Speech Speed Awareness System Slows Down Native Speaker's Talk

Tomoo Inoue(✉) and Wei Liao

University of Tsukuba, Tsukuba, Ibaraki 3058550, Japan
inoue@slis.tsukuba.ac.jp

Abstract. Conversation between native speakers and non-native speakers is not always easy. Non-native speakers sometimes feel hard to understand what native speakers talk because of their speaking speed. Even kind native speakers who speak slower in the beginning often get back to their "natural" speed unconsciously after a while. Although non-native speakers may ask a native speaker to repeat what they cannot catch, participants of the conversation would become uneasy for too often request of repeating, which may also reduce productivity of the conversation or the meeting. In this paper, we study how conversation between native and non-native speakers is conducted regarding speech rate, how the introduction of a speech speed awareness system could alleviate this communication problem. The system recognizes speech rate in real-time and makes participants aware when their speech rate is too fast for non-native speakers.

Keywords: Second language conversation support · Speech speed · Non-Native speaker support

1 Introduction

We have more chances of conversation in second languages than ever recently, but conversation between native speakers (NS) and non-native speakers (NNS) is not always easy. The conversation is often unbalanced when NS communicate with NNS.

The reasons of NNS listening and understanding difficulties in second language communication have been investigated in numerous studies. According to the investigation on the factors affecting NNS listening by Bloomfield et al. [1], NNS listening is affected by listener's factors such as working memory and fluency of language, and speaker's factors such as word length, complexity, pose and the speech rate. Speech rate (SR) is an element of the speaker's factors. It is known that it may cause difficulty in the NNS listening when the average SR of the NS becomes fast [2, 3]. However, these did not investigate the real-time SR change of NS. Also, a speech speed awareness system to alleviate this communication problem has been developed, which recognizes the speech rate in real-time and make the participants aware when the speech rate is too fast [4], but its effect has not been published yet.

The purpose of this research is to support conversation by NS and NNS. First, it was examined if increase in the SR of the NS causes understanding difficulty of the NNS in actual second language conversation. Then the proposed speech speed

H. Nakanishi et al. (Eds.): CRIWG+CollabTech 2019, LNCS 11677, pp. 159–171, 2019.
https://doi.org/10.1007/978-3-030-28011-6_11

awareness method, which notifies the NS (and NNS) when his/her SR is too fast, and which expects spontaneous SR adjustment by NS [4], was investigated by the WOZ system.

2 Related Work

2.1 Problems of NNS Listening in Second Language Conversation

In the context of second language conversations by NS and NNS, conversational imbalance is often seen due to NNS listening and understanding problems. According to Ikegami, the general process of listening is a series of processing in which the working memory receives the spoken speech information, instantaneous processing is performed, and lead to the understanding of the information. The series of information processing differs depending on NS and NNS. In the case of a native language, processing from sound perception to word/phrase recognition is performed automatically and unconsciously. However, when NNS processes a non-native language, the process is consciously performed and takes more time [5].

In an investigation of factors affecting NNS listening, Bloomfield et al. pointed out the listener element of working memory, language fluency, and the speaker elements of paragraph length, complexity, pause, and SR. It is also suggested that the SR, which interacts with other factors, may lead to difficulties in NNS listening [1].

Goh surveyed on NNS listening. It showed variety of reasons that NNS felt difficulty of understanding in conversation with NS. They are as follows: lack of vocabulary, cannot recognize the words you know, cannot imagine even if you listen to words, cannot keep up the flow of conversation, the next part is gone while thinking about the meaning of the previous words, appearance of unexpected words could lead to confusion, and if something went wrong they would not be able to understand the following parts, or they would forget the part they heard earlier while listening [6].

In this study, we conducted experimental research to investigate the SR change of NS and NNS incomprehension in second language conversation. We collected the video of the experiment as objective data, as well as the subjective comments like these previous studies.

Regarding the behaviors in second language conversation, when NNS are unable to catch the words of NS, they tend to ask for a repeating. But it becomes psychological burden for NNS if they ask many times [7]. In addition, NS are unconscious of their fast-talking even when they notice their partners are NNS [8]. They may speak slowly at first, but fasten their SR unconsciously. Also known is that SR tends to change dramatically in daily conversation [9]. Thus SR is an important factor.

2.2 Speech Rate and NNS Listening

SR is known to affect the listener's comprehension, which is an important factor especially in second language conversation.

