ACOUSTIC CUES FOR THE PERCEPTION OF ACCENT IN JAPANESE WHEN THE ACCENTED VOWEL IS DEVOICED

Hideo Mase

0. Introduction.

0.1. Phonologically or functionally speaking, Japanese accents are of so-called "register" type. There are only two contrastive registers, high(er) and low(er). In the standard dialect there is a distinction between words (or words with particles) containing a syllable with a high tone immediately followed by a syllable with a low tone, and words which contain no significant fall. In words of the former type the syllable with high tone is said to be accented. Both the place and the presence versus absence of a significant pitch drop are relevant. If, for example, a string of segmental phonemes constitutes three syllables we can distinguish three commutable accent units: (i) the first syllable is significantly high, (ii) the second syllable is significantly high, (iii) there is no syllable that is significantly high:

(i) /ó00/, (ii) /0ó0/, (iii) /000/ (0 = syllable*).

As there is only one significant pitch drop in a word the number of units that can be distinguished by accent depends on the number of syllables. If bisyllabic words are combined with a particle so that they form trisyllabic sequences, then obviously three different types can be distinguished as shown above, but if the particle is removed so that we get

(i) /ó0/, (ii) /0ó/, (iii) /00/

it is hardly possible to distinguish (ii) from (iii) though it is of course possible to distinguish (i) from the others). We

---

*) Long nasals and voiceless geminated consonants are not counted as syllables in this paper, although they form morae of their own.

may say that potentially there are three different types of bisyllabic words (three different underlying forms), but hardly more than two different realizations.

0.2. Phonetically speaking the pitch of the first syllable is usually lower than that of the second syllable whenever the accent is not on the first syllable. This is true of /oó/-words as well as /oo/-words. However, the pitch difference is greater, i.e. the second syllable is a little higher, in the former type than in the latter, especially if the words are spoken with emphasis (Neustupný (12 a), p. 84, Han (3), p. 112). This difference may serve to establish a phonological contrast between bisyllabic words of the two types for some speakers who either keep up or try to keep up the difference. But this is probably rather rare (cp. Neustupný). It is my impression that there is generally no difference between (ii) and (iii) when words of the two types are spoken isolatedly.

0.3. We shall consider briefly the major phonetic aspects of Japanese accentuation:

A. Pitch.

The most important cue for accent production and perception seems to be pitch variation. According to Han (3) "a pitch rise of 1.9 semitones is perceived as an emic change, while a drop of 1.9 semitones is often not interpreted as an emic drop. This means that native speakers of Japanese are sensitive to a smaller pitch rise than to an equivalent pitch drop, and use a wider range of actual frequency change in signalling emic pitch fall." (p. 109), "the pitch change of more than 2 semitones functions as an emic change." (p. 128). I think that native speakers will be more attentive to the pitch drop than to the pitch rise, as far as accent is concerned, because the difference between fall and no fall is phonologically contrastive. (Kawakami (5) interpretes the low pitched first syllable as a manifestation of an intonation.)

Kobayashi (6) says that the degree of fall is greater for words with accent on the first or second syllable than for those with accent on the third or later syllable. In the former case, the high pitch gives an impression of 'strong stress'.

This observation of his agrees to some extent with the characteristics of English stresses, though the latter language is not a tone language. In English fundamental frequency (both pitch level and pitch variation) is reported to play a greater part than intensity. (Fry (2), Lieberman (8), Morton & Jassem (10), etc.)

The observation by Kobayashi means that the pitch fall is gradually decreasing the farther it is from the beginning of the accent unit. A similar statement has been made by other investigators, for example, by Kawakami (op.cit.), but he does not give any quantitative data on pitch variation. Shohara (14) says that only 2 % of the phrases with auditorily level tone have acoustically level tone.

B. Intensity and duration.

As for intensity and duration, Han (op.cit.) says

> "the higher pitch is usually accompanied by a slight increase in intensity and duration. ...if we keep the pitch constant, the change of intensity and duration does not affect the meaning of the word: therefore the pitch is identified as the distinctive feature and the accompanying intensity and duration as redundant features." (p. 1o4).

On the other hand, Neustupný (12 b) points out the importance of the intensity factor in addition to the pitch factor. He says that the pitch and intensity factors are relatively independent of each other, and that sometimes the intensity factor determines the accent pattern, though in general "pitch seems to be more important than intensity" (p. 1). I think that his statement should be taken with some reservation. He gives special mention to Fig. 4 [a⌜na⌝ta] (the pitch curve is taken with a kymograph, but it seems that the intensity curve is from a high-speed level recorder or some other instrument) and Fig. 11 [a⌜ta⌝takana] (taken by means of a pitch recorder). For the evaluation of these examples, the following considerations may be relevant: (1) Though the pitch of the first syllable in such a word as [a⌜na⌝ta] is usually lower than that of the second syllable (cf. 0.2.), it may not be the high pitch of the first syllable in his Fig. 4 but the clear pitch fall between the second and third syllables that is decisive for the accent per-

ception. (2) There may be some influence of neighbouring consonants upon the fundamental frequency of the vowel, though Han (op. cit.) says

> "These differences are, however, minor and do not exceed 1 or 2 centiseconds in duration and 10 cycles per second in frequency. These are etic differences, and they are kept separate from the data pertinent to the structural analysis." (pp. 1o3-1o4).

House & Fairbanks, in their investigation of American English vowels (4), say

> "In the comparisons of voiced and voiceless consonant environments, vowels in voiced environments, with few exceptions, were longer in duration, lower in fundamental frequency, and greater in relative power." (p. 113).

In their article we find that the fundamental frequency of the vowel is considerably higher between [t]'s than between [n]'s, whereas the intensity of the vowel is considerably lower between [t]'s than between [n]'s. Accordingly, although I do not think the influence of the consonant upon the vowel is quite the same in English and Japanese, it is not improbable that the difference between the two consonants in Neustupný's examples, viz. [n] in Fig. 4 and [t] in Fig. 11 may have a certain influence upon the following vowel, whether the influence be subsidiary (or "etic" in Han's terminology) or not.

0.4. Devoicing of vowels.

Of the five Standard Japanese vowels /i,e,a,o,u/, /i/ and /u/ are often devoiced. Han (3, pp. 17-34) mentions four factors which cause devoicing of vowels:

(1) The intrinsic length of the five vowels: since /u/ is the shortest and /i/ the next shortest of the five vowels, /u/ is devoiced most easily and then /i/, while the other three vowels are rarely devoiced. (Both of them are, of course, high or close vowels.)

(2) Speech tempo: Except for slow speech tempo, the two vowels are easily devoiced. Slow tempo is that which is "used only when the speaker wishes to be precise, or when talking to a child or a foreigner, or in public speaking." (As for speech tempo, M. Temma (16) says in her study on the change in speech tempo within three generations that the first generation pro-

duces approximately 1o to 12 phonemes, 5 syllables, and 3 words per second, while the third generation speaks about 16 phonemes, 9 syllables, and 5.3 words per second. The third generation speaks 31 per cent, and the second generation 19 per cent faster than the first generation. This is the case of the Tokyo dialect, but almost the same tendency is seen in the Osaka and the Kyoto dialects.

(3) The pitch accent: vowels are easily devoiced except when they are 'accented'. Even if they are accented, they may sometimes be devoiced.

(4) The neighbouring consonants: the duration of the vowel is shorter between (generally also after) voiceless consonants than in voiced consonant environments. Of the voiceless consonants, [s] and [ʃ] give the strongest effect. (Nakano (11) reports that in fast speech voiced consonants may sometimes be devoiced, and then they can contribute to the devoicing of vowels, for example: [kab̥-to-tʃo·], [kab̥-ki-za] in the genuine Tokyo dialect, and [biʃi̥bi̥ʃi̥], [batʃi̥bat̥ʃi̥] (p. 8) (note that [b̥] = /b/, not /p/, [kab̥] = /kabu/).

A question is how the accent is perceived when the accented vowel is devoiced. (Sometimes phonological accent shifts occur, but this matter is out of concern here.) Han (3) says that "in this case where the vowel of the accented syllable is unvoiced, the very low pitch of the second syllable seems to signal the "virtual" high pitch of the preceding syllable." (p. 34).

Sugitow (15) says that in two-syllable words where the vowel of the first syllable is devoiced, the first syllable is perceived as having 'accent' when the pitch of the second vowel falls down very quickly, for example from 300 cps to 250 cps (a young female voice) after 1 centisecond. On the other hand the first syllable is perceived as not having 'accent' when the pitch of the second vowel goes down very slowly. (The informant is a young lady born in Osaka, but her mother was born in Tokyo. Accordingly, we cannot expect that this phenomenon is necessarily the same in the Standard dialect.

1. Procedure.

The investigation summarized here was undertaken in order to find some acoustic cues for the accent perception when the accented vowel is devoiced.

1.1. Texts and subjects.

I chose three-syllable accent units for my analysis. The text (written in Japanese script) consists of words with segmental phonemes: /hasi/ or /kaki/ followed by a particle: /wa/, /ga/, /o/, /mo/ or /to/ with different accent placement. The words were given in different orders, and both in isolated position and in short sentences. Only the words followed by the particle /to/ were used. In this environment the vowel /i/ in /hasito/ and /kakito/ is devoiced in speech of normal tempo. Both /hasito/ and /kakito/ make three commutable units by the difference of accent placement, namely,

Acc. 1: /hásito/; /kákito/ ("chopstick (and)"; "oyster (and)"),
Acc. 2: /hasíto/; /kakíto/ ("bridge (and)"; "fence (and)"),
Acc. 3: /hasito/; /kakito/ ("edge (and)"; "persimmon (and)").

(/to/ is also used in the meaning "to".)
Accent 1 words have accent on the first syllable, Accent 2 words have it on the second, and Accent 3 words have no accent, or, in other words, no significant drop.

The main subject person was a middle aged professor in natural science, who was born and has been living in Tokyo (M.O.). Two other subjects were born in Tokyo and Kanagawa respectively, and have been living in Tokyo. Both of them have completed university courses some years ago (K.T., H.M.). M.O. recorded isolated words 11 times in two sessions and words in sentences 6 times. Two other subjects recorded only isolated words 2 times. A short pause was made after the recording of one text.

1.2. Recording.

The recording was done in the sound-treated studio of the Institute of Phonetics, University of Copenhagen, in July and August, 1968.

1.3. Control of material.

In the speech of the main subject M.O., some words were spoken with different accents from the intended ones by mistake. (In two other persons' speech there was no confusion, but one occurrence of /kakíto/ (Acc. 2) in K.T.'s record was spoken with a different particle.) Besides, some Accent 2 and Accent 3 words have been spoken with a voiced second vowel, which is quite natural because the second syllable is rather high-pitched (higher in pitch than the first syllable). Specimens in which it was not clear which accent had been used, were left out of the material.

In order to find whether the intended accents can be perceived by other persons, all the words were rearranged in different order and given as a listeners' test. The subjects (4 in all) who listened to the test words are others than those who spoke (except H.M.). Each subject listened two times with a time interval of a week. (But a Japanese phonetician who visited Copenhagen listened two times in one day.) The 'perceived accent' is counted positive if the agreement of the answers is 7/8 or 8/8.

Table 1-a.

hasito — perceived accents

intended accents		Acc. 1	Acc. 2	Acc. 3	?
Acc. 1	21	21			
Acc. 2	19		13 (+2)	1	3
Acc. 3	22		1	17 (+1)	3

kakito — perceived accents

intended accents		Acc. 1	Acc. 2	Acc. 3	?
Acc. 1	27	27			
Acc. 2	19		13 (+2)	1 (+1)	2
Acc. 3	16			14	2

(The numbers in parentheses refer to words with voiced second vowel.)

Of kakito (Accent 2) spoken by K.T. only one example could be used in the table above. H.M. kakito (Accent 2) was spoken with voiced second vowel in both instances.

