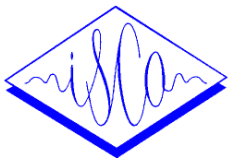

First International Conference on
Spoken Language Processing (ICSLP 90)
Kobe, Japan
November 18-22, 1990

PRODUCTION AND PERCEPTION OF THE ACCENT IN THE CONSECUTIVELY DEVOICED SYLLABLES IN TOKYO JAPANESE

Kikuo Maekawa

National Language Research Institute
Kita-ku Nishigaoka, Tokyo, 115 Japan

ABSTRACT

What happens to the phonological pitch accent of Tokyo Japanese when the accented syllable and all other syllables in the word are devoiced by the effect of phonetical vowel devoicing rule? The acoustic analysis and perception tests reveals that the accentedness of the devoiced syllable is perceived by the elevated pitch at the beginning of the voiced syllable immediately following the sequence of the devoiced syllables. The exact location of the accent in the devoiced syllables, however, can't be identified correctly.

I. INTRODUCTION

Tokyo Japanese is a language which has the system of phonological pitch accent or tone; also it possesses the phonetical rule of vowel devoicing. As the result,the language provides an interesting case of the interaction between phonology and phonetics.

Some descriptive studies of Tokyo Japanese note that the accent shifts to the following syllable when the accented syllable is devoiced. However, it is known that in the younger speakers' utterances the accent shift of the devoiced syllable does not occur[1].

The mechanism whereby we perceive the accent in a devoiced syllable has been extensively examined by Sugito[2][3]. Her study shows that both in Osaka and Tokyo the accentedness of a devoiced syllable is perceived not due to the acoustical characteristics of the devoiced segments --namely their intensity-- but due to the compensatory pitch modification in the following syllable: an elevated pitch beginning followed by a sharp fall.

There are, however, cases where Sugito's theory can't apply. The vowel devoicing rule of Tokyo Japanese operates whenever a close vowel is preceded and followed by voiceless consonants thus, producing a sequence of devoiced syllables. In fact, /shukuFuku/ (congratulation) can be pronounced either as shkuFku or as shkFku or even as shkFk[4].These pronunciations are not at all exceptional in the natural speech of present day Tokyo Japanese. The aim of this paper is to examine the mechanism of the realization of accent in the consecutively devoiced syllables.

II. PRODUCTION EXPERIMENT

2.1 Material

The production experiment is conducted using the nonsense phonemic sequences of /Fusitu/ and /Fuzizu/. Hereafter these sequences will be notated as /FST/ and /FZZ/.(Attentions should be paid to the facts that in Japanese phonemic sequence /si/ is phonetically realized as shi and /tu/ is realized as tsu. Also /zu/, the voiced counterpart of /tu/, is realized either as zu or as dzu.) All the three syllables of /FST/ can be devoiced when it is uttered in the carrier sentence "kore-o ...to ii-masu" (We call this ...). /FZZ/ is used for the comparison purposes.

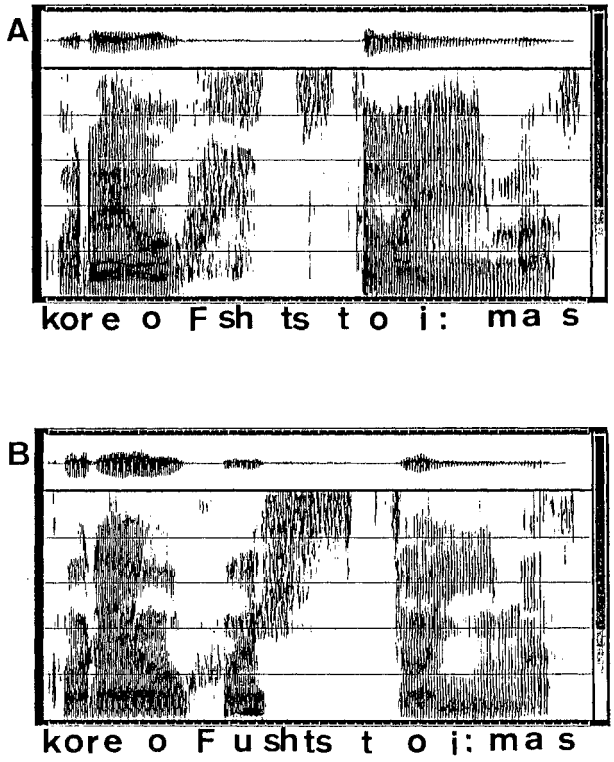

Figure 1 Wide band spectrograms (5KHz band) of the consecutive devoicing of test word F'ST. Panel A is an utterance by YK where all three syllables were devoiced. Panel B is an utterance by KMt where only second and third syllables were devoiced.

