A Preprocessing Technique for Improving Speech Intelligibility in Reverberant Environments: The Effect of Steady-state Suppression on Elderly People

Yusuke Miyauchi [†], *Nao Hodoshima* [†], *Keiichi Yasu* [†], *Nahoko Hayashi* [†], *Takayuki Arai* [†] *and*

Mitsuko Shindo [‡]

† Department of Electrical and Electronics Engineering, Sophia University
‡ Research Center for Communication Disorders, Sophia University
7-1 Kioi-cho, Chiyoda-ku, Tokyo, 102-8554 Japan
E-mail address: m-yuusuk@sophia.ac.jp

Abstract

In a large auditorium, perceiving speech may become difficult. One reason that reverberation degrades speech intelligibility is the effect of overlap-masking (Bolt and MacDonald, 1949; Nabelek and Robinette, 1978). Reverberation is a more critical issue for elderly people to perceive speech than it is for young people (Fitzgibbons and Gordon-Salant, 1999). Arai *et al*. suppressed steady-state portions of speech which have more energy but are less crucial for speech perception, and confirmed promising results for improving speech intelligibility (Arai *et al.,* 2001, 2002). Hodoshima *et al*. conducted perceptual tests to confirm the effectiveness of steady-state suppression with several reverberation conditions, and obtained significant improvements with reverberation times of 0.7-1.3 s.

In this study, we conducted an experiment for evaluating steady-state suppression with fifty elderly people and found that there were significant improvements. Also, steady-state suppression yielded better improvements in speech intelligibility for elderly people than it did for young people.

1. Introduction

In a large auditorium like a lecture hall or church where reverberation time is relatively long, speech intelligibility is degraded and it becomes difficult for people to perceive and understand speech. One of the reasons reverberation degrades speech intelligibility is overlap-masking, where reverberation tails of previous portions of a sound affect subsequent segments [1, 2]. Although reverberation adds richness of sound for music, it makes speech more difficult to understand. When making a speech or giving a lecture in a hall designed for music where reverberation time is relatively long, we need some procedure to reduce the influence of reverberation.

As an electroacoustical approach for improving speech intelligibility in reverberant environments, there are two general approaches: post-processing and pre-processing. Post-processing methods such as inverse filtering [3] and modulation filtering [4] are applied to speech signals already released into a room and affected by reverberation. Pre-processing approaches, on the other hand, processes a speech signal before it is affected by reverberation. It is a method that reduces the influence of reverberation on the transmission path. Kusumoto *et al*. proposed the modulation filtering which enhances the important modulation frequency regions for speech perception [5]. Arai (2005) proposed a novel processing method which pads zero sequences into steady-state portions of speech [6]. Since pre-processing operates on a speech signal between a microphone and loudspeaker, this method can be used with a Public Address (PA) system.

To decrease the effect of overlap-masking, Arai *et al.* invented signal processing to lessen the energy of preceding portions of a speech signal beforehand so that the energy of reverberation components overlapping to a subsequent portion is decreased. Moreover, it is reported that while spectral transitions are crucial for syllable perception, vowel nuclei are not necessary for either vowel or syllable perception [9]. Therefore, Arai *et al.* suppressed the steady-state portions of speech in order to reduce the influence of overlap-masking since these portions of speech have more energy but are less crucial for speech perception [7, 8]. They confirmed promising results for improving speech intelligibility.

Hodoshima *et al*. [10-12] conducted perceptual tests in a diotic (the same stimulus was presented simultaneously to both the right and left earphones) environment with a set of artificial reverberation conditions, in which reverberation times were 0.4–1.3 s [11, 12] and in the largest lecture hall at Sophia University as a dichotic (different stimuli was presented to both the right and left earphones) environment in which reverberation time was 1.3 s [10] to confirm the effectiveness of Arai's technique [7, 8]. Significant improvements were obtained with reverberation times of 0.7–1.3 s.

As one gets older, audibility and auditory temporal resolution become degraded. It is reported that elderly people are greatly influenced by reverberation in speech intelligibility more than young normal hearing people are [13, 14].

