PROCEEDINGS of the
23rd International Congress on Acoustics

9 to 13 September 2019 in Aachen, Germany

Reverberation-induced speech improves intelligibility in reverberation: Effects of taker gender and speaking rate

Nao Hodoshima[1]

[1] Tokai University, Japan

ABSTRACT

Humans adjust their speech when speaking in noise to enhance their intelligibility (known as the Lombard effect). Speech spoken through noise (noise-induced speech) is generally more intelligible than that spoken in quiet when heard in noise. The author has reported that "reverberation-induced speech" also improves speech intelligibility in reverberation, although the masking of noise and reverberation differed. The present study investigates whether reverberation-induced speech intelligibility depends on the gender and speaking rate of talkers. Four talkers (two males and two females) spoke under quiet (Q) and reverberation (R) conditions. In R, the reverberant speech was fed back to the talkers via headphones. Then speaking rate was adjusted to 1.2 (fast), 1.0 (original), and 0.8 (slow). 18 participants performed word identification tests in reverberation (reverberation time of 3.6 s). The results showed that speech was significantly more intelligible under R than Q, male talkers were significantly more intelligible than females, and slow speech was significantly more intelligible than fast speech. And the improvement in speech intelligibility under R compared with Q was not significantly affected by the speaking rate. These results suggest that reverberation-induced speech regardless of the speaking rate might increase intelligibility of announcements in public spaces such as airports.

Keywords: Reverberation, Speech production, Lombard effect, Speech intelligibility, Speaking rate

1. INTRODUCTION

Noise and reverberation generally degrade speech intelligibility compared with quiet environments. In addition, speech intelligibility in noisy or reverberant environments is generally lower for older adults and people with hearing impairments when compared with young adults without hearing loss [1]. The population of older adults is rapidly increasing. For example, as of 2018, people older than 65 years comprised 27.7% of the total population in Japan [2]. Therefore, care must be taken when broadcasting announcements in public spaces (e.g., train stations and airports), especially in emergency situations.

During speech communication, humans have a natural tendency to modify their speech in order to make it more robust against noise. This is widely known as the Lombard effect [3]. When compared with speech spoken in quiet environments, intensity, duration, fundamental frequency, and first formant frequency are increased in speech spoken in noisy environments [4, 5]. In addition, speech spoken in noisy environments has higher word intelligibility than that spoken in quiet environments, as judged by young adults in noisy environments [3–5].

In reverberant environments, similar results have been observed, although with noise and reverberation masking patterns that are temporally and spectrally different (noise masks speech simultaneously, while overlap-masking occurs in reverberation [6]). When compared with speech spoken in quiet environments, the intensity, fundamental frequency, and first formant frequency have been found to be increased in speech spoken in reverberation [7]. When participants listen to speech spoken in reverberant environments, it has been reported to be more intelligible for young and older adults compared with speech spoken in a quiet environment [8, 9].

Speaking slowly generally increases speech intelligibility, especially for older adults, because cognitive aging might result from a general slowing of processing speed [10]. Time-compressed speech (with a time compression rate of 40%) results in lower intelligibility for older than for younger adults in noisy or reverberant environments [11]. Significant interaction has been observed between

[1] hodoshima@tokai-u.jp

speech expansion and pause expansion, and higher speech intelligibility has been obtained with a short time expansion (100 ms) compared with non-expanded speech when the pause between phrases was 300 and 400 ms for both young and older adults in a noisy environment [12]. Slowing the speaking rate (with an expansion rate of 120%) has been shown to improve speech intelligibility in a reverberant environment [13]. However, to our knowledge, no studies have investigated whether the speaking rate affects the intelligibility of speech spoken in a reverberant environment.

Speech intelligibility differs by gender depending on several different experimental conditions. In a previous study, female talkers (5 talkers) had higher correct word rate than male talkers (5 talkers) in a noisy environment, regardless of listener gender [14]. In addition, vowel intelligibility in a noisy environment varied widely among 41 talkers for normal and clear speech (clearly articulated speech), and female talkers showed a larger clear speech vowel intelligibility benefit than male talkers [15]. A male talker enunciating in either a normal or an urgent style resulted in a higher correct and faster identification rate in a noisy environment than a female talker [16]. Therefore, talker gender may affect the intelligibility of speech spoken in a reverberant environment.

