On Perception of Word-based Local Speech Rate in Japanese without Focusing Attention

Makoto HIROSHIGE, Kantaro SUZUKI, Kenji ARAKI and Koji TOCHINAI
Graduate School of Engineering, Hokkaido University
N13W8, kita-ku, Sapporo 060-8628 JAPAN
Tel: +81-11-706-6535 Fax: +81-11-706-6277 E-mail: hiro@media.eng.hokudai.ac.jp

ABSTRACT

Fundamental investigations about differential limen (DL) for word-based speech rate variations in Japanese are described. We carry out auditory tests with stimuli made by equally lengthening or shortening a duration of a word in a sentence. We set up the experiments to diffuse the subjects' focus of attention. When the focus is diffused, the DL value of acceleration increases in several cases. The obtained DLs are 18.9 msec/mora for acceleration and 26.5 msec/mora for deceleration.

1. INTRODUCTION

It is very important to introduce human factors even into speech communication between human and machine. In usual conversations, human speaker expresses various information using prosodic expressions simultaneously with phonemic expressions. Several studies of prosodic informations have been carried out being conscious of applications on speech recognition or speech synthesis [1][2][3]. While many of these studies discuss about pitch or power information for speech synthesis, we are studying about local speech rate variations aiming for recognition of speaker's intentional control.

It is said that Japanese speech has fewer speech rate variations than the other languages. However even in Japanese, natural conversations sometimes contain considerable amount of speech rate variations. When there is a distinct slow or fast part in a sequence of speech, such part may contain some strong intention of the speaker. To detect the local speech rate variations, it is necessary to know how much speech rate variation can be detected by human beings.

It is especially important to gather up several knowledge about the perception of word-based speech rate variations, because it seems that speaker's intention appears in words.

In previous study[4], we carried out auditory tests with speech stimuli that contain various amounts of word-based speech rate variation for the purpose of a fundamental investigation about differential limen (DL) for word-based speech rate variations. DL is the stimulus variation between just perceptible and just unperceptible. In this previous study, subjects were induced to concentrate to a particular single word, whose speech rate was controlled. In this circumstance, however, subjects tend to concentrate higher than usual conversations, and the obtained DL values were much smaller than we expected. To overcome these phenomena and to get DL values for usual conversations, several experiments are carried out with new settings in this report. In a setting of experiments, no information about rate-modified word is given to subjects, so that subjects hardly concentrate a particular word, i.e., subject's focus is diverted. In another setting, information of rate-modified word is given by an underline, but the contents of stimuli sentence are selected to be completely different each other. Comparisons between these settings are discussed.

In section 2, the method and the results of the previous experiments[4] with the focusing attention of the subjects are briefly introduced. In section 3, two types of new experiments to diffuse the subjects' focus of attention are described. In section 4, comparisons between the results of these experiments are discussed.

2. EXPERIMENT I

In this section, we briefly introduce the procedure and the results of our preceding study[4] that the subjects focus on a particular word when detecting the speech rate variation.

2.1 Stimuli

We use three sentences uttered by a male announcer and two sentences uttered by a female announcer as original speech materials to make stimuli for auditory test. the followings are sentences used as original speech materials: (A) *Haruninatte / TAUE / no / kisetsuga / yattekita* (Spring has come and the season of RICE PLANTING has arrived.) (B) *Kokunaisendesukara / Naritakuukou / dehanakute / HANEDAKUUKOU / desu* (It's a domestic airline, so the airport is HANEDA, not Narita.) (C) *Karenokitaimo / hairiyouto / ittara / MONOSUGOI / monodatta* (He did it with a great passion, almost INCREDIBLY.) The sentence (A), (B) and (C) are uttered by the male announcer. The sentence (A) and (B) are uttered by the female announcer. Stimuli are made by altering the speech rate of a particular word represented by the underlined letters in each sentence. There are 5 shortened stimuli in which the particular word's speech rate are adjusted to be +0.5, +1.0, +1.5, +2.0 and +2.5 mora/sec comparing the average rate respectively. There are also 5 lengthened stimuli in which the particular word's speech rates are adjusted -0.5, -1.0, -1.5, -2.0 and -2.5 mora/sec. Including the original stimuli (±0 more/sec), we prepared 11 stimuli. For simplicity, we equally lengthen or shorten the durations of the particular words in this report. SoundEdit16 is used for altering the speech rates of particular words and it can equally lengthen or shorten the durations of the words. Each stimuli sentence is repeated 10 times, so that we get 110 stimuli. Then we alter the presenting order of the 110 stimuli randomly, to get a stimuli set. There are 5 sentences, so that we have 5 stimuli sets.