In this study, we conduct a perceptual experiment with fifty elderly people to investigate the effect of steady-state suppression. We detail experimental results below and compare them with the results of previous studies [11, 12].

2. Experiment

2.1. Steady-state Suppression

We applied the same method as in [10-12] to suppress the steady-state portions of speech. Figure 1 shows a block diagram of our steady-state suppression technique. First, a speech signal which had been sampled with frequency of

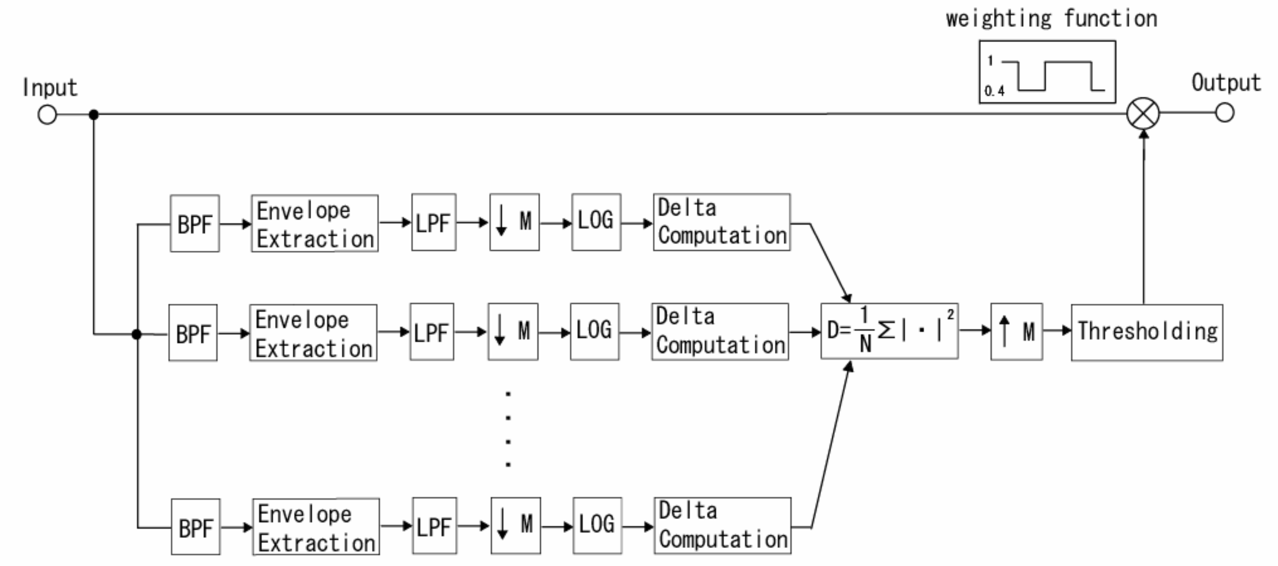

Figure 1: Block diagram of steady-state suppression

16kHz was split into 1/3-octave bands. In each band the temporal envelope was extracted. After the down-sampling, the regression coefficients were calculated from the five adjacent values of the time trajectory of the logarithmic envelope of a subband. Then the mean square of the regression coefficients, *D,* was calculated. We used the *D* parameter defined by Furui to measure the spectral transition [9]. After up-sampling, we defined a speech portion as steady-state when *D* was less than a certain threshold. Once a portion was considered as steady-state, the amplitude of the portion was suppressed. In this study, a suppression rate was 40% according to previous studies [7, 8, 10-12]. The correlation between suppression rates and reverberation time is discussed in [13].

2.2. Stimuli

The original speech samples consisted of fourteen nonsense Consonant-Vowel (CV) syllables embedded in a Japanese carrier phrase. The vowel was /a/ and the consonants were /p/, /t/, /k/, /b/, /d/, /g/, /s/, /ʃ/, /h/, /dz/, /dʒ/, /tʃ/, /m/ and /n/ (Table 1). They were obtained from the ATR speech database of Japanese. The CV syllables were selected from the monosyllable data set. The carrier phrase was a combination of two partial sentences taken from the sentence data set. We normalized the root-mean-square (RMS) energy in the CVs that have the same vowel, and then normalized the ratio of RMS in the carrier phrase relative to RMS energy in the CVs.

