lists patterns of subjective pitch for all possible accent types in words consisting of up to four morae, along with a sample word for each type. A circle (O) in the table represents a particular mora, a horizontal bar (‾) on top of a circle indicates a "high" subjective pitch associated with that mora, and a triangle (▷) indicates a one-mora particle ("ga" or "wa") whose existence serves, in some instances, to discriminate between two different types which are otherwise indistinguishable. The existence of a particle may also modify the pattern of subjective pitch in certain accent types.

The word accent system of the Tokyo dialect is characterized by the two constraints: 1) the subjective pitch invariably displays a transition, either upward or downward, at the end of the initial mora, and 2) no more than one downward pitch transition is allowed within a word. The type with a "high" last mora further divides itself into two depending on whether or not the high pitch is carried over to the following auxiliary. Thus the total number of accent types of n-mora words is n+1.

The word accent types of the Kinki dialects, on the other hand, are divided into two groups depending on whether the initial mora is "high" or "low." The "high-start" group is constrained by the fact that no more than one downward pitch transition is allowed within a word, resulting in a total

Table 1. Classification and comparison of word accent types in the Tokyo and Osaka dialects.

Dialect \ Mora		1	2	3	4
Tokyo		[e] (handle)	[ame] (candy)	[kodomo] (child)	[akatsuki] (dawn)
		[e] (picture)	[ame] (rain)	[midori] (green)	[ookami] (wolf)
			[haʃi] (bridge)	[anata] (you)	[magokoro] (sincerity)
				[tooge] (pass)	[sakazuki] (wine cup)
					[gokuraku] (paradise)
Osaka	"high-start"	[e] (handle)	[ame] (candy)	[kodomo] (child)	[akatsuki] (dawn)
		[ke] (hair)	[haʃi] (bridge)	[anata] (you)	[ookami] (wolf)
				[itʃizi] (one o'clock)	[magokoro] (sincerity)
					[gokuraku] (paradise)
	"low-start"	[e] (picture)	[asa] (hemp)	[tooge] (pass)	[gakkoo] (school)
			[ame] (rain)	[midori] (green)	[konban] (tonight)
					[irogami] (color paper)

of n accent types. The "low-start" group is further constrained by the fact that only one of the subsequent morae can be "high," resulting in a total of n-1 accent types. The "low-start" type which ends up with a "high" mora is peculiar in that a shift of the "high" mora occurs when followed by an auxiliary. The total number of accent types of n-mora words is thus 2n-1 in the Kinki dialects.

Although the binary notation is found to be valid for the accent types occurring in the Tokyo dialect as well as for those of the Osaka dialect for words containing three or more morae, the latter dialect possesses certain one-and two-mora word accent types which do not lend themselves to binary notation when pronounced in isolation; this is due to the fact that the subjective pitch is either noticeably rising (as indicated by $\acute{\circ}$) or noticeably falling (as indicated by $\grave{\circ}$) within a mora. It is also to be noted that these peculiar patterns of pitch change are lost when these words are combined with an auxiliary, as indicated in Table 1.

Thus the contemporary Kinki dialects, as represented by the Osaka dialect in Table 1, possess four accent types in two-mora words, i.e., those indicated by $\bar{\circ}\circ$, $\circ\grave{\circ}$, $\circ\bar{\circ}$, and $\overline{\circ\circ}$ (which shall hereafter be designated respectively as Types A, B, C, and D), and three accent types in one-mora words, i.e., those indicated by $\grave{\circ}$, $\acute{\circ}$, and $\bar{\circ}$ (which shall be designated respectively as Types A, C, and D on the basis of the similarity of their pitch contours with those of the corresponding types of two-mora words). A dictionary compiled toward the end of the Heian Period (Ruijū-myōgishō, ca. A.D. 1100), however, indicates the existence of still other accent types (characterized by a low flat pitch, to be indicated by $\circ\circ$ and $\circ$). Systematic correspondneces can of course be observed in the diachronic changes of accent types of the same dialect, as well as in the synchronic differences in accent types of different dialects, as shown by Table 2.

Table 2. Diachronic and synchronic variations of word accent types in the ancient and modern Kinki dialects and in the Tokyo dialect.