The number of words with devoiced second vowel is as follows:

Table 1-b.

hasito

	s.	w.	total
Acc. 1	6	15	21
Acc. 2	4	1o	14
Acc. 3	4	14	18

kakito

	s.	w.	total
Acc. 1	7	2o	27
Acc. 2	4	9	13
Acc. 3	4	11	15

(s. = words in sentences. w. = single words.)

1.4. Registration.

Mingographic curves were taken of all the words. Some spectrograms were also taken for the delimitation of segments (and in order to see the formant structures). Pitch contours of some words were also taken by means of a Schneider's Tonhöhenschreiber.

On the mingogram were registered (1) duplex oscillograms, (2) intensity curve (logarithmic, integration time 5 ms, HP-filtering with cutoff at 500 cps, (3) intensity curve (linear, integration time 10 ms, no filtering), and (4) fundamental frequency curve. As for the spectrograms, wide-band, narrow-band and section spectrograms were taken.

1.5. Methods of measurement.

All data were obtained from mingograms. Spectrograms and pitch curves made with the Tonhöhenschreiber were used only to control the mingographic curves.

Length of the vowel: the vowel begins at the point where regular vibrations are seen on the oscillogram. On the intensity curves, it begins at the rising point of intensity. The vowel ends at the point where the amplitude of vibrations, the intensity, and the fundamental frequency all go down abruptly. In most cases the beginning and end points of the vowel as defined above agree on the four curves within the accuracy of ± 0.25 cs. Consonants and devoiced vowels of the second syllable are measured according to the same principle. The duration of the open phases of consonants has been measured mainly on the logarithmic intensity curve. As for the intensity, I

have measured peak values only. The fundamental frequency curve shows a certain irregularity during the first 0 - 1 centiseconds (and sometimes up to 2 or 3 cs), so that the very beginning point is not measured. The accuracy of measurement is ± 0.25 dB for the intensity and ± 2.5 cps for the fundamental frequency, provided that the measuring scales are accurate.

2. Results.

The material is too small in size to be treated statistically; in most cases, therefore, all pertinent data are shown in figures and tables.

2.1. Words with voiced second vowel.

These have not yet been analyzed, but they will soon be, together with other words such as hana + particle and turu + particle. According to the traditional description one should expect

Acc. 1: high - low - low
Acc. 2: low - high - low
Acc. 3: low - high - high,

but a look at the curves for hasi and kaki with voiced second vowel before particles reveals the following tendencies:

Fundamental frequency:

Acc. 1: the second vowel is high falling rather than mid or low falling.

Acc. 2: the third vowel is not low but falling from mid to low or from high to low.

Acc. 3: the second vowel is often lower than in Accent 2 words (cp. 0.2.above).

Intensity:

Acc. 1: the first vowel is relatively stronger than in Accent 2 and Accent 3 words.

Acc. 1 & 2: the intensity of the third vowel decreases abruptly.

Acc. 3: the third vowel is relatively strong, or keeps a certain level of intensity longer than with Accent 1 & 2.

2.2. Words with unvoiced second vowel.

2.2.1. Length.

The subjects of my experiment belong to the second generation and to the generation between the second and the third generation. The words spoken by three subjects are between 30 and 50 centiseconds in length. This means that the subjects spoke the words with a normal speech tempo where devoicing of vowels occurs (cp. 0.4. above).

The duration of words is considerably different among different recordings of the same person and among different persons. The mean value of segments and the percentual distribution of segments are given in Table 2 and Figs. 2 and 3, respectively.

The open phase of medial [t] is usually between 1.5 - 2.5 cs in length. It seems that there is no great difference in the duration of the open phase according to word accentuation or speaker. The open phase of initial [k] in kakito shows rather interesting results. The open phase is shorter before vowels with greater intensity and longer duration (e.g. the first vowel of kákito (Acc. 1)) than before the vowels in the syllable ka of Accent 2 and Accent 3 words, which are shorter in duration and weaker in intensity. If we measure the combined length of open phase + vowel, the difference among three differently accented words with identical segmental structure (i.e. kakito) becomes smaller (see Fig. 1). The reason may be that the first vowel occurs between two voiceless consonants (i.e. k - k), so that when the vowel is unaccented, the tendency of devoicing of the vowel is greater for Accent 2 and Accent 3 words than for Accent 1 words. (NB.: It is quite rare that the vowel /a/ is devoiced completely. Cf. 0.4.)

As for vowel length, the first vowel of Accent 1 words tends to be a little longer than those of Accent 2 and Accent 3 words. It may mean that the accented vowel has a slightly longer duration than the unaccented one. The duration of the third vowel does not seem to differ so much. In hasito the third vowel is often shortest in Accent 1 words (hásito), and often longest in Accent 3 words (hasito). In kakito the third

Table 2. Average Length.

hasito	Accent type	MO. sentence	MO. word-1	MO. word-2	KT. word	HM. word
whole length minus initial consonant	1	4o.4	4o.8	34.6	51.3	45.3
	n.	(6)	(6)	(5)	(2)	(2)
	2	39.1		32.7	48.o	45.8
	n.	(4)		(6)	(2)	(2)
	3	4o.6	42.1	32.5	5o.o	48.8
	n.	(4)	(6)	(3)	(2)	(2)
[a]	1	8.o	7.8	6.7	9.o	8.3
	2	5.9		5.6	5.8	6.5
	3	5.8	6.5	4.8	7.3	7.3
[ʃi̤]	1	11.9	14.7	11.3	16.8	18.3
	2	11.9		1o.5	17.5	18.3
	3	13.o	14.7	1o.8	15.8	17.5
[t]	1	1o.4	11.5	11.o	18.5	13.o
	2	9.9		9.6	15.o	14.3
	3	11.3	11.4	1o.2	16.o	14.8
[o]	1	1o.1	6.8	5.6	7.o	5.8
	2	11.6		7.o	9.8	6.8
	3	1o.6	9.4	6.6	11.o	9.3

kakito						
whole length minus initial consonant	1	43.1	44.6	36.1	5o.o	43.o
	n.	(7)	(6)	(8)	(2)	(2)
	2	38.8	41.5	32.9	5o.o	
	n.	(4)	(4)	(4)	(1)	
	3	38.8	44.8	36.o	5o.5	43.8
	n.	(4)	(6)	(1)	(2)	(2)
[a]	1	8.o	7.5	6.6	7.o	5.8
	2	5.9	6.o	5.6	7.o	
	3	5.9	6.1	5.5	5.3	4.5
[ki̤]	1	13.1	16.3	13.1	18.o	16.3
	2	13.9	16.4	1o.5	16.o	
	3	12.7	16.8	12.5	2o.o	17.5
[t]	1	1o.9	13.1	11.3	15.5	14.8
	2	9.9	1o.8	11.1	17.o	
	3	1o.2	12.o	11.o	14.3	12.3
[o]	1	11.2	7.6	5.1	9.5	6.3
	2	9.1	8.4	5.6	1o.o	
	3	1o.o	9.8	7.o	11.o	9.5

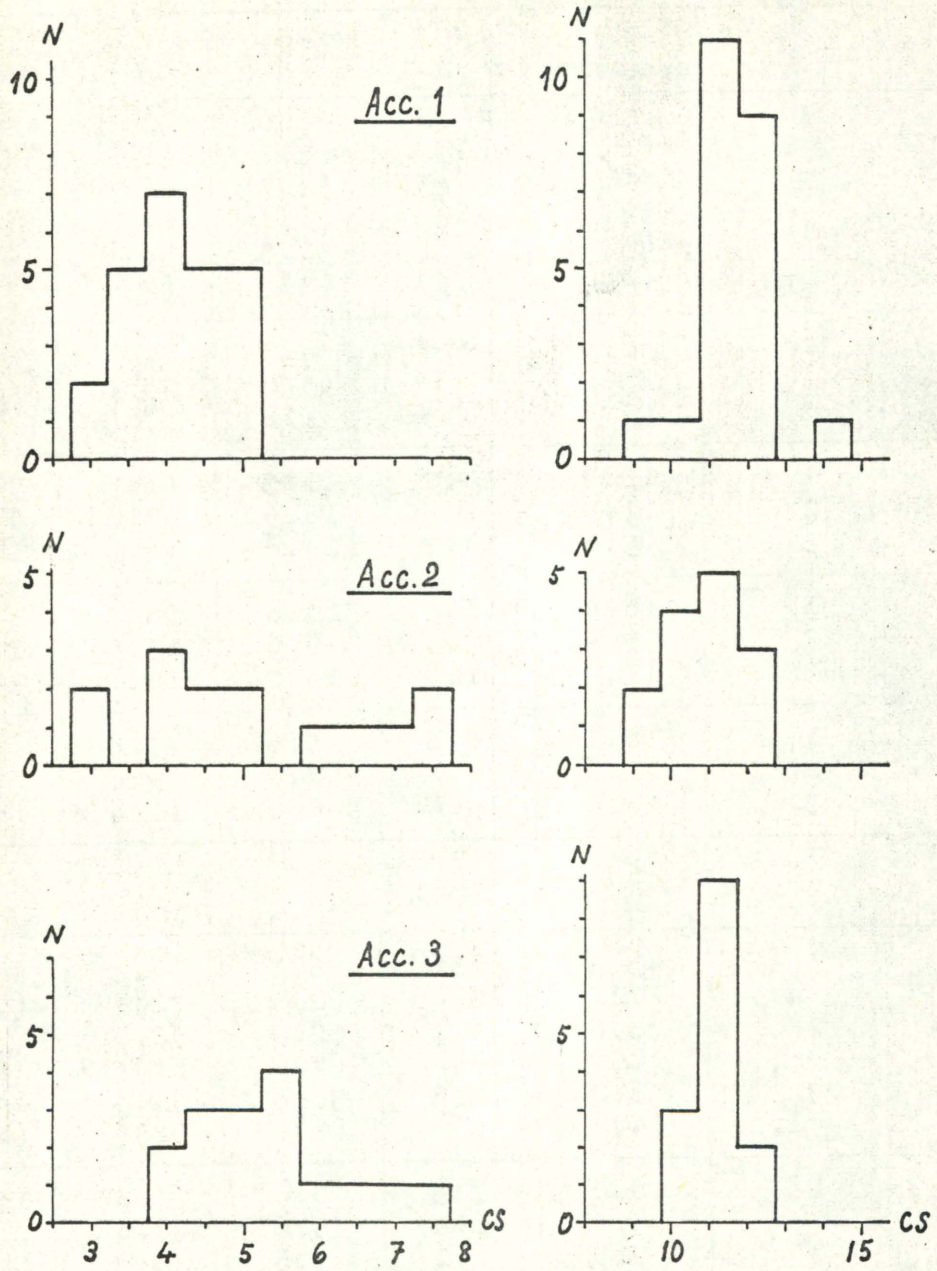

Fig. 1. Lengths of open phase of initial k (left) and of open phase of k + 1st vowel (right).