2.2 Procedure

Auditory tests are carried out with each of the stimuli sets. Each of the stimuli is presented through a headphone in an anechoic room. The subjects are 5 males who are all native speakers of Japanese. Subjects are asked to hear the difference between speech rate of the particular underlined word and speech rate of the whole sentence. If a subject feels that the particular word is faster or slower than the whole sentence, he selects the answer "fast" or "slow", otherwise "same". Auditory tests are carried out for all subjects together. Until all subjects complete selecting the answers to one stimulus, the next stimulus is no presented. One stimulus can be listen to only once. Several stimuli are prepared to rehearse listening to a stimulus and answering before each auditory test.

2.3 Results

According to the subjects' introspective reports, they usually judge using the preceding parts of the particular words as standards. Thus we regard speech rates of the preceding parts of the particular words as standard stimuli. Then we measure DL for word-based speech rate variations using the Pauli's equations.

The mean DLs obtained by the experiment I (Exp.I) are shown in Table 1. The DL values are expressed by mora duration (msec/mora) for convenience to compare with the results of other researches. In the condition of Exp.I, the all sentences in a stimuli set have same contents, and presented repeatedly. Moreover, there is an underline on the rate-modified portion in each stimuli sentence. Thus the subjects are considered to concentrate strongly to the rate-modified portion. Thus, the obtained value of DLs seems to be smaller than we expected, especially in the case of acceleration. To get DLs used in natural conversation, the subjects' focus of attention should be diffused.

(a) DL values for acceleration [msec/mora]

Sentence	(A)	(B)	(C)	(D)	(E)
Averaged DL within subjects	-6.48	-3.46	-5.88	-9.20	-7.24
Standard Deviation	2.19	2.87	3.89	3.23	1.49

(b) DL values for deceleration [msec/mora]

Sentence	(A)	(B)	(C)	(D)	(E)
Averaged DL within subjects	17.68	21.76	18.52	16.5	24.76
Standard Deviation	5.48	5.08	3.85	2.59	7.38

Table 1: The DL values obtained by Exp.I

3. EXPERIMENT II

In this chapter, we set up auditory tests for DL of speech rate change in which subjects can not focus on particular portion of the stimuli sentences. This condition can be considered as nearer condition to natural conversation than that of the Exp.I.

For the experiment II (Exp.II), two conditions are prepared. One is that there is NO underline on the sentence in the answer sheet, so that the subject decide the portion of rate variation completely freely(Exp.II-a). In Exp.II-a, stimuli sentences are all different each other. The other condition is that there is underline on a particular word on the sentence to lead subjects' concentration(Exp.II-b). The stimuli sentences are the same as the Exp.II-a. The Exp.II-b is for comparisons with Exp.II-a.

3.1 Stimuli

3.1.1 Recording of the Original Speech

To diffuse subjects' focus, all stimuli sentences are selected to be different each other. If there are same sentences in a stimuli set, subjects may remember the result of the previous same sentence, and may pay attention to the portion at which the subject felt variations of speech rate in the previous stimuli. If stimuli sentences are long, it becomes difficult to decide the standard stimuli, which is, in the case of this research, the portion of the sentence which is uttered in 'normal' speech rate and used as a standard when the subjects compare speech rate with the other portion of the speech. Thus, the all stimuli sentences are selected to contain about 3 words only. As same as the experiment I, the original recorded speech should be uttered as calmly as possible, not to contain large variation of speech rate.

For these reasons, we newly recorded 60 sentences uttered by male professional announcer. Then we select 21 sentences which contains small variations of speech rate.

Examples of the stimuli sentences are shown in Table 2.

1st word	/	2nd word	/	3rd word
Sonokaishawa	/	*shakkinde*	/	*kurushindeiru.*
The company is suffering from the debt.				
Konojishowa	/	*keitaini*	/	*benrideesu.*
This dictionary is good for carring.				