Four differently accented test words including the unaccented one can be derived from /FST/ and /FZZ/. In order to refer to each of the eight test words, the notations like FST, FZ'Z or FST' will be used in the following discussion: these stand respectively for the unaccented version of /FST/, the version of /FZZ/ with accented second syllable and the version of /FST/ with accented third syllable. Each of the eight test words thus derived are recorded twelve times in a sound proof room.

Two male native speakers of Tokyo Japanese (YK 42 years old; KMt 28 years old) and a male speaker who is not Tokyo native but has spent most of his life in Tokyo since he was 14(TA 33 years old) took part in the experiment. YK and KMt seem to have felt much difficulty in producing the accent on the third syllable.

The 288 recorded utterances(8 words x 12 repetitions x 3 speakers) were sampled at 10KHz and digitized with 16 bits accuracy. They are then fed to the LPC f0 tracking program by Imagawa and Kiritani[5]. In the following discussion, we will specifically refer to the f0 values measured at two distinct time points of the f0 contour: the peak f0 value at the beginning of the /to/ syllable in the carrier sentence(P) and the f0 value at the end of the whole sentence(V). (See figure 2.)

2.2 Relation between Accent and P-value

As for the utterances made by YK and TA, all phonological syllables of /FST/ were completely devoiced (See panel A of fig.1). In the utterances of KMt,however, the first syllable of /FST/ was not devoiced in most of the utterances(See panel B of fig.1). Complete devoicing over the three syllables happened three times in FS'T utterances and once in FST' utterances.

Table 1 shows the mean P-values as the function of the accent location. The P-values of the test words from /FST/ sequence are significantly higher than their /FZZ/ counterparts when the accent is located in the first or the second syllable. This elevation of P-value can be considered to have been caused by the devoicing of the accented syllable, because no significant difference is detected between the mean P-value of /FST/ and that of /FZZ/ when they were not accented.

There is another evidence to support this view: the difference between the mean P-value of F'ST and that of F'ZZ by KMt, who did not devoice the accented syllable of F'ST,is not significant. Panel B of figure 2 is an example of the f0 contour of F'ST uttered by KMt. The lack of the compensatory P-elevation is evident when compared to the contour by YK in panel A.

Finally there is the tendency that mean P-value of the accented /FST/ words becomes higher as the accent moves rightward. This suggests the possibility that P-value is the phonetical cue to the perception of the accent location. We will examine this possibility in Section 3.1.

2.3 Relation between Accent and V-value

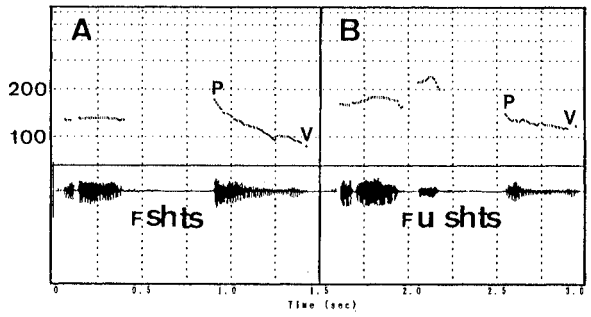

Figure 2 F0 contours of test word F'ST with and without the devoicing of accented syllable. Panel A and B correspond to the same panels in figure 1. Frequency scaling is logarithmic.

SPEAKER	ACCENT LOCATION IN TEST WORDS	MEAN P-VALUE±S.D. WITH/FST/, WITH/FZZ/		t-TEST (DEGR. FREEDOM)
YK	Unaccented	154± 6.9	151± 7.0	t= 1.14(22) ns
	First Syllable	176±10.1	126± 5.1	t=14.50(22)***
	Second Syllable	195± 7.2	143±12.9	t=12.17(22)***
	Third Syllable	200± 8.4	198±19.6	t= 0.22(22)***
KMt	Unaccented	176± 8.1	177± 9.8	t= 0.37(22) ns
	First Syllable	144± 8.4	142± 9.4	t= 0.41(22) ns
	Second Syllable	171±12.4	146± 5.5	t= 6.46(22)***
	Third Syllable	177±17.5	180±11.2	t= 0.54(22) ns
TA	Unaccented	124± 2.6	126± 3.2	t= 1.80(22) ns
	First Syllable	121± 4.8	112±10.2	t= 2.62(22) *
	Second Syllable	157± 8.2	112± 6.5	t=14.71(22)***
	Third Syllable	157±10.9	130± 6.0	t= 7.50(22)***

Table 1 Relation of mean P-value and devoicing. Differences between mean P-values of /FST/ words and that of /FZZ/ words were tested. Significance of t is indicated by *** p<.001; ** p<.01; * p<.05; ns p>.05.

