euronoise

Effects of training, style, and rate of speaking on speech perception of young people in reverberation

N. Hodoshima[a], T. Arai[a] and K. Kurisu[b]

[a]Dept. of Information and Communication Sciences, Sophia University, 7-1 Kiyoi-cho, Chiyoda-ku, 102-8554 Tokyo, Japan
[b]TOA Corporation, 2-1 Takamatsu-cho, Takarazuka, 665-0043 Hyogo, Japan
n-hodosh@sophia.ac.jp

Because of the difficulty of listening to speech in reverberation (e.g., at train stations), we need to find characteristics of intelligible speech sounds that are appropriate for announcements by spoken messages over loudspeakers in public spaces. This study investigated the effects of training (seven talkers who have received speech training or not), style (conversational/clear) and rate (normal/slow) of speaking on speech perception of young people in simulated reverberant environments. The talkers were instructed to speak nonsense words embedded within a carrier sentence clearly or normally in an anechoic room, and listening tests were carried out with young people in simulated reverberant environments. Results showed that correct rates significantly differed among the talkers, but no difference in correct rates was found between the two speaking rates, and conversational speech had significantly higher correct rates than clear speech. Casual inspections of the stimuli indicate that vowels are enhanced as well as consonants in clear speech so that clear speech had lower correct rates than conversational speech. This difference may be due to increased reverberant masking in clear speech compared to that in conversational speech. [Work supported by Sophia University Open Research Center.]

1 Introduction

Reverberation and/or noise sometimes make it difficult to listen to speech sounds over loudspeakers in public spaces such as train stations or airports. Several researches have reported approaches to improve speech intelligibility by reducing the effect of reverberation and/or noise such as an electroacoustical approach and an approach that focuses on speech production. An example of an electroacoustical approach is preprocessing, which processes speech signals before radiating them from loudspeakers [e.g., 1-3]. On the other hand, the effect of speaking style and rate has been investigated on the speech production side. It has been reported that clear speech had higher word intelligibility than conversational speech for people with normal hearing and with hearing impairments in noise [4] and in noise and reverberation [5]. Slowed speaking rate had higher word intelligibility than normal speaking rate for young and elderly people in noise [6].

The goal of this study is to find characteristics of speech materials that are intelligible in public spaces where sound reinforcement systems are used for speech transmission. Therefore, we focus on relatively severe reverberant conditions (reverberation time is longer than about 1.0 s) compared to those used in other studies on speech production (e.g., reverberation times of 0.18 and 0.6 s were used in [5]). We tested the effects of speech training, speaking style and speaking rate on speech perception of elderly people under the reverberant conditions that simulated such public spaces [7], and the results showed no effects of speech training, clear speech and slowed speaking rate. In order to study whether the tendency reported in [7] appears only for elderly people or not, the current study carried out a listening test for young people with the same stimuli used in [7], and compares the results of young people with those of elderly people [7].

2 Listening test

2.1 Participants

Twenty-one young people (6 males and 15 females, aged 23 years old on average) participated in this listening test. Their air-conduction thresholds were below 20dBHL from 125 to 8000 Hz.

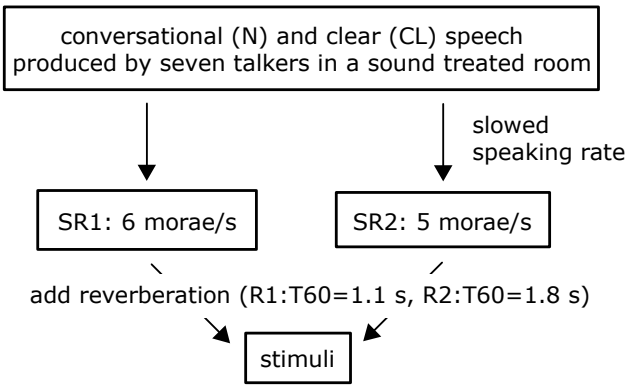

Fig.1 Stimuli used in the listening test.