Griffiths conducted an experiment of second language listening. NNS listened the text recorded at three levels of SR of fast (200 wpm), normal (150 wpm), and slow (100 wpm), and were tested their comprehension. The text recorded by fast SR significantly

reduced the comprehension of NNS, but normal and slow SR had no significant difference on the influence on NNS listening comprehension [2]. Zhao prepared listening materials in normal SR, those in different SR, and those that NNS could freely control their SR, let NNS listen to each listening material, and investigated SR and NNS listening comprehension. It was shown that NNS often adjusted to a slower SR than normal SR when he/she could control it, resulting in better comprehension [3].

In the context of second language learning, normal and slow SR for NNS are effective for speech comprehension. However, it is not true that the slower the SR the better the listening comprehension. There is an optimal SR by NNS that helps understanding. Hayati divided 62 English learners into 2 groups, and distributed natural SR listening materials and slow SR listening materials to each group. The learners participated in English classes with the same content. They were given a pre-test before listening and a post-test two weeks later to measure their listening comprehension. As a result, both groups improved their listening comprehension, but the group receiving the natural SR listening material showed more progress [10].

Also, in a study on the effects of SR and noise on NNS listening, Jones et al., prepared a fast SR (155 wpm) and a normal SR (178 wpm) instructions of a banking product by using a text-to-speech system. A comparative study was performed using those and also those with a background noise. As a result, it was found that NNS listening comprehension had little to do with the noise. It was also shown that the understanding rate of the content decreased in the fast SR instructions [11].

Matsuura et al. investigated the influence of SR on NNS listening in an accented second language. It was shown that Japanese learners of English learn more efficiently by listening at a slower SR of accented English than by listening at the usual SR [12].

2.3 Second Language Conversation Support

Various methods and systems have been proposed to support conversation in a second language.

Inoue et al. proposed an NNS conversation support method that considered the burden of the NNS in conversation. During the second language conversation including the NNS in a remote setting, the NS inputs keywords and key phrases of the conversation from the keyboard and shares them on the NNS screen. The method was shown to have the effect of increasing understanding of the conversation and increasing the participants' mutual understanding [13].

Okamoto et al. developed a face-to-face cross-cultural communication support system to compensate for the differences in knowledge brought by the cultural background of different countries. This was a system that presented related information of the nouns in a conversation, which was retrieved by the Web. From the evaluation experiment, it was shown that presenting images, explanatory information in the native language, related nouns, and related images about the nouns spoken during the conversation to the user may support the dialogue [14]. However, due to the accuracy of speech recognition, the accuracy of the presented information may not be sufficient.

Fukushima et al. developed an interface that allowed different language debaters to input words in various languages, which were translated into a common language of the discussion in real time by a multi-lingual translation server [15]. Although it could

improve the accuracy of the discussion, it needed extra workers for text input besides discussion participants.

This research focuses the SR of NS in second language conversation. and supports it by a method that The proposed system detects the SR of NS in real time, and when the SR is judged to be too fast, notifies it to the participants as an awareness sign.

2.4 Conversation Support by Speech Rate Conversion

A SR conversion systems have been developed to solve the listening problem of the fast SR voice. Fujitsu developed a SR conversion technology that detected the voice interval and the silent interval, and extended the voice interval while shortening the silent interval, so that the SR of the receiving voice could be reduced. As a result of evaluation experiment on participants of all ages, it was confirmed that it increased the ease of listening regardless of the listener's age when the SR of the receiving voice is fast [16]. Kiyoyama et al. proposed a SR conversion method based on expansion and contraction including silent sections, and developed a small SR converter aimed at improving voice broadcasting service for the elderly [17].

A change in SR has the effect of improving the ease of listening in general, but SR conversion could cause time lags in free conversation scenes, and may cause a sense of discomfort in synchronization. In order to examine the influence of the difference between the speaker video and the speech on the word intelligibility, Tsumura et al. prepared a pair word stimulus of one phoneme difference in each mora out of 4 mora words, conducted a word intelligibility experiment under the conditions of "video only," "voice only," and "audio + video;" and examined whether audio-visual integration occurred at each mora position. As a result, when word speech was expanded to 400 ms or more and video with normal speed was added, it was found that the word intelligibility might be affected [18].

