Effect of talker variability on speech perception by elderly people in reverberation

NAO HODOSHIMA AND TAKAYUKI ARAI

Department of Electrical and Electronics Engineering, Sophia University, 7-1 Kioicho, Chiyodaku, Tokyo, 102-8554, Japan

It has been reported that elderly people have much more difficulty in perceiving speech in reverberation compared to young people with normal hearing. This study investigated how characteristics of talker's speech affected speech perception of elderly people in reverberation to find speech materials that are easier to hear for elderly people in reverberation. In order to simulate various characteristics of talkers' speech, sentences were produced by people who have received speech training or not with different speaking rates and styles. Stimuli were prepared by convolving the sentences with impulse responses from rooms, and nonsense word identification test was carried out with Japanese elderly people in a diotic listening condition. The results of this study would provide some characteristics of speech materials that are robust to reverberation for elderly people, and those characteristics would be particularly useful in the situation where perfect speech communication is required such as listening to a speech alarm.

INTRODUCTION

Reverberation makes speech perception difficult, especially for elderly people and people with hearing impairments compared to young people with normal hearing (Nábělek and Robinson, 1982). With the increase in the population of elderly people, there is a growing need for achieving appropriate speech communication of elderly people by using speech materials that are easier to hear for elderly people in reverberation and/or by reducing the effect of reverberation in public spaces.

Several studies has been reported to improve speech intelligibility in reverberation and/or noise: one focuses on a speech signal such as an electroacoutical approach and an other focuses on speech production such as clear speech and slowed speaking rate. An example of an electroacoustical approach is a preprocessing technique, which processes a speech signal before it is radiated from loudspeakers (Arai *et al.*, 2002; Hodoshima *et al.*, 2006). When we focus on speech production, clear speech had higher word intelligibility than normal speech for people with hearing impairments in noise and reverberation (Payton *et al.*, 1994; Caissie *et al.*, 2005). Slowed speaking rate had higher word intelligibility than normal speaking rate for elderly people in noise (Sommers, 1997). However, few researches have been done on the effects of clear speech and slowed speaking rate for elderly people in reverberation.

The goal of this study is to find speech materials (e.g., a speech alarm and an announcement in train stations) which are easier to hear for elderly people in reverberation.

Auditory signal processing in hearing-impaired listeners. 1st International Symposium on Auditory and Audiological Research (ISAAR 2007). T. Dau, J. M. Buchholz, J. M. Harte, T. U. Christiansen (Eds.). ISBN: 87-990013-1-4. Print: Centertryk A/S.

This paper investigated the effects of speech training, clear speech and slowed speaking rate on elderly people in reverberation. We have four hypotheses: 1) less reverberation is more intelligible than much reverberation, 2) speech of talkers who have received speech training (trained talkers) is more intelligible than speech of those who have not (untrained talkers), 3) clear speech is more intelligible than normal speech, and 4) slowed speaking rate is more intelligible than normal speaking rate. To test these hypotheses, a listening test was carried out with elderly people in reverberation.

LISTENING TEST

Participants

21 elderly people (six males, 15 females and 74 years old on average) participated in this listening test. Table 1 shows air-conduction thresholds of the participants. None of them wore hearing aids.

Frequency (Hz)	125	250	500	1k	2k	4k	8k
dBHL	27	24	23	25	30	36	60

Table 1: Air-conduction thresholds of the participants.

Stimuli

The speech materials consisted of nonsense 20 Japanese vowel-consonant-vowel (VCV) words as targets embedded in a Japanese carrier phrase. The same vowel was used as the initial and the final vowels. All possible VCV combinations were selected from /p, t, k, b, d, g, s, S, h, z, Z,m, n/ and /a, i/ excluding those that do not meet Japanese phonotactics. Four trained talkers (T1-4: two males, two females and 28 years old on average) and three untrained talkers (T5-7: one male, two females and 23 years old on average) produced each speech material in normal speaking style (N) and clear speaking style (CL). All talkers were instructed to speak the both speaking styles in the same speaking rate (SR1: an average speaking rate was six mora/s). Trained talkers are professional announcers or have been to an announcer school for a year and more. All talkers had no articulation disorders. The recording was made using a DAT recorder (SONY, TDC-D10) at a sampling frequency of 16000Hz with a microphone (SONY, ECM-MS967) in a sound treated room.

Two speaking rates were prepared: original (SR1) and slow (SR2: an average speaking rate was five mora/s). Slowed speaking rate was manipulated by PRAAT [Praat Homepage].

