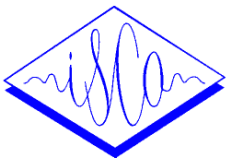

2nd International Conference on
Spoken Language Processing (ICSLP 92)
Banff, Alberta, Canada
October 12-16, 1992

ACCEPTABILITY AND DISCRIMINATION THRESHOLD FOR DISTORTION OF SEGMENTAL DURATION IN JAPANESE WORDS

Hiroaki KATO*, Minoru TSUZAKI*, and Yoshinori SAGISAKA**

*ATR Human Infomation Processing Research Laboratories
**ATR Interpreting Telephony Research Laboratories
Hikaridai, Seikacho, Kyoto, 619-02 Japan
E-mail:kato@atr-hr.atr.co.jp

ABSTRACT

Acceptability of temporal naturalness and temporal discrimination threshold were measured for various vowel segments in isolated words by modifying original segmental durations. A large size perceptual experiment using 1462 stimuli of 70 segments revealed that word acceptability is affected by the segment attributes and context such as vowel color, position in a word and, accent. An additional experiment showed that the above factors also affect discrimination threshold consistently.

10.21437/ICSLP.1992-163

I. INTRODUCTION

Acoustical characteristics of Japanese segmental duration have been extensively studied in the field of speech synthesis [1,2]. In these studies, the effect of the following control factors have been quantitatively confirmed: 1) vowel color, 2) adjacent phonemes, 3) position in a word, 4) mora count in a word, and 5) speaking rate. Although several pioneering studies have addressed the issue of perceptual duration for the speech segment, they used small sets of stimuli to observe perceptual characteristics.

Sato[3] showed that the acceptable deviation of vowel length in the first moraic segment was smaller than that of the other moraic segments using four words which started with the same phoneme sequence. (/saka/, /sakana/, /sakanaya/, /sakanaya-san/) Klatt[4] also reported that the positional factor would affect perceptual sensitivity to the durational modification for the segment, /deal/ , which appeared at different positions among seven sentences.

In this study, to investigate whether the positional effect is robust enough or not, we measured the perceptual sensitivity for the durational modification with a large set of samples. Through such an experiment, we can test whether the positional effect can be observed across several contexts and whether there are interactions with other factors. We included the factor of vowel color which was combined with the positional effect in a factorial way because it is a main factor found in acoustical analysis. It enables us to enlarge the stimulus variation enough for reliable analysis. We also included the factors of pitch accent and F0 contour which possibly influence a perceptual sensitivity for temporal aspect of speech, although they have not been considered as the control factors of segmental duration in Japanese.

II. ACCEPTABILITY

The main purpose of this experiment is to investigate the effect of vowel color and position in a word for the perceptual sensitivity using a large set of samples. We adopted the magnitude estimation of acceptability for the durational modification.

Although the factor of the pitch accent does not affect Japanese segmental duration control at all [2], it is an open question whether this factor would affect perceptual characteristics such as acceptability. Thus we included this factor in addition to the factor of vowel color and position in a word.

2.1 · Method

Design

According to the statistical analysis of acoustic characteristics, vowel /a/ is longer in duration than vowel /i/ and the vowel in the first mora is shorter than that in the third mora [5]. We chose /a/-/i/ contrast for the factor of vowel color and first-third contrast for the factor of position in a word. In addition, we chose contrast between the segment with high-tone and low-tone, for the factor of pitch accent.

Stimuli

The word samples were chosen from the large-scale Japanese speech database constructed at ATR [6]. These words consist of four mora words excluding the samples with vowel successions or geminated consonants which may affect the perception of temporal regularities observed in open syllable successions. We selected 70 segments from 63 sample words. Table 1 shows the distribution of the segment characteristics concerning the factors mentioned above. Figure 1 shows the distribution of vowel duration, which was defined by the manual labelling in the selected segments.

Table 1. The number of segments for each cell of the conditions in the experiment of acceptability.

	1st mora	3rd mora	total
/a/	21 (6)*	22 (15)	43 (21)
/i/	14 (0)	13 (13)	27 (13)
total	35 (6)	35 (28)	70 (34)

*Values in round brackets are numbers of high-tone segments.