Therefore, this research does not apply SR conversion technology. Instead, we try to support listening of the conversation by the method that the speaker spontaneously speaks slowly by informing when the SR of NS is too fast.

3 Experiment on NNS Comprehension of NS Speech

We conducted a face-to-face conversation experiment using NS and NNS as a pair. During a 10-minute free conversation, when the NNS felt difficulty in the NS utterance, he/she pressed the hand-held button to record the time.

It has been pointed out that the complexity and pause of the words and the SR may cause listening difficulties in NNS, but when NS and NNS talk, the real-time SR change of NS is NNS It is unclear what kind of influence it has on understanding.

3.1 Participants

The participants were 15 pairs of Japanese NS (Japanese students) and Japanese NNS (international students) who met for the first time, for a total of 30 participants.

NS was a person who used Japanese as a native language, and NNS could speak Japanese (N2 or more) but not so fluent as NS, who sometimes felt incomprehension.

NNS was asked to make self-introduction and daily conversation with NS for about 5 min, and his/her Japanese level was confirmed based on the behavior when the NNS is troubled [19], asking the NS to listen back during the conversation, silencing to the NS utterance, or making a troubled expression or gesture.

3.2 Apparatus and Setup of the Experiment

The layout and the setup during the experiment are shown in Figs. 1 and 2. NS and NNS sat face to face at a distance of 120 cm.

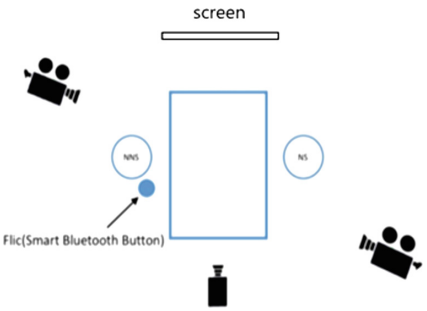

Fig. 1. Layout of the experiment.

Fig. 2. Snapshot of the experiment.

In order to keep track of when NNS felt misunderstanding, we handed Shortcut Labs' Bluetooth-connected push-button device Flic only to NNS (Fig. 3). By using the provided software development kit, the Rasberry Pi 3Model B with Bluetooth connection can record the time when Flic was pressed in a text file.

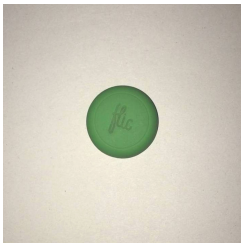

Fig. 3. Bluetooth-connected push-button device Flic.

3.3 Procedure of the Experiment

NNS was told to press once when NNS was incomprehensible in the conversation with NS, and handed the hand-held button. NNS held the button under the desk so NS would not notice. NS wore a microphone and recorded his speech. The experiment was videotaped. The pair talked freely for 10 min. After that, NNS and the experimenter watched the video recording of the conversation, and confirmed the timing of pressing the button and the reason.

3.4 Collected Data

The acquired data are the video and audio recording of the experiment, and the time log of the unrecognized utterance by the NNS button.

Time and Reason of Unrecognized Utterance. The experimenter with the NNS himself/herself, while playing back the video shot, confirmed to the NNS whether the timing recorded in the text file with the hand-held button matched the timing of the unrecognition, and corrected any mistake. We also identified the reasons for each of the incomprehension (based on the surveys of Bloomfield et al. [1] and Goh [6]). An excerpt is shown in Fig. 4.

#	#	Logged time	Incomprehension part	Reason
Pair13	1	16:06:47	"Stratum"	Unknown word
	2	16:07:09	"Information-related"	Too fast speech
	3	16:07:36	"…vertical…"	Too fast speech
	4	16:07:47	"Would you like it with your mouth"	Too fast speech
	5	16:07:56	"Sichuan"	Late recognition
	6	16:09:21	"Regent"	Unknown word
	7	16:09:34	"I want you to get stuck"	Unknown word/phrase
	8	16:10:46	"Is swimming in your school classes"	Too fast speech

Fig. 4. An excerpt from the sheet of time and reason of incomprehension in conversation.

3.5 Processing of the Collected Data

Labeling NS Utterance and NNS Unrecognition. For the experimental video data with a total of 150 min of speech from 10 min per the pair, the NS speech segment and the NS speech segment where NNS had incomprehension were labeled in the following procedure.