SPEAKER	WORDS	MEAN V-VALUE±S.D. (N.CASE) UNACCENTED(N),	ACCENTED(N)	t-TEST (DEGR. FREEDOM)
YK	/FST/	110±3.5(12)	94±2.2 (36)	t=19.06(46)***
	/FZZ/	101±4.5(12)	93±5.6 (36)	t= 4.34(46)***
KMt	/FST/	122±7.4(12)	103±10.0(33)	t= 6.14(43)***
	/FZZ/	147±5.7(12)	101±12.0(36)	t=12.86(46)***
TA	/FST/	104±2.8(12)	70±6.0 (35)	t=26.40(41)***
	/FZZ/	99±4.2(12)	67±7.9 (23)	t=12.74(33)***

Table 2 Relation of V-value and accentedness of the utterance. Sometimes the v-values could not be measured because of heavy creaky voice.

Table 2 shows the mean V-value as the function of the accentedness of the sentence. The theory of catathesis proposed by Pierrehumbert & Beckmann predicts that every tonal peak becomes lower to the right of the phonological accent[6]. This prediction is congruent with the results obtained here. V-value is significantly lower in the accented sentence than in the unaccented sentence even if the accented syllable is phonetically voiceless.

III. PERCEPTION EXPERIMENTS

3.1 Perception of Accent Location

In the following sections the perceptual aspect of the problem will be examined. The first question to be answered is whether the Tokyo Japanese speakers can perceive the location of the accent in the consecutively devoiced syllables.

A representative utterance of FST, F'ST, FS'T and FST' were selected from the utterances of YK. Each of them were presented ten times in random order to five listeners. The five listeners were Tokyo native linguists who knew nothing about the aim of the experiment. They were instructed to judge the accentedness of the utterance and, if the utterance was perceived as accented, to estimate the location of the accent in the test word.

Table 3 shows the pooled responses by the five listeners. Listeners perceived correctly the accentedness even when the accented syllables were phonetically voiceless. However,the estimated accent location of the accented utterances was incorrect; a clear tendency to perceive the accent on the third syllable can be discerned. The conclusion drawn from this is that systematic variation of P-value in table 1 can't be the phonetical cue to the perception of the accent location.

3.2 Phonetical Cue for Accentedness

The last experiment is concerned with the phonetical cues for the accentedness. There are two possibilities: the accentedness is perceived either by the pitch value at the time point P or by the sentence final pitch value V. The relative contributions of P-value and V-value were evaluated by splicing the utterances used in the previous experiment at five different time points.(See figure 3):

A: kore-o Fusitu#
B: kore-o Fusitu-to#
C: kore-o Fusitu-to i#
D: kore-o Fusitu-to ii#
E: kore-o Fusitu-to ii-masu#

Time point A excludes both P and V. Time points B to D include P but not V. Time point E include both P and V. (Note because the sentence final syllable /su/ is regularly devoiced in Tokyo Japanese, the V value we have been dealing with is the value at the end of /ma/.) The splicing of the utterance was carried out digitally on the computer.

The twenty stimuli(4 test words x 5 splicing points) were presented ten times in random order to the same five listeners. They were instructed to judge whether the stimulus they heard was accented or unaccented. Their pooled responses are shown in Table 4. Each cell

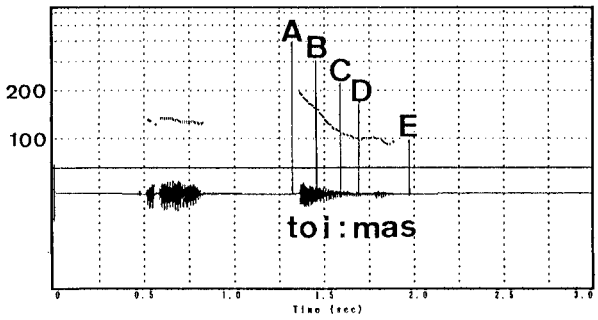

Figure 3 Location of the splicing time points. F0 contour of the FST' sample by YK is used as example.

STIMULUS PRESENTED	PERCEIVED AS FST	F'ST	FS'T	FST'
FST	47	0	2	1
F'ST	4	10	13	23
FS'T	0	2	10	38
FST'	1	3	12	34

Table 3 Perceived accent location in the consecutively devoiced syllables. Pooled responses of five listeners. Each stimulus was presented ten times in random order.

