ENHANCING SPEECH PROCESSING OF JAPANESE LEARNERS OF ENGLISH UTILIZING TIME-SCALE EXPANTION WITH CONSTANT PITCH

Kaoru Tomita-Nakayama, Kazuo Nakayama* and Masayuki Misaki***

*Yamagata University, Yamagata 990-8650, Japan,
**Matsushita Electrical Industries Co. Ltd., Kadoma 571-8501, Japan

ABSTRACT

This study demonstrated that the time-scale expansion of speech with the constant pitch (henceforth, expanded speech) enhanced the speech recognition of Japanese learners of English in contrast to the previous studies in which the time-scale expanded speeches did not contribute to the speech recognition chiefly because of the severe distortion of original speech and pitch change [1]. Experiments were administered with the stimuli of original normal speech and the corresponding expanded speech. These results showed that the expanded speech stimuli were intelligible to many of the subjects. The hypotheses are that the expanded speech enhances a listener's speech processing and it also enables the listener to call into play virtual memory capacity for an on-line speech processing, which are more apparent in a longer stimulus. The expanded speech worked well with most subjects. Another prescriptions should be prepared for the rest of the subjects for whom the expanded speech was not solely much effective.

1. INTRODUCTION

Performance of auditory speech processing is determined by various factors, one of which is speech rate. Matching of the transmission rates between a talker and a listener is not always established. Generally, it is made possible by the effort of a listener who tries to adjust his rate of speech communication to that of the other party [1]. Most foreign language learners frequently feel that a speech they are listening to is too fast for them to understand. They even find that incoming sound sequences are inseparable, which is thought to make lexical access more difficult. An adult native speaker of English will be able to respond immediately that a statement like "You can start a fire with a match." is true, while "People eat through their noses." is false. Therefore the degree to which a foreign language learner approximates to the native speaker skill in evaluating such statements might enable us to assess his or her ability of auditory processing [2]. With recent development of speech technology, various methods have been developed to enhance auditory speech processing [3].

2. PRINCIPLES BEHIND THE EXPERIMENTS

2.1. Sequential processing

In listening to the speech sounds, the flow of speech prohibits the listeners to stick to one point or to review the previous phrases and forces them to process the speech sequentially. The foreign language learners who are not familiar with the word order of the language have difficulty in this way of processing.

A listener's processing time plays an important role in auditory speech processing of the foreign language. There are several factors that would affect his or her processing time, speech rate, articulation clarity, the length of a sentence, and so on.

2.2. Speech rate and articulation clarity

Slow speech is less than 160 words per minute (henceforth, [wpm]), normal speech ranges from 160 to 190 [wpm], moderately fast speech from 190 to 210 and fast speech is more than 210 [4]. When the matching of the transmission rates between a talker and a listener are not established, the listener feels understanding the other is fairly difficult. Besides the matching of the transmission rates between talkers and listeners, the distortion of the sounds in a fast speech deteriorates the listener's comprehension, especially in a foreign language speech recognition.

2.3. Sentence length

The sentence length can be measured by duration [ms], the number of words and syllables. In the present study, it is measured by the number of words, because the word is considered as a basis for processing speech. The effort has been made to avert the multisyllabic words to minimize the difference of the sentence length counted by the number of words and syllables. In general, the longer the sentence is, the more difficult for the listener to process it.

2.4. Time-scale modification of speech

Time-scale modification of speech (henceforce, TSM) can mean both expansion and compression. Although effects of compression on foreign language learning have been studied [5], we will focus here on time-scale expansion. Apparently, speech rate can be easily altered by varying the rotating speed of a disk player or tape recorder, it is accompanied by a pitch

change in proportion to the change of rate, and thus the loss of speech intelligibility is inevitable. In order to solve this problem, various methods to modify the time-scale have been proposed and perceptual experiments have been reported. Among them, some studies are focused on speech recognition of a foreign language [6]. The time-scale modification method developed by Misaki, M. and his co-workers is able to solve the fundamental problems associated with these conventional time-scale modification methods in that the expanded speech still keeps almost the same pitch as the original speech. In addition, the distortion caused by time-scale modification is quite small [1].