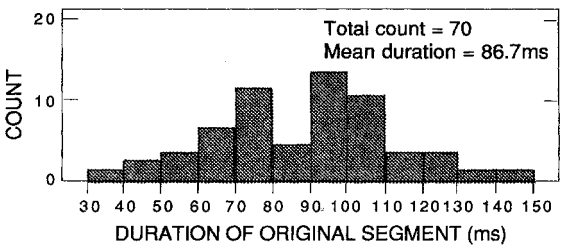

Figure 1. The distribution of duration of the original segments in the experiment of acceptability .

The word stimuli were synthesized using the selected word samples uttered by one male announcer by means of the LMA cepstral analysis synthesis technique [7]. The speech quality by this technique was natural enough for subjects to evaluate the acceptability. The duration of vowel portion in the target segments were varied from -50ms to +50ms in 5ms-step. In total, 1462 stimuli were prepared.

Procedure

Word acceptability for each stimulus was evaluated with the following procedure. Subjects were asked the score of acceptability for the modification in duration for each of randomly presented word stimulus. The acceptability scores corresponded to the following seven subjective categories.

3: very unnatural
2: unnatural
1: rather unnatural
0: undeclared
-1: rather natural
-2: natural
-3: very natural

Subject

Seven adult female subjects with normal hearing participated in the experiment.

2.2 Results and discussion

Feature parameters for acceptability

Mean evaluation scores over subjects were plotted as a function of change in duration of target segment for all stimuli (Figure 2a). A parabolic curve was chosen as a good approximation for the averaged scores as shown in Figure 2b. In this parabolic curve, the axis of the curve corresponds to the center of the acceptability range. The second order polynomial coefficient reflecting the sharpness of the curve corresponds to the acceptability range. This curve will be referred as "the acceptability curve" hereafter in this paper. An extreme example illustrating the difference of the acceptabilities are shown in Figure 3.

Axis shift of the acceptability curve

Because we adopted a synthesis technique based on the natural utterance, the original duration of the target segment was not equal. Thus we include the factor of original duration in the analysis. Four way analysis of variance was performed for the axis shift of the acceptability curve with vowel color, position, accent, and original duration as the main factors. Only the factor of original duration turned out to be significant [F(1,63)=16.6, p<0.0001]. The axis shift was correlated with the original duration as shown in Figure 4a. It means the longer the original duration , the shorter the preferred duration, and vice versa.

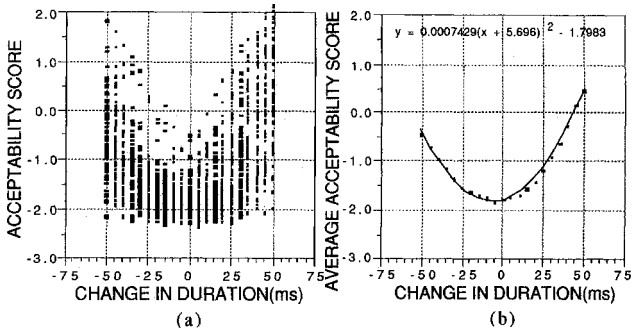

Figure 2. The acceptability score as a function of change in duration: (a) scores for each stimulus, (b) scatter plot of averaged scores and parabolic curve fitting.

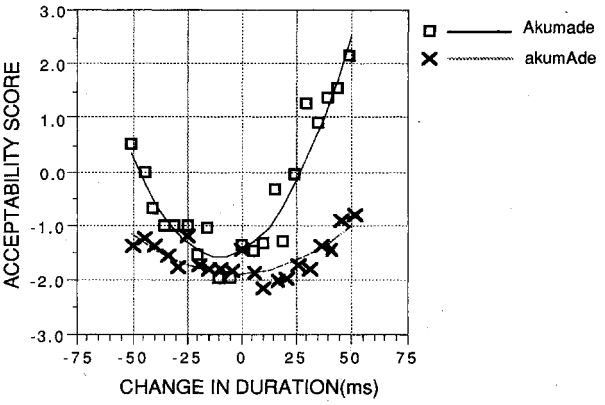

Figure 3. An example illustrating the difference of the acceptabilities. The acceptability scores and the acceptability curves for two segments in one word, i.e. the first high-tone segment and the third low-tone segment in the word "AkumAde" (to the bitter end)

