ACOUSTICAL LETTER

Steady-state suppression for improving syllable identification in reverberant environments: A case study in an elderly person

Nao Hodoshima*, Yusuke Miyauchi, Keiichi Yasu and Takayuki Arai

Department of Electrical and Electronics Engineering, Sophia University, 7–1 Kioi-cho, Chiyoda-ku, Tokyo, 102–8554 Japan

(*Received 20 June 2006, Accepted for publication 30 August 2006*)

Keywords: Speech enhancement, Reverberation, Elderly people, Speech intelligibility, Overlap-masking, Steady-state suppression

PACS number: 43.72.Ew, 43.66.Dc, 43.71.Lz, 43.38.Tj [doi:10.1250/ast.28.53]

1. Introduction

Elderly people with or without hearing loss [1,2], hearing-impaired people [3,4] and non-native listeners [5,6] have more difficulty understanding speech in reverberant environments, for example in large auditoria, than young native listeners with normal hearing. As the elderly population continues to grow around the world, ensuring clear speech perception by reducing the effects of reverberation in public spaces will become a more important issue, particularly in situations where safety is a consideration, for example when fire alarms are given verbally.

There are a number of ways to reduce the effects of reverberation and noise to improve speech intelligibility for elderly people. As an example of listening systems, Nábělek and Donahue [7] found that the word identification scores of elderly people in tests delivered through an assistive listening system were higher than corresponding scores for tests delivered through a public address system that was installed in a reverberant environment. Gordon-Salant [8,9] reported that increasing the energy of consonants relative to that of vowels improved consonant recognition in noise for both elderly people with hearing impairment and those with normal hearing. Clearly articulated speech might be effective for elderly people since Payton *et al.* [10] determined that clearly articulated speech is more intelligible than conversational speech for people with hearing impairment in situations with reverberation and noise.

Arai *et al.* [11,12] proposed the steady-state suppression technique as an approach for processing speech signals before they are passed through loudspeakers. This technique suppresses the steady-state portions of speech that are not so important for syllable perception [13] in order to reduce the effect of "overlap-masking", which is thought to be the main reason why reverberation reduces speech intelligibility [14,15]. Hodoshima *et al.* [16,17] investigated the effect of steady-state suppression on syllable identification for people with normal hearing under various reverberant conditions. They found that steady-state suppression significantly improved consonant identification under diotic and dichotic listening environments where reverberation time (RT) was artificially modified from 0.7 to 1.3 s [16,17].

The purpose of the present case study was to explore the effect of steady-state suppression on speech intelligibility for an elderly person under several reverberation conditions. Our participant was an elderly person with presbycusis, and the study was conducted under diotic listening conditions, with RTs of 0.7, 1.0 and 1.2 s.

2. Listening test

2.1. Participant

The participant of this listening test was a 65-year-old male, who was a native speaker of Japanese. Table 1 shows the air- and bone-conduction thresholds for this participant. The participant did not wear a hearing aid. An interview with the participant revealed that he was in general good health. He had neither history of otologic disease nor exposure to unusual levels of noise.

2.2. Stimuli

The original speech samples used in this study consisted of 24 nonsense Consonant-Vowel (CV) syllables embedded in a Japanese carrier phrase, which was the same as that used by Hodoshima *et al.* [16]. The vowels were /a, i/, and the consonants were /p, t, k, b, d, g, s, ʃ, h, ʥ, ʤ, ʧ, m, n/. All possible CV combinations were selected, excluding those that do not meet Japanese phonotactics (/ti/ and /di/). The original speech samples were obtained from the ATR Japanese speech database. The speaker was a 40-year-old male.

Two versions of the speech materials were used in this study, original (unprocessed) and steady-state suppressed (processed), as used by Arai *et al.* [11,12]. The steady-state suppression method calculates the D parameter to detect spectral transitions in a speech signal, where D is essentially similar to a parameter proposed in [13]. In the present paper, D is calculated as the mean square of the regression coefficients for each time trajectory of the logarithmic envelope of a subband of an input signal. A portion of speech is defined as steady-state when D is less than a specified threshold, and the amplitude of portions that are seemed to be steady-state is multiplied by a factor of 0.4, giving a 40% suppression rate.

Processed and unprocessed speech materials were reproduced using three reverberant conditions: RTs of 0.7, 1.0 and 1.2 s. The reverberant conditions were represented by multiplying exponential decays by an impulse response measured in Hamming Hall in Tokyo, as described by Hodoshima *et al.*

*e-mail: n-hodosh@sophia.ac.jp

Table 1 Air- and bone-conduction thresholds for the participant in dB HL (top: right ear; bottom: left ear).

Frequency (Hz)	125	250	500	1,000	2,000	4,000	8,000
Air conduction	10 10	10 5	15 10	30 30	40 30	30 40	35 55
Bone conduction		10 5	10 10	25 30	35 30	30 35	

[18]. The RTs used are the averages derived from the early decay time (EDT) of center frequencies of 0.5, 1, and 2 kHz of the 1-octave bandpassed impulse response.