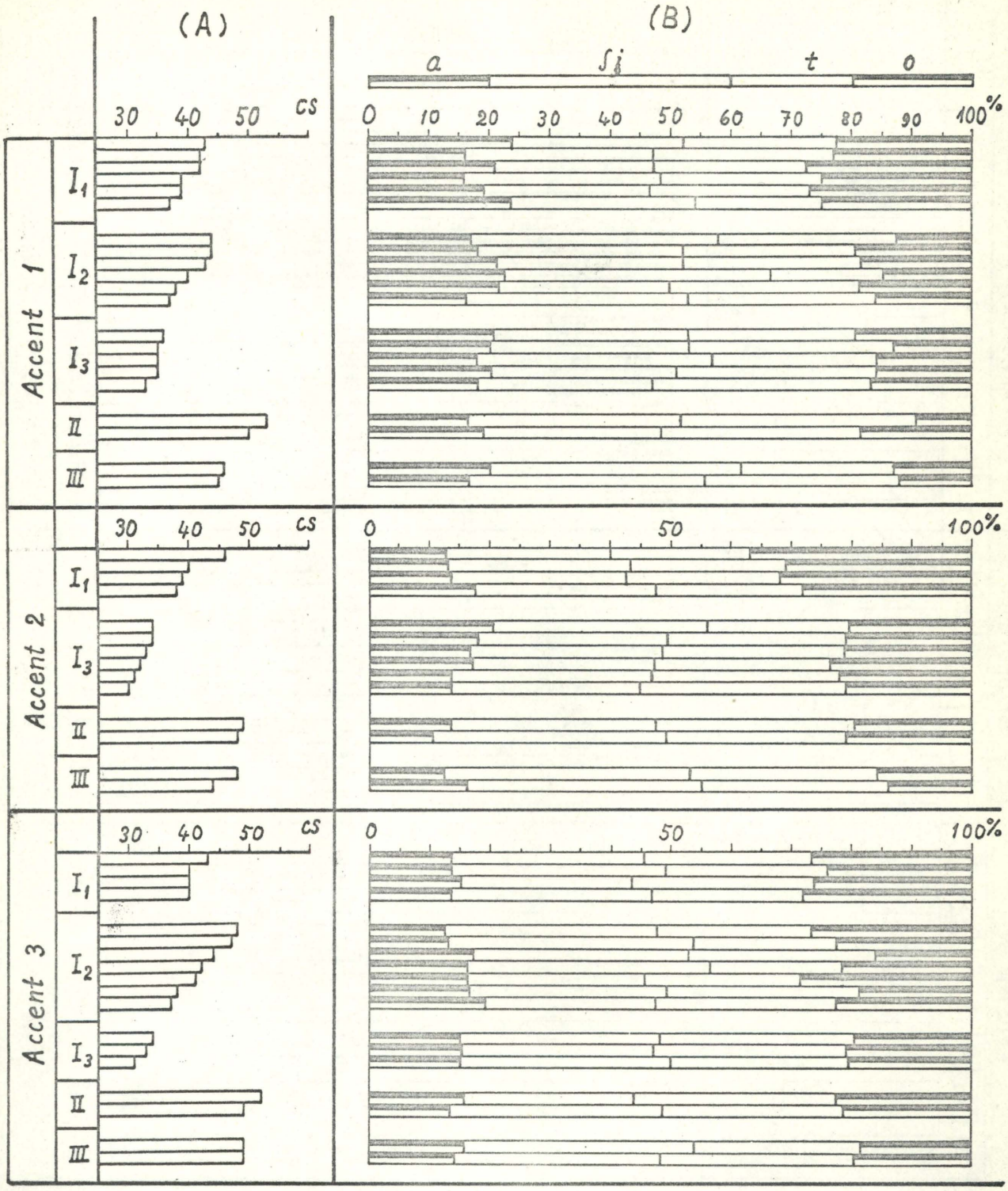

Fig. 2. Length (hasito)

(A) Length of the whole word (without initial h).

(B) Percentage length of segments in the word.

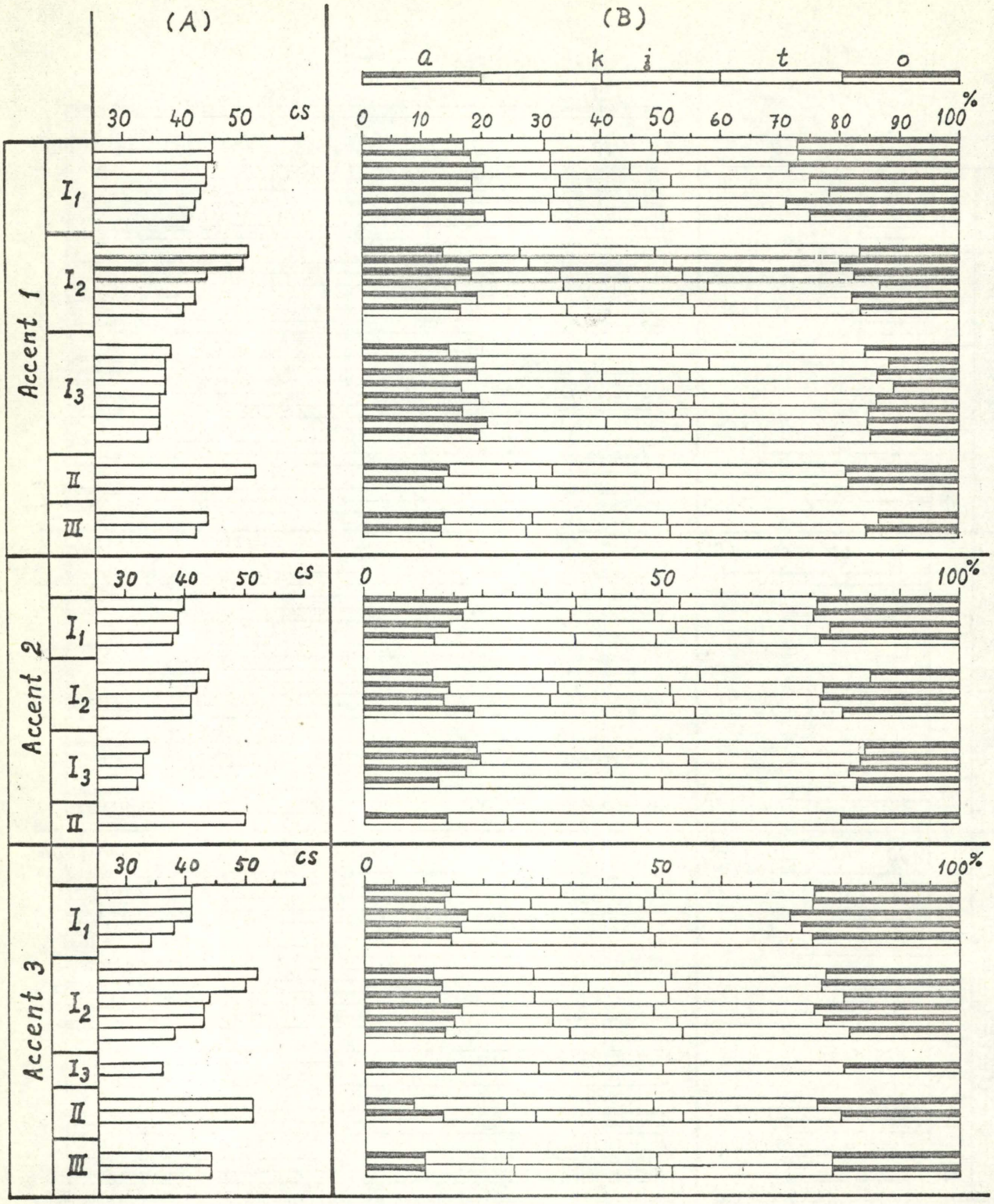

Fig. 3. Length (kakito)

(A) Length of the whole word (without initial k).

(B) Percentage length of segments in the word (the beginning of the syllable ki is here defined as the point of explosion).

vowel of Accent 1 (kákito) and Accent 2 (kakíto) words is often shorter than that of Accent 3 words (kakito), as one might expect. The difference, however, is not great. The duration of the second syllable does not seem to show any significant difference, either.

On the whole, what seems rather constant is that the accented first vowel of Accent 1 words (hásito, kákito) is a little longer than the unaccented first vowel of Accent 2 and Accent 3 words.

2.2.2. Intensity (see Figs. 4-6).

The intensity of the accented vowel of Accent 1 words (hásito, kákito) is a little higher than that of Accent 2 and Accent 3 words. Roughly speaking, the difference amounts to some 6 dB. The weaker vowel of Accent 2 and Accent 3 words tends to have its intensity peak in the first half of its duration, and the intensity falls down quickly, but not always. The relation between the first and the third vowels seems a more stable feature. As is seen in Figs. 4 and 5, the relative intensity of the third vowel compared to the first vowel of Accent 1 words is smaller compared with those of Accent 2 and Accent 3 words. (I should like to emphasize here that this is not a comparison between the inherent intensity of /a/ and /o/, but a comparison of the relation between the first and the third vowels in each type of words.) This difference in the intensity relation between the first and the third vowels, together with the duration factor, may contribute to some extent to distinguish accent placements.

The intensity of the second syllable is almost the same in all types of words. The position of the intensity peak of [ʃi̥] in hasito varies according to which part is stronger, the [ʃ] part or the [i̥] part. On the whole, I have not found any constant difference in peak distribution among the three types of words, but one can observe a slight tendency for the peak in Accent 1 words to come earlier and fall down more quickly than in Accent 2 and Accent 3 words. The intensity peak of the second syllable (either the consonant part or the devoiced vowel part), when compared to that of the first vowel, is lowest in

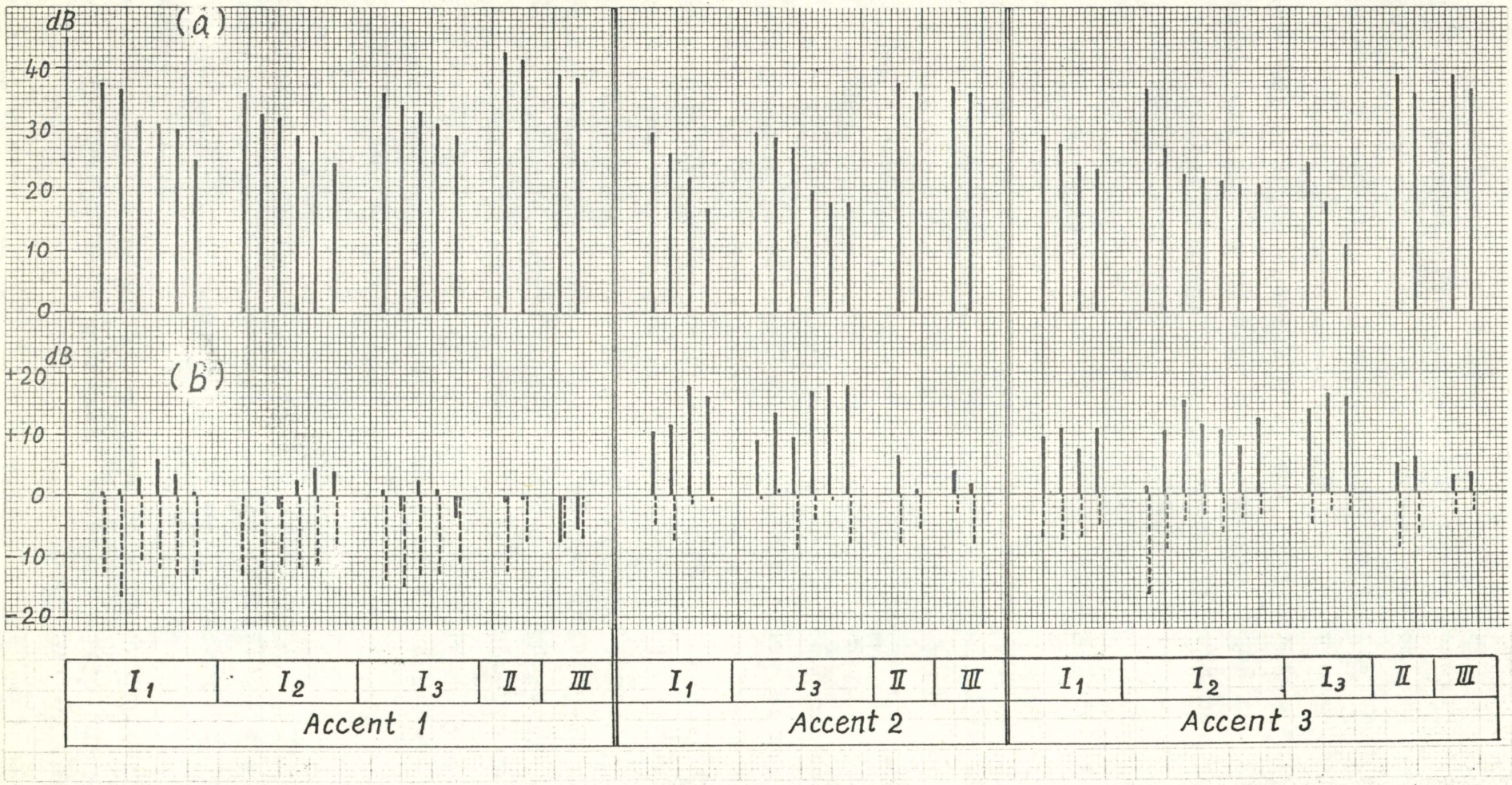

Fig. 4. Intensity (Peak Value) of <u>hasito</u>

(A) Intensity of 1st vowel.

