On Differential Limen of Word-based Local Speech Rate Variation in Japanese Expressed by Duration Ratio

Makoto HIROSHIGE, Kenji ARAKI and Koji TOCHINAI
Graduate School of Engineering
Hokkaido University, Sapporo, JAPAN
hiro@media.eng.hokudai.ac.jp

Abstract

Fundamental studies about differential limen (DL) for word-based speech rate variations in Japanese are described. In our previous study, the DLs are expressed by subtractive difference of mora duration. In this report, however, to fit the expression for various global speech rate, the DLs are expressed by variation ratio of mora duration. We carry out auditory tests with stimuli made by equally lengthening or shortening a duration of a word in a sentence. The subjects' focus of attention is diffused to get DLs that are used in the normal natural conversations. The obtained DLs are approximately 0.85 for acceleration and 1.18 for deceleration in variation ratio of mora duration.

1. Introduction

Recently, with progression of human interfaces of computers, the importance of humanized 'warm' communication is highly recognized. Precise investigation for the mechanism of humanized 'warm' communication and its simulation by machine are of course important for human-machine communication, and also important for human-human communication because of intermediation of machine between human-human communication.

In the area of speech communication, introduction of human factors also becomes one of major issues among speech researchers and engineers. 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.

We already carried out several auditory tests with rate-modified speech stimuli for the purpose of a fundamental investigation about differential limen (DL) for word-based speech rate variations[4][5]. DL is the stimulus variation between just perceptible and just unperceptible. To obtain the DL values used in our usual conversation, subjects' focus of attention is managed to defuse in our auditory tests[5]. In the previous study[5], however, the DLs are expressed in the absolute time value ([msec]) of subtractive difference of mora duration. Considering our usual experiences, our cognitive responses will be different for the same absolute value of subtractive difference of mora duration between the cases that the global speech rate is fast (i.e., averaged mora duration is short) and slow (i.e., averaged mora duration is long). It is considered that a variation ratio may be better expression for the DL of the rate variation.

In this report, we try to obtain the DL value of mora duration expressed in the variation ratio. Our auditory test in this report is, as similar to the previous study[5], managed to measure word-based rate variation under the condition that the subjects do not highly concentrate to any portion of speech stimuli. The number of subjects is increased to be 10 persons (in [5], subjects are 5 persons).

2. Auditory tests

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.

From the results of our previous study[5], it is essential

to select random contents of stimuli sentences to defuse subjects attention. Even if an underline is given to a transcribed text in the answer form to indicate the portion that will be compared with other portions of the stimuli, subjects' concentration do not highly increase[5]. Thus, in this report, a set of underlined stimuli with random contents is prepared.

2.1 Recording of the original speech

As mentioned above, 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. The original recorded speech should be uttered as calmly as possible, not to contain large variation of speech rate.

From these points, we use originally recorded 60 sentences uttered by male professional announcer. Then we select 21 sentences which contains small variations of speech rate (these data are the same as [5]).

Examples of the stimuli sentences are shown in Table 1.

Table 1: *Examples of the stimuli sentences*

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

2.2 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"). 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 only the durations of the words without modification of power and f0. 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 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 rate modification designs is shown in Figure 1. An example of waveform, power, f0 and mora duration curve before and after the modification is shown in Figure 2.

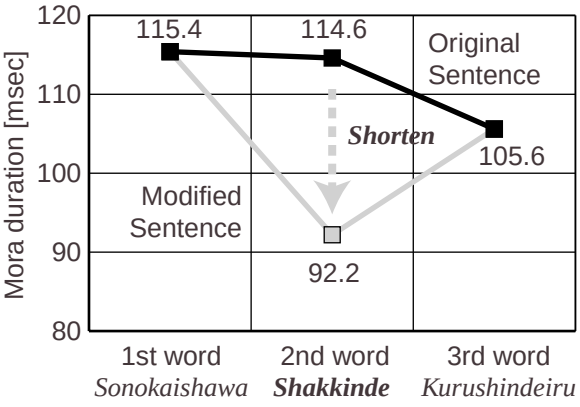

Figure 1: *Rate modification design*

2.3 Procedure

In the auditory tests, subjects are 10 males who are all native speakers of Japanese. A single session of the auditory tests are carried out with single stimuli set mentioned in 2.2. Totally 3 sessions are carried out using different stimuli set to the same 10 subjects. The 3 sessions carried out consecutively at once.