	Word		Ancient Kyoto	Contemporary Kyoto & Osaka	Contemporary Tokyo
Two-mora	[hana]	(nose)	$\overline{\circ\circ}$	$\overline{\circ\circ}$	$\circ\bar{\circ}$ ($\circ\overline{\circ\triangleright}$)
	[oto]	(sound)	$\bar{\circ}\circ$ }	$\bar{\circ}\circ$	$\circ\bar{\circ}$ ($\circ\bar{\circ}\triangleright$)
	[hana]	(flower)	$\circ\circ$ }		
	[sora]	(sky)	$\circ\bar{\circ}$ ($\circ\circ\bar{\triangleright}$)	$\circ\bar{\circ}$ ($\circ\circ\bar{\triangleright}$) }	$\bar{\circ}\circ$ ($\bar{\circ}\circ\triangleright$)
	[koe]	(voice)	$\circ\grave{\circ}$ ($\circ\bar{\circ}\triangleright$)	$\circ\grave{\circ}$ ($\circ\bar{\circ}\triangleright$) }	
One-mora	[e]	(handle)	$\bar{\circ}$	$\bar{\circ}$ }	$\circ$ ($\circ\bar{\triangleright}$)
	[na]	(name)	$\grave{\circ}$	$\acute{\circ}$ }	
	[ha]	(tooth)	$\acute{\circ}$	$\grave{\circ}$ }	$\bar{\circ}$ ($\bar{\circ}\triangleright$)
	[e]	(picture)	$\circ$	$\acute{\circ}$ }	

Although attempts have been made by several researchers [8-10] in the acoustic analysis and interpretation of these peculiar pitch patterns found in one- and two-mora words of the *Kinki* dialects, their results have not been conclusive due to a lack both of accuracy in measurement and of a means for extracting pertinent features from observed pitch contour. The present study was thus conducted on the peculiar accent types of the *Kinki* dialects to test the applicability of methods of analysis and feature extraction of pitch contours which have been proposed and proved to be valid for the word accent types of the *Tokyo* dialect.

III. Analysis of Pitch Contours of Two-mora Words of the Osaka Dialect [11]

3.1 The speech materials

In order to separate influences of linguistic factors such as dialect and accent type from those of speaker idiosyncrasy, the speech materials for analysis should be taken from utterances of one speaker, preferably one well versed not only in the dialect under study, but also in other dialects with which the former dialect is to be compared. Since certain phonemes such as voiceless and voiced plosives are known to exert local but appreciable influences upon the fundamental frequency of the glottal source, the words to be analyzed should avoid those consonants and should preferably consist of the same phoneme sequence pronounced in all possible accent types. Furthermore, it is desireable that all words tested should be meaningful, since an important role of prosody is to transmit differences in meaning which could not be conveyed by the phonemic constituents alone. The selection of speakers and sample words was based on these considerations, although it was not possible to fulfill all these requirements.

A 46-year-old male and a 24-year-old female speaker of the *Osaka* dialect supplied the speech materials. Both speakers were natives of *Osaka* and spent almost their entire lives in Osaka and its vicinity. The male speaker also had a strong command of the *Tokyo* dialect.

The speech materials used for analysis were:

(1) The two-mora word [ame] pronounced in all four possible accent types of the *Osaka* dialect. These words are specific in that only Type B ([amè], rain) and Type D ([$\overline{\text{ame}}$], candy) are meaningful in the *Osaka* dialect, whereas Type A ([$\overline{\text{a}}$me], rain) and Type C ([a$\overline{\text{me}}$], candy) are the only meaningful words in the *Tokyo* dialect. Utterances of Type A and Type C, which are nonsense in the *Osaka* dialect, however, were produced without being aware of the *Tokyo* dialect.

(2) The two-mora word [imi] pronounced in all four accent types of the *Osaka* dialect.

(3) The one-mora word [e] pronounced in all three accent types of the *Osaka* dialect.

The words were arranged in random order and written in *hiragana* (a system of syllabic letters) along with symbols indicating their respective accent types. They were read at intervals of approximately two seconds and recorded in an anechoic chamber. Next, they were digitized upon playback at a rate of 10 kHz with an accuracy of 8 bits, and stored on magnetic tape of a digital computer for subsequent analysis.