(B) ——— 3rd vowel minus 1st vowel,

----- [ʃi:] minus 1st vowel.

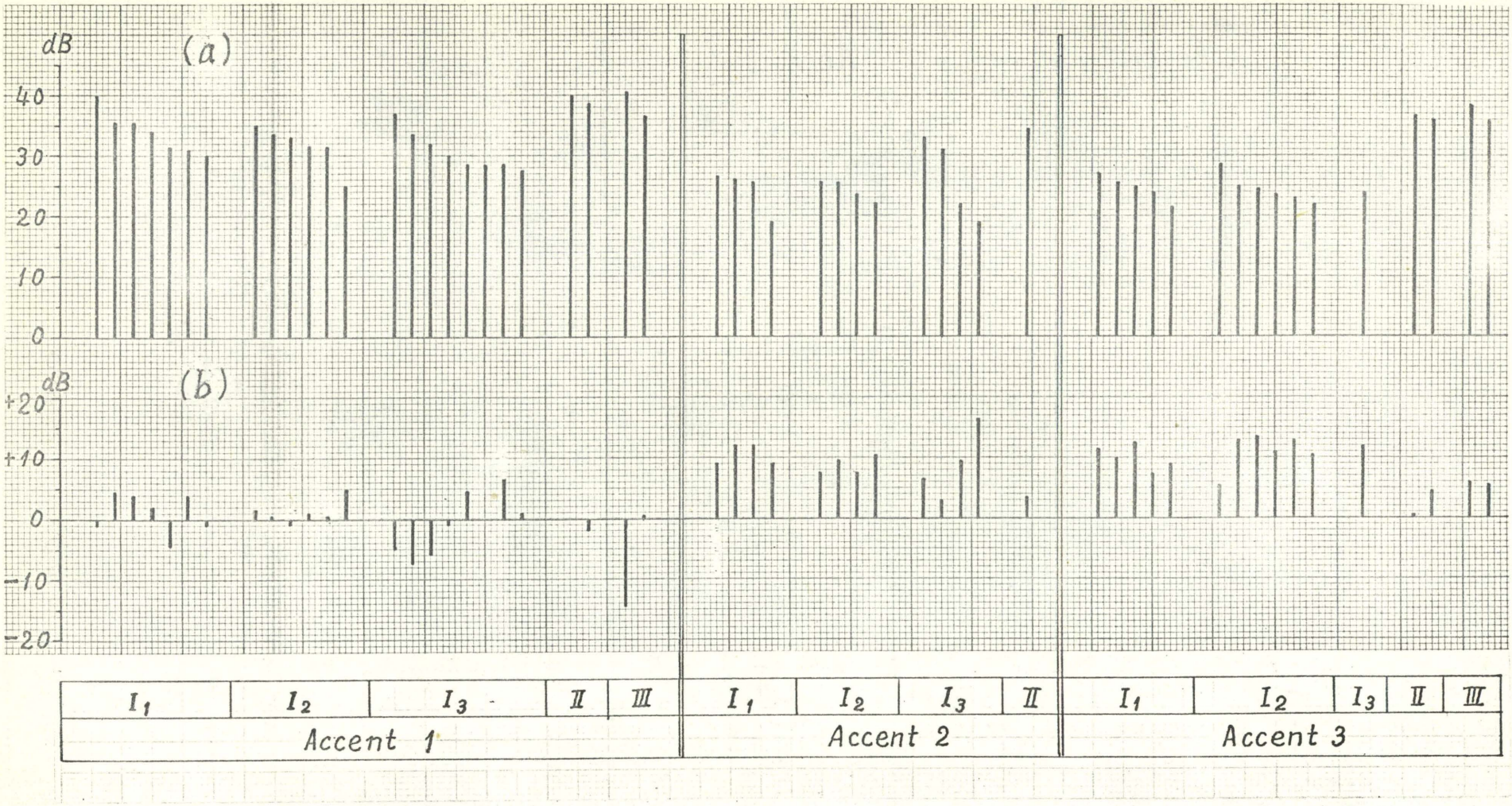

Fig. 5. Intensity (Peak Value) of kakito

(A) Intensity of 1st vowel.

(B) 3rd vowel minus 1st vowel.

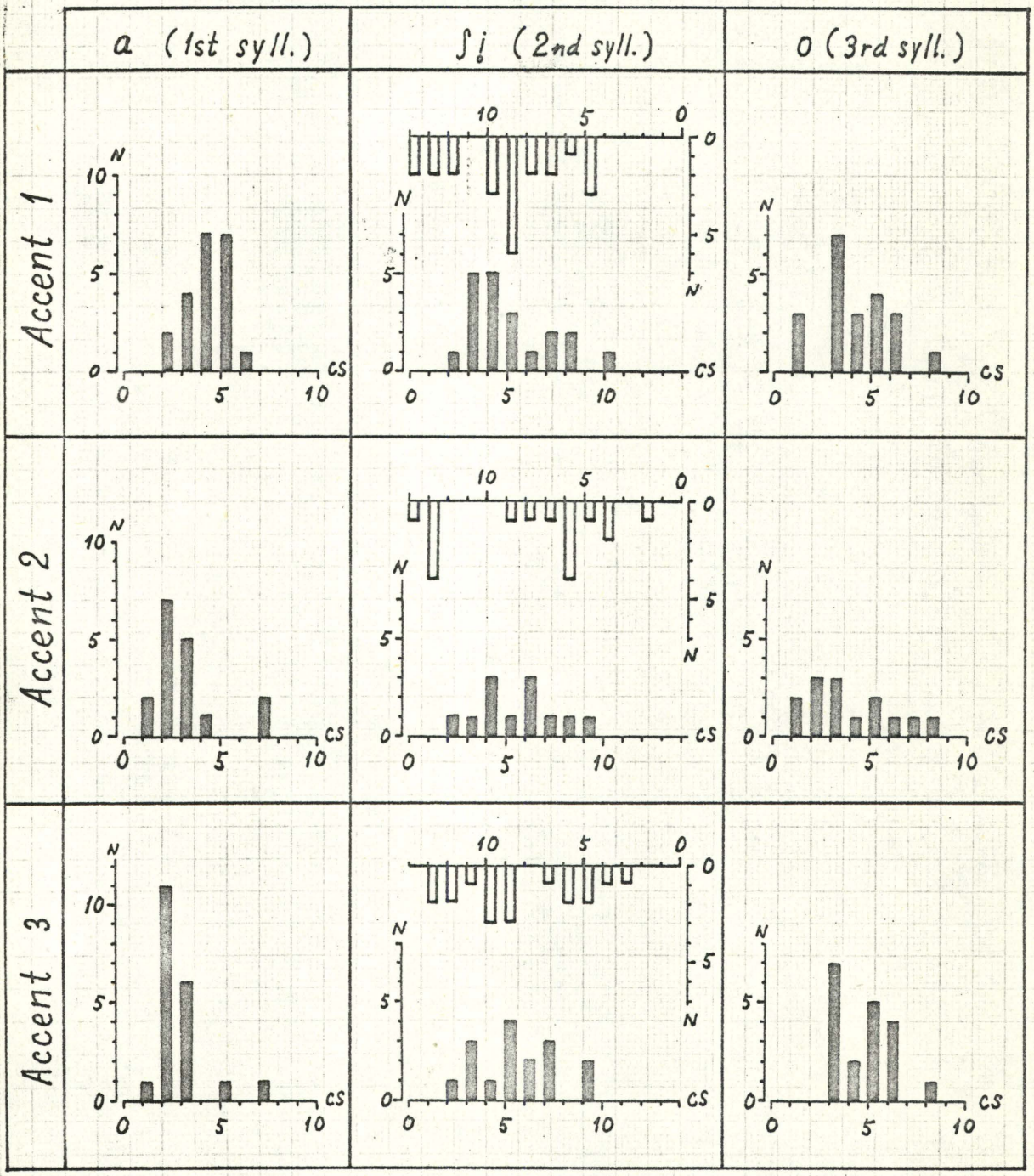

Fig. 6. Position of intensity peak (hasito)

Distance from the beginning (▬) and the end (▭).

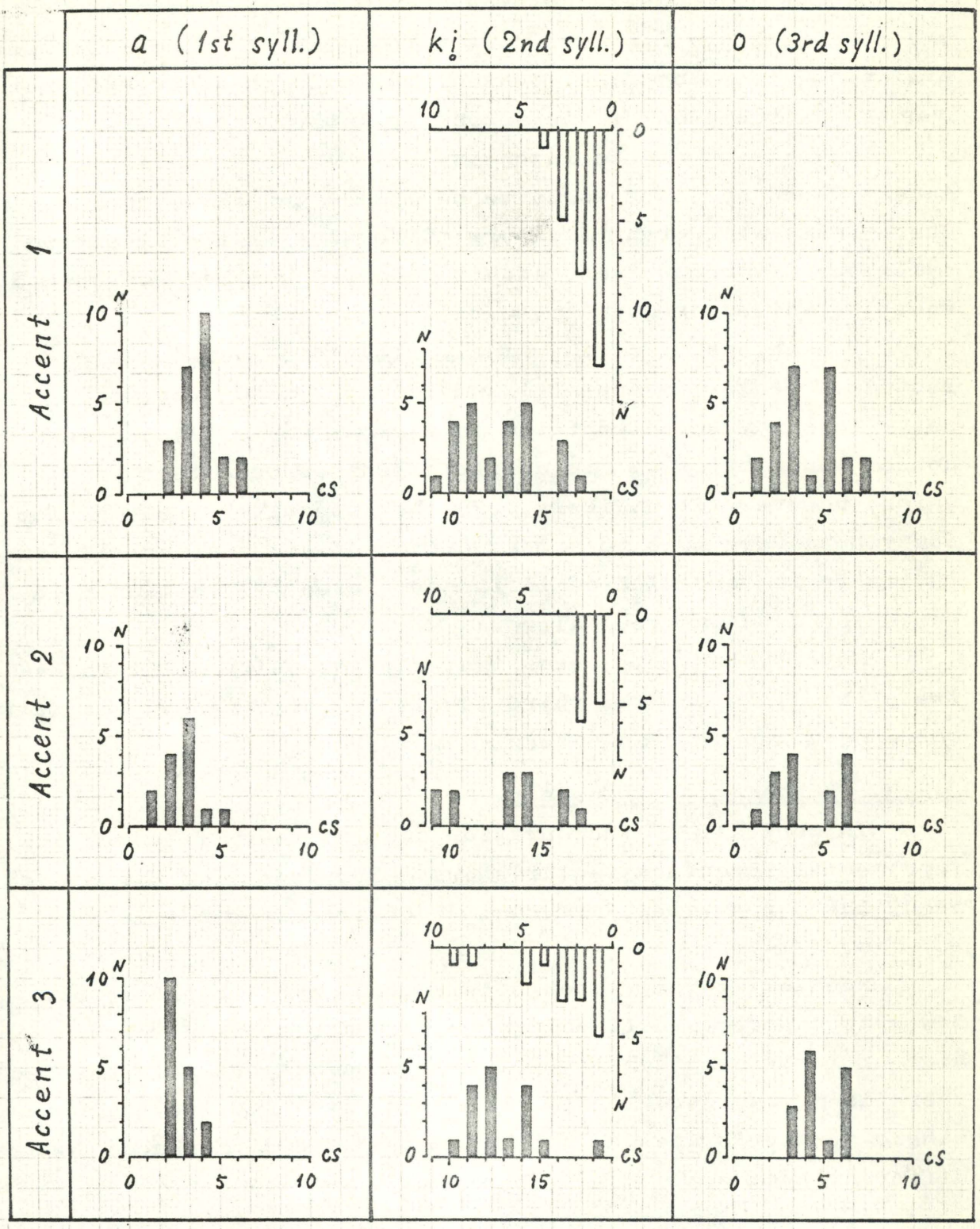

Fig. 7. Position of intensity peak (kakito)
Distance from the beginning (▬) and the end (▭).