Two reverberant conditions were prepared: an impulse response measured in a multiple-purpose hall (IR1: reverberation time of 1.1 s) and an impulse response which was derived from by changing an exponential decay of IR1 (IR2: reverberation time of 1.8 s). Reverberation times are derived from early decay time at the center frequencies of 0.5, 1, and 2 kHz of an 1-octave bandpassed impulse response.

A total of 1120 stimuli (seven talkers x two speaking styles x two speaking rates x two reverberant conditions x 20 speech materials) were prepared. The A-weighted energy was set equal for speech materials. See Table 2 for conditions used in this study.

talker	speaking style	speaking rate	reverberation time
trained (T1-4)	normal (N)	original (SR1)	1.1 s (R1)
untrained (T5-7)	clear (CL)	slow (SR2)	1.8 s (R2)

Table 2: Conditions used in this study.

Procedure

The listening test was carried out in a sound treated room. Before starting a main session which presents the 320 stimuli for two talkers, participants had practice trials to become familiar with the procedure. The sound level was adjusted to a comfortable level for each participant during the practice session, and the level was maintained throughout the main session.

In any given trial, a stimulus was presented over headphones diotically (STAX, SR-303). Participants were instructed to write down the target word in Kana orthography they heard on an answer sheet from the 20 VCVs used in the listening test. Stimuli were randomly presented for each participant.

RESULTS

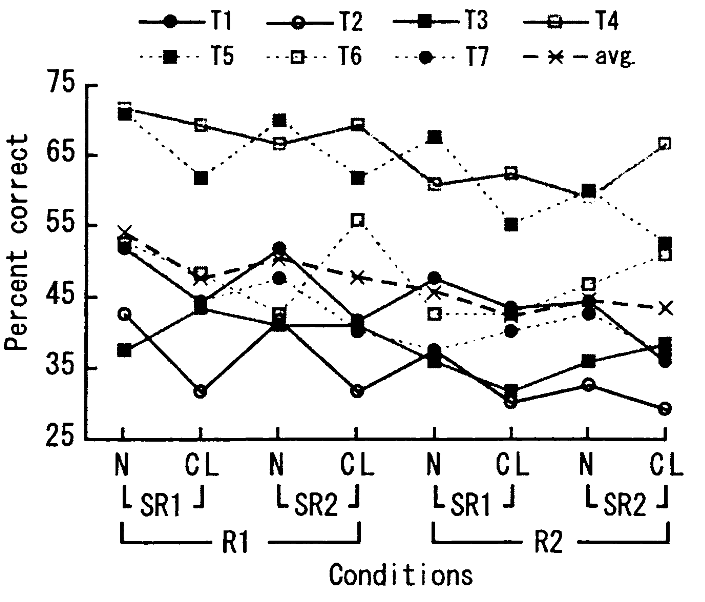

Fig. 1: Result of the listening test (T1-7: talkers, avg.: average of talkers, N: normal speaking style, CL: clear speaking style, SR1: original speaking rate, SR2: slowed speaking rate, R1: reverberation time of 1.1 s and R2: reverberation time of 1.8 s).

Figure 1 shows the result of the listening test. A mixed ANOVA was carried out with talkers as a nonrepeated variable, speaking style (normal and clear), speaking rate (original and slow) and reverberation (reverberation times of 1.1 s and 1.8 s) as repeated variables, and a mean percent correct as a dependent variable. Results showed that the mean percent correct significantly differed across talkers [$F(1,35) = 586.36$, $p < 0.01$]. Pairwise comparisons using t-test showed significant differences [$p < 0.50$] between T2 and T4, T2 and T5, T3 and T4, and T4 and T7. The mean percent correct was higher for the shorter reverberation time than for the longer reverberation time [$F(1,35) = 56.32$, $p < 0.01$]. The mean percent correct was higher for normal speaking style than for clear speaking style [$F(1,35) = 81.48$, $p = 0.07$]. In addition to these main effects, a significant interaction was observed between speaking style and talker [$F(6,35) = 3.16$, $p = 0.01$], with the difference between the mean percent corrects for normal and clear speaking style being different among talkers.

DISCUSSION

Longer reverberation time had lower correct rate (R1: 50.0% and R2: 44.0%), and therefore hypothesis 1) was supported. This was consistent with the previous research (Hodoshima *et al.*, 2006), which studied the effect of preprocessing on young people in reverberation.