3.2 Methods of analysis and feature extraction of pitch contours

The analysis of word pitch contours consisted of two stages: (1) extraction of the short-time fundamental frequency of speech (i. e., pitch extraction), and (2) extraction of essential acoustic parameters from the extracted pitch contours.

The method of pitch extraction was based on short-time autocorrelation analysis of the speech signal with a time window whose width was as narrow as a few fundamental periods. This was supplemented by a technique for suppressing subsidiary peaks in the short-time autocorrelation function which are due to formant structure, while retaining only those peaks produced by the periodicity of the glottal source. Since the value of this modified autocorrelation function at its greatest peak served as the index of signal periodicity, detection of periodicity and extraction of the fundamental period could be performed simultaneously by detecting the greatest peak exceeding a pre-determined threshold and by finding the time delay corresponding to the peak. Once the periodicity was detected and the short-time fundamental period extracted, the value was utilized to facilitate detection of the next period directly from the speech waveform. Fundamental periods were thus extracted pitch-synchronously but converted to their corresponding fundamental frequencies, which were then interpolated to produce a pitch contour uniformly sampled at intervals of 12.8 msec for the sake of further processing.

Because of various smoothing processes which intervene between the intended linguistic units and their physical realizations, the observed pitch contours are far from discretely binary; rather, they display a smooth rise and decay in the vicinity of the accented morae, superposed on a base line which initially rises and then gradually decays toward the end of a word. A model of the process of conversion from the linguistic units to the observed pitch contour is therefore indispensable for the quantitative analysis and interpretation of the observed pitch contours.

The model adopted in the present study was proposed by one of the authors and has proved to be capable of closely approximating observed pitch contours of isolated words in the Tokyo dialect. It is based on the following assumptions:

(1) Each type of word accent can be uniquely characterized by a logarithmic pitch contour.

(2) Commands for the actualization of linguistic units such as voicing and accent take the form of binary inputs.

(3) Control mechanisms for voicing and accent are separate, and each can be approximated by a second-order linear system converting the respective binary command into the control signal.

(4 These control signals are combined to control a non-linear mechanism of glottal oscillations, whose output is the voice fundamental frequency. The control characteristic of this final stage can be approximated by an exponential characteristic with hysteresis.

Figure 1 shows the functional model constructed on the basis of these assumptions. It consists of three elements: the voicing control mechanism, the accent control mechanism, and the control characteristics of glottal

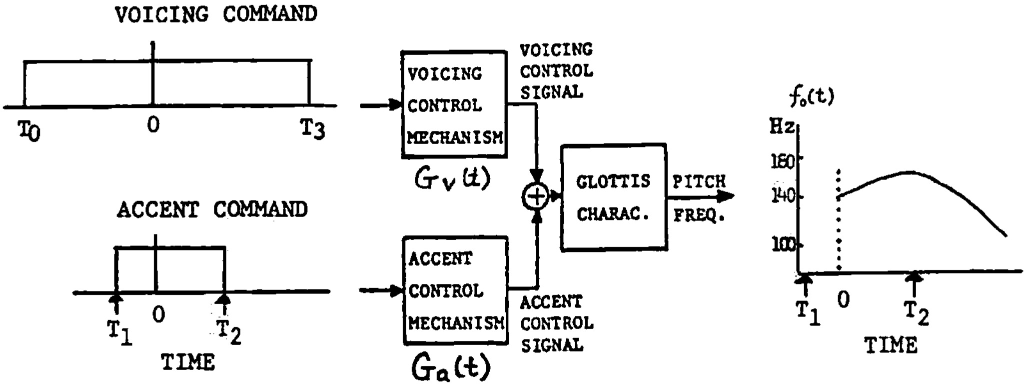

Fig. 1. A functional model for the process of generating a pitch contour from voicing and accent commands.