Subjects are asked to hear the stimuli sentences, and if they feel a variation of speech rate between the underlined word and the other portion of the stimuli, they are requested to describe "fast" or "slow". If they do not feel any rate variation in the stimuli, they can select the answer "same".

3. Results

We select the preceding parts of the underlined word as standard stimuli. 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.

In this report, variation ratio is used to express the DL. The variation ratio is calculated by dividing an averaged mora duration of the underlined word by an averaged mora duration of the standard stimuli mentioned above.

After the auditory test, we get a variation ratio and

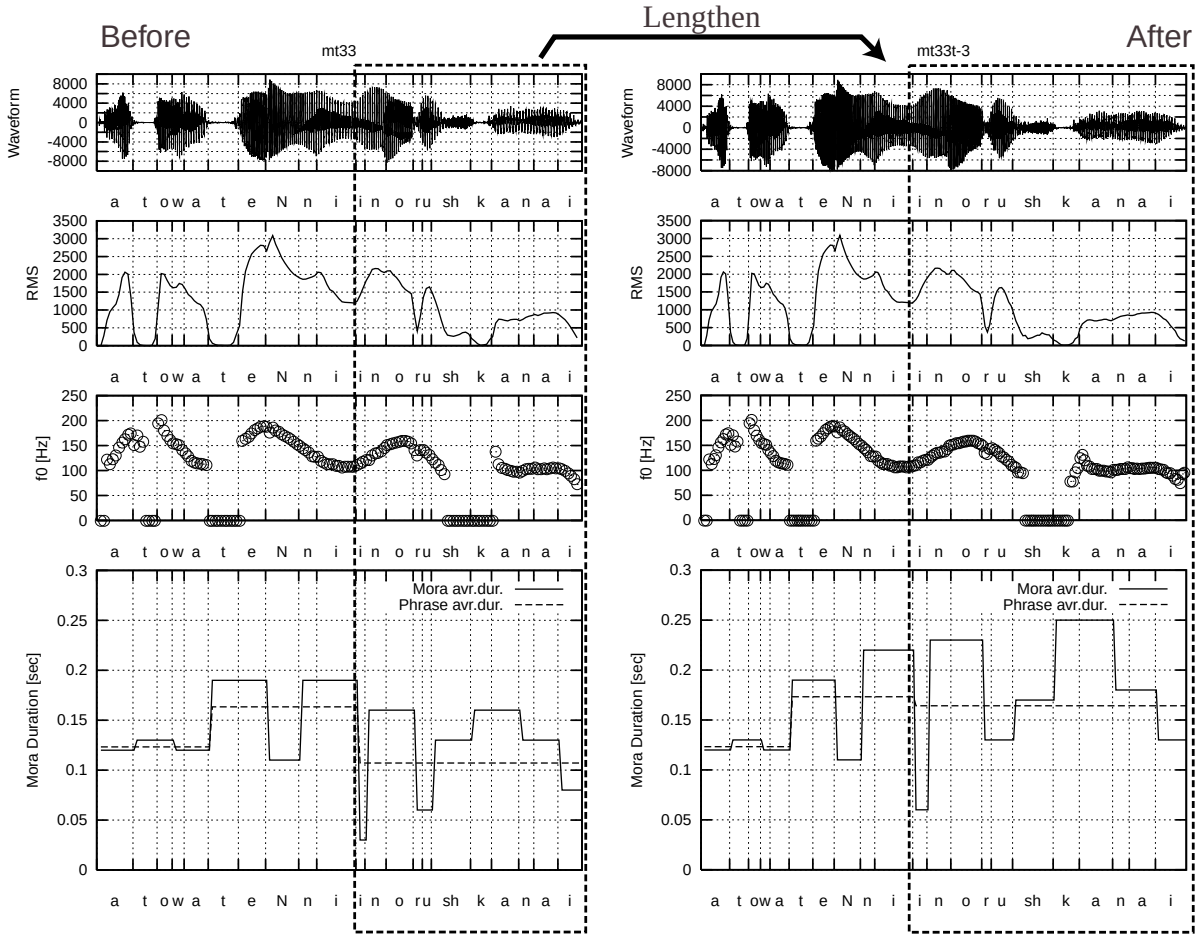

Figure 2: *Examples of waveform, power, f0 and mora durations of the stimuli before and after modification*

numbers of subjects who answered ‘fast’, ‘same’ and ‘slow’ for each stimuli sample.

The results of the auditory test are shown in Figure 3. In the all 9 figures in the Figure 3, the horizontal axis shows the variation ratio. A single ‘x’ mark shows a result for a single stimulus. Upper row of 3 figures shows the percentage who answered ‘fast’. Middle row of 3 figures shows the percentage who answered ‘same’, and lower row of 3 figures shows the percentage of ‘slow’. Left column of 3 figures shows the results of the case the first word is underlined. Middle column of 3 figures are for the second word, and right column of 3 figures are for the third word.

As described in 2.2, the stimuli set are made using subtractive difference of mora rate [mora/sec], so that the stimuli are not placed orderly on the horizontal axis of Fig.3, i.e., the axis of the variation ratio calculated using mora duration. Thus the Pauli’s equations can not be exactly used to find DL values. In this report, for simplicity, we estimate approximate DL values by manual interpolation of Fig.3. The DL values are estimated by the approximate point that the percentage trend curve crosses 50 percent point. The estimated DL values are shown in Table 2, and also shown by thick arrows in Fig.3. According to Table 2, the estimated DL value for acceleration is about 0.85 and about 1.18 for deceleration in variation ratio.

Table 2: *Estimated DL values*

	1st word	2nd word	3rd word	Avr.
Accel.	0.89	0.83	0.83	0.85
Decel.	1.15	1.20	1.20	1.18

4. Comparison with the results of our previous study

In our previous study[5], the DL for acceleration is about -18.9 msec/mora, and about 26.5 msec/mora for deceleration, expressed by subtractive difference of the mora duration. Using the averaged mora duration of the standard stimuli which is 118 msec (stdev.= 8.7 msec) in the stimuli of this report, the DLs in ratio obtained in this report are approximately converted into the DLs in subtractive difference as follows:

