OSU Working Papers in Linguistics 44 (146-165)

Is there 'dephrasing' of the accentual phrase in Japanese?*

Kikuo Maekawa
kikuo@tansei.cc.u-tokyo.ac.jp

Abstract: An experiment was carried out in order to examine two putative cases of 'dephrasing' of the accentual phrase in Japanese. The result revealed that it was possible to detect the accentedness of seemingly 'dephrased' accentual phrases in most of the cases. Although f0 contours of the accentual phrases in question are quasi-linear, we can detect their accentedness through the statistical examination of slope and intercept values of the regression lines fitted to the f0 contours.

1. Introduction

1.1. *Accentual phrases and their dephrasing*

In this paper, I will examine the notion of 'dephrasing' of the accentual phrase in Standard (Tokyo) Japanese. Experimental evidence will indicate that, contrary to current standard theoretical assumptions, it is possible to detect the accentedness of seemingly 'dephrased' accentual phrases. In this section I will review briefly the current theoretical assumptions about 'dephrasing'.

Standard Japanese has the system of pitch accent. Japanese accent differs considerably from the pitch accent of languages like English (and many other European languages) in that it is specified almost exclusively at the level of the lexicon. The accentual phrase of Japanese is usually defined as the domain of two independent phonological/phonetic events. It is the domain of at most one accent and it is also the domain of the phrase initial rise which demarcates the boundary of two successive accentual phrases. Because accent is paradigmatically contrastive in Japanese, Japanese accentual phrases can be classified into accented and unaccented ones according to their accentedness. In its physical realization, an accented accentual phrase is characterized by a sharp fall of fundamental frequency (f0) at the designated accent location, whereas an unaccented accentual phrase

*This study was carried out during my stay at OSU. I would like to express my gratitude to Mary Beckman who made my stay possible and provided many invaluable comments to this and other studies that I conducted at OSU. My gratitude goes also to Osamu Fujimura and Masashi Sawada who gave me various comments in the course of the study. My stay in the U.S. was supported by the Ministry of Education, Japan.

does not possess such a fall, and is instead characterized by a relatively gradual pitch fall spanning the whole time portion after its initial rise.

An important assumption which virtually all phonological analyses of Japanese intonation hold is that two or more accentual phrases can be dephrased and merged into one thereby deleting all accents except the left-most one (e.g. McCawley, 1968; Poser, 1984). Under this standard assumption, the effect of dephrasing can be shown schematically as in Figure 1. Dephrasing of unaccented plus unaccented accentual phrases results in one long unaccented accentual phrase (Case 1 of Figure 1), whereas if at least one of the component phrases is accented the resulting phrase is a long accented accentual phrase (Cases 2-4). It is important to note that case 4 differs considerably from all the rest because in this case dephrasing implies the tonal deletion of the second accent. This is one of the rare instances in Japanese where a lexical accent is deleted post-lexically, and it is this case that will be examined in this paper. Two more points are to be noted here. First, the standard assumption predicts that the accentual phrase resulting from the dephrasing of accented and unaccented accentual phrases (Case 3) and the one resulting from the dephrasing of two accented accentual phrases (Case 4) are exactly the same. Second, when two accented accentual phrases are concatenated (i.e. not dephrased), the phonetic prominence of second accent (most notably in terms of its peak height) is usually reduced by the influence of the preceding accent. This effect is known as downstep or catathesis. (Poser, 1984; Pierrehumbert & Beckman, 1988).

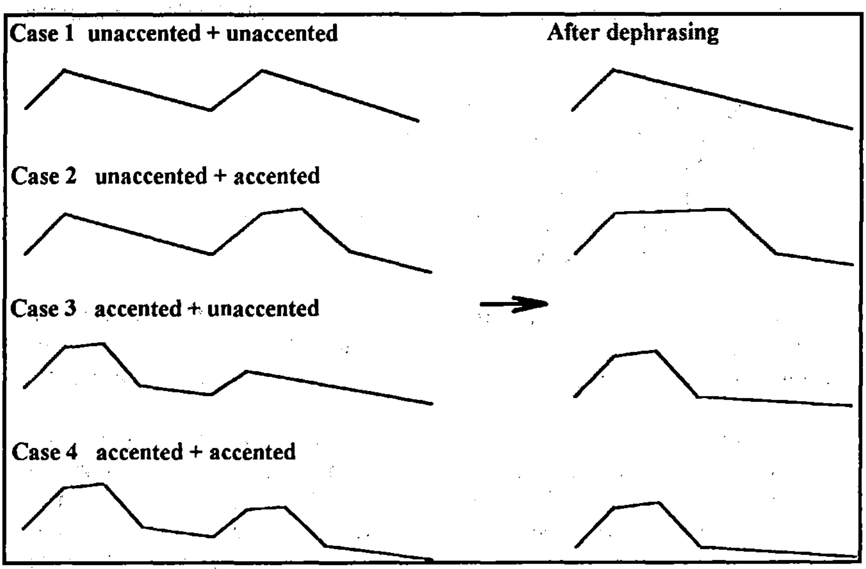


Figure 1. Schematic representation of the effect of dephrasing.

1.2. *Reported instances of 'dephrasing'*

There are two instances of dephrasing mentioned in recent experimental phonological studies of Japanese intonation which will have direct relevance to the present experiment. The first one is dephrasing occurring in post-focus position; when an accented accentual phrase is emphasized due to some pragmatic/discourse factor, the accentual phrase which immediately follows the focused accented accentual phrase seems to be dephrased quite often. Beckman & Pierrehumbert (1986: 265) compared two f0 tracks of (a) *UMA'I mame'wa arimase'n* 'There are no tasty beans' and (b) *UMA'I amewa arimse'n* 'There is no tasty candy' and concluded that "the tonal contours are identical because the grouping has deleted the lexical accent in *mame'*." (Capitals show emphasis and an apostrophe denotes the location of accent.) In this case, the effect of dephrasing can be shown as in (1a) and (1b), where []$_\alpha$ denotes an accentual phrase.

(1a) [uma'i]$_\alpha$ [mame'-wa]$_\alpha$ [ari-mase'-n]$_\alpha$
tasty beans-TOPIC be-POLITE-NEG
=> [uma'i mame-wa]$_\alpha$ [arimase'n]$_\alpha$
(1b) [uma'i]$_\alpha$ [ame-wa]$_\alpha$ [ari-mase'-n]$_\alpha$
tasty candy-TOPIC be-POLITE-NEG
=> [uma'i ame-wa]$_\alpha$ [arimase'n]$_\alpha$.

In this respect, my own study reported in Maekawa (1991) is to be noted, since it seems to be the only study that examined the perceptual aspect of this issue. It dealt with the distinction of quasi-homophonic WH and Yes-No intonation, as shown in (1c) and (1d).