oscillations. Characteristics of the voicing control mechanism, whose output is responsible for the initial rise and the subsequent gradual decay of the fundamental frequency, are represented by the step response function

$$G_v(t) = [A_v \alpha t \exp(-\alpha t)]\, u(t), \tag{1}$$

while characteristics of the accent control mechanism, whose output is responsible for the smooth rise and decay of the fundamental frequency at the accented morae of a specific accent type, are represented by the step response function

$$G_a(t) = [A_a\{1-(1+\beta t)\exp(-\beta t)\}]\, u(t). \tag{2}$$

The third mechanism, which is responsible for the generation of the pitch contour from the combined control signals, is characterized by an exponential nonlinearity given by

$$f_0(x) = \begin{cases} f_{min}\, e^x, & \text{for } x > x_1 \text{ if } \dot{x} > 0, \text{ and} \\ & \text{for } x > x_2 \text{ if } \dot{x} < 0, \\ 0, & \text{otherwise,} \end{cases} \tag{3}$$

where x represents the combined control signals, and $f_0(x)$ represents the instantaneous fundamental frequency determined by x. Denoting the instants of onset and offset of the voicing command respectively by T_0 and T_3, and the instants of onset and offset of the accent command respectively by T_1 and T_2, the fundamental frequency at time t can be given by

$$f_0(t) = f_{min} \exp[G_v(t-T_0)+G_a(t-T_1)-G_a(t-T_2)-G_v(t-T_3)]. \tag{4}$$

The model allows one to quantify the characterisitics of a given pitch contour in terms of its parameters, estimated by the method of successive approximation known as Analysis-by-Synthesis. While parameters such as Av, Aa, α, β, and Fmin are more or less specific to the individual glottal mechanism as well as its manner of control, and thus are subject to individual differences, the timing parameters of voicing and accent commands, i.e., T_0, T_1, T_2, and T_3, are specific to accent types and remain fairly stable against individual differences.

3.3 Results of analysis of pitch contours

Figure 2 shows examples of Analysis-by-Synthesis of pitch contours for each of the accent types of [ame] uttered by the male speaker. A "+" symbol indicates the pitch frequency interpolated at intervals of 12.8 msec; the solid curve represents the best approximation by the model, and the stepwise waveform indicates schematically the timing of the binary accent command. The results clearly indicate that the model is not only valid for Types A and C which have their counterparts in the Tokyo dialect, but also for Types B and D which are specific to the Osaka dialect. It should especially be noted that the model is capable of closely approximating the peculiar pitch contour of Type B, which is characterized by a rapid downward

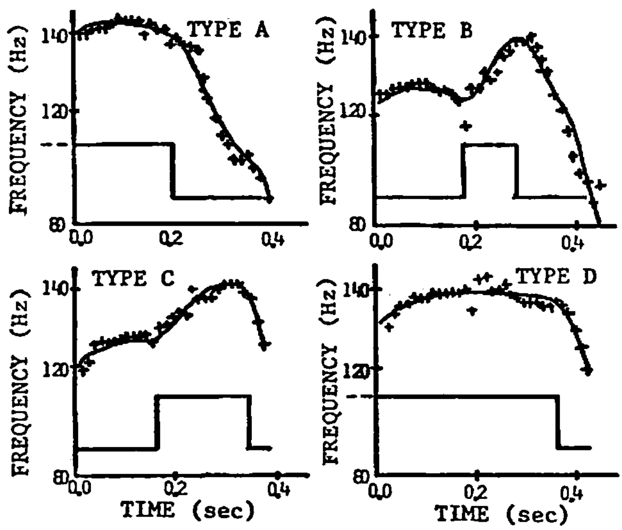

Fig. 2. Analysis-by-Synthesis of pitch contours of four accent types of two-mora words [ame] together with extracted accent command.

pitch transition within the second mora. Parameters obtained from the analysis of six utterance samples of each accent type of [ame] are averaged and listed in Table 3. Although different accent types are found to be most clearly characterized by the timing of the accent command, i.e., by T_1 and

Table 3. Parameters of pitch contours of word accent Types A to D obtained by Analysis-by-Synthesis. Mean of six measurements for each type.

Parameters / Accent Types	Fmin Hz	Voicing Parameters				Accent Parameters			
		T_0 sec	T_3 sec	Av	α sec^{-1}	T_1 sec	T_2 sec	Aa	β sec^{-1}
Type A [āme] (rain) Tokyo	76	-0.25	0.33	1.18	4.30	-0.14	0.19	0.41	20.5
Type B [amè] (rain) Osaka	79	-0.19	0.39	1.45	4.40	0.16	0.28	0.46	21.0
Type C [amē] (candy) Tokyo	77	-0.20	0.37	1.38	3.10	0.17	0.35	0.33	23.0
Type D [$\overline{ame}$] (candy) Osaka	73	-0.23	0.37	1.15	2.70	-0.24	0.36	0.37	13.0

T_2 , other parameters are also found to vary from one accent type to another. For example, the rate of response to voicing command (α) is seen to be smaller in Types C and D involving an upward pitch transition, as compared to its value in Types A and B, which involve a downward pitch transition. The rate of response to the accent command (β), on the other hand, is particularly small in Type D as compared to the other types, contributing to the overall flatness of the pitch contour.