Table 2: Examples of the stimuli sentences in Exp.II

3.1.2 Making Up the Stimuli

We select 7 sentences among the 21 sentences, then modify the speech rate of the first word to have the following modification amounts of speech rate from original: +3, +2, +1, 0(original), -1, -2 and -3 mora/sec, respectively ("the 1st word case"). Then, we select 7 sentences among the remainder 14 sentences, and modify the rate of the second word by the same way ("the 2nd word case"). For the remainder 7 sentences, we modified the

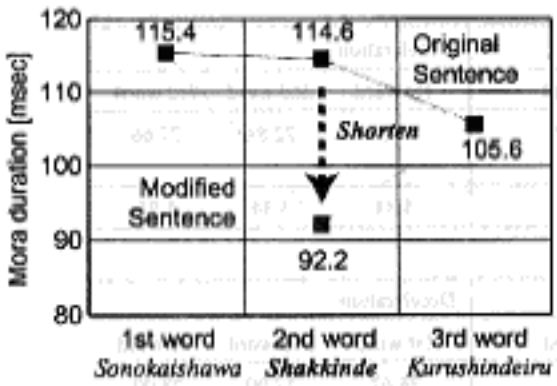

Figure 1: Modification of rate in Exp.II

third (the last) word ("the 3rd word case"). Gathering all 21 sentences, we get a set of stimuli. Selecting different word for modification for all 21 sentences, we can get 2 more sets of stimuli. An example of the modification of rate is shown in Figure 1.

3.2 Procedure

There are two different conditions, i.e., Exp.II-a and Exp.II-b.

3.2.1 Procedure of Experiment II-a

The experiment II-a (Exp.II-a) is designed to diffuse subjects' attention completely. There is no underline in the sentence in the answer sheets. A single session of the auditory tests are carried out with single stimuli set mentioned in 3.1.2. Totally 3 sessions are carried out using different stimuli set to the same 5 subjects. The 5 subjects are the same persons as the Exp.I. Between the 3 sessions, there are enough interval (several months) so that the subjects forget the contents of the previous auditory test.

Subjects are asked to hear the stimuli sentences, and if they feel a variation of speech rate in the stimuli sentence, they are requested to give an underline where they feel rate variation, and describe "fast" or "slow". If they do not feel any rate variation in the stimuli, they can select the answer "same".

Each stimuli sentence in the stimuli set are presented only once in random order. Thus total number of presented stimuli is 21 in a single session. Auditory tests are carried out for all subjects together. Until all subjects complete selecting the answers to one stimulus, the next stimulus is no presented. Several stimuli are prepared to rehearse listening to a stimulus and answering before each auditory test.

3.2.2 Procedure of Experiment II-b

The experiment II-b (Exp.II-b) is designed for comparison with Exp.II-a. The main difference is that there is underline on the rate-modified word in the answer sheet. Subjects are 5 males who are all native speakers of Japanese, but they are different persons with the Exp.II-a and Exp.I. The stimuli sentences, the construction of the stimuli set, and the order of presentation are the same as Exp.II-a. Since there are already underlines, the subject need not select the portion of the speech rate variation. They can concentrate on the underlined word, and only select the answer "fast", "slow" or "same". The difference between Exp.I is the construction of stimuli. In the Exp.I, the same sentences with different rate modifications are presented at once, while in the Exp.II-b, stimuli sentences have all different contents each other.

In the Exp.II-b, the 3 stimuli sets are presented consecutively at one session. This is since the subject need not forget the contents of the previous stimuli set in this case.

3.3 Results

As same as Exp.I, we select the preceding parts of the underlined word as standard stimuli. (In Exp.II-a, the underlines are given by the subjects.) In case the second word is underlined, the speech rate of the first word is used as standard. In case of the third word, an averaged rate within the first and the second words is used. In case the first word is underlined, an averaged rate within the second and the third word is used as standard.

Experimental results of Exp.II-a and Exp.II-b are shown in Table 3. The DL values of stimuli are expressed by mora duration value (msec/mora).

4. COMPARISONS OF THE RESULTS AND DISCUSSIONS

4.1 Comparison between Exp.I and Exp.II-a

Conditions of the experiments are considerably different between Exp.I and Exp.II-a, so that direct comparisons are difficult. In this report, we try to get preliminary information by the comparisons with t-test.

We select comparison pairs by the location of the underlined word. The sentences (A) and (D) in Exp.I have an underlined word in the middle of the sentence, so that the results with the (A) and (D) in Exp.I are compared with the results with "the 2nd word case" in Exp.II-a. The sentences (B), (C) and (E) have an underlined word near the end of the sentence, so that they are compared with "the 3rd word case" in Exp.II-a.