(1c) na'ni-ga mi-e'-ru
what-NOM see-POTENTIAL-IMPERFECT
'What can you see?'
(1d) na'nika mi-e'-ru
Something see-POTENTIAL-IMPERFECT
'Can you see anything?'

A perception test using LPC synthetic intonation showed that the best approximation to the WH-question intonation could be obtained when the intonation of the accented predicate was approximated by a linear f0 contour, such as in Case 4 of Figure 1.

Although it was very natural to regard the result as perceptual support for the above mentioned standard assumption of dephrasing, I hesitated to do so, since my intuition as a native speaker of the language dissuaded me from treating the intonation of the WH question as an instance of a dephrased accentual phrase. It seemed to me perfectly possible to hear an accent in the linear portion of the intonation, even in the synthetic stimuli.[1] The experiment which will be presented

[1] The conclusion of Maekawa (1991) was that the intonational difference between WH and Yes-

in section 2.1 below will address this problem and I will propose a solution to it.

The second instance of dephrasing mentioned in the recent experimental phonological literature is the one represented by Poser's(1984: 142ff) treatment of auxiliary verb *mi'ru.* He treated the intonational difference between the two instances of *yonde miru* ('try reading' or 'read and see') as a difference in accentual phrasing (his 'minor phrase').

(1e) [yo'n-de mi-ru]$_\alpha$
read-GER see
'try reading'

(1f) [yo'n-de]$_\alpha$ [mi'-ru]$_\alpha$
read-GER see
'read and see'

According to Poser, *mi'ru* in (1e), which functions as an auxiliary verb, is dephrased by the principle of *minimal minor phrase formation* which states roughly that the prosodic word consisting of a content word and any following function word should form a minor phrase. By this account *mi'ru* in (1f) is not dephrased since this is a full verb. Poser's account is not specific to auxiliary verbs; the same account is used for case markers and other particles as well as conjunctions and the copula (see Poser, 1984: 148).

Poser's account of auxiliary verbs was criticized in Kubozono (1993). Based on the visual inspection of larger amount of f0 data[2], Kubozono found that there was a considerable amount of variation in the physical realization of the accent in auxiliary verbs as well as auxiliaries like *daro'o* 'perhaps' and two morae long particles like *ma'de* 'until' and *yo'ri* 'rather than'.[3] Kubozono classified the allophonic variants observed in his study into the three types shown in Figure 2.

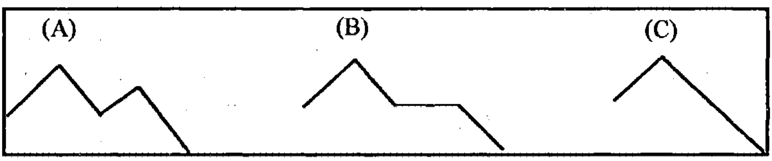


Figure 2. Classification of the realizations of two consecutive accented accentual phrases. Adopted from Kubozono (1993: 104).

Kubozono's type A is the full realization of the auxiliary verb (or other particles) as an independent accentual phrase (his 'minor phrase'), with the second accent (on the auxiliary) downstepped. Type B is what Kubozono calls 'total downstep' following the terminology used by Poser, who adopted the term from Pierrehumbert's (1980) analysis of the English H*L pitch accent. Type C is the

No questions should be attributed to a difference in intermediate phrasing.

[2] Poser's f0 data was taken from only one speaker, while Kubozono analyzed four.

[3] Auxiliaries like *daro'o* are syntactically different from auxiliary verbs primarily in that they do not inflect, and secondarily because they don't have full-verb counterparts.

expected form of dephrasing as predicted by Poser. Kubozono's conclusion is that apparent 'dephrasing' of auxiliary verbs and other particles is not the case of phonological dephrasing (his 'accent reduction') but it is rather the weakening of an accent caused by the effect of downstep at the phonetic level. It is important to note that Kubozono states at the end of his discussion that it is not unrealistic to suppose that type C "represents an extreme case of downstep, a case where the second accented element is totally reduced by the preceding accented element ---*so much so that neither the phrase-initial F0 rise nor the accent of the second element is now realized as the phonetic output.*" (p.113, italics mine.)

Kubozono's account is in accordance with native speakers' intuition; in fact the same view was presented by Japanese linguists much earlier back in late 1960s based fully on impressionistic observations. (see Wada 1969 among others, and also Fujimura 1991 for relevant issues.) However, it seems that Kubozono did not explore the full possibility of his proposal. If we follow his view and push it one step further, couldn't we expect that some remaining trace of the accent is to be found even in the type C realization, whose accent Kubozono regarded to be completely unrealized? This point will be examined in section 2.2.

2. Experiment

In this section I will show the results of two experiments on the dephrasing of accentual phrases in Japanese. One of them was concerned with dephrasing in post-focus position and the other one was concerned with dephrasing of auxiliary verbs. The experimental material used here is a part of larger data set which awaits complete analysis. At present, two native speakers of Tokyo Japanese plus the author have taken part in the recording, and a few more speakers are expected to be recorded. Both of the speakers —abbreviated as YO and NF below— are female graduate students of the Ohio State University in their early thirties, and the author (KM) is a male in his mid thirties. YO has a fair amount of knowledge in linguistics and applied linguistics. NF is majoring in education and has no specialized knowledge of linguistics. Neither of them were told the purpose of the experiment. Both took part in the recording twice and read the material described below in slightly different ways.

The data set consisted of thirty-four sentences; each one of them was either a question or an answer to a previously uttered question. Each pair of related question and answer was printed on a separate index card and the cards were randomized every time before new recording session started. Every speaker read each card at least ten times. Recording was conducted in a sound booth in a quasi-conversational setting, i.e. a dummy speaker —actually KM, the author— took part in the recording as an interlocutor, and the target speaker was asked to read the pairs of question/answer sentences interchanging his/her role as questioner and respondent with that of dummy speaker. This setting was devised after the first recordings of YO and NF in order to elicit as natural utterances as possible, since in their first recordings sentences prepared for post-focus dephrasing showed considerable variation ranging from Kubozono's type A through C, of which only type C is relevant for the purpose of the current experiment. Judging from the impressionistic as well as quantitative criterion described below, the new setting was an effective one. All data were taken from the second recordings.

2.1. *'Dephrasing' in post-focus position*

Two pairs of question/answer sentences were used for this experiment. The WH words in the questions are supposed to be focused in their physical realization. The question mark at the end of sentence stands for the final rise for question rendition.