Figure 3 shows six samples each of the four accent types uttered 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 voicing of the utterance. Also shown on the axes are the mean segment boundaries of all the utterances, thus serving as a reference in the timing of the accent command. The figure indicates that the sign of T_1 relative to the voice onset corresponds to differences between the "high-start" types and the "low-start" types. In the "low-start" types (Types B and C), T_1 is found to occur approximately at the onset of the consonant [m]. The value of T_2, on the other hand, corresponds to the "high/low" distinction of the second mora and occurs at the onset of the nasal consonant [m] in Type A, while in Type B T_2 is found to occur approximately at the segment boundary between [m] and [e], thus giving rise to a noticeable downward transition of pitch within the second mora, a feature not found in the other types.

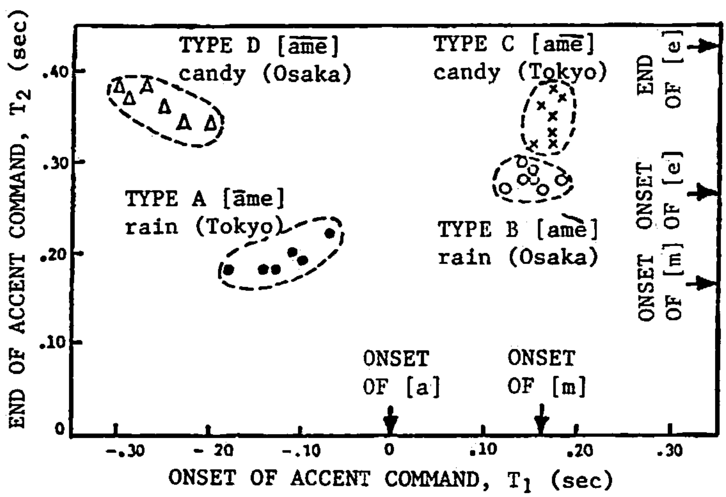

Fig. 3. Representation of four word accent types of [ame] in terms of onset (T_1) and end (T_2) of the accent command.

IV. Synthesis and Perception of Two-mora Words in the Osaka Dialect

4.1 Identification tests of accent types of two-mora words [11]

Although the results of the foregoing analysis indicated that the four accent types of two-mora words in the Osaka dialect are primarily characterized by the timing of the accent command relative to the segmental features of an utterance, thus suggesting that the timing would thereby play an important role in perception, its perceptual relevance has yet to be confirmed by listening tests using synthetic speech stimuli, which would separate influences of other factors such as duration and intensity that are also found to vary with the accent type.

In the first experiment, synthetic words [ame] with the identical formant pattern but with a variety of pitch contours were thus generated by digital computer simulation of a series-type terminal-analog speech synthesizer. The pitch contours were generated by the same model as used in the analysis, but by systematically varying the timing parameters of the accent command, i.e., T_1 and T_2. The parameters used in the synthesis of the forty stimuli are listed in Table 4. Parameters not shown in the table were held constant at their appropriate values: the values for Fmin, Av, and Aa were held at 75 Hz, 1.20, and 0.40, respectively. Stimuli No. 1, 11, 21 and 31 were selected on the basis of preliminary listening tests as typical samples of the four accent types, forming a quadrangle on the T_1-T_2 plane. Other stimuli were selected at those points on the sides of the quadrangle which would divide each side into 10 equal intervals of equal length. Although α and β were held respectively at 4.0 and 20.0 for Stimuli No. 1 to No. 21, both were reduced respectively to 3.0 and 15.0 in Stimulus No. 31, according to the results of analysis of Type D pitch contour. Values of α and β for stimuli No. 22 to No. 30 as well as for stimuli No. 32 to No. 40 were then determined by linear interpolation. The stimuli were synthesized by using only the vowel section of the synthesis program, with control of the fundamental frequency (F_0), frequencies of the first three formants (F_1, F_2, and F_3), and intensity of the buzz source (A_b). The intervocalic nasal [m] was also generated by controlling these parameters. Formant transitions between segments were linear with a transition time of 25 msec, and the durations of [a], [m], and [e] were held respectively at 140, 100, and 140 msec, assuming the mid-point of the linear transition to be the segment boundary.