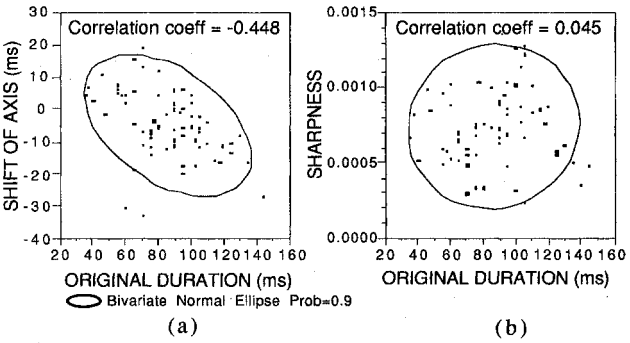

Figure 4. The effects of the original duration of segments: (a) a negative correlation with the shift of axis of the acceptability curve, (b) no correlation with the sharpness of the curve.

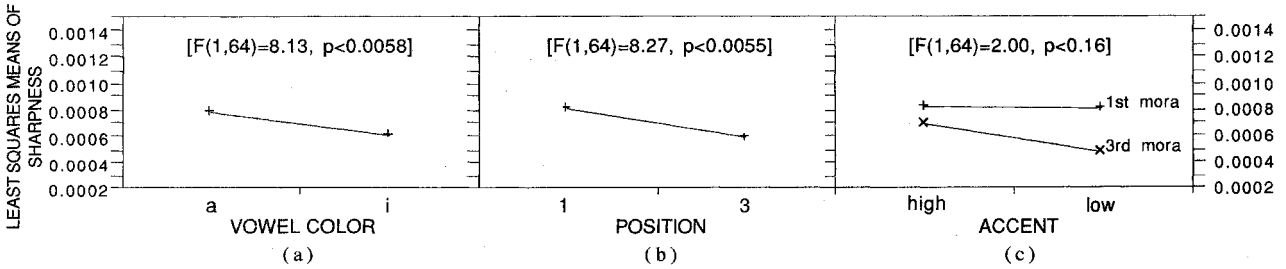

Figure 5. The effects on the sharpness of the acceptability curve by the three factors. The sharpness is higher (a) for the vowel /a/ than for /i/, (b) for the 1st mora than for the 3rd mora, (c) for the high-tone segment than for the low-tone segments in the 3rd mora.

Sharpness of the acceptability curve

As shown in Figure 5, panels a-c, the influences of the factors of vowel color, position, and accent were found on the sharpness of curve. Also the correlation coefficient between the original duration and the sharpness of curve was almost zero as shown in Figure 4b.

The factors of both vowel color and position in a word were statistically significant. The interaction between these two factors was not significant. As for the factor of accent, the tendency that the sharpness is higher for the high-tone segments than for the low-tone segments was observed only at the third moraic position. An ad hoc test for the subset constituted from the third mora segments was performed, which supported this tendency [$F(1,32)=4.06$, $p<0.0523$].

Estimation of acceptable range using the acceptability curve

Acceptability thresholds can be calculated from the acceptability curve if one value of the acceptability score could be given. The horizontal range of the curve at the value is the range between acceptability thresholds which can be estimated as the sum of the acceptability thresholds for the two direction i.e. lengthening and shortening. The axis of curve points to the center of the range between acceptability thresholds. The effects of the factors described in this study reflect the range between acceptability thresholds in proportion to square root of the sharpness.

The experiment for the acceptability showed two of the main factors based on acoustical analysis would also affect the perceptual evaluation. However this experiment is not enough to conclude that these two factors, the vowel color and the position, actually affect the perceptual "sensitivity" which would be measured by discrimination thresholds. Evaluation of acceptability may be influenced more by rather "cognitive" process than perceptual process. Thus, it is possible that these two factors play their roles only in the "cognitive" stage rather than in the "perceptual" stage.

In the second experiment, we measured the discrimination thresholds for the durational modification at two positions in two Japanese words.

III. DISCRIMINATION THRESHOLD

3.1 Method

Design

Adding to the factors of vowel color and position in a word, the factor of F0 contour was included. When a segmental duration is modified, the F0 contour of the segment is also modified. This difference could be a cue for discrimination. We chose a contrast between the segment with natural F0 and flattened F0 as the third factor.

Stimuli

Two words were chosen from the data base in the same manner as the first experiment. The following set of four segments were employed.

/i/ in the 1st and 3rd mora of "shinagire" (sold out).

/a/ in the 1st and 3rd mora of "nameraka" (state of being smooth).