(2a) ho'n-o da're-ga yo'n-de-iru'-no?
book-OBJ who-NOM read-GER-PROG-Q
'Who is reading the book?'
(2b) ho'n-wa yu'mi-ga yo'n-de-iru
book-TOPIC Yumi-NOM read-GER-PROG
'It is Yumi who is reading the book.'
(3a) ke'n-o da're-ga yon-de-iru'-no?
Ken-OBJ who-NOM call-GER-PROG-Q
Who is calling Ken?
(3b) ke'n-wa yu'mi-ga yon-de-iru
Ken-TOPIC Yumi-NOM call-GER-PROG
'It is Yumi who is calling Ken.'

The crucial difference between (2) and (3) is in the accentedness of the predicate verb stems. Under the standard assumption, the accent of the verb in (2a) should be deleted if dephrasing takes place between the second and the third constituents of the sentence due to the intrinsic focus of the WH-word. Hence the resulting intonation contour is expected to be the same as the one observed for (3a), which is under the same effect of dephrasing. The expected prosodic structure shared by (2a) and (3a) under the effect of dephrasing is $[\text{da'rega yondeiruno}]_{\alpha}$.[4]

Figure 3 shows typical f0 contours of the relevant portions of (2a) and (3a) as uttered by speaker YO.[5] Two vertical lines in each panel are set to the beginning and the end of the verb stem defined by spectrographic investigation. The left edge of the stem was defined as the mid point of the second formant transition from /j/ to /o/, and the right edge was defined as the beginning of the vowel /e/ of the gerundial /de/.[6] At first glance, comparison of the two f0 contours seems to provide good support for the standard assumption of dephrasing since neither a phrase initial rise nor a local f0 fall due to the accent can be seen in the delimited time portion. For both accented and unaccented verb stems, the whole time portion is associated with a quasi-linear f0 contour which constitutes a part of simple

---

4 Note it is assumed that the accents at the end of /iru/ in (2) and (3) are both deleted under the effect of dephrasing.

5 All f0 tracking was done by the 'formant' program of ESPS with the framelength of 10ms.

6 I have excluded the segment of /j/ from the acoustic definition of verb stem because in some speakers' f0 contours the L of bitonal HL accent was realized later in time and found itself in the time segment of /j/.

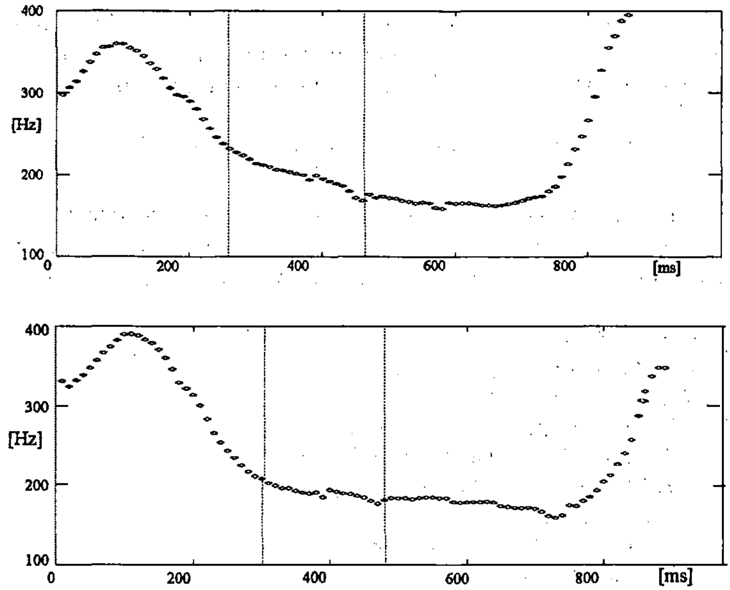


Figure 3. Typical f0 contours of (2a) and (3a) by YO. Two vertical lines on each panel correspond to the left and right edges of verb stem defined by acoustic segmentation. Part of the f0 contours preceding WH-words were omitted. The top panel shows (2a) *da'rega yo'ndeiruno?* The bottom panel shows (3a) *da'rega yondeiruno?*.

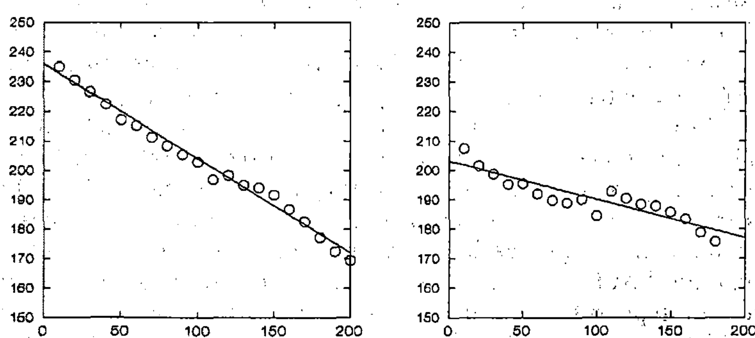


Figure 4. Example of line fitting. Lines were fitted to the f0 data points extracted from the verb stems of (2a) and (3a) by the least-squares method. Here, the same utterances as in Figure 3 were used for examples. The left panel is the accented verb stem (2a). The right panel is unaccented (3a). For each panel, the abscissa stands for time in ms, and the ordinate stands for f0 in hertz.

Table 1: RMS prediction error of (2a) and (3a) for speaker YO

Sentence	Verb stem	N	Mean	SD	*t*-test (two-tailed)
(2a)	Accented	172	2.197	1.926	*t* = -1.586, (D.F.=341.2)
(3a)	Unaccented	178	2.556	2.299	P=0.114

Table 2: RMS prediction error of (2a) and (3a) for speaker KM

Sentence	Verb stem	N	Mean	SD	*t*-test (two-tailed)
(2a)	Accented	191	1.864	1.817	*t* = -1.733, (D.F.=374.3)
(3a)	Unaccented	193	2.213	2.121	P=0.084

Table 3: RMS prediction error of (2a) and (3a) for speaker NF

Sentence	Verb stem	N	Mean	SD	*t*-test (two-tailed)
(2a)	Accented	186	4.824	4.565	*t* = 1.453, (D.F.=370.4)
(3a)	Unaccented	188	4.155	4.324	P=0.147

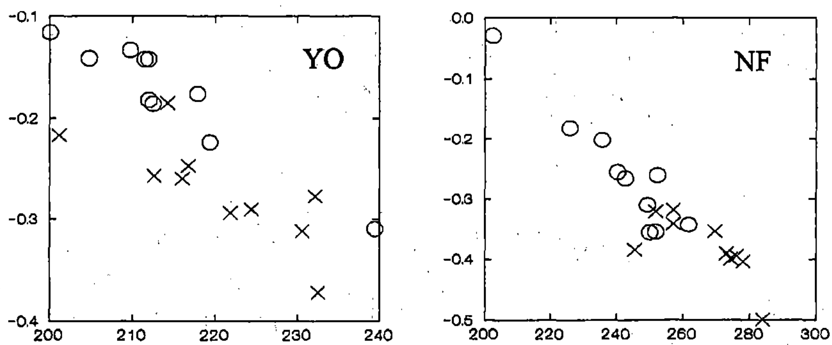
Table 4: Means and SDs of accented and unaccented data clouds

		Mean slope (SD)			Mean intercept (SD)	
Speaker	N	Accented (2a)	Unaccented(3a)	N	Accented (2a)	Unaccented (3a)
YO	10	-0.271 (0.052)	-0.173 (0.057)	10	220.221 (10.017)	213.905 (10.565)
KM	10	-0.119 (0.044)	0.005 (0.047)	10	116.050 (7.903)	105.055 (4.691)
NF	10	-0.381 (0.053)	-0.253 (0.100)	10	266.652 (12.806)	241.182 (16.856)

Table 5: Statistical test of the data shown in Table 4.