Correct rate was different among talkers (T1: 45.0%, T2: 34.6%, T3: 38.0%, T4: 65.7%, T5: 62.4%, T6: 47.7% and T7: 42.5%), and there was no difference in correct rate between trained and untrained talkers. Therefore, hypothesis 2) was not supported. One possible reason is a limited counterbalance of talker-participant combinations because each listener heard two of seven talkers.

Normal speech had higher correct rate than clear speech (N: 48.7% and CL: 45.3%), and therefore hypothesis 3) was not supported. This was inconsistent with the previous researches (Payton *et al.*, 1994; Caissie *et al.*, 2005), which studied the effect of clear speech on people with hearing impairments in noise and reverberation. Less benefit of clear speech was observed under severe reverberant conditions because features of clear speech, e.g. release of stop bursts (Picheny *et al.*, 1986), are masked by long reverberation tail. Severe reverberant conditions may also made participants difficult to obtain information of target words from reverberant acoustic signals alone because no top-down information was available. Clear speech might be intelligible for elderly people when we use shorter reverberation time as well as target words with high familiarity.

There was no difference in correct rate between original and slowed speaking rates (SR1: 47.4% and SR2: 46.5%), therefore hypothesis 4) was not supported. This was inconsistent with the previous research (Sommers, 1997), which studied the effect of slowed speaking rate on elderly people in noise. Correct rate decreases due to increased reverberant masking by simply speaking slowly. Speech intelligibility in reverberation may improve by applying a signal processing (e.g. preprocessing) after slowing speaking rate of a speech signal (Arai *et al.*, 2007).

CONCLUSIONS

Speech intelligibility of elderly people in reverberation differed among talkers, but speech intelligibility did not improved by speech training, clear speech and slowed speaking rate. The main difference between this study and the relevant previous studies (Payton *et al.*, 1994; Sommers, 1997; Caissie *et al.*, 2005) was that no top-down information was available in this study while top-down processing was available in the previous studies. And this difference may affect the effect of clear speech and slowed speaking rate on speech perception of elderly people. Future research would conduct an acoustic analysis of the speech materials used in this study and compare results of elderly people and young people to find characteristics of speech signals that has higher correct rate for elderly people in reverberation.

ACKNOWLEDGMENTS

This research was supported by Grants-in-Aid for Scientific Research (A-2, 16203041) from JSPS and by Sophia University Open Research Center from MEXT. Authors appreciate Hideki Tachibana, Kanako Ueno and Sakae Yokoyama at the University of Tokyo (at the time) for offering impulse response data, and participants from Chiyoda-City Silver Human Resources Center.

REFERENCES

Arai, T., Kinoshita, K., Hodoshima, N., Kusumoto, A., and Kitamura, T. (**2002**). "Effects of suppressing steady-state portions of speech on intelligibility in reverberant environments," Acoust. Sci. Tech., **23**, 229-232.

Arai, T., Nakata, Y., Hodoshima, N., and Kurisu, K. (**2007**). "Slow speech with steady-state suppression to improve intelligibility in reverberant environments," Acoust. Sci. Tech., **28**, 282-285.

Caissie, R., Campbell, M. M., Frenette, W. L., Scott, L., Howell, I., and Roy, A. (**2005**). "Clear speech for adults with a hearing loss: Does intervention with communication partners make a difference," J. Am. Acad. Audiol., **16**, 157-171.

Hodoshima, N., Arai, T., Kusumoto, A., and Kinoshita, K. (**2006**). "Improving syllable identification by a preprocessing method reducing overlap-masking in reverberant environments," J. Acoust. Soc. Am., **119**, 4055-4064.

Nábělek, A. K., and Robinson, P. K. (**1982**). "Monaural and binaural speech perception in reverberation for listeners of various ages," J. Acoust. Soc. Am., **71**, 1242-1248.

Payton, K. L., Uchanski, R. M., and Braida, L. D. (**1994**). "Intelligibility of conversational and clear speech in noise and reverberation for listeners with normal and impaired hearing," J. Acoust. Soc. Am., **95**, 1581-1592.

Picheny, M. A., Durlach, N. L., and Briada, L. D. (**1986**). "Speaking clearly for the hard of hearing II," J. Speech Hear. Res., **29**, 434-446.

Praat Homepage (Version 4.3.04): http://www.praat.org.

Sommers, M. S. (**1997**). "Stimulus variability and spoken word recognition. II. The effects of age and hearing impairment," J. Acoust. Soc. Am., **101**, 2278-2288.