The experiment reported in Nakayama, K. et al. 1998, 1998 in press employed stimuli of original and expanded speech [3, 7], TSM ratio of 1.50 (henceforth, TSM 1.50). The experiments reported here employed the expanded speech (TSM 2.00), which is considered to expand the processing time which we thought to play an important role. In addition, longer stimuli were employed than those in the previous studies [3, 7]. Although we must admit that articulation (clarity) decreases when TSM ratio is larger than 1.50, we still believe that the subjects would better understand speech whose ratio is larger than 1.50 or 2.00, when people want to understand the speech better.

3. EXPERIMENT 1

3.1. Screening test

All subjects received a screening test with five stimuli in the methods employed in the experiment. Those who understood the procedures and responded more than or equal to 60 percent correctly were qualified as the subjects of the experiment. Throughout the test, each subject tuned the volume with which they like to listen to the stimuli.

3.2. Purpose

The experiment was designed to demonstrate the enhancement observed in the expanded speech stimuli. There is one hypothesis to be investigated in experiment 1, that the expanded speech enhances a listener's speech processing in English.

3.3. Methods

3.3.1 Subjects

Subjects were 17 university students (12 males, 5 females).

3.3.2. Stimuli

Thirty-six simple English sentences consisting of five to thirteen words were chosen for the experiment. They can be classified into two types of statements: true and false statements. Consider the following examples.

True statement: You can start a fire with a match.

False statement: People eat through their noses.

Syntactic structures of these sentences were restricted to those taught in junior high schools and in part taught in high schools. They are four kinds of stimulus: original speech of true/false statement and corresponding expanded speech of true/false statement. Expanded speech was TSM 2.00.

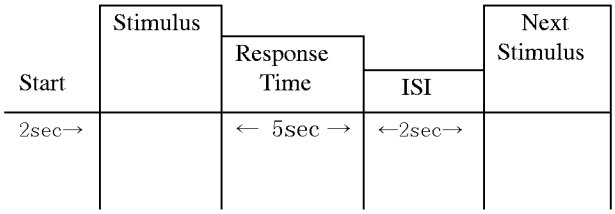
	Original	**TSM 2.00**
Average length [sec]	2.75	5.50
Average speech rate [wpm]	200.0	100.0

Table 1: Comparison of original/expanded speech

The subjects listen to the sentence of the same meaning twice, one is the original speech and the other is the expanded speech. Repetition effects were not observed as in Nakayama et al. 1998 in press [7], however, the order of stimuli is counterbalanced among the original speech and expanded speech stimuli.

3.3.3 Procedures

A subject was asked to judge that the statement was true or false. S/he was seated in front of a computer display and wore a headphone. When s/he was ready, s/he clicked the start button on the display. Then after two seconds, the stimulus was presented binaurally. It was presented only once. S/he listened to it and clicked a circle (meaning true) or a cross (meaning false) within five seconds. The stimulus is presented in the following sequence.



ISI stands for inter-stimulus interval

3.4. Results

Out of 289, the sum of correct responses to the original speech and those to the expanded speech were 193 and 199, respectively. The statistical test performed on the number of the correct responses of each stimuli showed no significant difference ($t(6)=0.61$, $p>0.05$).

Among eight students showed better performance with the expanded speech, two of them were conspicuous as shown in the following figures.