Accent 1 words. The intensity of [ki̥] in kakito was sometimes too weak to be measured precisely. As a whole, it holds true that the peak often occurs just before the end of the second syllable. (The intensity peak of the open phase of k is not counted as the peak of the second syllable.)

We can expect the intensity of the second syllable of Accent 2 and Accent 3 words to be almost the same, because there is no noteworthy auditory difference in the first two syllables of Accent 2 and Accent 3 words, whether the second vowel be voiced or not (cp. 0.2.).

The intensity of the final, i.e. third vowel was expected to be low in Accent 1 and Accent 2 words, but it was not. This may be interpreted in various ways: (1) the inherent sonority of [o] is a little greater than that of [a] which, however, is unlikely, (2) the intensity level may not contribute very much to accent differentiation after all, and (3) there is some kind of influence caused by the manner in which the text was read aloud.

Nevertheless, it seems that the intensity relation between the first and the third vowels can distinguish Accent 1 words from Accent 2 and Accent 3 words.

2.2.3. Fundamental frequency.

In contrast to the duration and intensity factors, the fundamental frequency factor shows a very positive difference among the three types of words. A clear difference is seen in Figs. 8-14.

Needless to say, though the pitch level, high(er) or low(er), has certain fixed ranges in which it is perceived as high or low, the pitch level is not absolute but relative, so there is a considerable difference in pitch level even in one single speaker's speech, not to mention the differences among different speakers.

If I may jump to the conclusion or result, I find the following tendency to distinguish the three types of words by pitch:

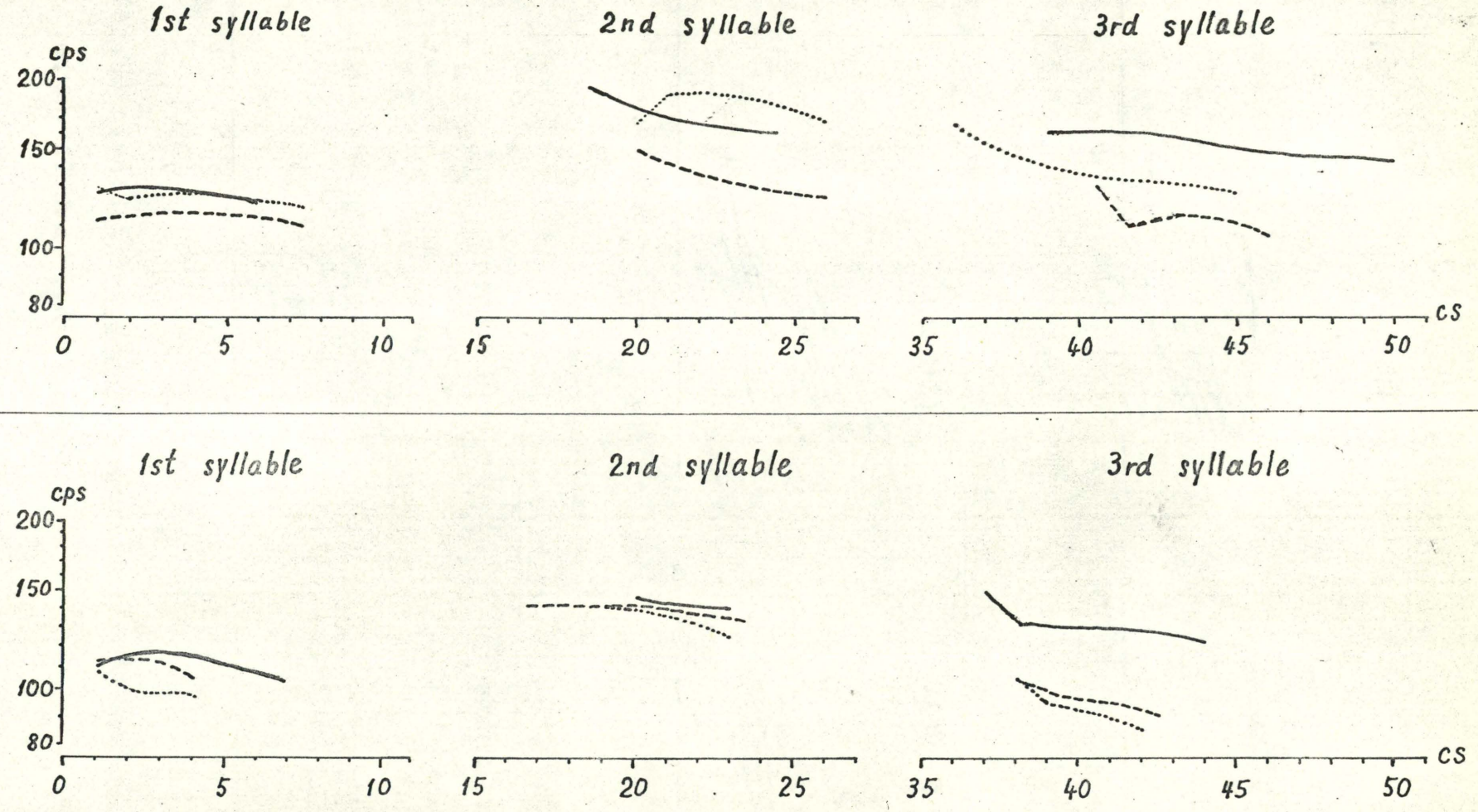

Fig. 8. Pitch contours of words with voiced 2nd vowel.

Upper curve: hasito ....., ------- Acc. 2 (M.O.), —— Acc. 3 (M.O.)

Lower curve: kakito ....., ------ Acc. 2 (H.M.), —— Acc. 3 (M.O.)

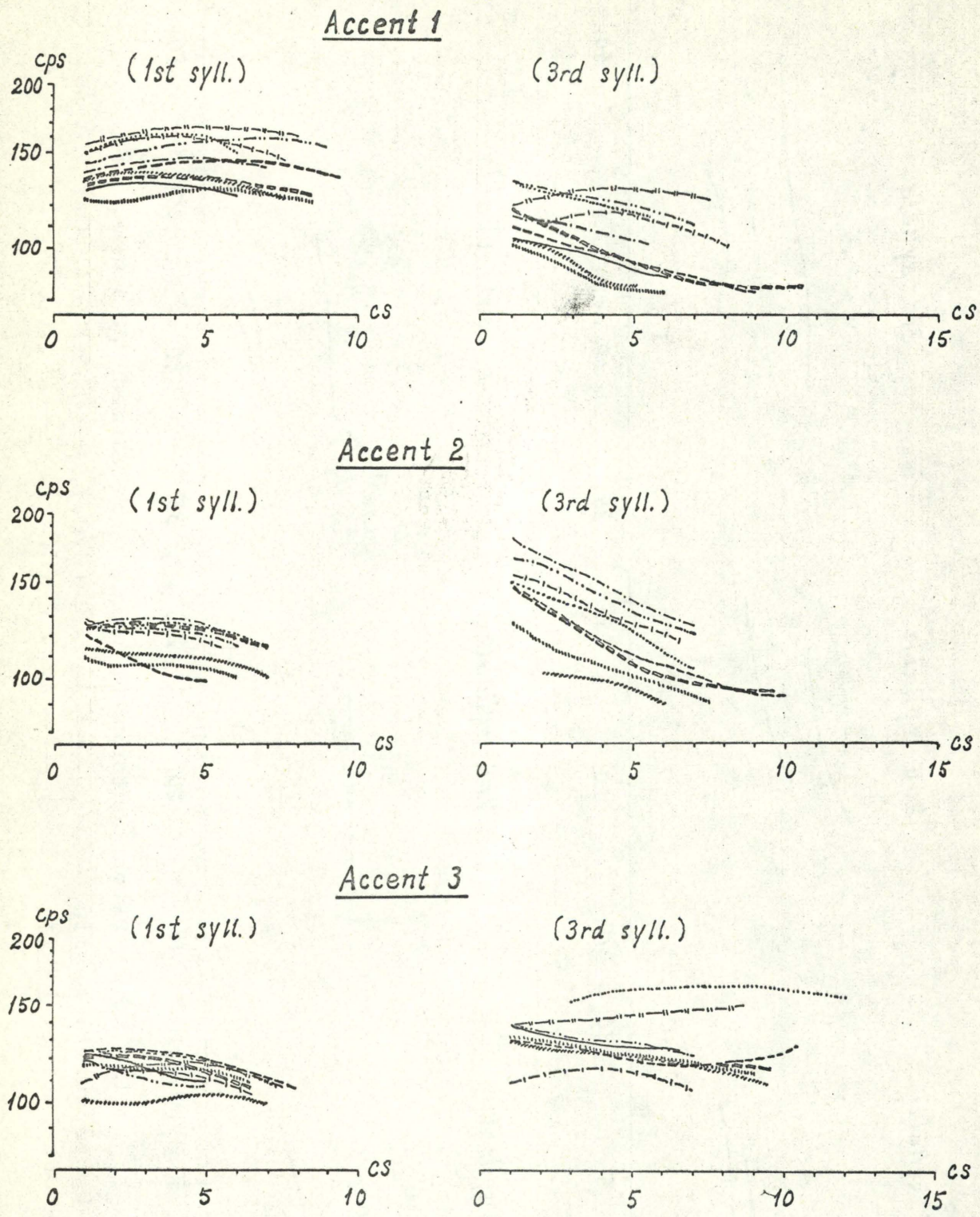

Fig. 9. Pitch contours of hasito (isolated words)

▪▪▪▪; ---- = K.T., ////////; ..... = H.M., others = M.O.