The stimuli consist of two conditions: the original signals (the speech samples above) or the processed signals were convoluted with two impulse responses. The impulse responses we use were measured in the largest lecture hall at Sophia University (reverberation time is 1.3 s) and in Hamming Hall, Higashi-Yamato City, Tokyo (reverberation time is 1.1 s when a reflection board was not used). Reverberation time is defined as the time the decay curve of the impulse response decrease to 60 dB below from its original level. We used Early Decay Time (EDT), which is the time it took for the first 10 dB down of the decay curve, and we multiplied it by six to estimate the reverberation time. For the latter impulse response, we modified its reverberation time to 1.0 s by multiplying an exponential decay according to the previous studies [11, 12].

Table 1: 14 Consonant-Vowel (CV) syllables

	Voiceless C + Vowel	Voiced C + Vowel
Stop C + Vowel	/pa/ /ta/ /ka/	/ba/ /da/ /ga/
Fricative C + Vowel	/sa/ /ʃa/ /ha/	
Affricate C + Vowel	/tʃa/	/dza/ /dʒa/
Nasal C + Vowel		/ma/ /na/

2.3. Subjects

Fifty elderly subjects (21 males and 29 females, ages 56 to 90, average age of 67.9) participated in the experiment. All were native speakers of Japanese.

To see the difference of the effect of steady-state suppression between groups, we define elderly normal hearing people as a normal hearing group and people with gradually-sloping hearing loss as a presbycusis group. We divided subjects into the two groups based on their audiogram. There were 16 people with gradually-sloping hearing loss whose ages are between 57-90 (average age of 70.4) and 22 normal hearing people whose ages are between 56-75 (average age of 65.8). Figures 2 and 3 show the audiograms of each group.

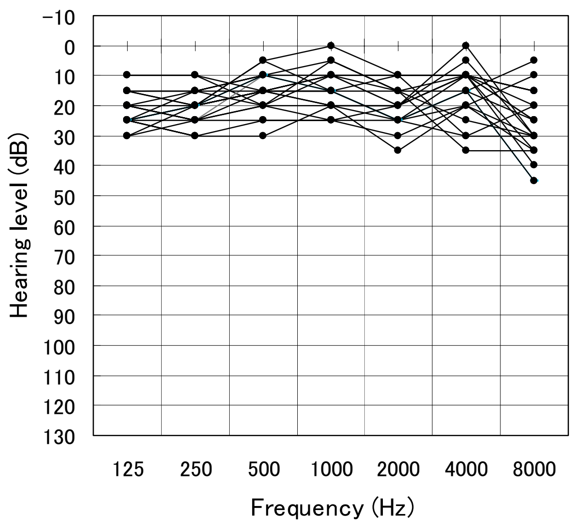

Figure 2: Audiograms of normal hearing group

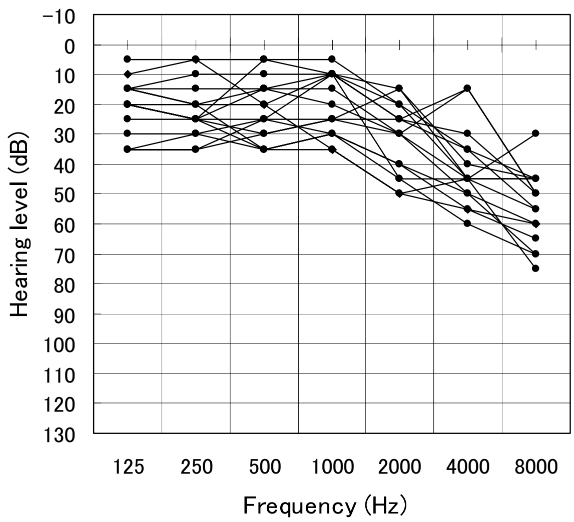

Figure 3: Audiograms of presbycusis group

2.4. Procedure

The experiment was conducted in a soundproof room using a computer. The stimuli were presented through headphones (STAX SR-303) connected to the computer. The sound level was adjusted to each subject's comfort level before the experiment and was kept constant. A stimulus was presented at each trial. Then subjects were forced to write one of fourteen CVs provided to them in Kana orthography. When they wrote a CV, the next stimulus was presented by clicking a "next" button on the screen with a mouse. For each subject, 56 stimuli were presented randomly (2 reverberation conditions x 14 CVs x 2 processing conditions). The experiment was carried out at each subject's pace.