STIMULUS PRESENTED	SPLICING TIME POINTS A	B	C	D	E
FST	42	42	35	40	43
F'ST	7	39	46	47	47
FS'T	8	44	45	48	48
FST'	6	48	47	48	48

Table 4 Perceived accentedness of the spliced speech. Each cell represents the number of responses the accentedness of the presented stimulus is perceived correctly. Pooled responses of five listeners. Each stimulus was presented ten times in random order.

of the table represents the number of correct responses i.e. the number of cases where the accentedness of the presented stimulus was judged correctly.

The stimuli made of unaccented test word FST were judged correctly as unaccented irrespective of the splicing time points. On the other hand,the perception of the accentedness of accented test words was influenced by the splicing time points. They were perceived as unaccented when there was no following voiced segments (point A). This is congruent with Sugito's view that the acoustic characteristics of the voiceless segments, voiceless vowels and/or consonants, is not the cue for the perception of accentedness.

When P-point (syllable /to/) was included, the percentage of correct responses went up to about 80% or more(Point B). Extra sequence of voiced syllables increases the correct response to some extent in the cases of F'ST and FS'T. However, the number of correct responses did not change in the case of FST' because no room was left for the contribution from the extra syllables(Points C,D and E).

These findings confirm the view that P-value is the primary phonetical cue for accentedness, and the contribution of V-value can be neglected. Although the contribution of sharp f0 fall after P-point can not be neglected,it isn't a primary cue neither, since the contribution of extra voiced syllables is smaller than that of the P-value. Accentedness of the utterance containing accented consecutively devoiced syllable is perceived by the pitch elevation at the beginning of the first voiced syllable after the voiceless segments.

IV. CONCLUDING REMARKS

The experiments reported in this paper have revealed that the accent of the devoiced syllable can be acoustically realized through the compensatory local pitch elevation in the following voiced syllable even when the voiced syllable is pulled apart by the intervening devoiced syllables. However the exact location of the accent in the sequence of the devoiced syllables can not be perceived correctly. In one words, part of the linguistic information contained in the accent can be lost under the effect of devoicing. This discovery will provide an interesting topic for the discussion of the constraints of consecutive devoicing.

The important problem of how these facts are to be formalized in the rules of Tokyo Japanese phonology and/or phonetics is left untouched in this paper. This is to be done in future studies, though it won't be an easy task at all. The formulation should be able to distinguish, among others, older speakers' accent shifting type devoicing from the accent preserving type devoicing of younger speakers,which has been discussed in this paper. What follows is a rough sketch of a possible formulation.

In accent shifting devoicing, the H tone of the combined HL accent tone is shifted one syllable to the right of the devoiced accented syllable,and the L tone of HL--a kind of floating tone-- is realized at the right edge of the syllable next to the newly accented syllable. In accent preserving devoicing,on the other hand, the accent H is moved rightward to the first voiced syllable, and the L is realized at the right edge of the same syllable, thus making the syllable phonetically falling. An implication of this crude formulation is that the compensation of accent in devoiced syllable can't be a purely phonetical process. Rather,it seems to concern with some kind of tonal relinking occurring near the surface. The elevation of P-value, on the other hand, will be interpreted to be caused by a phonetical rule operating specifically upon the syllable linked with accent H and accompanying an L on its right edge on time scale. The adequacy of this hypothesis should be examined by further studies, needless to say.

ACKNOWLEDGMENTS

The author is very grateful to Takako Ayusawa of NLRI for her comments on an earlier version of this paper. Part of this study was supported by the Grant-in-Aid for Scientific Research on Priority Areas, the Ministry of Education,Culture and Science, Japan.

REFERENCES

[1]K.Akinaga,"On'into akusentotono kankeino hoosoku," Appendix to Meikai nihongo akusento jiten. Tokyo: Sanseido, 1981.

[2]M.Sugito,Nihongo akusento no kenkyuu.Tokyo: Sanseidoo,1982.

[3]M.Sugito and H.Hirose "Production and perception of accented devoiced vowels in Japanese," Ann. Bull.RILP 22,Univ.Tokyo(1988).

[4]K.Maekawa,"Boinno museika," in Nihongono onsei on'in:joo, Vol.2 of Kooza nihongoto nihongo-kyooiku. Tokyo:Meijishoin, 1989.

[5]H.Imagawa and S.Kiritani,"High-speed speech analysis system using a personal computer with DSP and its applications to pronunciation training," Ann.Bull.RILP,23,Univ.Tokyo(1989).

[6]J.B.Pierrehumbert and M.Beckman,Japanese tone structure, MIT Press,1988.