Improving speech intelligibility by steady-state suppression as pre-processing in small to medium sized halls

Nao Hodoshima[1], Takayuki Arai[1], Tsuyoshi Inoue[1], Keisuke Kinoshita[1] and Akiko Kusumoto[2]

[1] Department of Electrical and Electronics Engineering
Sophia University, Tokyo, Japan
n-hodosh@sophia.ac.jp

[2] Department of Veterans Affairs
Portland VA Medical Center, OR 97207, USA

Abstract

One of the reasons that reverberation degrades speech intelligibility is the effect of overlap-masking, in which segments of an acoustic signal are affected by reverberation components of previous segments [Bolt et al., 1949]. To reduce the overlap-masking, Arai et al. suppressed steady-state portions having more energy, but which are less crucial for speech perception, and confirmed promising results for improving speech intelligibility [Arai et al., 2002]. Our goal is to provide a pre-processing filter for each auditorium. To explore the relationship between the effect of a pre-processing filter and reverberation conditions, we conducted a perceptual test with steady-state suppression under various reverberation conditions. The results showed that processed stimuli performed better than unprocessed ones and clear improvements were observed for reverberation conditions of 0.8 - 1.0s. We certified that steady-state suppression was an effective pre-processing method for improving speech intelligibility under reverberant conditions and proved the effect of overlap-masking.

1. Introduction

In a large auditorium, perceiving speech is often difficult. This is due to reverberation that is caused by a superposition of reflected sounds with various delays and amplitudes. Reverberation is important for music as it provides rich sounds, but it degrades speech intelligibility. Because reverberation tails affect subsequent segments, an acoustic signal of one segment is masked by the reverberation components of the previous portion, and this effect of overlap-masking degrades speech intelligibility [1] [2].

There are several general approaches for improving speech intelligibility in reverberant environments: microphone array, post-processing and pre-processing. Microphone array takes advantage of spatial information about sound source and assures the direction of the desired signal. Thus it can enhance the desired sound source by reducing noise and reverberation [3] [4].

A post-processing method is applied to a speech signal already released into a room and affected by reverberation. As an example of a post-processing approach, inverse filtering [5] [6] and modulation filtering [7] [8] are used. Minimum phase inverse filtering is a dereverberation technique which applies inverse filtering, supposing that the impulse response of a room has a minimum phase. Modulation filtering alters the modulation spectrum of a signal, which is derived from a frequency analysis of the temporal envelope of the band-passed signal.

In the pre-processing approach, a speech signal is processed between a microphone and loudspeaker. Since pre-processing operates on a speech signal between a microphone and loudspeaker, this method can be used with a Public Address (PA) system. Langhans and Strube applied the same technique in pre-processing as they used in post-processing, but no clear improvement was found [7]. It has been discovered that the important modulation frequency of a signal for speech perception is around 4Hz. It has also been confirmed that the peak of the modulation spectrum shifts to the lower modulation frequency and the modulation index is reduced as the acoustic signal is reverberated [9]. Thus, Kusumoto et al. enhanced this particular frequency region in their application of a modulation filter [10]. They showed promising results for improving speech intelligibility.

In their pre-processing approach, Arai et al. suppressed the steady-state portions of speech in order to reduce the influence of overlap-masking caused by reverberation, and they obtained clear improvements [11]. To decrease the effect of overlap-masking, one might think to lessen the energy of preceding portions beforehand so that the energy of reverberation components overlapping to a subsequent portion is attenuated. However, it is reported that while spectral

transition is crucial for syllable perception, vowel nuclei are not necessary for either vowel or syllable perception [12]. Therefore, Arai et al. suppressed the steady-state portions, as these portions of speech have more energy but are less crucial for speech perception, in order to reduce the effect of overlap-masking caused by reverberation tails of previous portions [11]. They confirmed promising results for improving speech intelligibility.

The purpose of our study is to explore the effect of steady-state suppression under various reverberant conditions. Our ultimate goal is to provide a filter for pre-processing which is suitable for an individual auditorium having a distinct reverberation time. In order to achieve this, we need to better understand the relationship between the effect of a pre-processing filter and reverberation condition. However, it is difficult to examine the effect of pre-processing by varying both parameters simultaneously. Thus, we altered only the reverberation condition and explored the effects on a single filter as reverberation condition varied. To explore this relationship, a perceptual experiment has been carried out with the steady-state suppression under reverberation times of 0.9 - 1.3s [13]. The results show that the steady-state suppression with a specific suppression rate prevented the degradation of speech intelligibility within a certain range of reverberation time. The results also show clear improvements for relatively shorter reverberation conditions within the range of 0.9 - 1.3s. In order to investigate the effect of the steady-state suppression at shorter reverberation conditions than those in [13], we conduct a perceptual test with reverberation times of 0.4 - 1.0s.