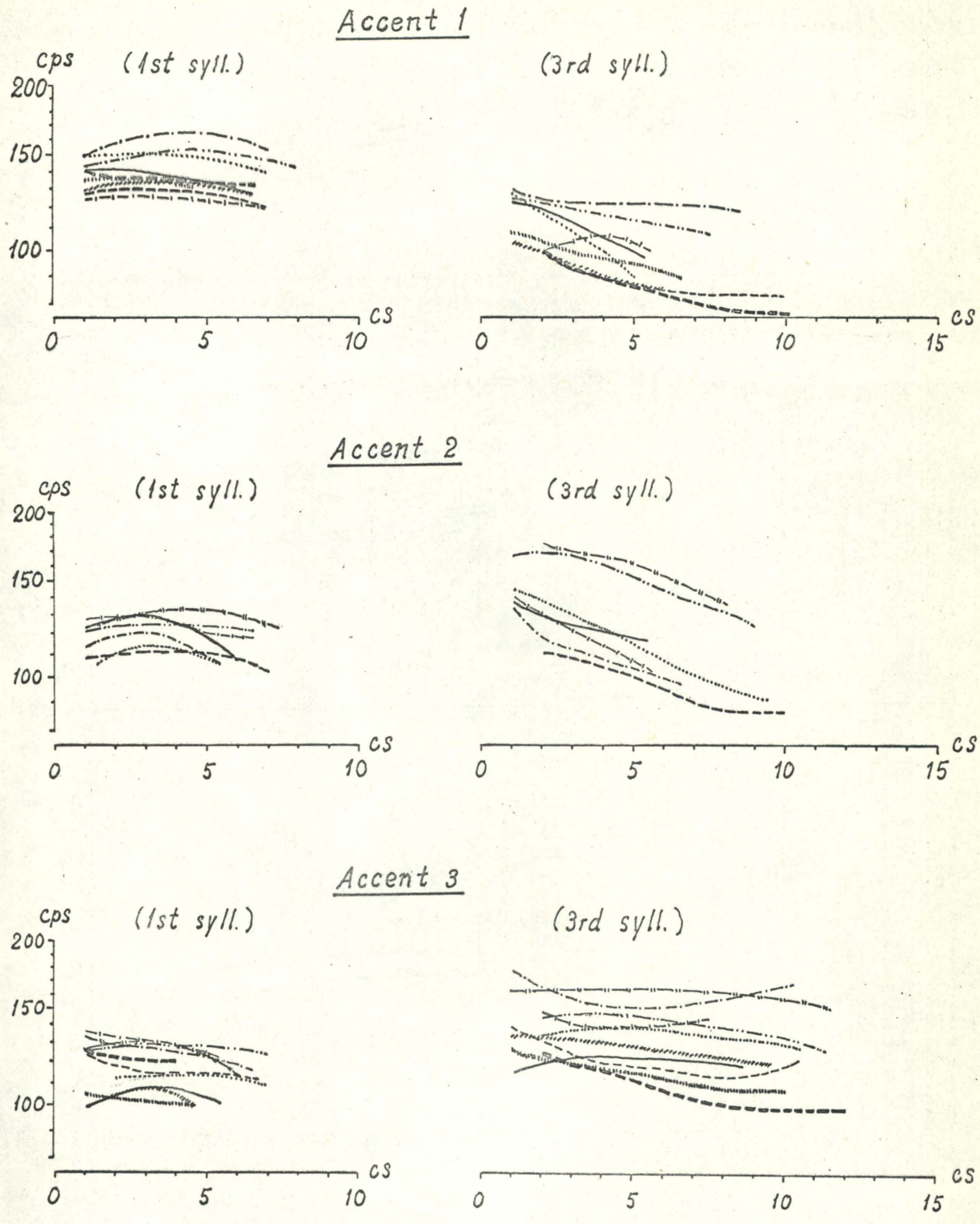

Fig. 1o. Pitch contours of kakito (isolated words)
▫▫▫▫▫; ----- = K.T., ‴‴‴‴‴; ⋯⋯⋯ = H.M., others = M.O.

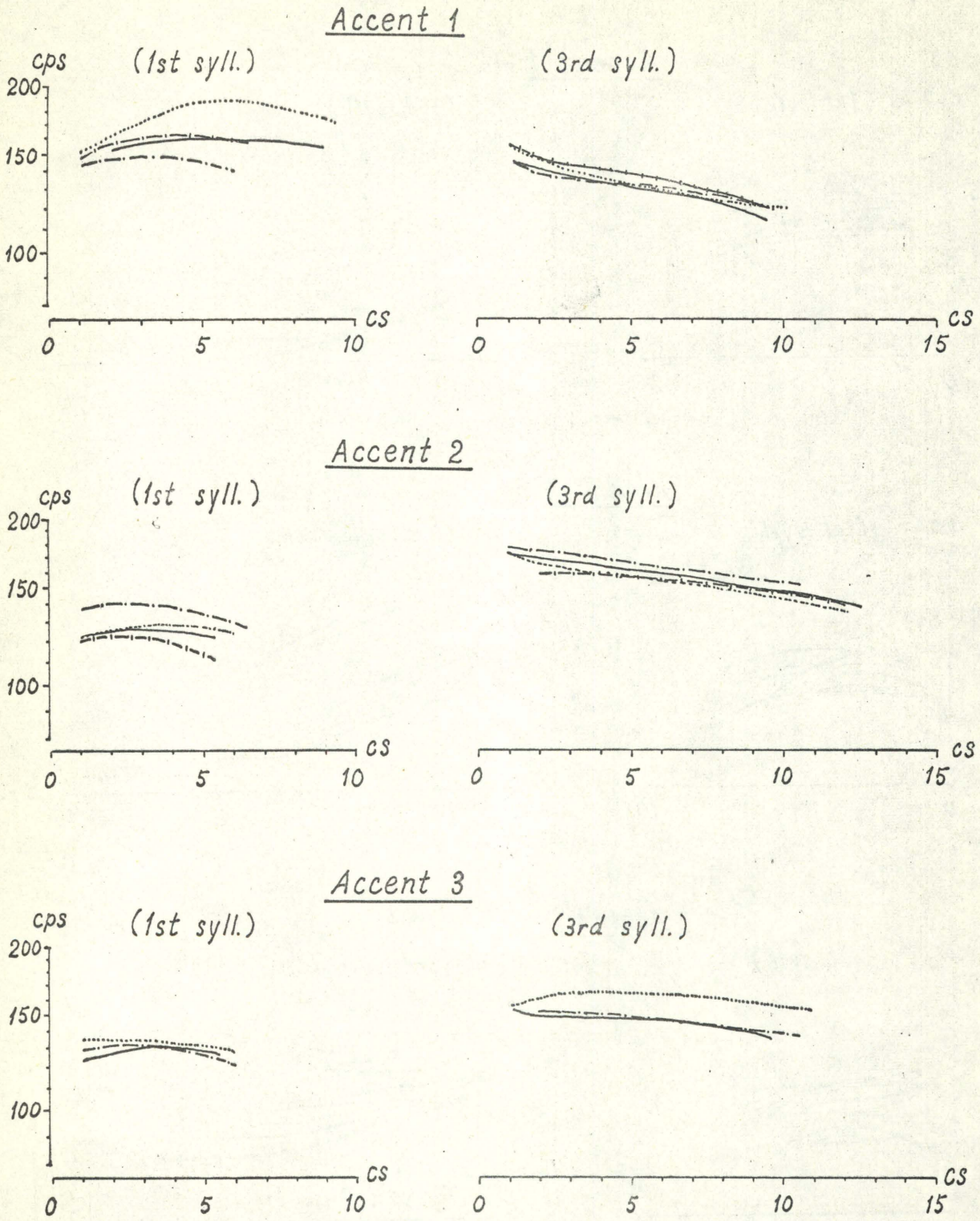

Fig. 11. Pitch contours of hasito (M.O. in sentences).

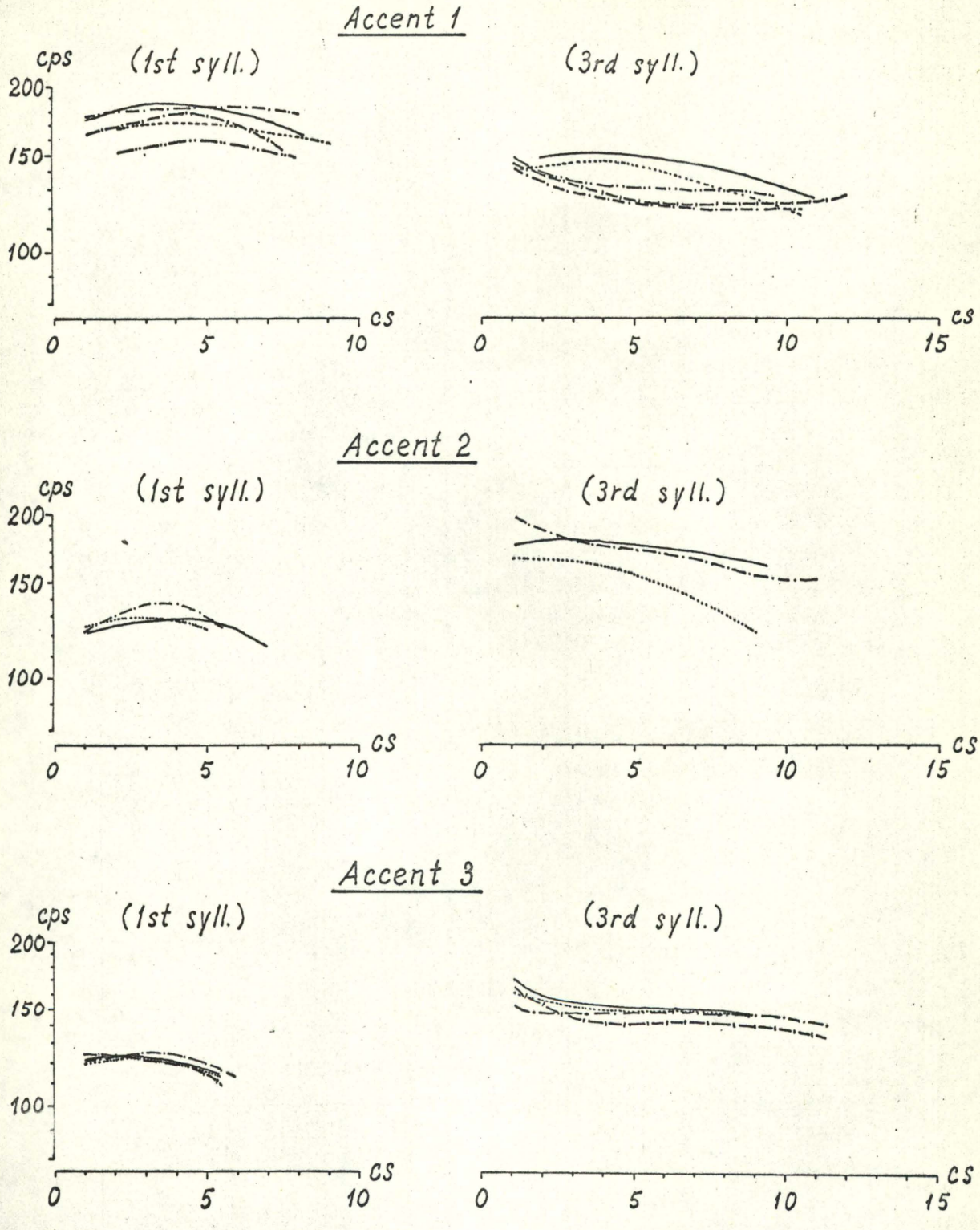

Fig. 12. Pitch contours of kakito (M.O. in sentences).