The reverberant stimuli comprised a total of 144 stimuli (three reverberant conditions × 24 CV syllables × unprocessed/processed), which we call "one session" in this paper. Additionally, 48 non-reverberant stimuli (24 CV syllables × 2 processing conditions) were also prepared.

2.3. Procedure

The participant participated in 20 sessions of the listening test, with a maximum of two sessions per day. Before starting the first session, the participant underwent practice trials to become familiar with the procedure. During each session, the 144 reverberant stimuli were presented randomly. After the 20 sessions were finished, the 48 non-reverberant stimuli were presented randomly in a single session.

The computer-controlled listening test was conducted in a sound-treated room. The stimuli were presented over headphones (STAX SR-303). The sound level was adjusted to a comfortable level for the participant before the trials began, and the comfortable level was kept constant throughout the test. In each trial, a stimulus was presented, after which a PC monitor displayed 24 CV syllables in Japanese kana orthography. While the stimuli were being presented to the participant, the kana representing the 24 CV syllables were not visible. The participant was instructed to use the mouse to click on the CV syllable on the monitor corresponding to the syllable he heard. Once a CV syllable was selected, the next trial was presented.

A score representing the mean percentage of correct answers provided by the participant for each of the different reverberant and processing conditions was calculated. The scores for the different reverberant conditions are an average of the 20 sessions, whereas the score for the non-reverberant stimuli is just the score for a single test. A 2 × 3 ANOVA was carried out with RT and processing as repeated variables. *T*-tests were used as a post hoc analysis.

3. Experimental results and discussion

The square symbols in Fig. 1 show the participant's mean scores for the different reverberant and processing conditions. An ANOVA showed that processed stimuli had a higher score than unprocessed stimuli [$F(1, 19) = 4.23,\ p = 0.05$]. The participant's mean score was significantly higher for processed stimuli than for unprocessed stimuli for RTs of 0.7 s [$t(19) = -2.64,\ p = 0.016$] and 1.0 s [$t(19) = -2.27,\ p = 0.035$]. The percent correct also reliably differed across RTs [$F(2, 38) = 16.33,\ p < 0.01$]. Multiple comparison tests

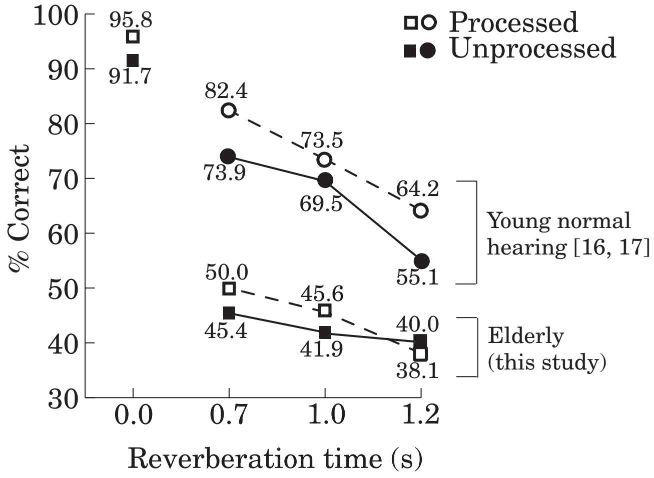

Fig. 1 Mean percentage of syllables correctly identified by participants in a perception test using unprocessed (filled symbols) and steady-state suppressed (open symbols) stimuli for four different reverberation conditions. Squares represent the mean scores of the elderly participant involved in the current study; circles represent the mean scores of young participants with normal hearing involved in a previous study [16].

showed significant differences [$p < 0.05$] between all RT pairs, indicating that the percent correct was higher for RT of 0.4 s, 0.7 s and 1.1 s, in this order. In addition to these main effects, a significant interaction was observed between RT and processing [$F(2, 38) = 4.47,\ p = 0.018$]. These results show that steady-state suppression improved syllable identification for this participant with presbycusis with RTs of 0.7 s and 1.0 s. There was no significant difference between scores for the unprocessed and processed stimuli with an RT of 1.2 s, showing that steady-state suppression has little effect on syllable identification with an RT of 1.2 s.

The circular symbols in Fig. 1 represent the scores obtained in previous studies by 22 young participants with normal hearing [16] for reference. In those tests, the same speech material and experimental conditions were used as in the present study. It can be seen that the scores of the elderly participant were lower than those of the young participants with normal hearing in all conditions. The differences between the mean score of the elderly participant and those of the young participants with normal hearing for the unprocessed stimuli (28.5% with an RT of 0.7 s, 27.6% with an RT of 1.0 s and 15.1% with an RT of 1.2 s) are in accordance with the results of a study by Helfer and Huntley [2], in which it was found that young people with normal hearing had higher syllable identification scores than those of elderly people with hearing loss (21% in average) at an RT of 0.93 s (the average of the RTs of 0.6, 0.9 and 1.3 s used in that study).