2. Perceptual Experiment

2.1. Reverberant conditions

The artificial impulse responses h_n were created as Eq. (1) to obtain the desired reverberation conditions [14]:

$$h_n(t) = e^{-t/\tau} h_o(t) \qquad (1)$$

where τ is a time constant. The original impulse response h_o used for this study was measured in the Hamming Hall, Higashi-Yamato City, Tokyo (A reflection board was not used.). Thus, we can obtain the desired reverberation time as a function of τ. Table 1 shows the set of reverberation conditions used in our experiment.

Reverberation time is defined as the time the decay curve of the impulse response decrease 60 dB from steady state. We used Early Decay Time (EDT), which is the time it took for 10 dB of reverberation decay, and we multiplied that by six, to extrapolate the reverberation time.

Table 1: Reverberation conditions used in the experiment

Impulse response	r1	r2	r3	r4	r5
RT (s)	0.4	0.6	0.8	0.9	1.0

2.2. Steady-state Suppression

We applied the same method as in [11] to suppress the steady-state portions of speech. This signal processing calculates the *D* parameter to measure a spectral transition [12] and defines a speech portion as steady-state when *D* is less than a certain threshold. *D* is basically same as the parameter proposed by Furui [12]; in this paper, we used *D* as the mean square of the regression coefficients for each time trajectory of the logarithmic envelope of a subband. Once a portion is considered steady-state, the amplitude of the portion is multiplied by the factor 0.4 (a suppression rate of 40%).

2.3. Stimuli

The original signals consisted of nonsense Consonant-Vowel (CV) syllables embedded in a Japanese carrier phrase. The twenty-four CVs used in the experiment are shown in Table 2. The original speech samples were obtained from the ATR Speech Database of Japanese. The CV syllables were selected from the monosyllable data set. The carrier phrase is a combination of two partial sentences taken from a sentence data set. The beginning position of the target vowel was adjusted to 150ms from the end of the pre-target carrier phrase to control the amount of energy overlapping to the target from the previous portion. We used 150ms because mean durations of Japanese syllables are between 150-200ms.

The stimuli consisted of four conditions: the original signals (Org), the processed signals (Proc), the original signals with reverberation (Org_rev) and the processed signals with reverberation (Proc_rev).

2.4. Subjects

Twenty-two normal hearing subjects (11 males and 11 females, ages 19 to 27) participated in the experiment. All were native speakers of Japanese.

2.5. Procedure

The experiment, controlled by a computer, was conducted in a soundproof room. The stimuli were presented with headphones (STAX SR-303), and the sound level was adjusted to each subject's comfort level. In the experiment, a stimulus was presented at each trial. Then 24 CVs in Kana orthography were

Table 2: CVs used in the experiment

	Voiceless consonants +Vowels	Voiced consonants +Vowels
Stops + Vowels	/pa/ /ta/ /ka/ /pi/ /ki/	/ba/ /da/ /ga/ /bi/ /gi/
Fricatives + Vowels	/sa/ /ʃa/ /ha/ /ʃi/ /hi/	
Affricates + Vowels	/tʃa/ /tʃi/	/dza/ /dʒa/ /dʒi/
Nasals + Vowels		/ma/ /na/ /mi/ /ni/

shown on the PC screen. Subjects were forced to choose one of 24 CVs by clicking a button on the PC screen with a mouse. For each subject, 288 stimuli were presented randomly (5 reverberation conditions x 24 CVs x 2 processing conditions + 24 CVs x 2 processing conditions).

3. Experimental Results

The mean percent correct for each reverberation and processing condition is shown in Table 3. A 2 x 5 ANOVA for repeated measures was performed, confirming significant main effects of processing ($p < 0.001$), impulse response ($p < 0.001$) and interaction ($p = 0.006$). For the comparison of means between processing, a t-test was performed for each impulse response. A significant difference was obtained for the r3-r4 conditions [r3: Org_rev (73.9%), Proc_rev (82.4%), $p < 0.001$; r4: Org_rev (70.1%), Proc_rev (79.2%), $p < 0.001$]. The mean percent correct for each processing condition without reverberation is shown in Table 4.