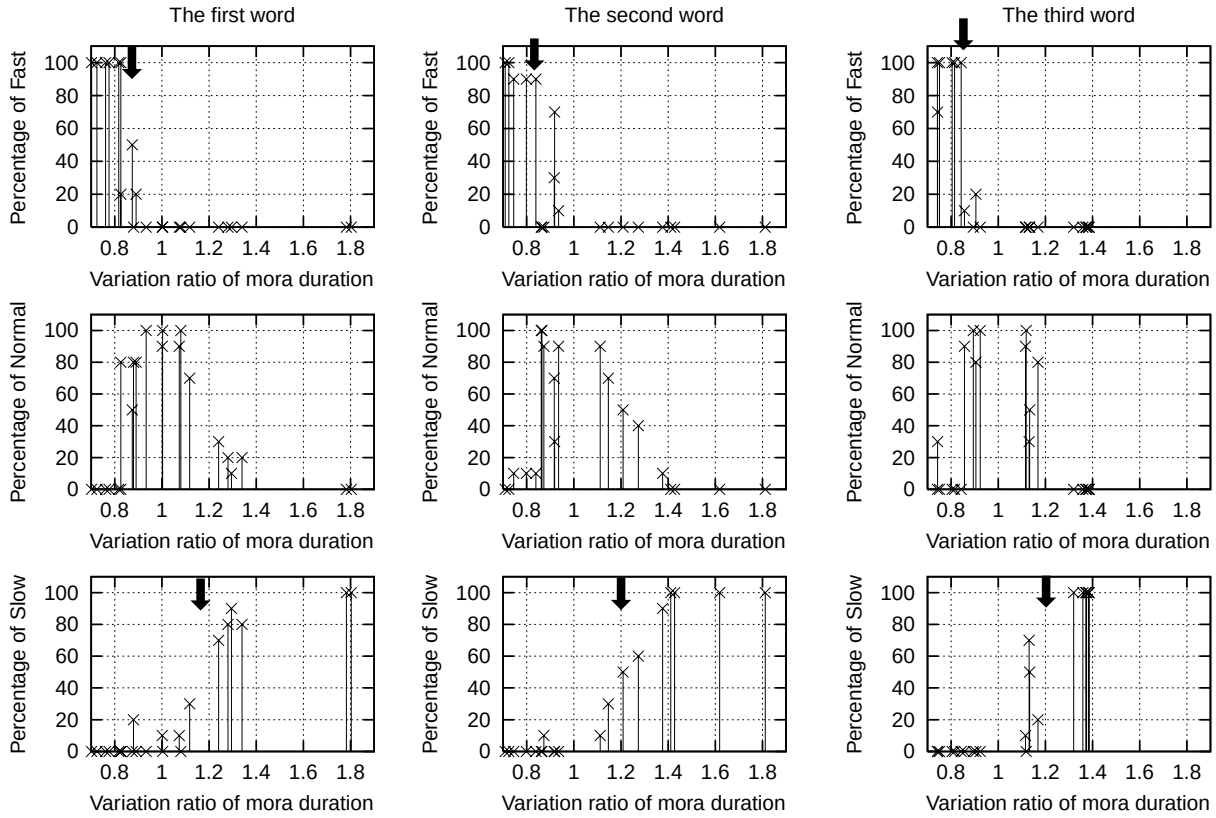

Figure 3: *Experimental results of the auditory test : Thick arrows show the estimated DLs.*

Accel. : (118msec x 0.85) - 118msec = -17.7 msec
Decel. : (118msec x 1.18) - 118msec = 21.2 msec

Thus the results in this report roughly agree with the results of our previous study.

5. Conclusions

In this report, fundamental investigations about differential limen (DL) for word-based speech rate variations have been described. We have tried to express the DLs in variation ratio of mora duration. We have carried out auditory tests with a setting diffusing the subjects' focus of attention. The obtained DLs are approximately -0.85 for acceleration and 1.18 for deceleration in variation ratio. These values agree with DL values obtained in our previous study expressed in subtractive difference of mora duration.

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. The amount of modification of stimuli should be refined to fit the expression of DL in variation ratio so that we can get more accurate DL values by exact calculation. It is better to prepare several kinds of speech samples which have various global speech rate to confirm the appropriateness of the ratio expression of DLs.

6. References

[1] A.W.F.Huggins, "Just noticeable differences for segment duration in natural speech" J.Acoust.Soc.Am., vol.51(4), pp.1270-1278, 1972.

[2] S.Kobayashi and S.Kitazawa, "Factors Concerning Paralinguistic Feature Identification in Natural Dialogue" Technical Report of IEICE of Japan, SP98-1, pp.1-8, 1998.

[3] S.Ohno and H.Fujisaki, H.Taguchi and N.Watanabe, "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, no.1-7-11, pp.201-202, 1983.

[4] K.Suzuki, K.Takamaru, M.Hiroshige and K.Tochinai, "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, pp.13-16, 1999.

[5] M.Hiroshige, K.Suzuki, K.Araki and K.Tochinai, "On perception of word-based local speech rate in Japanese without focusing attention" Proc.of ICSLP2000, vol.III, pp.255-258, 2000.