2.2 Stimuli

The speech materials were the same as those used in [7]. They consisted of 20 nonsense Japanese vowel-consonant-vowel (V_1CV_2) words as targets embedded in a Japanese carrier phrase. All possible 20 V_1CV_2 combinations were selected from /p, t, k, b, d, g, s, ʃ, h, z, ʒ, m, n/ and /a, i/ excluding those that do not meet Japanese phonotactics. The same vowel was used as V_1 and V_2.

We used four talkers who have received speech training (T1-T4: two males and two females, aged 28 years old on average) and three talkers who have not received any speech training (T5-T7: one male and two females, aged 23 years old on average). All talkers had no articulation and hearing disorders. They produced speech sounds in conversational (N) and clear (CL) speaking styles in the same speaking rate (SR1: six morae/s on average). The recording was made using a Digital Audio Tape recorder (SONY, TDC-D10) at a sampling frequency of 16000 Hz with a microphone (SONY, ECM-MS967) in a sound treated room.

Two speaking rates were used: original (SR1) and slow (SR2: five morae/s on average) manipulated by the Praat software using the Pitch-synchronous Overlap and Add method [8].

Two reverberant conditions were used: 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).

A total of 1120 stimuli (7 talkers x 2 speaking styles x 2 speaking rates x 2 reverberant conditions x 20 speech

materials) were used. The A-weighted energy was set equal for the speech materials. See Fig. 1 for the stimuli used in the listening test.

2.3 Procedure

The listening test was carried out in a sound treated room. Each participant listened to 320 stimuli that correspond to two talkers. Before starting a main session, each participant had six 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 diotically over headphones (STAX, SR-303). Then participants were instructed to select a VCV they heard from the 20 CVCs that were used in the listening test displayed on a computer monitor. Stimuli were randomly presented for each participant.

3 Results

Figure 2 shows the result of the listening test, and Figure 3 shows results of correct rate of each talker, speaking style, speaking rate and reverberation. A mixed ANOVA was carried out with talkers as a nonrepeated variable, speaking style (N and CL), speaking rate (SR1 and SR2) and reverberation (R1 and R2) as repeated variables, and a mean percent correct response (correct rate) as a dependent variable. Results showed that the correct rate significantly differed across talkers [$F(6,35) = 6.984$, $p < 0.01$]. Pairwise comparisons using t-test showed significant differences [$p < 0.50$] between T1 and T2, T1 and T3, T2 and T4, T3 and T4, and T4 and T7. Conversational speaking style had a higher correct rate than clear speaking style [$F(1,35) = 27.581$, $p < 0.01$]. The shorter reverberation time had a higher correct rate than the longer reverberation time [$F(1,35) = 8.187$, $p = 0.007$]. Significant interactions between talker and reverberation [$F(6,35) = 8.294$, $p < 0.01$] and among talker, speaking rate, speaking style and reverberation [$F(6,35) = 2.656$, $p = 0.031$] were observed.

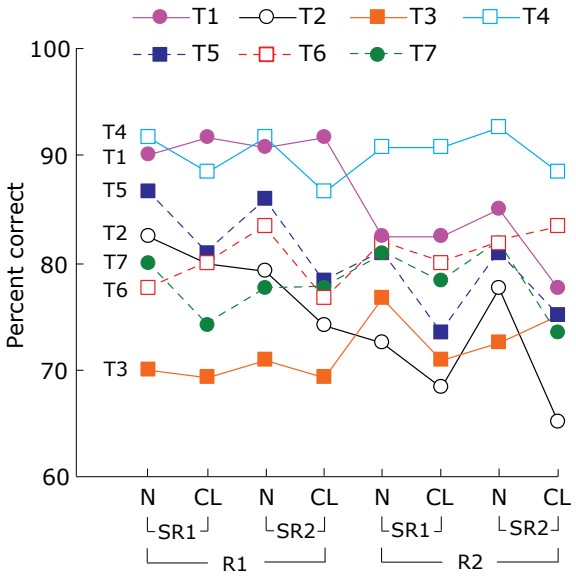

Fig.2 Result of the listening test (T: talkers, N and CL: speaking style, SR: speaking rate and R: reverberation).