(1) Using the multimodal data annotation tool ELAN, we cut out the speech segments divided by silence of more than 200 ms.
(2) After automated processing by ELAN, manual inspection and correction were performed. Non-verbal speech segments such as laughter, cough and sigh were excluded from the speech segment.
(3) In order to analyze the speech rate of NS, only the speech segments of NS were left.
(4) Every segment of NS speech was transcribed.
(5) We identified and labeled NS speech segments that NNS could not catch (Fig. 5).

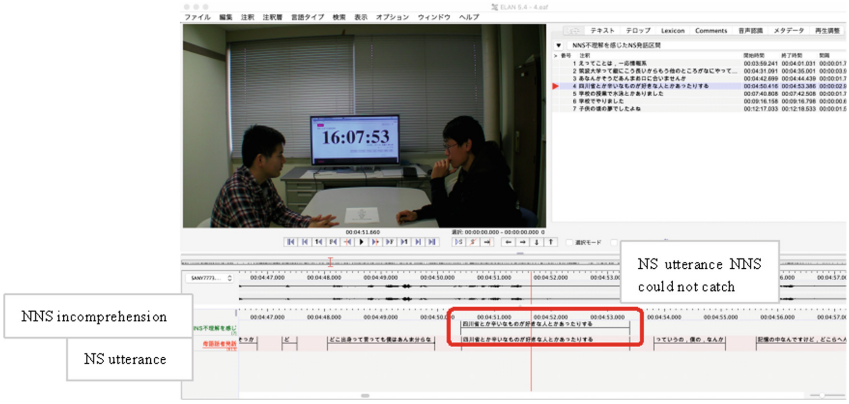

Fig. 5. NS speech segments that NNS could not catch were labeled.

Calculation of SR of NS. Although the SR to listen was constant in the conventional study of SR of NS and NNS understanding, the actual speech does not always go at the same speed. Marushima measured speed changes during speech and showed that the speed of natural speech was constantly changing [9]. Therefore, in order to investigate the relationship between SR and the degree of comprehension of the utterance content, it is considered necessary to investigate the SR in the vicinity where the incomprehension arose and the change thereof.

In the experiment, it was found that the button press of the NNS occurred during or immediately after the incomprehensible speech (Fig. 6). Therefore, with regard to the SR of NS at the time of incomprehension, we calculated two SRs in the speech segment that became incomprehensible and the speech segment immediately before that, and examined the change. The SR was defined using the number of syllables per second in the speech segment.

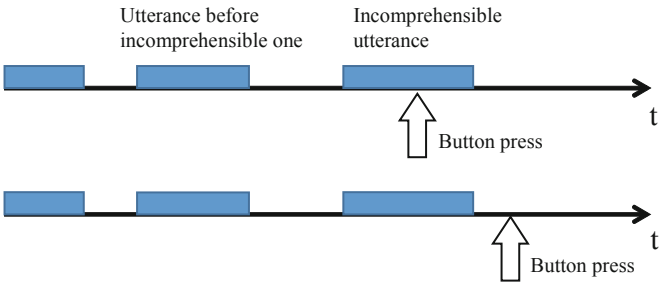

Fig. 6. Button press of the NNS occurred during or immediately after the incomprehensible speech.

3.6 Result

The Reasons of NNS Incomprehension. The reasons of NNS incomprehension of NS speech are shown in Table 1.

Table 1. The reasons of NNS incomprehension of NS speech.

Factor	Reason	# Occurred
Hearer's	Lack of vocabulary	37
	Slow recognition and response to the comprehensible words	25
	Could not keep up with the flow of conversation	2
	Stuck in the previous word	2
Speaker's	Too fast SR	18
	Loudness, crispness	5
Composite	Includes too fast SR	6
	Not include too fast SR	7
Other	Did not hear	1
	Do not know	12
Total		115

From the total of 115 incomprehensible speech segments in 150 min of conversation, 24 cases were because of too fast SR of NS, which comprises 21% of all the incomprehensible cases.

Therefore alleviating the problem of fast SR of NS will contribute to better NNS comprehension.

Change of SR of NS. The change of SR of NS when NNS faced incomprehension was analyzed. Seventy-five cases were used from the total of 115 cases. The cases by the reasons clearly unrelated to the SR, which were "unknown vocabulary" and "did not hear," were excluded. The cases that could not transcribe were also excluded. The SR of the speech segment that became incomprehensible and that of the speech segment

immediately before that were shown in Fig. 7. The SR of NS speech that NNS could not hear was an average of 7.85 syllables/second (SD = 1.96), and the SR of NS speech just before its incomprehension was an average of 7.15 syllables/second (SD = 2.14), showing the significant difference ($t(74) = 2.231$, $p < .05$). This shows that the increase of SR of NS during conversation actually occurs when NNS faces incomprehension of NS speech.