Speaker	T^2	F (DF)	P	*t* (DF) P of slope	*t* (DF) P of intercept
YO	1.906	16.199 (2,17)	0.000	-4.020 (17.8) 0.001	1.372 (17.9) 0.187
KM	2.013	17.106 (2,17)	0.000	-6.019 (17.9) 0.000	-3.783 (17.9) 0.002
NF	1.906	6.972 (2,17)	0.006	-3.575 (13.6) 0.003	3.805 (16.8) 0.001

linear interpolation between the accentual L of preceding accent (of *da're*) and the beginning of the final rise for question rendition which is usually realized on the last mora. However, since this interpretation is based fully on visual inspection, which can not be free from subjectivity, we must seek for a more reliable quantitative method of analysis. The method adopted here is a line fitting by means of the least-squares principle. This method computes a line which represents the overall trend of given data points by minimizing the prediction error. Specifically, when the prediction error is defined as the distance between the line and a data point on the ordinate, it is equivalent to the computation of a regression line in statistics, which is the case for the present experiment. Figure 4 shows examples of lines fitted to accented and unaccented verb stems respectively. In conducting these and other computations of line fitting, the first data point was assigned to the time value of 10ms in order for the comparison across utterances to be possible. Tables 1, 2 and 3 show the means and the standard deviations (SD) of root-mean-square (RMS) prediction error averaged over the accentedness of the verb stem. The mean difference between accented and unaccented verb stems was not significant for all three speakers. Note also that RMS values are greater for unaccented verb stems than for accented verb stems for two out of three speakers.



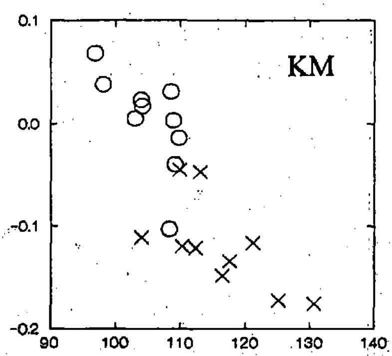


Figure 5. Intercept vs. slope scatter plot of the lines fitted to (2a) and (3a). Upper left = Speaker YO. Upper right = NF. Bottom = KM. In each panel, the abscissa stands for intercept and the ordinate for slope. Cross and circle plot symbols stand for accented and unaccented data points respectively.

I see this as evidence showing that the local pitch fall due to the accent was absent in this context.[7]

The results we have obtained so far seem to show nothing contradictory to the standard assumption of dephrasing. However, the linearity of intonation does not directly assure the occurrence of 'dephrasing', since Tables 1, 2 and 3 do not provide full information concerning the exact shape of f0 contours. Figure 5 shows scatter plots of slope vs. intercept values of lines fitted to the observed f0 contours of the three speakers.[8] The data points are classified according to the accentedness of the verb stems. The figure shows clearly that there is a difference between the lines fitted to the accented verb stem and those fitted to the unaccented verb stem. The former lines are characterized by larger intercept values and smaller slope values (in terms of their absolute values). Table 4 shows means and SDs of accented and unaccented data clouds, and the results of statistical tests are summarized in Table 5.

The difference of two-dimensional means between the accented and the unaccented data clouds was tested by Hotteling's T^2 statistic, which follows the F distribution under certain mathematical transformation (Green, 1978).[9] Results of a (univariate) t-test on slope and intercept are shown in Table 5 also. All T^2 values are statistically significant at least at the 0.01 level. Results of t-test confirm the separability of data clouds on each dimension with the sole exception of the intercept of speaker YO. These statistics show clearly that there is a systematic difference of f0 contour between the two verb stems which is related to the difference of the accentedness of the two lexical items. If we stick to the view that the f0 contours exemplified by the upper panel of Figure 3 are an instance of dephrasing, it is impossible to provide a linguistic explanation for the observed physical difference.[10]

2.2. *'Dephrasing' of auxiliary verbs*

The second experiment dealt with the following sentences. Phonological contrasts between (4) and (5) are attributable to the choice of accented or unaccented auxiliary verbs, i.e. /mi'ru/ or /iru/.

(4a) na'ni-o   no'n-de-mi'ru-n-desu-ka?
what-OBJ   drink-GER-try-COMP-copula-Q
'What will (you) try drinking?'

---

[7] The means and SDs of speaker NF are considerably greater than the values of YO and KM. In fact, some of her utterances of (2a) and (3a) would be likely to fall under Kubozono's category B.

[8] As is known in analytic geometry, the shape of a line can be parameterized perfectly by its slope and intercept.

[9] Hotelling's T^2 is a common diagnostic in multivariate statistical analysis. See Maekawa (1989) for the application of this statistic for the study of vowel merger.

[10] Needless to say, there is a semantic difference between the two verb stems. However, there seems to be no linguistically plausible explanation relating the semantic difference between 'to read' and 'to call' to the observed difference of f0 contours.

(4b) no'n-de-mi'ru-no-wa amerika-no bi'iru-desu
drink-GER-try-COMP-TOPIC Amerika-GEN beer-copula
'(I) will try drinking American beer.'

(5a) na'ni-o no'n-de-iru-n-desu-ka?
what-OBJ drink-GER-ing-COMP-copula-Q
'What are (you) drinking?'

(5b) no'n-de-iru-no-wa amerika-no-bi'iru-desu
drink-GER-PROG-COMP-TOPIC America-GEN-beer-copula
'(I'm) drinking American beer.'

Some native speakers of Tokyo Japanese might feel that the complementizer (COMP) *no* in (4b) and (5b) and the auxiliary verb *iru* in (5a) are accented.[11] Following this intuition, the representation of these sentences would be as follows.