The synthesis procedure was the same as in the first experiment except that the modification range is from -60ms to +60ms and the step is 2.5ms. Another set were synthesized as the stimuli with flattened F0 contour in which F0 values were fixed to the mean F0 of the target segments in the original utterances.

Table 2. Mean and standard deviation of the measured discrimination thresholds for each target segment.

Segment	Vowel	Position	F0	DT+ (ms)	DT- (ms)	DT (ms)
shInagire	/i/	1	natural	25.4(4.4)*	26.8(10.2)	52.2(7.9)
			flat	40.8(13.5)	30.4(6.7)	69.0(13.6)
shinagIre	/i/	3	natural	36.9(8.2)	30.7(14.5)	67.6(14.4)
			flat	40.5(12.8)	37.3(18.6)	77.8(20.5)
nAmeraka	/a/	1	natural	22.9(5.5)	18.7(9.1)	40.4(11.6)
			flat	27.9(9.7)	22.1(12.3)	51.1(9.0)
namerAka	/a/	3	natural	28.9(5.3)	37.9(12.0)	67.7(15.9)
			flat	36.0(7.1)	40.6(13.0)	75.0(16.9)

DT+ : Mean discrimination threshold for lengthening direction.
DT- : Mean discrimination threshold for shortening direction.
DT : Mean discrimination threshold range (= (DT+) + (DT-)).
*Values in round brackets are standard deviations.

Procedure

Discrimination thresholds were measured by the up-down paradigm with two response alternatives; "same" or "different". Subjects were presented two stimuli which differed only in the duration of one of the four moraic segments. Four series of stimulus pairs were randomly presented to prevent prediction of the position of the target segment. As a check, trials with physically identical pairs were inserted occasionally, to prevent too short an estimation of the threshold. Each series was tested with both a natural and a flattened F0 contour.

Subject

Eight adult female subjects including the subjects in the first experiments participated in the experiment.

3.2 Result and discussion

Table 2 shows the mean and the standard deviation of the measured discrimination thresholds for each segment.

Effect of vowel color, position, and F0 contour

The following effects on the range between discrimination thresholds (DT) were confirmed by the analysis of variance with vowel color, position in a word, and F0 contour.

(1) vowel color: DT(/a/) < DT(/i/) [$F(1,51)=25.5$, $p<0.000$]
(2) position: DT(1st) < DT(3rd) [$F(1,51)=4.72$, $p<0.0343$]
(3) F0 contour: DT(natural F0) < DT(flattened F0) [$F(1,51)=9.08$, $p<0.004$]

No interaction among the three factors was significant. The effects of the first two factors are consistent with the effects observed in the experiment of acceptability.

Relationship between discrimination threshold and acceptability

We also measured acceptability for the same stimuli employed in the discrimination threshold measurement. Correlation analyses were performed on the results obtained from the two measurements. A positive correlation was found between the center shift of the range between discrimination thresholds and the axis shift of acceptability curve which stands for the center shift of the acceptability range. Moreover a negative correlation were found between the range between discrimination thresholds and the sharpness of the

acceptability curve which stands for the narrowness of the acceptability range. (See Figure 6a,b.)

These findings suggest that the subjects evaluated acceptability based on the perceptual sensitivity. The fact that the vowel color and the position in a word affected the discrimination threshold, supports the hypothesis that these two factors function at a very early stage of perception.

IV. GENERAL DISCUSSION

Effect of position: is it a common effect?

In this study, the effect of position on the acceptability and the discrimination threshold is consistent with previous findings [3, 4]. That is the perceptual sensitivity for durational modification is higher at the first position than at other positions. However, it is still an open question whether such a positional effect is common enough in the perception of temporal aspect, because a phenomenon inconsistent with this effect has been found in experiments using non-speech sounds. Hirsh[8] reported that the temporal discriminability for the intervals in tone-burst successions was worst in initial positions and best in final positions when each interval was set to 50ms. Further studies would be required to investigate what difference exists between information processing for speech and for non-speech sounds.

Effect of vowel color: why high discriminability for /a/ not for /i/?

It was predicted from Weber's law that discrimination thresholds would be shorter for vowel /i/ than for vowel /a/ because the inherent duration of /i/ is shorter than that of /a/. However the reverse was the case for both the range between discrimination thresholds and the acceptability range.

The following two statements possibly explain it.