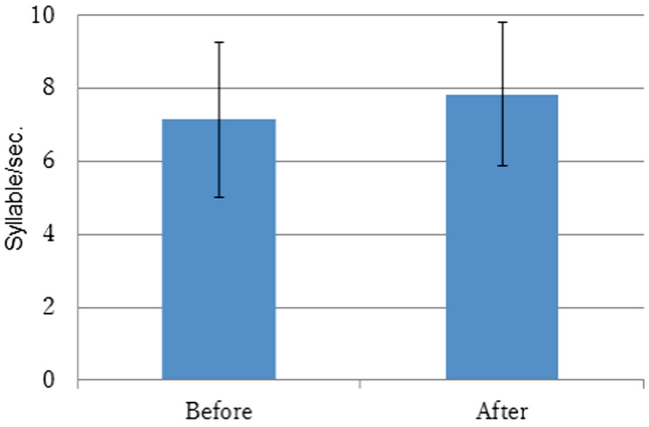

Fig. 7. Change of NS speech speed when NNS incomprehension occurred.

4 An Experiment of Speech Speed Awareness Method

4.1 Speech Speed Awareness Method

Considering the influence of SR in the second language conversation that NS fast-forward is one of the factors that make NNS' comprehension difficult, a method to inform NS fast-forward in real time and encourage spontaneous adjustment of NS has been proposed [4]. In this method, the NNS is assumed those who can make conversation to some degree with the second language used, but who is not as fluent as NS. It is often the case in international business settings in recent years, and thus this assumption is realistic.

4.2 Speech Speed Awareness System by WOZ

The Speech Speed Awareness system by WOZ method was designed, and its effect was evaluated. The flowchart diagram of the system is shown in Fig. 8. To study the effect of the method, WOZ system was used to avoid the influence on the result of other factors than the method such as accuracy of speech recognition.

The system displayed the red screen with the text "TOO FAST" when SR of NS becomes faster than the preset speed that NNS felt comfortable. Its display turned green with the text "GOOD" when SR of NS becomes slower.

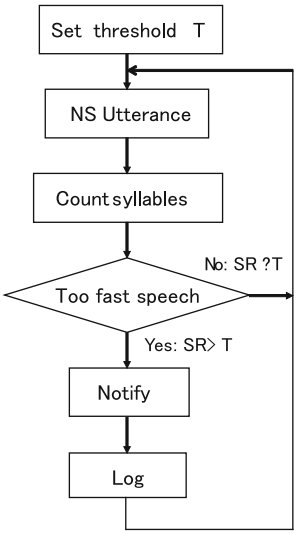

Fig. 8. Flowchart of the system.

4.3 Participants

The participants were 20 pairs of Japanese NS (Japanese students) and Japanese NNS (international students) who met for the first time, for a total of 40 participants.

NS was a person who used Japanese as a native language, and NNS could speak Japanese (N2 or more) but not so fluent as NS, who sometimes felt incomprehension.

NNS was asked to make self-introduction and daily conversation with NS for about 5 min, and his/her Japanese level was confirmed based on the behavior when the NNS is troubled [19], asking the NS to listen back during the conversation, silencing to the NS utterance, or making a troubled expression or gesture.

4.4 Procedure

NNS prepared the threshold SR of notification by the system, by listening sample speeches of various speed and choosing comfortable one. NS wore a microphone and recorded his/her speech. The experiment was videotaped. Each NS and NNS pair talked 5 min. The experiment is shown in Fig. 9. The acquired data were the video and audio recording of the experiment.

4.5 Processing of the Collected Data

For the experiment video data with a total of 100 min of speech from 5 min per a pair, the NS speech segment and the screen notification section by the system were labeled in the following procedure.

(1) Using the multimodal data annotation tool ELAN, we cut out the speech segments divided by silence of more than 200 ms.