In order to determine the criterion adopted in the perception of word accent types, identification tests were conducted on the four stimulus continua represented by the four sides of the above-mentioned quadrangle. Each test consisted of a semi-random list of 110 items, containing 10 each of the 11 stimuli, preceded and followed by 5 dummy stimuli, presented at intervals of 4 sec for written response. The results were analyzed individually to detect the category boundary between the two accent types on each of the stimulus continua.

The subjects were five speakers of the Osaka dialect and two speakers of the Tokyo dialect. Speakers of the Osaka dialect were instructed to identify the accent types without regard to the meaning of Types B and D, while speakers of the Tokyo dialect were instructed to be aware of the difference of the two dialects. The two groups of subjects, however, were found not to be significantly different in the results of their identification tests.

Table 4. Parameters of synthetic word stimuli used for identification tests of accent types. Stimuli No. 1, 11, 21 and 31 respectively correspond to typical accents of Types A, B, C, and D.

Stimulus Number	1	2	3	4	5	6	7	8	9	10
T_1	-.140	-.112	-.084	-.056	-.028	0.0	.028	.056	.084	.112
T_2	.16	.17	.18	.19	.20	.21	.22	.23	.24	.25
α	4.0	"	"	"	"	"	"	"	"	"
β	20.0	"	"	"	"	"	"	"	"	"
Stimulus Number	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
T_1	.140	"	"	"	"	"	"	"	"	"
T_2	.26	.27	.28	.29	.30	.31	.32	.33	.34	.35
α	4.0	"	"	"	"	"	"	"	"	"
β	20.0	"	"	"	"	"	"	"	"	"
Stimulus Number	**21**	**22**	**23**	**24**	**25**	**26**	**27**	**28**	**29**	**30**
T_1	.140	.112	.084	.056	.028	0.0	-.028	-.056	-.084	-.112
T_2	.36	"	"	"	"	"	"	"	"	"
α	4.0	3.9	3.8	3.7	3.6	3.5	3.4	3.3	3.2	3.1
β	20.0	19.5	19.0	18.5	18.0	17.5	17.0	16.5	16.0	15.5
Stimulus Number	**31**	**32**	**33**	**34**	**35**	**36**	**37**	**38**	**39**	**40**
T_1	-.140	"	"	"	"	"	"	"	"	"
T_2	.36	.34	.32	.30	.28	.26	.24	.22	.20	.18
α	3.0	3.1	3.2	3.3	3.4	3.5	3.6	3.7	3.8	3.9
β	15.0	15.5	16.0	16.5	17.0	17.5	18.0	18.5	19.0	19.5

Figure 4 shows the results of the identification tests in terms of the mean category boundary represented by a filled circle, with the range of distribution of individual boundaries indicated by a thick line segment.

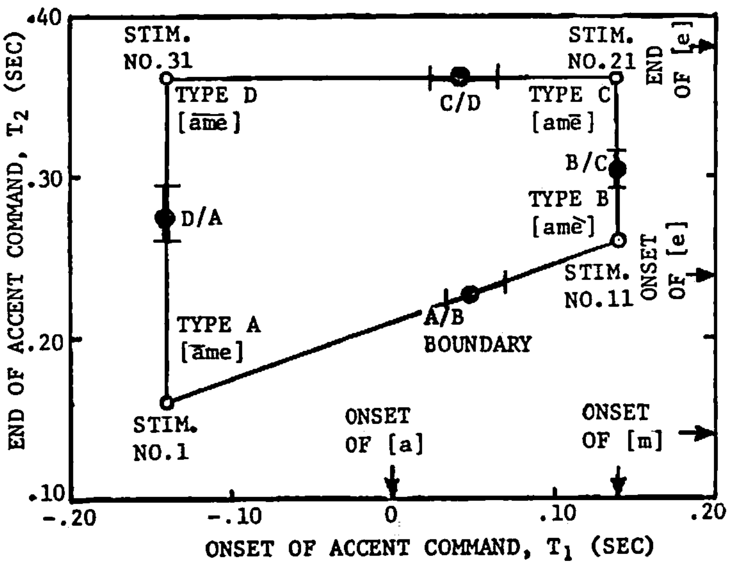

Fig. 4. Stimulus continua used in the identification tests of word accent types and category boundaries between accent types. Mean and range of distribution of boundaries in 7 subjects.