(4a') na'ni-o no'n-de-mi'ru-n-desu-ka?
(4b') no'n-de-mi'ru-no'-waamerika-no bi'iru-desu
(5a') na'ni-o no'n-de-iru'-n-desu-ka?
(5b') no'n-de-iru-no'-wa amerika-no bi'iru-desu.

This intuition is an important one because it is contradictory to the prediction provided by the standard assumption based on the 'dephrasing' of auxiliary verbs. In any case, supposing that post-focus dephrasing does not take place, the standard assumption predicts the following phrasings.

(4a") $[\text{na'ni-o}]_\alpha$ $[\text{no'n-de-miru-n-desu-ka}]_\alpha$
(4b") $[\text{no'n-de-miru-no-wa}]_\alpha$ $[\text{amerika-no bi'iru-desu}]_\alpha$
(5a") $[\text{na'ni-o}]_\alpha$ $[\text{no'n-de-iru-n-desu-ka}]_\alpha$
(5b") $[\text{no'n-de-iru-no-wa}]_\alpha$ $[\text{amerika-no bi'iru-desu}]_\alpha$

According to this prediction, the f0 contours of the second accentual phrases of (4a") and (5a") on the one hand, and the f0 contours of the first accentual phrases of (4b") and (5b") on the other should be exactly the same.

Basically the same procedure of data analysis as that used in the previous experiment was applied for this data set. However, there are difficulties inherent in this data set. For one thing, it is difficult to find a reliable acoustic segmental boundary between /e/ and /i/ of /no'ndeiru/ in (5a,b) since what we observe on spectrograms is continuous transition of vocalic formants from /e/ to /i/. Secondly, and more importantly, the lengths of auxiliary verbs are different in terms of the number of their component segments. Although they are of the same length from a moraic point of view, there is in fact substantial difference in the acoustic duration of /mi'ru/ and /iru/. These points can be critical for the purpose of the present experiment because different acoustic segmentations result directly in different intercept values which play a central role in the data analysis. Since there could be

[11] It is probable that some natives 'feel' an accent before the complementizer /n/ in (4a). In this case we have *na'ni-o no'n-de-mi'ru'-n-desu-ka* instead of (4a') above.

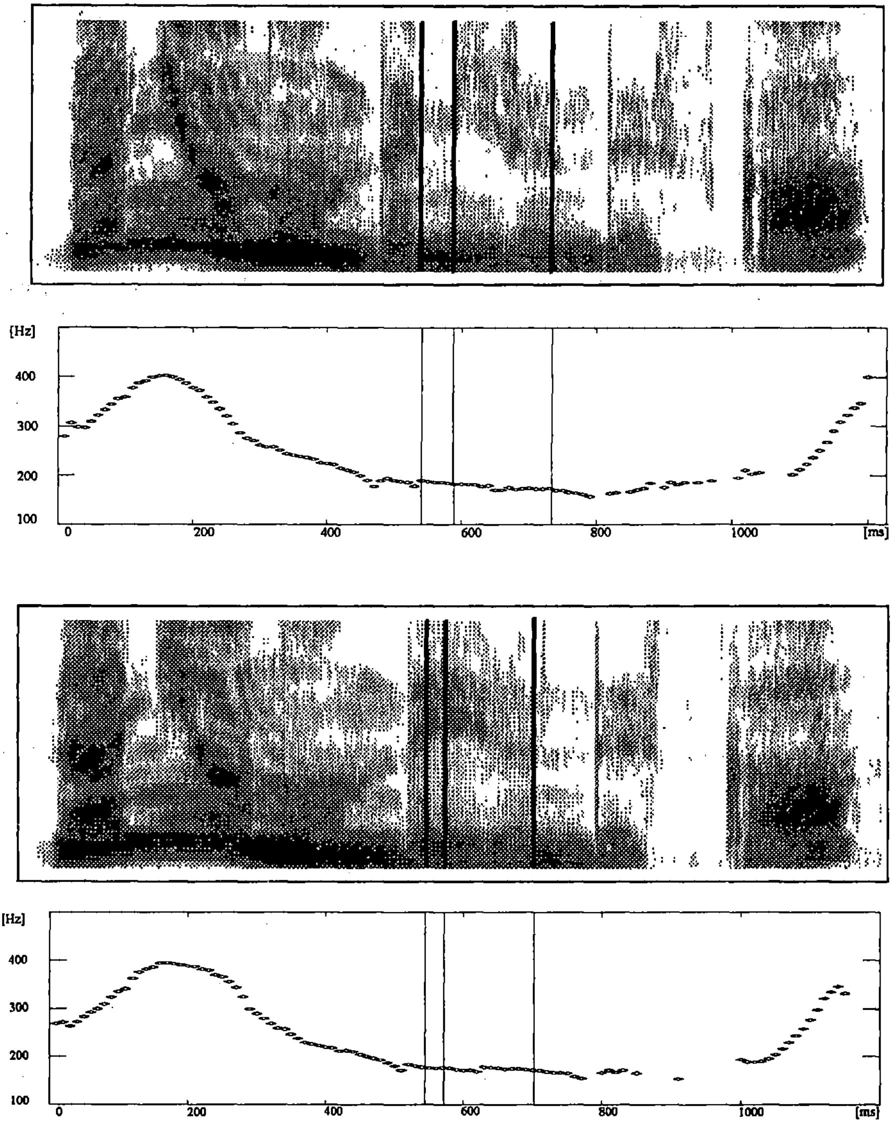


Figure 6. Wide-band spectrograms (4kHz) and f0 contours of (4a) and (5a) by YO. The upper two panels are for (4a). Vertical lines correspond to the beginning of /m/, the end of /m/ and the end of /u/ from left to right. The lower panels are for (5a). Lines correspond to the mid point of F2 transition, the time of the highest F2 and the end of the vowel /u/. These criteria were applied to (4b) and (5b), too.

no perfect solution to these problems, I had to make do with applying several distinct segmentation criteria for each sentence. For /mi'ru/, lines were fitted for both the entire region of /miru/ and its sub-portion omitting the initial consonant /m/. Similarly, for /iru/, the boundary between /e/ of gerundial /de/ and /i/ of the auxiliary verb was defined either as the time point where the F2 transition reaches its maximum or as the mid point of the F2 transition from /e/ to /i/.[12] The acoustic duration of /iru/ given by the latter segmentation is relatively longer than the one given by the former definition. The following abbreviation will be used in the rest of the paper to refer to these segmentation criteria. Segmentation of /mi'ru/: Entire Region = ER, Excluding /m/ = EM. Segmentation of /iru/: Time of Highest F2 = HF, Mid Point of transition = MP. Figure 6 shows example segmentations