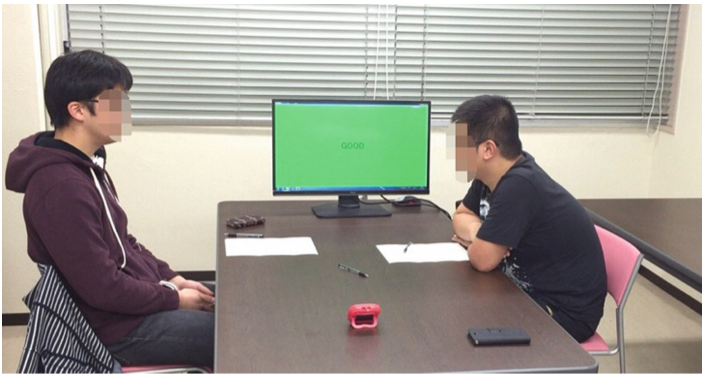

Fig. 9. Snapshot of the WOZ system experiment.

(2) After automated processing by ELAN, manual inspection and correction were performed. Non-verbal speech segments such as laughter, cough and sigh were excluded from the speech segment.
(3) In order to analyze the SR of NS, only the speech segments of NS were left.
(4) We identified and labeled the screen notification sections.
(5) NS speech segments before and after the screen notification were transcribed (Fig. 10).

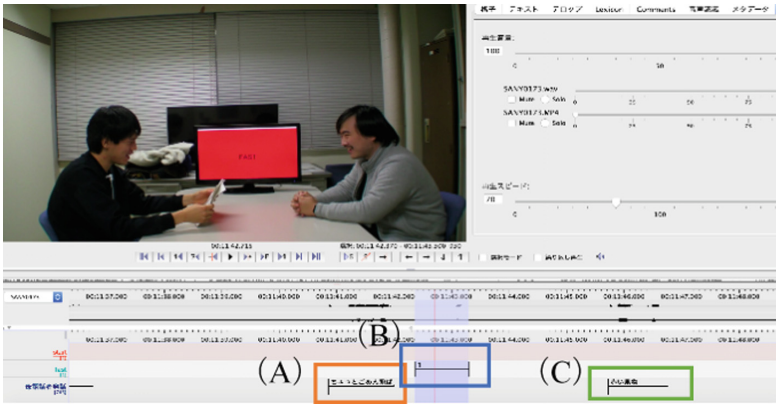

Fig. 10. Labeling a screen notification section and NS speech segments before and after that. (A) NS speech before the screen notification, (B) The screen notification, (C) NS speech after the screen notification.

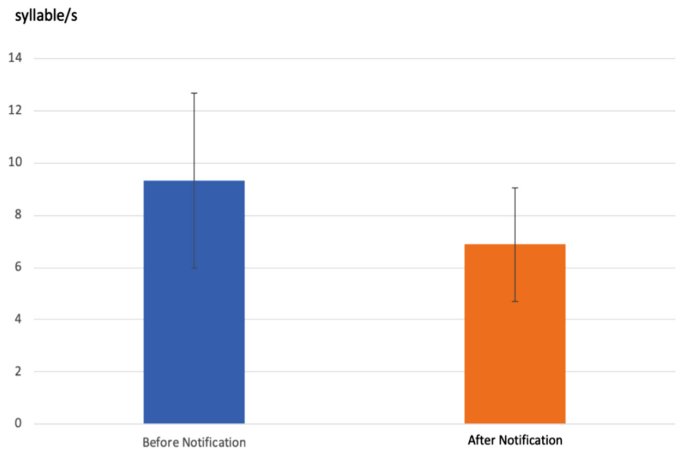

Fig. 11. Change in NS speech speed by the screen notification of too fast NS speech.

4.6 Result

The change in SR of NS by the screen notification of too fast NS speech was analyzed. Fifty-three cases were used from the total of 63 cases in the total of 100 min. conversation. The cases that the experimenter (wizard) made wrong judgement of SR over threshold were excluded. The SR of the speech segment immediately before the screen notification and that of the speech segment immediately after that were shown in Fig. 11. The SR of NS speech immediately before the screen notification was an average of 9.37 syllables/second (SD = 3.35), and the SR of NS speech immediately after the screen notification was an average of 6.87 syllables/second (SD = 2.18), showing the significant difference ($Z = -5.759$, $p = 0.000 < .05$). This shows that the notification of too fast speech of NS actually prompted NS slow down.

It was proved that Speech Speed Awareness system by WOZ was effective in reducing SR of NS when it became too fast for NNS.