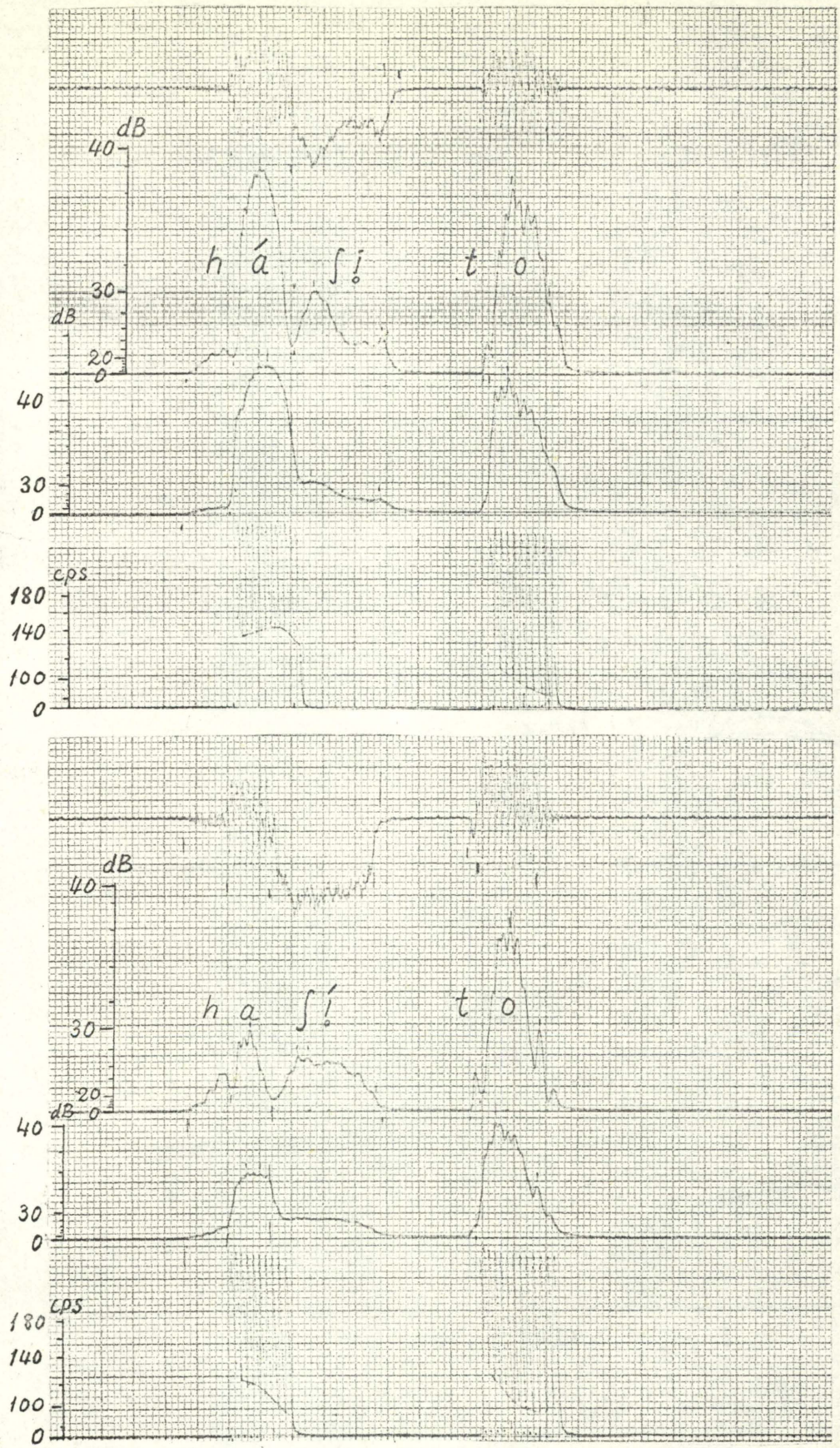

Fig. 13-1. Mingograms of <u>hásito</u> and <u>hasíto</u>
(K.T., isolated words)

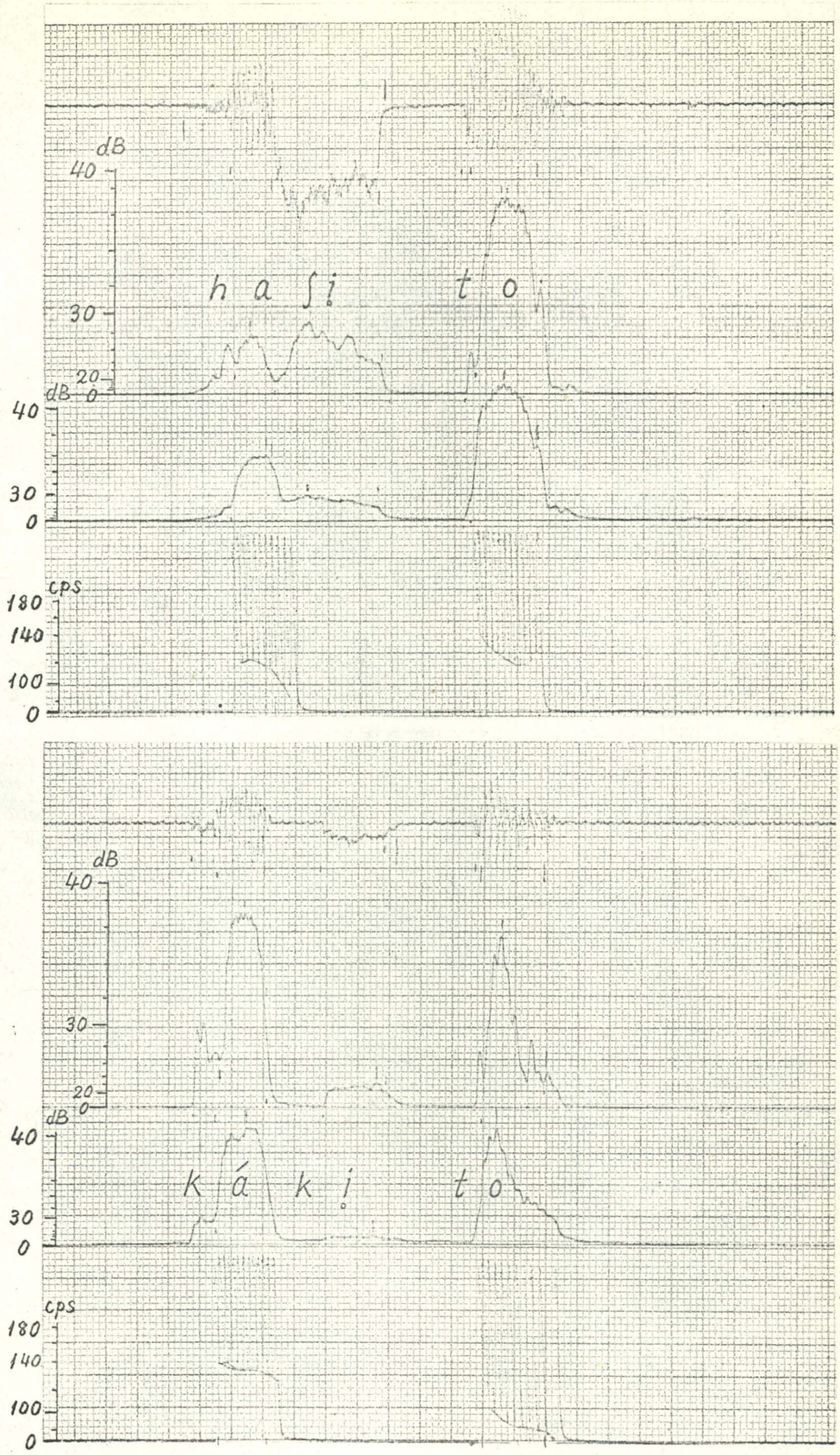

Fig. 13-2. Mingograms of <u>hasito</u> and <u>kákito</u>
(K.T., isolated words)

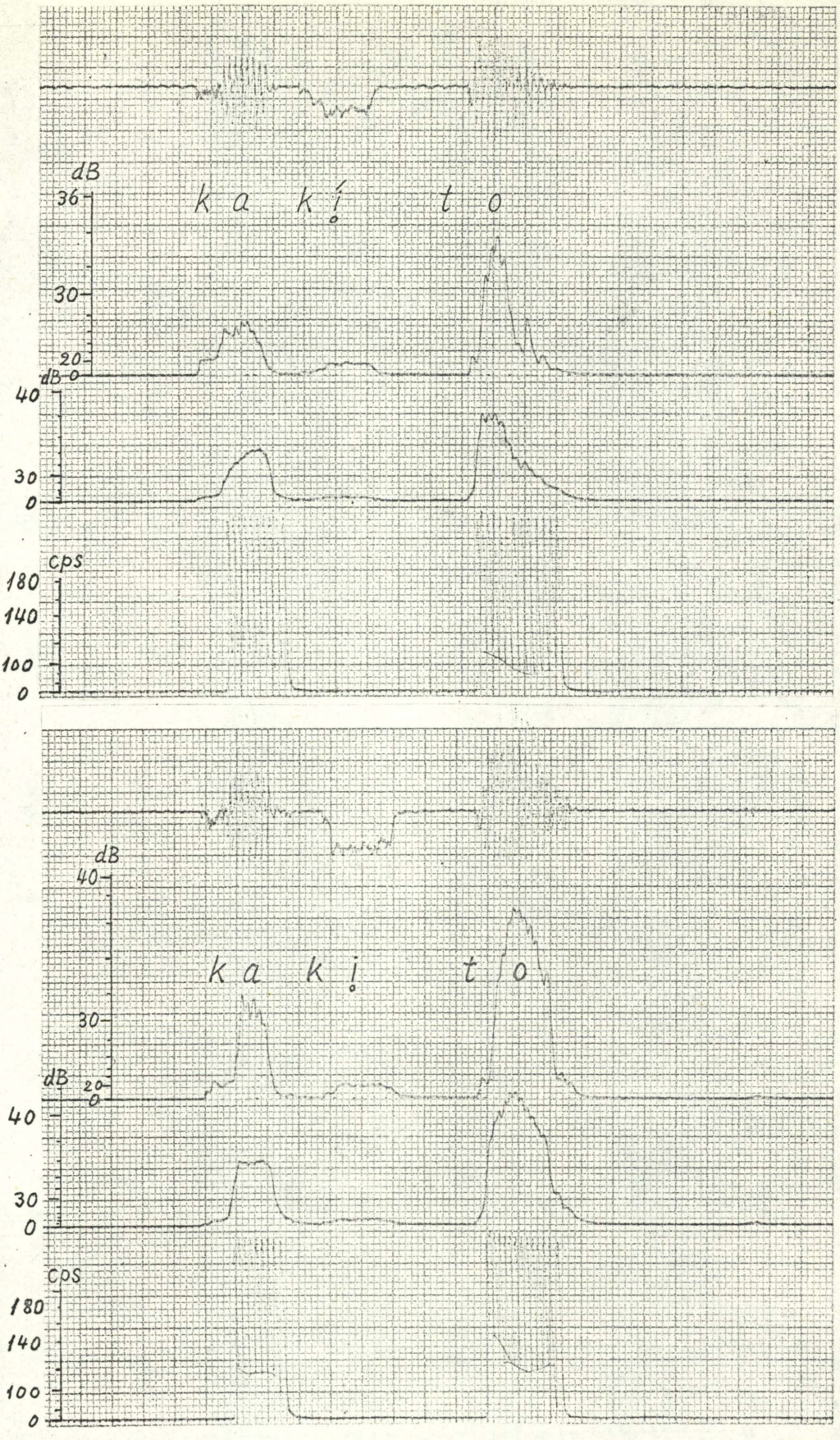

Fig. 13-3. Mingograms of kakíto and kakito
(K.T., isolated words)