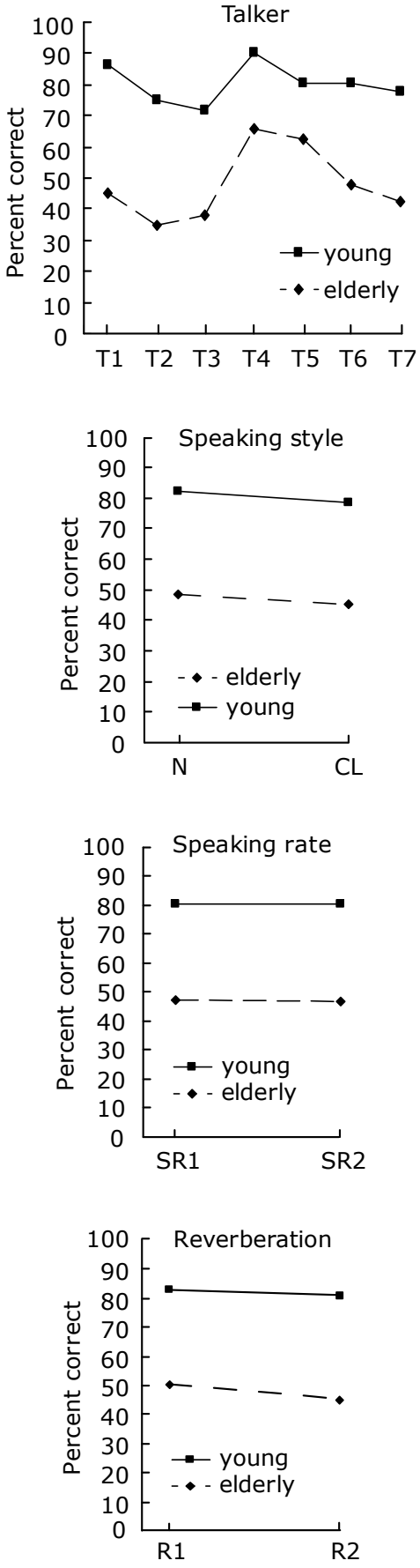

Fig.3 Result of talkers, speaking style, speaking rate and reverberation combined with the results of elderly people [7].

4 Discussion

Longer reverberation time had lower correct rate (R1: 81.3% and R2: 79.3%). This is consistent with the previous research [7].

Correct rate was different among talkers, and did not differ between talker groups who have received speech training or not. This is consistent with the previous research [7]. The results showed that talkers who are intelligible to young people are also intelligible to elderly people.

Conversational speech had higher correct rate than clear speech (N: 82.0% and CL: 78.6%). This is consistent with the previous research [7]. On the other hand, this is inconsistent with the other research [5]. One possible reason for less benefit of clear speech is that features of clear speech (e.g., release of English stop bursts as reported in [9]) are mostly masked by long reverberation tails because the current study used severe reverberant conditions than those used in [5] (reverberation times of 0.18 and 0.6 s). Another possible reason is that the amount of reverberant component of V1 which masks C was increased because vowels as well as consonants were stressed in clear speech. To study this possible reason, the V1 to C ratio in intensities for T2 and T5 were calculated in CL and N. The V1-C ratio of CL was greater than that of N by 5 dB or more in /aza/ for T2 and in /ata, ada/ for T5, and therefore casual inspection on the V1-C ratio showed that characteristics of clear speech seem to mostly vary in target utterances as well as talkers. Furthermore, characteristics of clear speech may be varied in environments where we record clear speech. The current study used similar instructions to talkers and recording environments to make clear speech stimuli to those in other researches [e.g., 5]. However, clear speech produced in reverberation might have much higher correct rate than clear speech produced in a sound treated room and then convolved with an impulse response because talkers probably adjust their speaking style to be intelligible in environments around the talkers, which is similar to the Lombard effect [10].