The "high-start"/"low-start" judgments are found to be dependent exclusively on a threshold of T_1, being approximately equal for Type A/ Type B judgment (at approximately 50 msec of T_1), and for Type C/Type D judgment (at approximately 45 msec of T_1). On the other hand, both Type B/ Type C judgment and Type A/Type D judgment are made on the T_2 continuum, but their criteria are found to be appreciably different. Individual differences in all the category boundaries are, however, quite small and indicate the stability of the four perceptual categories in all the subjects, as far as the subjects for the present investigation are concerned.

4.2 Relative timing of segment boundary and accent command [12]

Although the foregoing results indicated that perception of accent types is based on categorical judgment in the T_1-T_2 plane, identification of a particular accent type is considered to depend not on the absolute value of T_1 or T_2, but rather on the relative timing of the accent command and the segmental features of the word. A further investigation is thus necessary to find out which segment boundary serves as the crucial reference for the identification of a specific accent type. The answer to this question can be obtained by finding the influence of a shift in various segment boundaries upon the category boundary between accent types.

The first set of experiments was conducted to examine the effects of shifts in the timing of [m] on the perceptual category boundary between Type A and Type B. It consisted of three experiments:

(1) A backward shift of 30 msec in the [a-m] boundary.

(2) A backward shift of 30 msec in the [m-e] boundary.

(3) Simultaneous shifts of 30 msec both in [a-m] and [m-e] boundaries.

Except for these changes in the timing of segmental features, the stimuli used in these identification tests were the same as those in the original test of Type A /Type B identification. The subjects were the same 5 speakers of the Osaka dialect as in the original test, and the results were processed in the same way. The results indicated that the effect on accent type boundary increases as one goes from (1) to (2), and from (2) to (3). Analysis of these results can be interpreted in terms of a perceptual center of the intervocalic [m], whose influence on the category boundary on the T_1-T_2 plane is quite significant, as shown in Fig. 5.

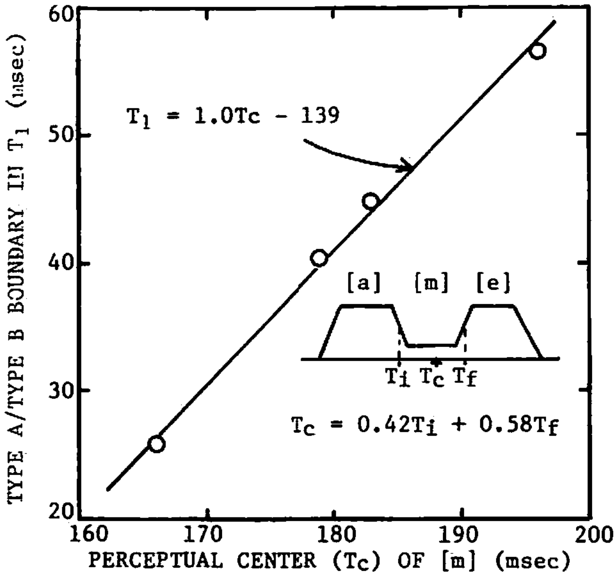

Fig. 5. Influence of timing of intervacalic [m] on perceptual boundary between accent Type A [āme] and Type B [amē].

While the effects of shifts in the timing of [m] have also been examined in other category boundaries, no appreciable change was observed in those boundaries, indicating that perception of a particular accent type is dependent on the timing of the accent command relative to a specific phonemic segment.

The second set of experiments was conducted to examine the effects of changes in duration of the final vowel [e] on the boundary between Type B and Type C. This set consisted of two experiments:

(1) A forward shift of end point of [e] by 30 msec.

(2) A backward shift of end point of [e] by 30 msec.