3. Results and Discussions

Figure 4 shows the result of this experiment. Bold line shows the mean percent correct of all subjects. A *t*-test indicated that the correct rate of the processed signal was significantly higher than that of the unprocessed signal in each reverberant environment (p<0.01).

3.1. Comparison of the results between the normal hearing and presbycusis groups

In both the normal hearing and presbycusis groups, the mean percent correct of the processed stimuli was higher than that of the unprocessed stimuli. A significant difference was obtained for both groups in each reverberant condition (Higashi-Yamato Hall: p<0.01; Sophia University lecture hall: p<0.05).

Comparing the results of the normal hearing and presbycusis group, the correct rate of the processed and unprocessed stimuli of the normal hearing group was higher than those of the presbycusis group. There were no differences of improvement between the reverberant conditions for presbycusis group. On the other hand, there is a large difference of improvement between the reverberant conditions for normal hearing group. Thus, we consider that the effect of steady-state suppression is different between the presbycusis and normal hearing group.

3.2. Comparison of the results between the young and the elderly

Figure 5 shows the mean percent correct of overall in this experiment and previous studies with young normal hearing people whose ages were between 18 and 27 [11, 12] in each reverberant environment. In previous studies [11, 12], the correct rate of the processed stimuli was significantly higher than that of the unprocessed stimuli in each reverberant environment. Additionally, we obtained significant improvements by steady-state suppression for elderly people with each reverberant condition in this experiment. Thus, it is confirmed that steady-state suppression improves speech intelligibility for elderly people as well as for young normal hearing people.

Comparing the results of the elderly people in this study and the young people in the previous studies [10], the mean percent corrects of the elderly people with the processed and unprocessed signal were lower than those of the young people in both reverberant environments. This result may support Gordon-Salant and Fitzgibbons [14, 15] who reported that the

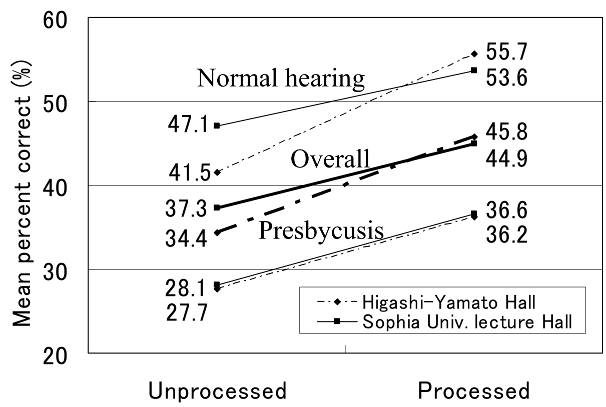

Figure 4: Mean percent corrects of normal hearing, presbycusis group and overall (%)

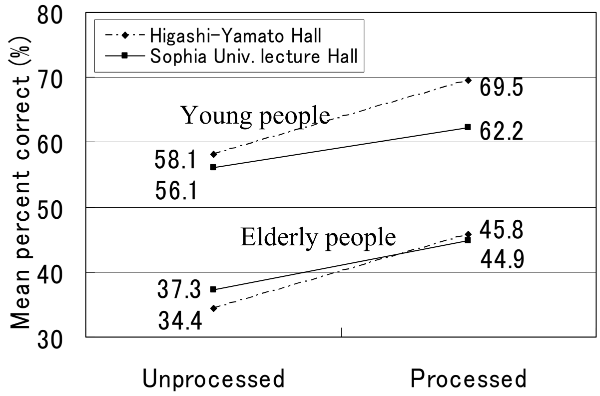

Figure 5: Mean percent corrects of young and elderly people (%)

audibility and auditory temporal resolution degrades with age and elderly people become greatly affected by reverberation. Elderly people tend to be influenced by reverberation more than young people, but the improvement by steady-state suppression in different reverberant conditions for the elderly people was equal (Higashi-Yamato Hall) to or greater (Sophia Univ. lecture Hall) than that for the young people. This result may suggest that steady-state suppression compensates for their auditory temporal processing problem.