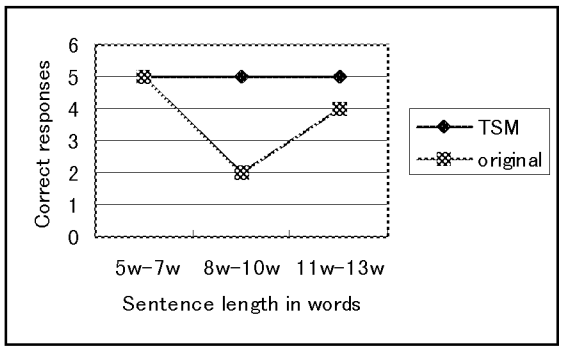


Figure 1: Correct responses of two kinds of stimuli

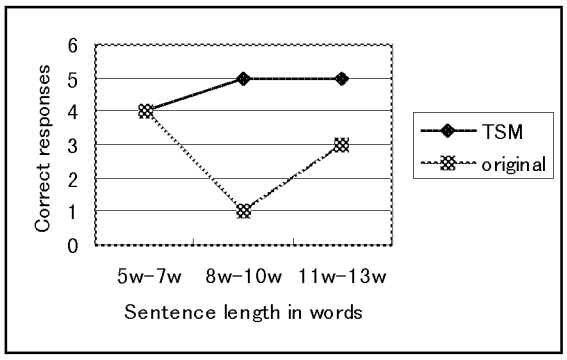


Figure 2: Correct responses of two kinds of stimuli

3.5. Discussion

The overall correct response increased with the expanded speech, compared with the corresponding original speech, though not statistically significant. The scores of the original speech tend to decrease as the sentences get longer. It is predicted that the scores of the expanded speech will be higher than those of the expanded speech. As shown in figure 1 and 2.

4. EXPERIMENT 2

4.1 Screening test

All the subjects received a screening test with the same procedure as the experiment 1. Those who understood the procedures and responded more than or equal to 60 percent correctly were qualified as the subjects of the experiment. Two of the subjects who scored less than 60 percent, however, were included in the experiment 2 to observe the performance of the elemental level (as far as auditory processing is concerned) language learners.

4.2 Purpose

There is one hypothesis to be investigated in experiment 2, that TSM enables the listener to share more processing time than original speech, which would be reflected in the scores of the expanded speech.

4.3 Methods

4.3.1 Subjects

Subjects were four graduate students (2 males, 2 females), and five university students (3 males, 2 females).

4.3.2. Stimuli

Forty-four simple English sentences consisting of 10 to 20 words were chosen for the experiment 2. Even though the most effort has been made to avert the multisyllabic words to minimize the difference of the sentence length counted by the number of the words and that of the syllables, the shortest sentence is 11 syllables and the longest sentence is 27 syllables. The informant was asked to keep the speech rate as constant as possible. The pauses longer than 200 [ms] are extracted. One of the characteristics of the read-out speech employed in experiment 2 was that "normal" phrasal intonations were controlled; the informant was asked to read each stimulus sentence without emphatic intonation. There are 44 stimulus. The original speech stimuli were expanded (TSM 2.00).

	Original	TSM 2.00
Average length [sec]	4.35	8.70
Average speech rate [wpm]	209.0	105.0

Table 2: Comparison of original/expanded speech

4.3.3 Procedures

The experiment 2 followed the same procedures as experiment 1.

4.4. Results

Out of 198, the sum of the correct responses to the original speech and those to the expanded speech were 112 and 120, respectively. Two statistical tests were performed. Differences in frequencies between the correct responses of the original speech and the expanded speech were not statistically significant ($t(8)=0.68$, $p>0.05$). Differences in frequencies between the correct responses of the original speech and the expanded speech both in the first presentation showed tendency ($t(8)=1.40$, $p<0.10$).

Five students showed better performance with the expanded speech, two of which are shown below.

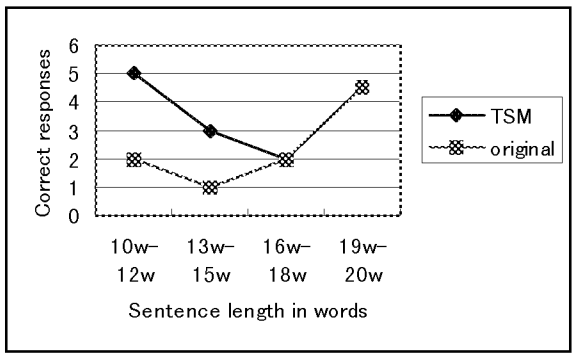


Figure 3: Correct responses of two stimuli

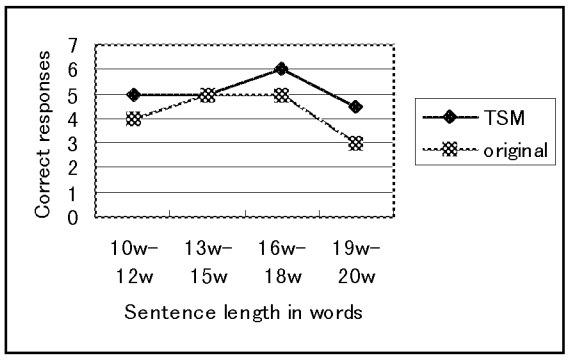


Figure 4: Correct responses of two kinds of stimuli

The scores of the original speech went down at the sentence length of 16 words to 18 words, and those of the expanded speech also fell. Still, scores on the expanded speech are higher than the original speech.