The results, shown in Fig. 6, indicate that the boundary between Type B and Type C, expressed in terms of T_2, shifts almost by the same amount as the

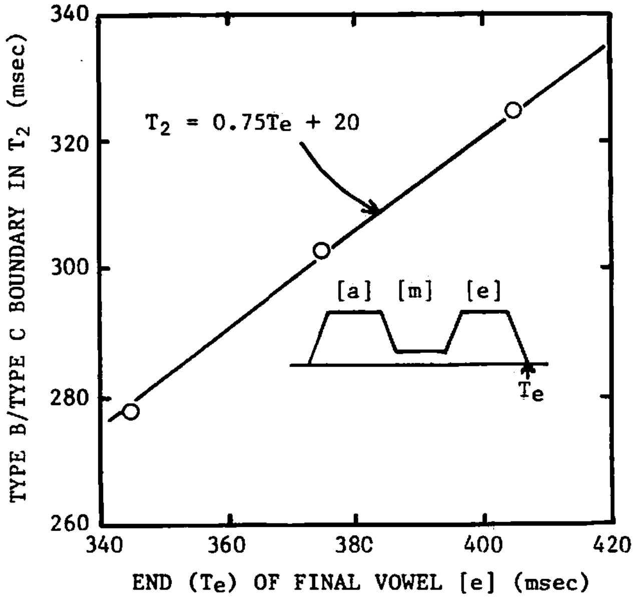

Fig. 6. Influence of duration of final vowel [e] on perceptual boundary between accent Type B [amè] and Type C [amē].

change in duration of the final vowel [e], thereby suggesting that Type B/ Type C identification is based mainly on the temporal relationship between the offset of accent command and the end of the utterance. The boundary between Type A and Type B, on the other hand, remains almost unaffected by changes in [e]-duration.

V Conclusions

In order to investigate into the acoustic characteristics which play a major role in the generation as well as in the perception of word accent types in Japanese, pitch contours of all accent types found in two-mora words of the Osaka dialect have been analyzed. Distinctions between accent types have been found to be most clearly represented by the timing of the onset and offset of the underlying accent command extracted from a measured pitch contour. Perceptual importance of these parameters has been confirmed by identification tests using synthetic speech stimuli with a variety of pitch contours. Results of further perceptual experiments 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.

References

1. Fujisaki, H. (1960), "Automatic Extraction of Fundamental Period of Speech by Autocorrelation Analysis and Peak Detection," J. Acoust. Soc. Am., 32, 1518.
2. Fujisaki, H. and Y. Tanabe (1973), "A Time-Domain Technique for Pitch Extraction of Speech," J. Acoust. Soc. Japan, 29, 418-419.
3. 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.
4. Fujisaki, H. and H. Sudo (1970), "Models for the Word and Sentence Pitch Contours of Japanese," Annual Report of the Engineering Research Institute, Faculty of Engineering, University of Tokyo, 29, 215-221.
5. Kindaichi, H. (1943), "A Historical Study of Japanese Word Accent," in Lectures on Japanese Word Accent, pp. 107-209, Shunyodo Publishing Co., Tokyo. (in Japanese)
6. Yamada, B. (1892), Nihon Dai Jiten (A Complete Dictionary of Japanese), Nihon Dai Jiten Publishing Co., Tokyo. (in Japanese)
7. Polivanov, E. D. (1928), Introduction to Linguistics for Students of Orientalism, Leningrad. (in Russian)
8. Inoue, O. (1931), "Two Kinds of Accent Recording Apparatus (Pitch Recorder)," Study of Sound, 4, 57-60. (in Japanese)
9. Ikeda, K. (1952), "The Word Accent System of a Kinki Dialect," in Essays on Japanese Accent, pp. 347-381, Hosei University Press, Tokyo. (in Japanese)
10. Sakuma, K. (1931), "Word Accent in the Kyoto Dialect," Study of Sound, 4, 61-77. (in Japanese)
11. Fujisaki, H., Y. Mitsui and M. Sugito (1974), "Analysis, Synthesis, and Perception of Accent Types of Two-Mora Words in Tokyo and Osaka Dialects," Transactions of the Committee on Speech Research, Acoust. Soc. Japan, S73-51.
12. Fujisaki, H. and M. Sugito (1974), "Characteristics of Accent Types and Their Perception," Transactions of the Committee on Speech Research, Acoust. Soc. Japan, S74-15.