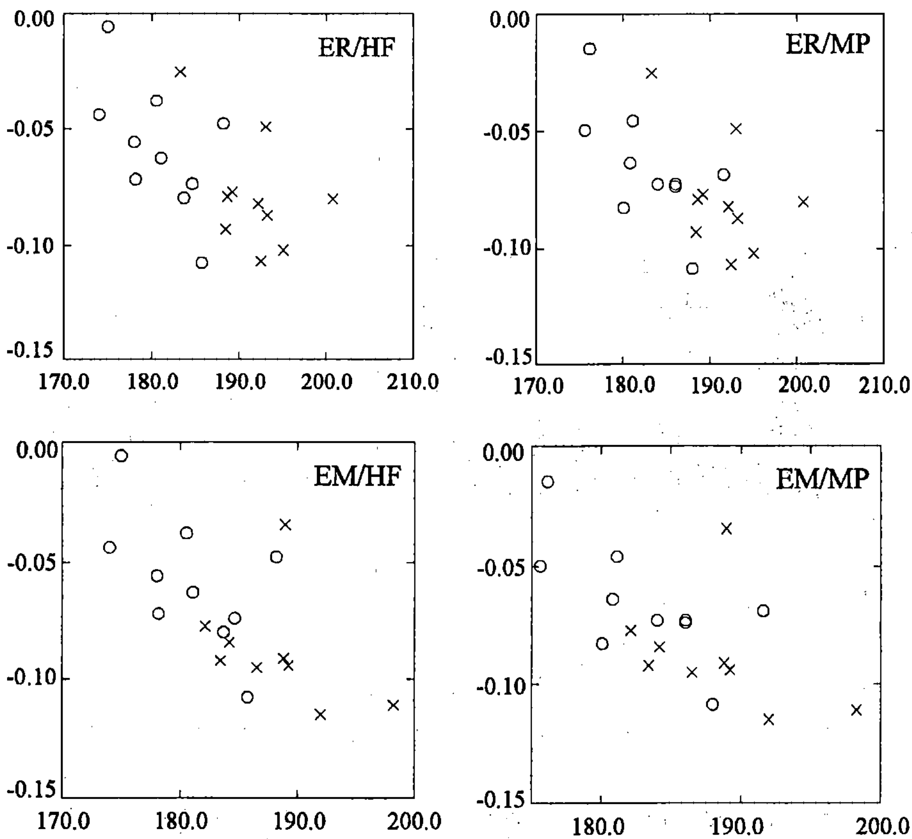


Figure 7. Scatter plot of lines fitted to auxiliary verbs of (4a) and (5a) uttered by YO. The abscissa shows intercept value and the ordinate shows slope. The combination of segmentation criteria is shown on upper right corner of each panel. Cross and circle plot symbols are used for accented and unaccented data respectively.

[12] The right edge of auxiliary verbs could be invariably defined as that of the vowel /u/.

based on these criteria. Note that given the monotonically decreasing f0 contours, it is predicted that the combination EM/MP disfavors the separation the most in terms of the intercept value since the left edge of /mi'ru/ assigned by criterion EM is relatively late in time (when compared to the edge obtained by ER) and the left edge of /iru/ assigned by MP is relatively early (when compared to HF).

Figure 7 shows the intercept vs. slope scatter plots of lines fitted to the auxiliary verbs in (4a) and (5a) uttered by speaker YO. Figure 8 shows the same for the auxiliary verbs in (4b) and (5b) uttered by the same speaker. Each panel in the figure stands for a specific combination of segmentation criteria. As can be seen from the figure, the separability of accented and unaccented data clouds differs according to the combinations of segmentation criteria. The same analysis was carried out for the utterances of speaker KM. As for speaker NF, only the auxiliary verbs in (4a) and (5a) were analyzed since in her speech phonetic realization of the auxiliary verbs in (4b) and (5b) showed considerable amount of variation through Kubozono's types A to C.

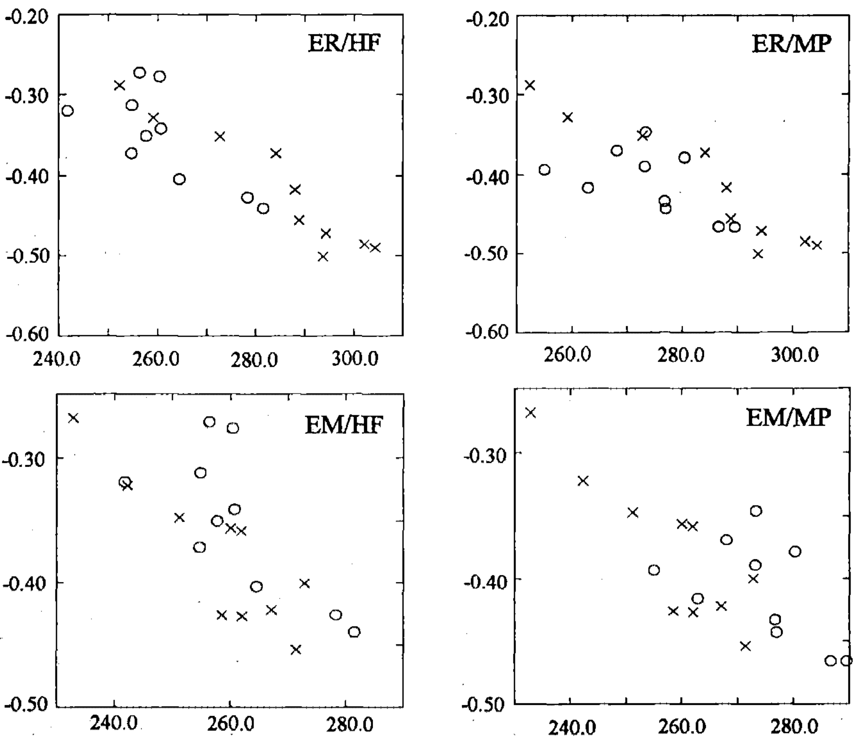


Figure 8. Scatter plots of lines fitted to the auxiliary verbs of (4b) and (5b) uttered by YO. Same manner of plotting as in Figure 7.

Table 6: Statistics of auxiliary verbs in (4a,b) and (5a,b) by Speaker YO.