When we compared processed and unprocessed speech, steady-state suppression improved syllable identification both for the young participants [16] and the elderly participant. The degree of improvement in perception produced by steady-state suppression (the difference between scores for the unprocessed and processed stimuli) for each reverberant condition was different from the elderly participant and young normal hearing participants in each reverberant condition. This might imply that the effect of steady-state suppression is participant to hearing loss and age-related hearing decline. However, it would be difficult for a direct comparison between two participant groups since the numbers of participants were completely different for the two groups. Therefore, in order to compare the two groups such as the degree of improvement by steady-state suppression, further studies are needed with elderly participants to match the numbers of young participants.

Without reverberation, the scores for the unprocessed and processed stimuli in the present study (in the elderly participant) were both above 90%. In a study by Hodoshima *et al.* [16], scores for young people with normal hearing were 97.0% and 97.2% for unprocessed and processed stimuli, respectively (using the same speech material and experimental conditions as used in the present study). The results of the present and previous studies indicate that it relatively easy for elderly people to identify syllables in situations without reverberation (relative to the reverberant conditions used in the present study). This result is agreement with a previous finding that age effects are usually not apparent in normal or hearing-impaired listeners in quiet environments [19]. The results for unprocessed and processed stimuli without reverberation were derived from an average for 24 syllables in a single test, so further tests are needed in order to confirm this tendency.

4. Conclusions

We conducted a case study in order to investigate the effect of steady-state suppression on syllable perception in an elderly person. Diotic listening tests were conducted with RTs of 0.7, 1.0 and 1.2 s and without reverberation. With RTs of 0.7 and 1.0 s, the participant achieved higher scores (percentage of correctly identified syllables) for steady-state suppressed signals than for unprocessed signals. This suggests that steady-state suppression would be effective in improving perception for elderly people in small- to medium-sized rooms, such as lecture halls. In order to arrive at general conclusions regarding the effect of steady-state suppression on the elderly, further studies should examine the effect of steady-state suppression on perception by elderly participants with and without hearing loss under various reverberant conditions.

Acknowledgements

This research was supported by Grants-in-Aids for Scientific Research (A-2, 16203041) and for JSPS Fellows (176911), both from the Japan Society for the Promotion of Science. The authors would like to thank Hideki Tachibana, Kanako Ueno and Sakae Yokoyama for offering the use of the impulse response data, and Chikashi Michimata and Hirofumi Kamata for their guidance regarding the statistical analysis.

References

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

[2] K. S. Helfer and R. A. Huntley, "Aging and consonant errors in reverberation and noise," *J. Acoust. Soc. Am.*, **90**, 1786–1796 (1991).

[3] A. K. Nábělek and J. M. Pickett, "Monaural and binaural speech perception through hearing aids under noise and reverberation," *J. Speech Hear. Res.*, **17**, 724–739 (1974).

[4] T. Finitzo-Hieber and T. Tillman, "Room acoustics effects on monosyllabic word discrimination ability for normal and hearing-impaired children," *J. Speech Hear. Res.*, **21**, 440–458 (1978).

[5] A. K. Nábělek and A. M. Donahue, "Perception of consonants in reverberation by native and non-native listeners," *J. Acoust. Soc. Am.*, **75**, 632–634 (1984).

[6] Y. Takata and A. K. Nábělek, "English consonant recognition in noise and in reverberation by Japanese and American listeners," *J. Acoust. Soc. Am.*, **88**, 663–666 (1990).

[7] A. K. Nábělek and A. M. Donahue, "Comparison of amplification systems in an auditorium," *J. Acoust. Soc. Am.*, **79**, 2078–2082 (1986).

[8] S. Gordon-Salant, "Recognition of natural and time/intensity altered CVs by young and elderly subjects with normal hearing," *J. Acoust. Soc. Am.*, **80**, 1599–1607 (1986).

[9] S. Gordon-Salant, "Effects of acoustic modification on consonant recognition by elderly hearing-impaired subjects," *J. Acoust. Soc. Am.*, **81**, 1199–1202 (1987).

[10] 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**, 1581–1592 (1994).

[11] 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.*, Vol. 1, pp. 449–450 (2001).

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

[13] S. Furui, "On the role of spectral transition for speech perception," *J. Acoust. Soc. Am.*, **80**, 1016–1025 (1986).

[14] R. H. Bolt and A. D. MacDonald, "Theory of speech masking by reverberation," *J. Acoust. Soc. Am.*, **21**, 577–580 (1949).

[15] A. K. Nábělek and L. Robinette, "Influence of precedence effect on word identification by normally hearing and hearing-impaired subjects," *J. Acoust. Soc. Am.*, **63**, 187–194 (1978).

[16] 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**, 4055–4064 (2006).

[17] N. Hodoshima, T. Goto, N. Ohata, T. Inoue and T. Arai, "The effect of pre-processing approach for improving speech intelligibility in a hall: Comparison between diotic and dichotic listening conditions," *Acoust. Sci. & Tech.*, **26**, 212–214 (2005).

[18] N. Hodoshima, T. Arai and A. Kusumoto, "Enhancing temporal dynamics of speech to improve intelligibility in reverberant environments," *Proc. Forum Acusticum*, Sevilla (2002).

[19] S. Gordon-Salant, "Age-related differences in speech recognition performance as a function of test format and paradigm," *Ear Hear.*, **8**, 277–282 (1987).