4.5. Discussion

The overall correct response increased with the expanded speech compared with the corresponding original speech stimuli though not statistically significant. It is a coherent tendency observed in the previous experiments [3, 7] and experiments 1 and 2. The number of the correct responses tend to increase statistically ($t(8)=1.40$, $p<0.10$) with the expanded speech of the first presentation compared with the corresponding original speech stimuli of the first presentation.

Most of the subjects including those who obtained high scores preferred the expanded speech for the test taking occasions. In the interview after the experiment, some subjects reported that the expanded speech is intelligible but it took more time to normalize the speech sound, that is, to access the right word in their mental lexicon.

5. CONCLUSION

Our goal in this experiment was to compare the subjects' processing of two types of the stimuli: the original speech of moderately fast speech and the expanded speech of very slow speech. Intelligibility, including familiarity of a certain initial section of a sentence seems to greatly influence subjects' success in understanding the stimuli. The methods can be extended to sentence intelligibility test in a foreign language, which can avoid difficulty of evaluation. They can also be extended to a part of auditory rehabilitation program of sentence comprehension. Thus the experimental studies described here would lead to a computer-based testing system with high validity and reliability. Although true/false judgement of an utterance is an important part of interpretation, it isn't all.

This study shows some advantages of speech technology. A speaker can slow down the speech when asked. But in this slow speech, prosodic features become degenerate while clarity in articulation appears to be intact. In addition, it is usually difficult for the informant to speak at the speech rate as they are asked to. With TSM, an utterance can be gradually converted into fast or slow one. The series of experiments with intersectional study should be carried out to determine the best rate of the expansion to enhance the listener to process foreign language speech.

Acknowledgements

We would like to acknowledge Professor Kiritani, S. of the University of Tokyo, Japan for his invariable comments as well as giving us permission to use the facilities of the university. Thanks also goes to Ms. Reiko-Akahane Yamada of ATR for her comment and encouragement, to Prof. Donna Erickson of Ohio State University who was kind enough to read the stimuli. when she visited ATR, Japan. This research is partially supported by the Grant-in–Aid for exploratory Research (09878038) by Mombusho, the Japan Ministry of Education Science, Sports and Culture.

6. REFERENCES

1. Suzuki, R. and M. Misaki 1992 Time-Scale Modification of Speech Signal Using Cross-Correlation Functions, *IEEE Transactions on Consumer Electronics*, August, 357-358.

2. Emery, P.G. 1980 "Evaluating spoken English." *ELT Journal*, 34, 2, Jan.,96-98.

3. Nakayama, K., Tomita-Nakayama, K. and M. Misaki 1998 "Enhancing speech perception of Japanese learners of English utilizing time-scale modification of speech and related techniques." *Speech Technology in Language Learning*. KTH, 123-126.

4. Pimsleur, P. et al. 1977 "Speech Rate and Listening Comprehension," Burt, M. et al. eds. *Viewpoints on English as a Second Language: In Honor of James E Alatis*. New York: Regents, 27-34.

5. Conrad, R. 1989 "The Effects of Time-Compressed Speech on Native and EFL Listening Comprehension" *Studies in Second Language Acquisition Research*, 11, 1-16.

6. Kawai, H. and S. Yamamoto 1995 "An Evaluation of Speech Rate Conversion Technique Applied to English Speech," *Proc. ASJ '95*, Sep. 265-266.

7. Nakayama, K., Tomita, K. and M. Misaki 1998 in press "Expanded speech processing for on-line processing of English sentence: speech recognition and sentence length" *Congress Proc. ASJ.*