Sentence	Aux. verb	Criterion	N	Mean (SD) of intercept	Mean (SD) of slope
4a	/ mi'ru /	ER	9	192.540 (3.838)	-0.084 (0.017)
4a	/ mi'ru /	EM	9	188.163 (4.957)	-0.088 (0.024)
5a	/ iru /	HF	10	180.900 (4.670)	-0.058 (0.028)
5a	/ iru /	MP	10	182.949 (5.128)	-0.065 (0.025)
4b	/ mi'ru /	ER	10	283.991 (17.447)	-0.416 (0.076)
4b	/ mi'ru /	EM	10	258.049 (12.645)	-0.378 (0.058)
5b	/ iru /	HF	10	261.094 (11.609)	-0.350 (0.059)
5b	/ iru /	MP	10	274.276 (10.409)	-0.409 (0.041)

Table 7: Statistical tests of the data in Table 6. Speaker YO.
t-tests are two-tailed.*

Sentences	Criteria	T^2	F (DF) P	*t*-test on intercept	*t*-test on slope
4a / 5a	ER/HF	1.545	F=13.129 (DF 2,16) P=0.001	t=5.140 (DF 18.0) P=0.000	t=-1.731 (DF 17.7) P=0.101
4a / 5a	ER/MP	0.363	F= 3.089 (DF 2,16) P=0.003	t=3.958 (DF 17.8) P=0.001	t=-1.223 (DF 18.0) P=0.237
4a / 5a	EM/HF	0.706	F= 5.651 (DF 2,16) P=0.005	t=3.227 (DF 16.5) P=0.005	t=-2.576 (DF 17.0) P=0.020
4a / 5a	EM/MP	0.366	F= 2.927 (DF 2,16) P=0.083	t=2.252 (DF 16.9) P=0.038	t=-2.188 (DF 16.9) P=0.049
4b /5b	ER/HF	0.825	F= 7.013 (DF 2,17) P=0.006	t=3.455 (DF 15.7) P=0.003	t=-2.164 (DF 16.9) P=0.045
4b /5b	ER/MP	0.311	F= 2.814 (DF 2,17) P=0.088	t=1.512 (DF 14.7) P=0.152	t=-0.258 (DF 13.9) P=0.800
4b /5b	EM/HF	0.372	F= 3.164 (DF 2,17) P=0.068	t=-0.561 (DF 17.9) P=0.582	t=-1.077 (DF 18.0) P=0.295
4b /5b	EM/MP	0.658	F= 5.597 (DF 2,17) P=0.014	t=-3.133 (DF 17.4) P=0.006	t=1.374 (DF 16.4) P=0.188

*Shaded cells are significant at least at 0.05 level.

Table 8: Statistics of auxiliary verbs in (4a,b) and (5a,b) by speaker KM.

Sentence	Aux. verb	Criterion	N	Mean (SD) of intercept	Mean (SD) of slope
4a	/ mi'ru /	ER	10	104.632 (4.485)	-0.119 (0.025)
4a	/ mi'ru /	EM	10	97.265 (4.342)	-0.102 (0.028)
5a	/ iru /	HF	10	91.958 (2.618)	-0.090 (0.033)
5a	/ iru /	MP	10	97.342 (2.765)	-0.110 (0.028)
4b	/ mi'ru /	ER	10	148.322 (8.355)	-0.240 (0.043)
4b	/ mi'ru /	EM	10	130.282 (7.311)	-0.182 (0.055)
5b	/ iru /	HF	10	133.108 (8.267)	-0.213 (0.072)
5b	/ iru /	MP	10	145.295 (10.401)	-0.266 (0.064)

Table 9: Statistical tests of the data in Table 8. Speaker KM.
t-tests are two-tailed.*

Sentences	Criteria	T^2	F (DF) P	*t*-test on intercept	*t*-test on slope
4a / 5a	ER/HF	3.313	F=28.157 (DF 2,17) P=0.000	t=7.717 (DF 14.5) P=0.000	t=-2.153 (DF 16.8) P=0.046
4a / 5a	ER/MP	1.092	F=9.281 (DF 2,17) P=0.002	t=4.357 (DF 15.0) P=0.001	t=-0.764 (DF 17.8) P=0.455
4a / 5a	EM/HF	0.614	F=5.219 (DF 2,17) P=0.017	t=3.310 (DF 14.8) P=0.005	t=-0.882 (DF 17.4) P=0.390
4a / 5a	EM/MP	0.019	F=0.159 (DF 2,17) P=0.854	t=-0.854 (DF 15.3) P=0.963	t=0.572 (DF 18.0) P=0.574
4b /5b	ER/HF	1.723	F=14.645 (DF 2,17) P=0.000	t=4.093 (DF 18.0) P=0.001	t=-1.039 (DF 14.6) P=0.316
4b /5b	ER/MP	0.516	F=4.383 (DF 2,17) P=0.029	t=0.718 (DF 17.2) P=0.483	t=1.075 (DF 15.8) P=0.298
4b /5b	EM/HF	0.062	F=0.528 (DF 2,17) P=0.599	t=0.810 (DF 17.7) P=0.429	t=0.810 (DF 16.8) P=0.305
4b /5b	EM/MP	0.783	F=6.656 (DF 2,17) P=0.007	t=3.734 (DF 16.1) P=0.002	t=-3.162 (DF 17.7) P=0.005

*Shaded cells are significant at least at 0.05 level.

Table 10: Statistics of auxiliary verbs in (4b) and (5b) by speaker NF.

Sentence	Aux. verb	Criterion	N	Mean (SD) of intercept	Mean (SD) of slope
4b	/ mi'ru /	ER	10	347.872 (44.234)	-0.665 (0.131)
4b	/ mi'ru /	EM	10	310.807 (42.212)	-0.626 (0.161)
5b	/ iru /	HF	10	327.809 (39.887)	-0.739 (0.207)
5b	/ iru /	MP	10	339.522 (53.338)	-0.749 (0.194)

Table 11: Statistical tests of the data in Table 10. Speaker NF. t-tests are two-tailed.*

Sentences	Criteria	T^2	F (DF) P	*t*-test on intercept	*t*-test on slope
4b / 5b	ER/HF	0.979	F=8.318 (DF 2,17) P=0.003	t=1.065 (DF 17.8) P=0.301	t=0.952 (DF 15.2) P=0.356
4b / 5b	ER/MP	0.999	F=8.490 (DF 2,17) P=0.003	t=0.381 (DF 17.4) P=0.708	t=1.128 (DF 15.8) P=0.276
4b / 5b	EM/HF	0.145	F=1.231 (DF 2,17) P=0.317	t=-0.926 (DF 17.9) P=0.367	t=1.354 (DF 17.0) P=0.194
4b / 5b	EM/MP	0.141	F=1.198 (DF 2,17) P=0.326	t=-1.335 (DF 17.1) P=0.199	t=1.531 (DF 17.4) P=0.144

* Shaded cells are significant at least at 0.05 level.

Tables 6 and 7 show the results of statistical analysis of the utterances of speaker YO. In six out of eight combinations of criteria, statistical significance of at least at the 0.05 level is obtained at least in more than two statistical tests, the remaining two cases (ER/MP and EM/HF of (4b)/(5b)) are significant at the level of 0.1 in terms of their T^2 values. However, it should be pointed out that in the case of EM/MP of (4b)/(5b), the mean intercept value of the unaccented data is greater than that of the accented data. This is the reverse of what was observed in the previous experiment for all three speakers. This reversal could happen when the left edge of /iru/ assigned by the criterion MP is too early on the time axis and is penetrating into the time region of the accentual fall caused by the accent of *no'nde*. In fact, the temporal location of the accentual L of /no'nde/ seems to coincide sometimes with the beginning of /r/ of /iru/.