There was no difference in correct rate between original and slowed speaking rates (SR1: 80.4% and SR2: 80.1%). This is consistent with the previous research [7]. On the other hand, this was inconsistent with the other research [6], which tested the effect of slowed speaking rate in noise. The result showed that uniformly slowed speaking rate by software did not improve speech intelligibility under severe reverberant conditions. It is interesting to test the effect of slowed speaking rate in reverberation when talkers speaks slowly as was used in [6] or when a signal processing (e.g., preprocessing) is applied after slowing the speaking rate of speech signals [11].

5 Conclusions

Speech intelligibility of young people in reverberation differed among talkers and speaking styles. On the other hand, speech intelligibility did not differ in speech training and speaking rate, and less benefit of clear speech and slowed speaking rate was observed as was reported in [7]. A similar trends observed in both the current study on young people and the previous one on elderly people [7], indicates that young and elderly people use acoustical signal information in a same way when they listen to nonsense words of different talkers, speaking rate and speaking styles in reverberation, although an overall correct rate was lower for elderly people than for young people probably due to reduced temporal processing abilities of elderly people. The inconsistency with other studies [e.g., 5] may be due to differences in stimulus (the amount of reverberant masking overlapping to the first consonant of the targets increased more in the current study using VCV than in [5] using CVC) and/or reverberant conditions (severe reverberant conditions were used in the current study compared to the conditions in [5]). Future research would conduct detailed acoustic analyses of the stimuli to find characteristics of speech signals that are intelligible in reverberation.

Acknowledgments

This research was supported 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.

References

[1] T. Arai, K. Kinoshita, N. Hodoshima, A. Kusumoto, and T. Kitamura, "Effects of suppressing steady-state portions of speech on intelligibility in reverberant environments", *Proc. Autumn Meet. Acoust. Soc. Jpn*., 1, 449-450 (2001) (in Japanese)

[2] T. Arai, K. Kinoshita, N. Hodoshima, A. Kusumoto, and T. Kitamura, "Effects of suppressing steady-state portions of speech on intelligibility in reverberant environments", *Acoust. Sci. Tech*., 23(4), 229-232 (2002)

[3] N. Hodoshima, T. Arai, A. Kusumoto, and K. Kinoshita, "Improving syllable identification by a preprocessing method reducing overlap-masking in reverberant environments", *J. Acoust. Soc. Am*., 119(6), 4055-4064 (2006)

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

[5] K. L. Payton, R. M. Uchanski, and L. D. Braida, "Intelligibility of conversational and clear speech in noise and reverberation for listeners with normal and impaired hearing", *J. Acoust. Soc. Am*., 95(3), 1581-1592 (1994)

[6] M. S. Sommers, "Stimulus variability and spoken word recognition. II. The effects of age and hearing impairment", *J. Acoust. Soc. Am*., 101(4), 2278-2288 (1997)

[7] N. Hodoshima and T. Arai, "Effect of talker variability on speech perception by elderly people in

reverberation", *Proc. International Symposium on Auditory and Audiological Research* (2007)

[8] Praat Homepage (Version 5.0.20): http://www.praat.org

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

[10] J-C, Junqua, "The Lombard reflex and its role on human listeners and automatic speech recognizers", *J. Acoust. Soc. Am*., 93(1), 510-524 (1993)

[11] T. Arai, Y. Nakata, N. Hodoshima, and K. Kurisu, K "Slow speech with steady-state suppression to improve intelligibility in reverberant environments", *Acoust. Sci. Tech*., 28(4), 282-285 (2007)