The aim of the present study was to investigate whether the speech intelligibility of speech spoken in reverberation depends on the talker and speaking rate, with the goal of making spoken announcements in public spaces more intelligible. We would like to accomplish this by modifying the input of broadcasting announcements from loudspeakers (e.g., broadcasting speech spoken in noise or reverberation), such that architectural acoustical and/or electroacoustical treatments are not required in public spaces.

2. EXPERIMENT

2.1 Participants

The participants were 18 native speakers of Japanese aged 21–24 years with self-reported normal hearing.

2.2 Stimuli

Four native speakers of Japanese (two males [M1, M2] and two females [F1, F2], aged 21–22 years) who self-reported no hearing or voice disorders served as talkers.

The speech materials consisted of 50 target words embedded in a carrier sentence. The target words were four morae (a phonological syllable-like unit in Japanese) that were selected from a database of familiarity-controlled Japanese word lists (FW03) [17]. Word familiarity in this study was chosen between 2.5 and 4.0 on a 7-point scale (1 for the least familiar and 7 for the most familiar).

The speech materials were recorded on a computer through a microphone (SHURE KSM141; condenser, cardioid) and a digital audio interface (TASCAM US-144MKII) in a sound-treated room (see Figure 1) at a sampling frequency of 44,100 Hz. Two speaking conditions were used in the recording: quiet and reverberation. Under the reverberation condition, talker utterances were convolved by an impulse response recorded in a church, and reverberant sounds were fed to the talkers through headphones (SENNHEISER HDA200; dynamic, closed circumaural type). The playback level was set to –10 dB relative to the speaking level at the talker's ears. The reverberation time was 3.6 s for octave bands from 125 to 4000 Hz. The recording was controlled by Adobe Audition 3.0 (the delay time caused by the software was less than a few ms). The talkers were asked to imagine that their speech was being broadcast into a public space with the same room acoustics they were experiencing through the headphones, and instructed to speak as clearly as possible.

After the recording was finished, one carrier sentence was chosen, and the target words with fixed preceding and following pauses were embedded in the carrier sentence for each speaking condition and each talker. This was done to control the effect of overlap-masking on the target words. Next, the intensity ratio of the carrier sentence relative to the target word was normalized across each speaking condition and speaker.

Then, the speaking rate was changed to slow (70% of the original speaking rate) and fast (120% of the original speaking rate) using the Praat software, which applies a pitch-synchronous-overlap-add (PSOLA) method for time-scale modification. The PSOLA method can modify the speaking rate without changing the fundamental or formant frequencies of the original speech [18].

Finally, the concatenated speech sounds were convolved with the impulse response, which was the same as that used in the recording. The overall intensity of the stimuli was normalized across speaking conditions and talkers. The total number of stimuli was 1200 (two speaking conditions × three

speaking rates × four talkers × 50 sentences). Table 1 shows the experimental conditions used in the listening test.

2.3 Procedures

The listening test was carried out in a sound-treated room. Stimuli were presented to each participant diotically over headphones (SENNHEISER HDA 200) through a digital audio interface (TASCAM US-144MKII) connected to a computer. Two practice trials were held to familiarize the participants with the experimental procedure.

The playback level was adjusted to each participant's comfort level. In each trial, a stimulus was presented once, and the participants were instructed to write down what they heard as a target word on their answer sheets. For each participant, 48 stimuli (two speaking conditions × three speaking rates × two talkers × four sentences) were presented randomly. The target word and listening condition combinations were counter-balanced across the participants.

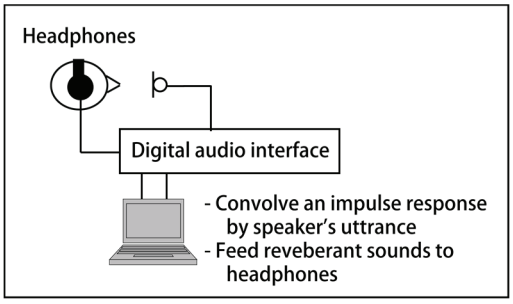

Figure 1 – Recording setup

Table 1 – Experimental conditions

Speaking rate	Speaking condition	
	Quiet	Reverberation
Original	Q	R
Fast	QF	RF
Slow	QS	RS

3. RESULTS AND DISCUSSION

Figure 2 shows the mean percent correct mora rates of the target words for each condition and talker. Statistical analyses were carried out using IBM SPSS Statistics. A mixed analysis of variance was carried out with the speaking condition (Q and R) and speaking rate (original, fast and slow) as repeated variables, the talkers as the between-subjects variable, and the correct mora rate as the dependent variable.

