Effect of Speech Rate on the Intelligibility of Easy Japanese Speech under Reverberant Environment

Jiahe Wei
Graduate School of Engineering
Tohoku University
Sendai, Japan

Yuya Chiba
Graduate School of Engineering
Tohoku University
Sendai, Japan

Takashi Nose
Graduate School of Engineering
Tohoku University
Sendai, Japan

Akinori Ito
Graduate School of Engineering
Tohoku University
Sendai, Japan

***Abstract*—A lot of foreigners live in Japan nowadays. Under a special situation such as a disaster, it is important to find an easy way to inform them of various information in Japanese. This study evaluates the effect of variation in speech rate and Speech Transmission Index on intelligibility of Japanese synthesized speech to non-native listeners of Japanese. Non-native speakers of Japanese were assigned to listen to twenty-five synthesized speech related to disaster prevention under five speech rates: 320, 360, 400, 440, 480 (mora/min) and five reverberant environment. All of the speech were synthesized from twenty-five sentence with almost the same difficulty. The experiment showed that speech rate did not affect intelligibility of non-native listeners and the Speech Transmission Index affect intelligibility.**

Keywords- Easy Japanese; speech rate; Speech Tranmission index; speech synthesis; speech intelligibility

I. INTRODUCTION

With the process of economic globalization, now there are more and more people living in the foreign land. They leave their mother land to work, travel, settle down and so on. And also, this is happening in Japan. The number of medium and long term foreign residents in 2015 is about 2.12 million, and this number will continue to grow in the foreseeable future. The first and most important thing when they set foot on Japan is mastering the local language—Japanese.

In the daily life, the requirement for Japanese is not so high. It is fine that foreigners just know the daily languages. However, unlike other countries, Japan is in the earthquake zone. So it is much easier that earthquake, tsunami or volcanic eruption occurs in Japan. When the foreigners find themselves in a middle of a disaster, the method to obtain the necessary information for safety and survival in Japanese which you do not understand will be a big problem. To deal with the problem, there is an information system designed to boil down information available to all foreigners that will be necessary for survival in case of an emergency. This system is called "Easy Japanese"[1]. And to judge which Japanese sentence is easy, a system called YAsashii Nihongo SIen System (YANSIS) is developed, which can use several factors like the length of sentence, difficulty of words, number of foreign words to calculate the difficulty of sentences[2].

II. BACKGROUND

A. Research on the speech rate

When a disaster occurs, the most possible source of information foreigners can get is from television and radio. So the information that saves lives is not just some sentences wrote, it becomes to speech which contain variables like loudness, speech rate and pause.

There are several research about the effect on intelligibility of speech using Easy Japanese. Hafiyan did the research on the effect of speech rate and pause on perception of spoken Easy Japanese using speech synthesis [3]. He found adding pause in the right place of sentence can raise the intelligibility of speech. Then he used synthesized speech with speech rate from 200(mora/min) to 400(mora/min) to conduct experiment in order to find which speech rate have more intelligibility. However, the result of research shows the speech rate do not affect the speech intelligibility much.

The result on speech rate is a little contrary to our common sense, we often think the slower you speak, the easier others can hear the words. So I think maybe there are other factor that effect the intelligibility such as the difficulty of sentences, the proficiency of subjects, noise from environment.

So this time, I will use two factors speech rate and speech transmission quality to product experiment to find out if transmission quality can affect the speech intelligibility.

B. Introduce of Speech Transmission Index(STI)

In order to measure speech transmission quality, variables called STI is introduced [4]. The STI measures several physical characteristics of a transmission channel, to calculate the capability of transmission channel to carry speech signal.

TABLE I. STI AND SPEECH QUALITY

STI	0-0.3	0.3-0.45	0.45-0.6	0.6-0.75	0.75-1

Speech quality	BAD	POOR	FAIR	GOOD	EXCELLENT

Though there are many influence affecting transmission quality such as speech level, frequency response of the channel, non-linear distortion, quality of the sound reproduction equipment and so on, we can also use the impulse response of a reverberant room to calculate STI easily.

III. EXPERIMENT

From impulse response library [5], I downloaded several impulse responses under difficult environments, and test their STI respectively. Then from these responses I chosen four response (STI: 0.4, 0.56, 0.7, 0.82) and no response (STI: 1) as experiment variables. Add all five responses to the speech with five kinds of speech rate, with the combination of STI and speech rate, we will get twenty-five situations. Divide these situations to five groups, make sure every subject can hear all the STI and speech rate. The experimental order follows TABLE II, one color means the speech one subject can hear.