The first possibility is a difference in power. The power of vowel /a/ is inherently higher than that of /i/ in spoken Japanese [9]. It means that change in power could be larger in vowel /a/ than in /i/ even when same the changes occurred in duration. And such a difference in power can affect duration discriminability, because there is a trading relationship between duration and power in discrimination of sound whose duration is 200ms or less [10].

The second possibility is a devocalizing inclination. The vowel /i/ is often devocalized in a certain context of spoken Japanese while the devocalization of /a/ seldom occurs. In the acceptability evaluation, extremely small scores can be given in the shortening direction for the segments which incline to be devocalized. (For instance, some subjects responded with the acceptability score "-3", which corresponds to "very natural", even when the portion of vowel /i/ is shortened to zero.) Such small scores could flatten the acceptability curve.

V. CONCLUSION

The experiment which examined the acceptability to the durational modification showed that the acceptability score changes corresponding to the durational modification:

(1) more sensitive in case of the vowel /a/ than in case of vowel /i/,

(2) more sensitive in case of the first moraic segments than in case of the third moraic segments,

(3) more sensitive in case of the high-tone segments than in case of the low-tone segments when this contrast exists in the third moraic segment.

The result of the experiment which examined the discrimination threshold were consistent with the first two findings mentioned above.

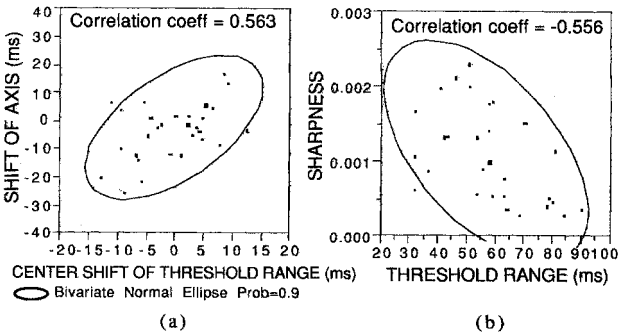

Figure 6. Correlation between acceptability and discrimination threshold: (a) a positive correlation between the axis shift of the acceptability curve and the center shift of the threshold range. (b) a negative correlation between the sharpness of curve and tthe threshold range.

REFERENCES

[1] Takeda,K., Sagisaka,Y., and Kuwabara,H., "On Sentence-level Factors Governing Segmental Duration in Japanese", J. Acoust. Soc. Am. 89, pp.2081-2087 (1989)

[2] Kaiki,N., Takeda,K., and Sagisaka,Y., "Linguistic Properties in the Control of Segmental Duration for Speech Synthesis", in *Talking Machines: Theories, Models, and Designs*, Bailly,G., and Sawallis,T., Eds., pp255-263, Elsevier Science Publish. (1992)

[3] Sato,H., "Segmental Duration and Timing Location in Speech" (in Japanese with English abstract), Proc. Trans. Committee on Speech Acoust. Soc. Jpn, S77-31 (1976)

[4] Klatt,D.H., "Linguistic Uses of Segmental Duration in English: Acoustic and Perceptual Evidence", J. Acoust. Soc. Am. 59, pp.1208-1221 (1976)

[5] Sagisaka,Y., "Prosody Controls for Japanese Speech Synthesis" (in Japanese), Ph.D. Dissertation, Waseda University (1985)

[6] Sagisaka,Y., Takeda,K. Abe,M. Katagiri,S., Umeda,T., and Kuwabara,H., "A Large-Scale Japanese Speech Database", Proc. ICSLP90, pp.1089-1092 (1990)

[7] Imai,S. and Kitamura,T., "Speech Analysis Synthesis System Using the Log Magnitude Approximation Filter" (in Japanese), J.Electron.Inf.Commun.Eng. 61-A, pp.527-534 (1978)

[8] Hirsh,I., Monahan,C., Grant,K., and Singh,P., "Studies in Auditory Timing: 1. Simple Patterns", Perception & Psychophysics 47, pp.215-226 (1990)

[9] Mimura,K., Kaiki,N., and Sagisaka,Y., "Analysis and Control of Temporal Patterns of Speech Power Using Statistical Methods" (in Japanese with English abstract), Proc. Trans. Committee on Speech Acoust. Soc. Jpn, SP91-4 (1991)

[10] Scharf,B., "Loudness," in *Handbook of Perception(Vol.IV), Hearing*, Carterette,E., and Friedman,M., Eds., pp.187-242, Academic Press (1978)