The main effect of the speaking condition was significant ($p < 0.01$). This result indicates that reverberation-induced speech was more intelligible than speech spoken in a quiet environment when the listeners listened in a reverberant environment, which is consistent with the results of previous studies [8, 9].

As expected, the main effect of the speaking rate was significant ($p < 0.01$), and a post hoc test revealed that the normal speaking rate had higher speech intelligibility than the fast speaking rate ($p = 0.003$), the slow speaking rate had higher speech intelligibility than the normal speaking rate ($p = 0.016$), and the slow speaking rate had higher speech intelligibility than the fast speaking rate ($p < 0.01$). These results are consistent with those of a previous study [13] (both of these studies carried out listening tests in a reverberant environment).

The talker effect was significant ($p < 0.01$), and a post hoc test revealed that F1 had a significantly lower correct rate than the other talkers ($p < 0.01$). These results suggested that the reverberant-induced speech was clearly dependent on the talkers. This talker-dependent intelligibility is consistent with normal speech [14], clear speech [15], and urgent speech [16], whereas the listening environment

was different (all previous studies [14–16] were conducted in a noisy environment, whereas the present study was carried out in a reverberant environment). When merging the results of the four talkers into gender, male talkers were more intelligible than female talkers ($p < 0.01$). However, the correct rates for the quiet condition for F1 appeared to be considerably lower than the other conditions and talkers. Therefore, it remains unclear whether the speech intelligibility of reverberation-induced speech depends on gender; future research should investigate this question further.

The interaction between the speaking condition and the speaking rate was not significant ($p = 0.066$). This result suggests that reverberation-induced speech increased speech intelligibility in a reverberant environment, regardless of the speaking rate.

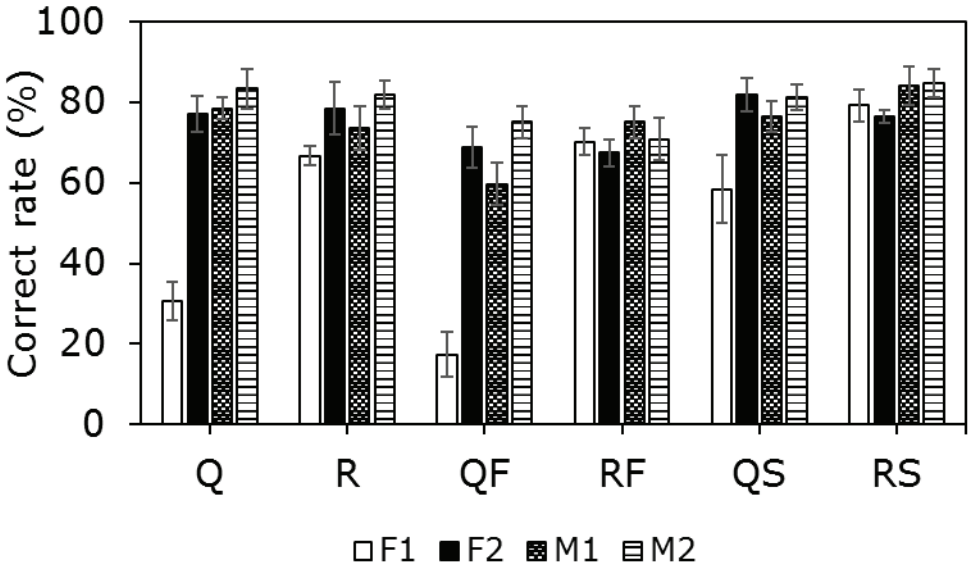

Figure 2 – Mean correct mora rate and standard error of target words for each condition and talker

4. CONCLUSIONS

The present study carried out a listening test to investigate whether the intelligibility of reverberation-induced speech differed in terms of the speaking rate and talkers. The results of listening tests in a reverberant environment obtained from 18 participants showed that reverberant-induced speech was significantly more intelligible than that spoken in a quiet environment. The results also showed that slow speech was significantly more intelligible than normal and fast speech. These results were consistent with those from previous studies. The results also showed that male talkers were significantly more intelligible than female talkers; however, the results also suggest that speech intelligibility is more talker-dependent. The improvement in speech intelligibility under the reverberation-induced speech condition compared with the quiet condition was not significantly affected by the speaking rate.