4. Conclusions

In this study, we investigated the effect of steady-state suppression for elderly people, and confirmed that steady-state suppression is effective for not only young people but also elderly people in reverberant conditions. The degree of improvement for elderly people was greater than that of young people. And the results of this study indicate that the effect of steady-state suppression is different according to the reverberant conditions and subjects.

For future works, we would like to conduct experiments in the actual environments (e.g. halls or auditoriums) with elderly people, and compare these results with this study and other previous studies [10-12] to investigate the effect of steady-state suppression.

5. Acknowledgements

We are grateful to Hideki Tachibana, Kanako Ueno and Sakae Yokoyama for providing the impulse response data. We thank Tsuyoshi Inoue of the Department from Electrical and Electronics Engineering at Sophia University for his support. We are grateful to every subject for their cooperation in this study. A part of this research was supported by Grants-in-Aid for Scientific Research (A-2, 16203041) from the Japan Society for the Promotion of Science.

6. References

[1] R. H. Bolt and A. D. MacDonald, “Theory of speech masking by reverberation,” J. Acoust. Soc. Am., 21, pp. 577-580, 1949.

[2] A. K. Nabelek and L. Robinette, “Influence of precedence effect on word identification by normally hearing and hearing-impaired subjects,” J. Acoust. Soc. Am., 63, pp. 187-194, 1978.

[3] S. T. Neely and J. B. Allen, “Invertibility of a room impulse response,” J. Acoust. Soc. Am., 66(1), pp. 165-169, 1979.

[4] T. Langhans, and H. W. Strube, “Speech enhancement by nonlinear multiband envelope filtering,” Proc. IEEE ICASSP, 7, pp. 156-159, 1982.

[5] A. Kusumoto, T. Arai, K. Kinoshita, N. Hodoshima and N. Vaughan, “Modulation enhancement of speech by a pre-processing algorithm for improving intelligibility in reverberant environments,” Speech Communication, 45, pp. 101-113, 2005.

[6] T. Arai, “Padding zero into steady-state portions of speech as a preprocess for improving intelligibility in reverberant environments,” Acoust. Sci. Tech, 26, 2005.

[7] 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, pp. 449-450, 2001 (in Japanese).

[8] 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, pp. 229-232, 2002.

[9] S. Furui, “On the role of spectral transition for speech perception,” J. Acoust. Soc. Am., 80(4), pp. 1016-1025, 1986.

[10] 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(2), pp. 212-214, 2005.

[11] N. Hodoshima, T. Inoue, T. Arai, A. Kusumoto and K. Kinoshita, “Suppressing steady-state portions of speech for improving intelligibility in various reverberant environments,” Acoust. Sci. Tech., 25(1), pp. 58-60, 2004.

[12] N. Hodoshima, T. Arai, T. Inoue, K. Kinoshita and A. Kusumoto, ‘‘Improving speech intelligibility by steady-state suppression as pre-processing in small to medium sized halls,” Proc. Eurospeech, pp. 1365–1368, 2003.

[13] N. Hodoshima and T. Arai, “Investigating an optimum suppression rate of steady-state portions of speech that improves intelligibility the most as a pre-processing approach in reverberant environments,” J. Acoust. Soc. Am., Vol. 118, 2005.

[14] P. J. Fitzgibbons and S. Gordon-Salant, “Age effects on discrimination of timing in auditory sequences,” J. Acoust. Soc. Am., pp. 1126-1134, 2004.

[15] S. Gordon-Salant and P. J. Fitzgibbons, “Profile of auditory temporal processing in older listeners,” J. Speech, Language, and Hearing Research, 42, pp. 300-311, 1999.