TABLE II. ORDER OF EXPERIMENT

STI	**Speech Rate**				
	320	360	400	440	480
1	a	e	d	c	b
0.82	b	a	e	d	c
0.7	c	b	a	e	d
0.56	d	c	b	a	e
0.4	e	d	c	b	a

For each subject, the speech played is random. After listening to a sentence, subject wrote down the sentence as they caught in Japanese.

IV. RESULT

A. Procedure of answers

The subject's answer is Japanese, and it contains kanji and hirakana. In order to facilitate the calculation, the original text and subject's answer are all converted to hirakana which is a Japanese syllabary can spell each Japanese word. Then I use Dynamic programming to calculate the longest common subsequence (LCS) between original text and answer. And score means the rate between length of longest common subsequence and length of original text.

$$\text{score} = \frac{L(lcs)}{L(original)} \quad (1)$$

The equation show the calculation of score, $L(lcs)$ denotes the length of longest common subsequence, and $L(original)$ denotes the length of original text.

B. Result

Fig. 1 shows the relationship between the STI and the average score of all subjects. When STI is 1 means no impulse response, the score is the highest show the best intelligibility, and when STI = 0.4 with a serious impulse response, the score is the lowest. And with the analysis of variance, the STI is shown a statistically significant ($p < 0.05$) with score.

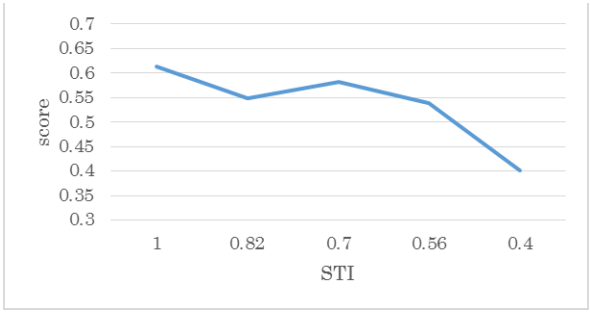

Figure 1. Relationship between STI and score

Fig. 2 shows the relationship between speech rate and average score of all subject. From Fig. 2 we cannot find out the relationship between every speech rate. And we cannot find statistically significant between speech rates.

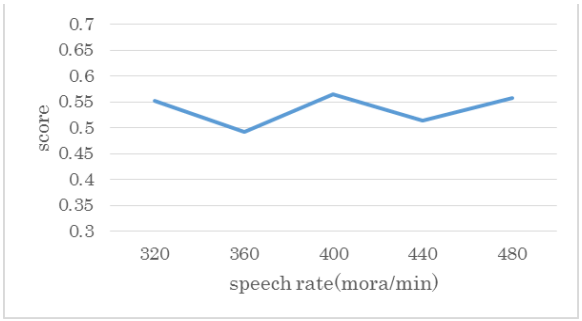

Figure 2. Relationship between speech rate and score

V. RESULT

In this research, I tried to examine the effect of speech rate and Speech Transmission Index to the intelligibility of Easy Japanese. I performed an experiment, and found out the worse transmission environment be, the worse intelligibility of speech be. However speech rate did not affect intelligibility of non-native listeners.

For the future research, we will analysis the answer with more details, such as analysis kanji instead of hirakana.

REFERENCES

[1] K.Sato, "Easy expressions save foreigners!" http://human.cc.hirosaki-u.ac.jp/kokugo/tagengoenglish.html

[2] A. Ito, A. Kashima, R. Maeda, Y. Mizuno, Y. Misono, M. Yoneda and K. Sato, "An Assistant System for Writing in Easy Japanese", Tohoku Section Joint Convention of Institutes of Electrical and Information Engineers, Japan, 2008, p. 299(in Japanese)

[3] H. Prafianto, T. Nose, Y. Chiba, A. Ito, K. Sato, "A Study on the Effect of Speech Rate on Perception of Spoken Easy Japanese Using Speech Synthesis", International Conference on Audio, Language and Image Processing (ICALIP), 2014

[4] H. J. M. Steeneken, T. Houtgast, "A physical method for measuring speech-transmission quality", Acoustical Society of America,1980

[5] The University of York, "Demo, download and share acoustic impulse responses for auralization" http://www.openairlib.net/