5 Conclusion

Conversation or meeting by NS and NNS could possibly have unique value considering the viewpoints brought in from such diversified participants. However, conversation between NS and NNS is not always easy in reality. NNS sometimes feel hard to understand NS talk because of their speech rate. In this paper, we study conversation between NS and NNS regarding SR of NS. The first experiment revealed that dynamic SR change of NS actually occurred and that it caused incomprehension of NNS. The second experiment indicated the introduction of a speech speed awareness system, which made aware of SR of NS to him/herself, could actually slow down SR of NS.

The effect by the implemented speech speed awareness system will be investigated in the future. Also, influence on NNS comprehension and mutual understanding between NS and NNS will be investigated.

References

1. Bloomfield, A., et al.: What makes listening difficult? Factors affecting second language listening comprehension. Maryland Univ College Park (2010)
2. Griffiths, R.: Speech rate and NNS comprehension: A preliminary study in time benefit analysis. Lang. Learn. **40**(3), 311–336 (1990)
3. Zhao, Y.: The effects of listeners' control of speech rate on second language comprehension. Appl. Linguist. **18**(1), 49–68 (1997)
4. Jing, Y., Tomoo, I.: A speech speed awareness system for non-native speakers. In: Proceedings of the 19th ACM Conference on Computer Supported Cooperative Work and Social Computing Companion (CSCW 2016 Companion), pp. 49–52 (2016)
5. Ikegami, M.: The relationship between the stage of development of student's listening comprehension skills and the effects of pauses and speech speed. Stud. Lang. Lit. **32**(1-1), 59–88 (2012)
6. Goh, C.C.: A cognitive perspective on language learners' listening comprehension problems. System **28**(1), 55–75 (2000)
7. Yanagimachi, T., Nagano, K., Enai, M., Baba, N.: Communication Problems and Solutions for International Students Speaking in Japanese. Jpn. Soc. Eng. Educ. **61**(4), 43–47 (2013)
8. Sakuma, A.: Various aspects of crosscultural communication: concerning elements other than language. Bull. Nagoya Bunri Univ. **3**, 13–21 (2003)
9. Marushima, A.: Perception of gradually changing speech rate. Res. Exp. Phon. Linguist. **4**, 1–21 (2012)
10. Hayati, A.: The effect of speech rate on listening comprehension of EFL learners. Creat. Educ. **1**, 107–114 (2010)
11. Jones, C., Berry, L., Stevens, C.: Synthesized speech intelligibility and persuasion: Speech rate and non-native listeners. Comput. Speech Lang. **21**(4), 641–651 (2007)
12. Matsuura, H., et al.: Accent and speech rate effects in English as a lingua franca. System **46**, 143–150 (2014)
13. Hanawa, H., Song, X., Inoue, T.,: Keyword generation by native speaker is quick and useful in conversation between native and non-native speaker. In: Proceedings of the 2017 IEEE 21st International Conference on Computer Supported Cooperative Work in Design (CSCWD 2017), pp. 145–150 (2017)
14. Okamoto, K., Yoshino, T.: Development and evaluation of face-to-face intercultural communication support system using related information of nouns in conversation. J. Inf. Process. **52**(3), 1213–1223 (2011)
15. Fukushima, T., Yoshino, T., Kita, C.: Development of non-native language user support system PaneLive at face-to-face discussion using common language. IEICE Trans. Inf. Syst. **J92-D**(6), 719–728 (2009)
16. Togawa, T., Otani, T., Suzuki, K.: Speech enhancement technology, speech rate control technology. J. Inst. Electron. Inf. Commun. Eng. **96**(11), 874–881 (2013)
17. Seiyama, N., Imai, A., Mishima, T., Takagi, T., Miyasaka, E.: Development of a high-quality real-time speech rate conversion system. Trans. Inst. Electron. Inf. Commun. Eng. **J84-D-II**(6), 918–926 (2001)
18. Tsumura, K., Tanaka, A., Sakamoto, S., Suzuki, Y.-i.: Effect of audio-visual asynchrony by speech-rate conversion on word recognition. IEICE Tech. Rep. **105**(479), 103–108 (2005)
19. Ozaki, A.: Use and avoidance of clarification requests by Brazilians in Japanese contact situations. Jpn. J. Lang. Soc. **4**(1), 81–90 (2001)