The comparison results are shown in Table 4. In all cases of mora shortening(i.e., acceleration of the rate), there is significant differences between Exp.I and Exp.II-a ($t(8)$=(in Table 4),$p<1\%$). In cases of mora lengthening(i.e., deceleration of the rate), only the case between (C) and "the 3rd word" has significant difference($t(8)=4.24$, $p<1\%$), but the other cases have no significant difference.

4.2 Comparison between Exp.II-a and Exp.II-b

In this case of comparison, conditions of the experiments are almost the same except for the subjects, so that we can investigate the difference by a direct comparison. In all cases both mora shortening and lengthening, there is no significant difference between Exp.II-a and Exp.II-b.

(a) DL values by Exp.II-a [msec/mora]

	Acceleration			Deceleration		
	1st word	2nd word	3rd word	1st word	2nd word	3rd word
Averaged DL within subjects	-19.86	-17.86	-19.02	29.14	22.84	27.66
Standard Deviation	2.33	3.29	1.49	5.84	3.34	1.95

(b) DL values by Exp.II-b [msec/mora]

	Acceleration			Deceleration		
	1st word	2nd word	3rd word	1st word	2nd word	3rd word
Averaged DL within subjects	-20.27	-16.50	-18.33	26.67	27.00	28.00
Standard Deviation	1.50	2.00	2.24	4.71	9.80	3.67

Table 3: The DL values obtained by Exp.II

4.4 Discussions

From results shown above, the following things can be considered. 1) When the subject's focus is diffused, the DLs for the variation of speech rate seems to become large in the case of acceleration. 2) For diffusing of the subject's focus, the randomize of the contents of the stimuli sentence seems to be more efficient than the underlining of a particular portion in the sentence. 3) In our experiments, DL for acceleration is derived as -18.9 msec/mora, and DL for deceleration is derived as 26.5 msec/mora (These DLs are averaged values in Exp.II-a).

5. CONCLUSIONS

In this report, fundamental investigations about differential limen (DL) for word-based speech rate variations have been described. We have carried out auditory tests with stimuli made by equally lengthening or shortening a duration of a word in a sentence. Especially, we have been managed to set up the experiments to diffuse the subjects' focus of attention, to get DLs that are used in the normal natural conversations.

The results we have obtained are as follows: We can diffuse the subjects' focus when the stimuli sentence has random contents. When the focus is diffused, the DL value of acceleration seems to increase. The obtained DLs are 18.9 msec/mora for acceleration and 26.5 msec/mora for deceleration.

Sentence of Exp.I	(A)	(B)	(C)	(D)	(E)
Word of Exp.II-a	2nd	3rd	3rd	2nd	3rd
t value (DL for accel.) df=8	5.76 (p<1%)	9.61 (p<1%)	6.31 (p<1%)	3.75 (p<1%)	11.18 (p<1%)
t value (DL for decel.) df=8	N.S.	N.S.	4.24 (p<1%)	N.S.	N.S.

Table 4: The results of t-test for Exp.I and Exp.II-a

In this report, we have modified the durations of the words to obtain the stimuli that have various speech rate variation. This operation may affect the natural rhythm of the word, so that further considerations for the modifying method are needed. To confirm the result in this report, further comparisons between the cases with various levels of the subjects' focus.

REFERENCES

1. Huggins, A.W.F., "Just noticeable differnces for segment duration in natural speech," *J.Acoust.Soc.Am.*, 51(4): 1270-1278, 1972.
2. Kobayashi, S. and Kitazawa, S., "Factors Concerning Paralinguistic Feature Identification in Natural Dialogue," *Technical Report of IEICE of Japan*, SP98-1: 1-8, 1998.
3. Ohno, S. and Fujisaki, H., Taguchi, H. and Watanabe, N. "A study of speech rate variations using the local speech rate — Analysis of influences of the average speech rate and word accent types —", *Proc. Spring Meeting of Acoust.Soc.Japan*, 1-7-11: 201-202, 1983
4. Suzuki, K., Takamaru, K., Hiroshige, M. and Tochinai, K. "A fundamental study on perception of word-based speech rate variations — Measurement of differential limen with several kinds of speech stimuli —," *Proc.of ITC-CSCC 99*: 13-16, 1999.