Since this was a preliminary study to determine the effects of reverberant-induced speech in different speaking rates and talkers, further research is needed with more talkers and a wider range of reverberant conditions. In addition, acoustical analyses of reverberant-induced speech should be carried out.

The results of the present study revealed that reverberation-induced speech might increase the intelligibility of announcements in public spaces such as airports, regardless of the speaking rate. Further research is needed to determine the appropriate combinations of the speaking rate and talker characteristics to improve further the intelligibility of emergency public-address announcements.

ACKNOWLEDGEMENTS

This work was supported by a Grant-in-Aid for Scientific Research (C) from the Japan Society for the Promotion of Science (18991829). We are grateful to Hideki Tachibana, Kanako Ueno, and Sakae

Yokoyama for providing the impulse response data, and to Hikaru Kawakami for carrying out the listening test.

REFERENCES

1. Nablek A. K. and Robinson P. K., “Monaural and binaural speech perception in reverberation for listeners of various ages”, J. Acoust. Soc. Am., 71, 1242-1248, 1982.
2. The Cabinet Office, “Annual Report on the Aging Society”, Japan, 2018.
3. Lane H. and Tranel B., “The Lombard sign and the role of hearing in speech”, J. Speech Hear. Res., 14, 677-709, 1971.
4. Van Summers W., Pisoni D. B., Bernacki R. H., Pedlow R. I. and Stokes M. A., “Effects of noise on speech production: Acoustics and perceptual analysis”, J. Acoust. Soc. Am., 84, 917-928, 1988.
5. Junqua J. C., “The Lombard reflex and its role on human listeners and automatic speech recognizers”, J. Acoust. Soc. Am., 93, 510-524, 1993.
6. Hodoshima N., Arai T. and Kurisu K., “Speaker variabilities of speech in noise and reverberation”, IEICE Technical Report, SP2009-69, 43-48, 2009. (in Japanese)
7. Nabelek, A. K., Letowski, T. R. and Tucker, F. M., ‘‘Reverberant overlap- and self-masking in consonant identification’’, J. Acoust. Soc. Am., 86, 1259–1265, 1989.
8. Hodoshima N., Arai T. and Kurisu K., “Intelligibility of speech spoken in noise and reverberation”, Proc. International Congress on Acoustics (paper ID: 663), 2010.
9. Hodoshima N., Arai T. and Kurisu K., “Intelligibility of speech spoken in noise/reverberation for older adults in reverberant environments”, Proc. Interspeech (paper ID: P6a.06), 2012.
10. Salthouse, T. A., “Theoretical perspectives on cognitive aging”, Erlbaum, Hillsdale, NJ, 1991.
11. Gordon-Salant, S. and Fitzgibbons, P. J. “Recognition of multiply degraded speech by young and elderly listeners”, J. Speech Hear. Res., 38, 1150-1156, 1995.
12. Tanaka, A., Sakamoto, S. and Suzuki, Y. “Effects of pause duration and speech rate on sentence intelligibility in younger and older adult listeners”, Acoust. Sci. Tech., 32(6), 264-267, 2011.
13. Arai, T., Nakata, Y., Hodoshima, N., and Kurisu, K. “Decreasing speaking-rate with steady-state suppression to improve speech intelligibility in reverberant environments.” Acoust. Sci. Tech, 28(4), 282-285, 2007.
14. Yoho, S. E., Borrie, S. A., Barrett, T. S. and Whittaker, D. B. “Are there sex effects for speech intelligibility in American English? Examining the influence of talker, listener, and methodology”, Atten Percept Psycho, 81, 558–570, 2019.
15. Ferguson, S. H. “Talker differences in clear and conversational speech: Vowel intelligibility for normal-hearing listeners”, J. Acoust. Soc. Am., 116(4), 2365-2373, 2004.
16. Arrabito, G. R. “Effects of talker sex and voice style of verbal cockpit warnings on performance”, Hum Factors, 51(1), 3-20, 2009.
17. Amano S., Kondo T., Sakamoto S. and Suzuki Y., “Familiarity-controlled word lists 2003 (FW03)”, The Speech Resources Consortium, National Institute of Informatics in Japan, 2006.
18. Moulines, E. and Charpentier, F., “Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones”, Speech Commun, 9, 453–467, 1990.