The result of speaker KM shown in Tables 8 and 9 is basically the same as that for YO. Six out of eight combinations were significant at least at the 0.05 level in terms of their T^2 values. The remaining cases, EM/MP of (4a)/(5a) and EM/HF of (4b)/(5b), are not significant in any of the three tests at any level, however. The segmentation by the criteria EM/MP causes the same reversal of parameter values as in YO's utterances, but here it is not only the relation between (4b) and (5b) but also that of (4a) and (5a) that is reversed. The result of speaker NF is shown in

Tables 10 and 11. Here again, the combination of EM/MR caused the same reversal, although in her speech the measurement provided by this combination did not show significance at any level.

3. Discussion and concluding remarks

Concerning the 'dephrasing' in post-focus position, there seems to be almost no room for controversy about the detectability of the accentedness of seemingly 'dephrased' accentual phrases. Nearly perfect separability of data clouds on the slope vs. intercept plane shown in Figure 5 provides strong support for the interpretation that the f0 contour of accented accentual phrase like the one shown in the upper panel of Figure 3 above is not an instance of dephrasing. Therefore the two contours of Figure 3 can not be identical from a phonological point of view. No matter how its physical manifestation differs from the canonical form, the accent of the seemingly 'dephrased' accented accentual phrase is still there. In this sense, it is appropriate to call this kind of contour a *degenerate* accentual phrase and distinguish it from the 'true' dephrasing in which the accentual contrast can not be detected anymore. In this respect, the intuitive judgment that the accent in the predicate of sentence (1c) can be perceived even when the predicate is associated with linear f0 contour is not surprising at all, since the linear synthetic intonation used in Maekawa (1991) had a relatively high intercept value and showed relatively steep declination. Perception of the accent in the linear f0 contour should not be treated as a perceptual illusion due to speakers' internalized knowledge of the accentedness.

When compared to the clarity in the result of the first experiment, the result of the second experiment leaves much room for controversy. However, the results of statistical tests showed arguably that it was possible to detect the accentedness of auxiliary verbs at least in the speech of speaker YO, and therefore this speaker showed no 'dephrasing' of auxiliary verbs. As for the other two speakers, although it is possible to reject the 'dephrasing' view based on the results of statistical tests which showed significant differences in half of the case examined, it would be safer to eschew any decisive conclusion at this time. (The data set of an experiment currently underway involves sentences similar to (4) and (5) in which the unaccented auxiliary verb *yaru* ('do something for someone') is used instead of *mi'ru* and *iru*. Comparison of *mi'ru* and *yaru* would contribute to the clarification of the problem since the segmental lengths of the two auxiliary verbs are the same. But in this case, we must face the difference in vocalic intrinsic pitch caused by the difference of the vowel height of the auxiliary verbs; this was why I analyzed the pair *mi'ru* and *iru* in the previous experiment.)

To sum up, although the experimental evidence presented in this paper should be reinforced by analyzing a wider range of data and more informants, it suggests strongly the view that contours hitherto regarded as 'dephrased' are not dephrased. At the same time, the data presented in this paper show convincingly that it is risky to rely solely upon the visual, hence subjective, inspection of f0 data, especially the visual detection of the peak features.

Finally, I would like to point out two issues left unresolved by the present study. First, it remains as an open question how to treat the 'dephrasing' of unaccented accentual phrases, i.e. the treatment of the f0 contours like the ones shown in the lower panels of Figure 3 or Figure 6.[13] From a theoretical point of view, to regard these contours as dephrased causes less problem than the dephrasing of an accented accentual phrase because the dephrasing of an unaccented accentual phrase does not involve accent deletion. Intuitively, however, I feel that I perceive the existence of an accentual phrase boundary when the slope value of the quasi-linear —hence seemingly 'dephrased'— unaccented accentual phrase is very close to or higher than zero. In the course of the present study, I have tried to screen out the true cases of dephrasing by examining the normal probability plot of extracted slope values of sentences (3a,b) uttered by KM; his slope values seemed to be split into two distinct distributions on both sides of the zero point. But I could not arrive at any clear conclusion due mainly to the limitation of the number of data. I think, however, that this method is worth further effort.

The second problem is a more important one. At present it is unclear whether degenerate accented accentual phrases have the same phonological representation as their fully realized counterparts. It seems to me that there is no *a priori* reason to suppose that degeneration of accentual phrase is a pure phonetic process as supposed by Kubozono. As far as I am aware of, one way to examine this problem involves the quantitative comparison between the effect of downstep caused by a fully realized accentual phrase and the one caused by a degenerate accentual phrase. Fragmentary analysis of relevant material contained in the current data set suggests that there is difference between the two types of accentual phrases. This problem should be discussed in a separate paper.

References

Beckman, M. and Pierrehumbert, J. (1986) Intonational structure in Japanese and English. *Phonology Yearbook*, **3**, 255-309.

Fujimura, O. (1991) Remarks on phrasing and prosodic attachment. In *Interdisciplinary approaches to language* (C. Georgopoulos & R. Ishihara, editors), pp.207- 216, Netherlands, Kluwer Academic Publishers.

Green P.E.(1978) *Analyzing multivariate data*. Hinsdale, IL: The Dryden Press.

Kubozono, H. (1993) *The organization of Japanese prosody*. Tokyo: Kuroshio Publishers.

Maekawa, K. (1989) Statistical tests for the study of vowel merger. *Quantitative Linguistics*, **39**, 200-219, Bochm.

Maekawa, K. (1991) Perception of intonational characteristics of WH and NON-WH questions in Tokyo Japanese. In *Proceedings 12th International Congress on Phonetic Sciences*, Aix-en-Provence, **4**, 202-205.

McCawley, J. (1968) *The phonological component of a grammar of Japanese*. The Hague: Mouton.

[13] This problem was pointed out to me by Jennifer Venditti (personal communication).

Pierrehumbert, J. (1980) *The phonology and phonetics of English intonation*. Ph.D. dissertation, MIT.
Pierrehumbert J. and M. Beckman, (1988). *Japanese tone structure*. Linguistic Inquiry Monograph Series, 15, Cambridge, MA: MIT Press.
Poser, W. (1984) *The phonetics and phonology of tone and intonation in Japanese*. Ph.D. dissertation, MIT.
Wada, M. (1969) Ji-no akusento (Accent of particles), *Kokugokenkyuu*, **29**.