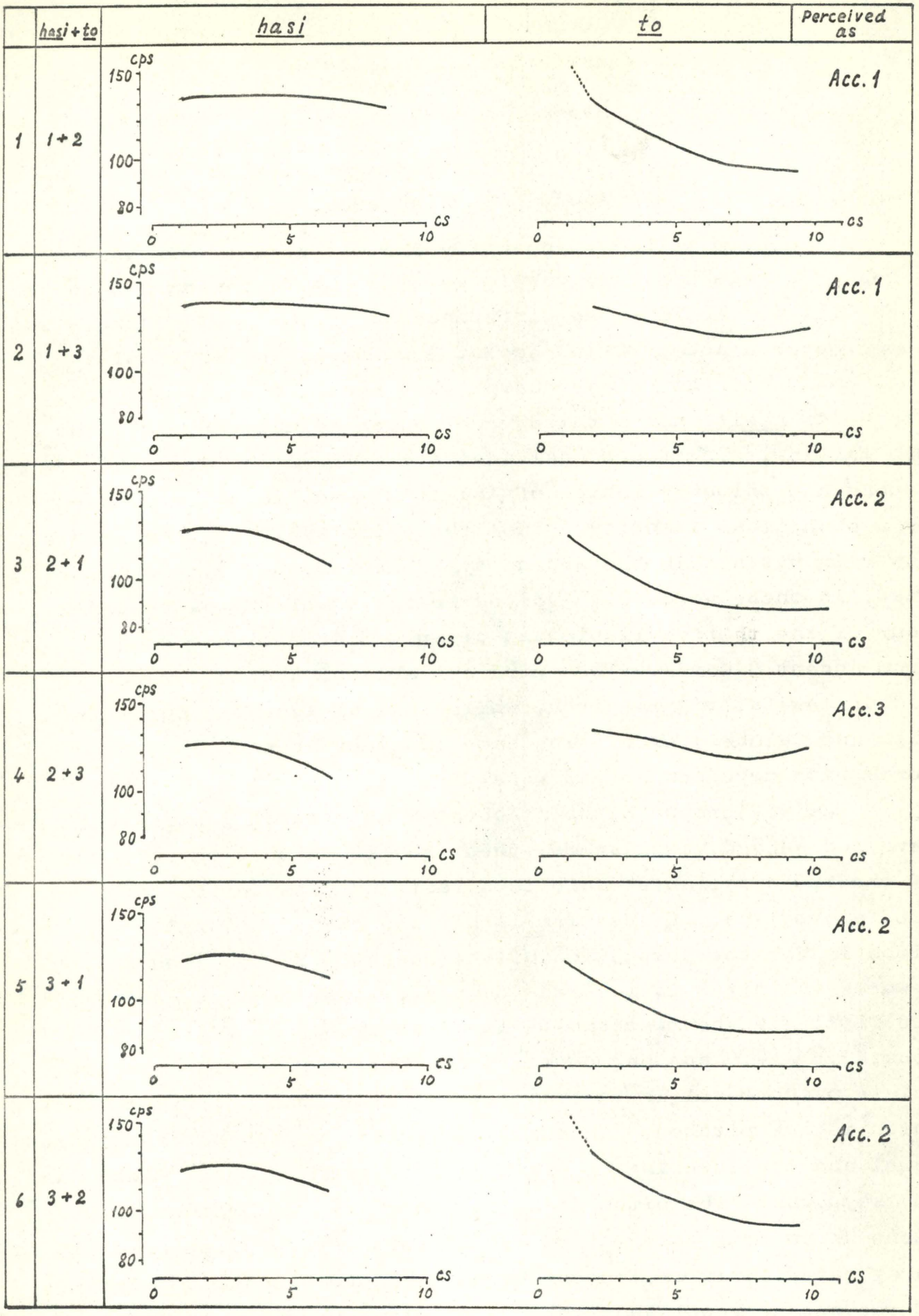

Fig. 14. Perception of accent in words after tape-cutting and recombination.

	1st syllable	3rd syllable
Acc. 1:	high	low falling;
Acc. 2:	low	high (or mid) falling, sometimes, low falling);
Acc. 3:	low	high even.

A problem is how to quantify the difference. In the words with voiced second vowel, we find the following tendencies (see Fig. 8): there is no difference in pitch level of the first vowel between Accent 2 and Accent 3 words in these examples (but cp. 2.1.). Among the examples of Accent 2 words one instance of hasito has a third vowel which is not very low, but mid falling. The functional identity of level between the second and third syllables of Accent 3 words is not reflected by a phonetical identity, since the third syllable is a little lower in pitch than the second syllable (cp. 0.2.). In any case, in these words with voiced second vowel the pitch contour of the third syllable can clearly distinguish Accent 2 from Accent 3 words. The third syllable of Accent 2 words is mid or low falling in pitch, while that of Accent 3 words is high and (almost) even. The same tendency is seen in the words with devoiced second vowel.

The difference between Accent 2 and Accent 3 words with devoiced second vowel seems, then, to lie in that the pitch of the third syllable exhibits a fast and considerable drop in the former type. A question is how important this falling pitch is for the perception of the accent. In isolated Accent 2 words the pitch is lowered about 5 cps per centisecond within the first five centiseconds. In a few cases the falling rate is about 3.5 - 4.0 cps per centisecond, but in this case the absolute pitch level is the same as or lower than the first vowel which is low pitched. Accent 3 words have a high pitched third vowel and the lowering of pitch is less than 10 cps. Mieko Han (3) says that "the pitch change of more than 2 semitones functions as an emic change" (p. 128). This may be quite crucial. In my material it is true of words with and without devoicing of the second vowel that the pitch change within the third syllable of Accent 2 words exceeds two semitones, and besides the

pitch ends at a lower level than that of the first syllable. In some cases, the third vowel has low falling pitch. As for the Accent 3 words, all the third vowels have even pitch contours. Their pitch levels are, except for one instance of hasito, higher than those of the first syllable.

As mentioned just above, the pitch contour of the third vowel is similar in words with voiced second vowel and those with devoiced second vowel. (The amount of pitch drop is a little greater in the latter case.) Then, a question is raised: which is more important for the accent perception, the pitch contour of the accented syllable, or that of the syllable following after the accented syllable? However, it is evident that the relation between the accented syllable and the immediately following syllable is most important. Words with devoiced vowel should probably be considered a special case (see, however, 3.2. below).

Accent 1 words do not seem to pose any problem. The accented first vowel is high in pitch, as well as longer in duration and greater in intensity, than the third vowel.

A little more problematic is the pitch contour of the words spoken in a sentence. The general tendency is the same as for isolated words. The pitch contours of Accent 1 words are just as clear as in the case of isolated words. The problem is that there is no great apparent difference between Accent 2 and Accent 3 words, especially in the case of hasito. This problem has not yet been completely solved. The formant structure of these words may give some cues, but minute spectrographic analyses have not yet been done.

2.4. Tapecutting experiment.

Three examples of hasito with different accents were chosen. They have almost the same intensity level and duration of segments. They were cut on the tape between hasi and to at a point 5 centiseconds before the explosion of t and treated in three alternative ways: (a) the final to was removed, (b) a 50 cs pause was inserted between hasi and to, and (c) different parts of hasi and to were combined. The results of a small

perception test (three subjects) show that

a) when there is no final syllable Accent 1 is easy to perceive, while there is no difference between the parts of hasi of Accent 2 and Accent 3,

b) when there is an internal pause Accent 1 is again easy to perceive. Accent 2 and Accent 3 are also perceived correctly. (It seems that the timing between syllables does not influence accent perception in Japanese. However, it is a matter of course that the information given by the third syllable is lost if the distance between the second and the third syllables is too great, which fact may be reflected in the case of (a).

c) when the parts are recombined at random the high pitch of the first syllable is the cue for the perception of Accent 1 (cf. (a)) which is further confirmed by the low falling pitch of the third syllable (see Fig. 14, curves 1 & 2). The falling pitch of the third syllable is a cue for Accent 2 (see Fig. 14, curves 3, 5 & 6), and the (almost) even and high pitch of the third syllable in addition to the lower pitch of the first syllable is a cue for Accent 3 (Fig. 14, curve 4). In this small experiment the perception of Accent 3 worked only with a combination of the hasi of Accent 2 and to of Accent 3. The result obtained here shows the same tendency as the above mentioned results of the whole experiment, but this should be confirmed further by tests including several words.

3. Final remarks.

3.1. In summary, taking all the three acoustic factors together, the accents tend to show the following acoustic differences:

1. The pitch factor plays the greatest role of the three factors.
2. Intensity and duration may also be of some importance, especially for Accent 1 words.

The result that the pitch movement in the third syllable differentiates between Accent 2 and Accent 3 words is similar to the result obtained by Sugitow (15) in her investigation of the pitch movement of two-syllable words where the first vowel is devoiced, though she investigated another dialect. As for the degree of pitch movement and the pitch of the third vowel

as cues, the results here agree to a large extent with Han's statements.

As for the role of intensity for the accent perception, Neustupný (op.cit.) says that in some cases accent perception does not work without the help of the intensity factor. In my experiment, the case of hasito in the sentence might be explained as a compensation, if we do not take formant structure into account. In other cases, however, the pitch movement does not show any negative result for pitch. The pitch factor may be supported by the intensity factor, but not vice versa.

3.2. There are many unsolved matters, such as the evaluation of the area integral of intensity, the relation between acoustic and auditory level tone, etc. But these problems are of a more general phonetic character.

In this experiment formant structures were not analyzed. At the first look at spectrograms, it is hard to find the difference between three differently accented words. However, if minute spectrographic analyses are undertaken, other factors may be found which are relevant to accent perception. Meyer-Eppler (9) investigated whispered German tense vowels sung in "the first five tones of a diatonic scale", and found some extra noises for whispered /i,e,o/ and noises in addition to formant transitions for whispered /a,u/. I do not know whether such extra noises occur in Japanese devoiced /i/ and /u/, which make the intensity higher and bring about differences in pitch perception. It is said that there is a difference in quality between accented and unaccented vowels (Onishi (13), Koizumi (7), etc.). Koizumi says that a certain correlation between the accent and the tension of muscles has been found:

> "The high pitch accent shows itself in the tense form of lower vowel variants [ɑ,ɔ,ɛ] or higher ones [i,u]. The low pitch accent shows itself in the lax form of higher vowel variants [a,o,e] or of lower ones [I,U]." (p.9)

If this phenomenon is constant there may be some difference in devoiced vowels too, both acoustically and auditorily. That is, there may be some difference in formant structure and in auditory impression.

References:

(1) Bernard Bloch, "Studies in Colloquial Japanese IV, Phonemics", Language 26 (1950), pp. 86-125; Readings in Linguistics (I), ed. Joos (1958), pp. 329-348.

(2) D.B. Fry, "Experiments in the Perception of Stress", Language and Speech I (1958), pp. 126-152.

(3) Mieko Shimizu Han, Japanese Phonology, An Analysis Based upon Sound Spectrograms (Tokyo, 1962).

(4) A.S. House & Fairbanks, "The Influence of Consonant Environment upon the Secondary Acoustical Characteristics of Vowels", JASA 25 (1953), pp. 105-113.

(5) Shin Kawakami, "Japanese Accent, Observed by the Aid of Pitchgraph", The Study of Sounds 10, pp. 115-129.

(6) Ryuji Kobayashi, "Nihongo (Tōkyō hōgen) akusento no nijūsei ni tsuite" ("On the Duplex Nature of Pitch Accent in Japanese (Tokyo Dialect)"), Nihon onkyō-gakkai kōen ronbun-shū (1967), pp. 281-282.

(7) Tamotsu Koizumi, "Nihongo no akusento to boin no onka" ("On the Japanese Accent and Vowel Quality"), Bull. Phon. Soc. Japan 126 (1967), pp. 9-11.

(8) Philip Lieberman, "Some Acoustic Correlates of Word Stress in American English", JASA 32 (1960), pp. 451-454.

(9) W. Meyer-Eppler, "Realization of Prosodic Features in Whispered Speech", JASA 29 (1957), pp. 164-166.

(10) John Morton & Wiktor Jassem, "Acoustic Correlates of Stress", Language and Speech 8 (1965), pp. 159-181.

(11) Kazuo Nakano, "Shiin no yūseika to museika" ("Consonantal Vocalization and Devocalization"), Bull. Phon. Soc. Japan 125, pp. 7-8.

(12a) J.V. Neustupný, "Haná ("flower") to Hana ("nose") no hatsuon wa hatashite chigau ka" ("Is there any difference of pronunciation between /hana˥/ and /hana/ ?"), Gengo Seikatsu (Tokyo) 172, Part 1, pp. 80-87.

(12b) J.V. Neustupný, "Nihongo no akusento wa kōtei akusento ka" ("Is the Japanese Accent a Pitch Accent?"), Bull. Phon. Soc. Japan 121, pp. 1-7.

(13) Masao Onishi, "Onshitsu to onchō-suii no sōkankankei" ("Relativity of Sound Quality and Pitch Transition in the Japanese Language"), The Study of Sounds 13, pp. 51-81.

(14) Hide Helen Shohara, "Some Characteristics of Pitch Variations in Japanese", Lingua I (1948), pp. 405-409.

(15) Miyoko Sugitow, "Musei boin to akusento" ("Reciprocal Action between Japanese Accent and Devocalization"), Bull. Phon. Soc. Japan 123, pp. 4-7.

(16) Michoko Temma, "An Experimental Study on the Phonetic Changes in Japanese Speech", The Tsuda Review, No. 6, November 1961, Tsuda College, Tokyo, pp. 23-40.