Table 3: Mean percent correct in each condition with reverberation

Impulse Responses	r1	r2	r3	r4	r5
Org_rev (%)	90.7	84.3	73.9	70.1	69.5
Proc_rev (%)	92.8	86.6	82.4	79.2	73.7

Table 4: Mean percent correct in each condition without reverberation

Org (%)	97.0
Proc (%)	97.2

4. Discussions

The significant main effect of processing is that steady-state suppression improved speech intelligibility in reverberant environments. The interaction indicated that the effect of the steady-state suppression depended on reverberation time. The results show that correct rates for both Org and Proc were almost the same. This shows that speech intelligibility of processed stimuli was not degraded compared to stimuli in which steady-state portions were not suppressed. Therefore, our results confirm that the steady-state suppression is useful for improving speech intelligibility as a pre-processing method and that the effect of the steady-state suppression differed with respect to reverberation time.

The significant main effect of reverberation confirms that correct rates declined as reverberation time increased, regardless of processing. The results also show that the difference of correct rates between Proc_rev and Org_rev decreased as reverberation time shortened. The results indicate that 0.8 s is the lower limit of reverberation time in which we observe significant improvements with the suppression rate of 40 %.

In addition to the results that significant improvements were obtained with reverberation times of 0.8 - 0.9 s, a close to significant difference was obtained for r5 condition [Org_rev (69.5%), Proc_rev (73.7%), $p=0.057$]. The previous study [13] had tested the effect of the steady-state suppression with reverberation times of 0.9 - 1.3 s and significant improvements were observed with those of 0.9 - 1.2s. The results in this study and in [13] indicate that the steady-state suppression is effective for r5 condition because we think that the significant effect of processing may appear continuously in the certain range of reverberation time. Therefore, we concluded that the effect of the steady-state suppression with the suppression rate of 40 % improved speech intelligibility for reverberant times of 0.8 - 1.2s.

5. Conclusions

In this paper, we investigated the range of reverberant conditions for steady-state suppression [11] that improve speech intelligibility in reverberant environments by reducing the effect of overlap-masking. To explore the relationship between the steady-state suppression and several reverberation conditions, we conducted a perceptual test with a set of artificial reverberation conditions. The results showed that clear improvements were obtained with reverberation times of 0.8 – 1.0 s. We certify that steady-state suppression is an effective pre-processing method for improving speech intelligibility under reverberant conditions and

proves the effect of overlap-masking. We predict that the range of reverberant conditions in which clear improvements are observed may be different as we change the suppression rate of the steady-state portions. Thus, we would like to investigate the upper limit of reverberation time which prevents degrading speech intelligibility by the steady-state suppression when the suppression rate is increased. Also we would like to explore the effect of steady-state suppression in an actual hall.

6. Acknowledgements

We appreciate Hideki Tachibana, Kanako Ueno and Sakae Yokoyama for offering to use the impulse response data. Also, we thank the subjects who participated in our experiment.

7. 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] Y. Kaneda and J. Ohga, "Adaptive microphone-array system for noise reduction", IEEE Trans. Acoust. Speech and Signal Process., ASSP-34 (6), pp. 1391-1400, 1986.

[4] J. L. Flanagan, D. A. Berkley, G. W. Elko, J. E. West and M. M. Sondhi, "Autodirective microphone systems", Acoustica, 73 (2), pp. 58-71, 1991.

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

[6] M. Miyoshi and Y. Kaneda, "Inverse filtering of room acoustics", IEEE Trans. Acoust. Speech and Signal Process., 36(2), pp. 145-152, 1988.

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

[8] C. Avendano and H. Hermansky, "Study on the dereverberation of speech based on temporal envelope filtering", Proc. ICSLP, pp. 889-892, 1996.

[9] T. Houtgast and H. J. M. Steeneken, "A review of MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria", J. Acoust. Soc. Am., 77(3), pp. 1069-1077, 1985.

[10] A. Kusumoto, T. Arai, M. Takahashi and Y. Murahara, "Modulation enhancement of speech as a preprocessing for reverberant chambers with the hearing-impaired", Proc. IEEE ICASSP, pp. 933-936, 2000.

[11] T. Arai, K. Kinoshita, N. Hodoshima, A. Kusumoto and T. Kitamura, "Effects on suppressing steady-state portions of speech on intelligibility in reverberant environments", Acoustical Science and Technology, 23, pp. 229-232, 2002.

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

[13] N. Hodoshima, T. Inoue, T. Arai and A. Kusumoto, "Suppressing steady-state portions of speech for improving intelligibility in various reverberant environments", Proc. China-Japan Joint Conference on Acoustics, pp. 199-202, 2002.

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