Steady-state pre-processing for improving speech intelligibility in reverberant environments: Evaluation in a hall with an electrical reverberator

Nahoko Hayashi[1], Takayuki Arai[1], Nao Hodoshima[1], Yusuke Miyauchi[1] and Kiyohiro Kurisu[2]

[1] Department of Electrical and Electronics Engineering, Sophia University, Tokyo, Japan

n-hayash@sophia.ac.jp

[2] TOA Corporation, Kobe, Japan

Abstract

To improve speech intelligibility in reverberant environments, Arai et al. proposed the methods of "steady-state suppression" (Arai, 2002) and "steady-state zero-padding" (Arai, 2005) as a pre-processing method. We conducted two perceptual experiments to evaluate and compare these methods in a hall with an electrical reverberator. The advantage of using this electrical reverberator is that the same subjects are able to participate in the experiments with many different reverberant conditions (reverberation time was varied from 2.6-3.3s, in this study). As results, steady-state suppression showed the floor effect under these reverberant conditions, whereas steady-state zero-padding yielded significant improvements in terms of the speech intelligibility in the same reverberant conditions.

1. Introduction

It is difficult to perceive speech in a room with a long reverberation time because the speech intelligibility is degraded by reverberation. This is due to overlap-masking. In other words reverberant components of one segment mask the segments that follow [1]. In order to reduce overlap-masking, Arai et al. proposed "steady-state suppression" [2] and "steady-state zero-padding" [3]. "Steady-state suppression" is a form of processing to suppress the steady-state portions of speech which have more energy compared with transitions but are less crucial for speech perception. In previous studies, this processing was evaluated under reverberation times of 0.7 – 1.3s in a diotic condition using headphones and a reverberation time of 1.3s in a dichotic condition at a hall in Sophia University. Speech intelligibility showed an improvement of 4.4 – 9.1% and 9.1%, respectively [4-6]. Steady-state zero-padding is a method of separating syllables to reduce the amount of overlap-masking. This is based on the fact that slow speech is more intelligible in reverberant environments.

In the present study, we conducted two experiments. Experiment 1 introduced steady-state suppression into a real acoustic environment and we examined the relationship between the effects of steady-state suppression and the reverberant conditions. In Experiment 2, we evaluated the effect of steady-state zero-padding and compared the results with those of steady-state suppression.

Generally, when experiments are conducted in physically different reverberant environments, it is inconvenient for the same set of subjects to move between different halls. In the present study, the experiments were conducted in a single hall with an electrical reverberator which made it possible to present several reverberant conditions to the same subjects.

2. Methods

The algorithm for detecting steady-state portions in this study is the same as that proposed by Arai et al. [2]. In this technique, an original signal is first split into 1/3-octave subbands and the envelope is extracted in each subband. After down-sampling, the regression coefficients are calculated from the five adjacent values of the time trajectory of the logarithmic envelope of each subband. Then the mean square of the regression coefficients, D, is calculated. This parameter D is similar to what Furui proposed to measure the spectral transition [7]. After up-sampling, we define a portion of speech as 'steady-state' when D is less than a certain threshold.

We processed signals using these two methods: steady-state suppression and steady-state zero-padding. Once a portion of speech is considered to be steady-state, the amplitude of that portion is multiplied by a factor of 0.4 (a suppression rate of 40%) for steady-state suppression, as in [2,4,5,6]. In steady-state zero-padding, we padded zeros in the middle of each steady-state portion [3]. As the length of the padded zero sequence, or T_z is variable in this method, we tested two values: 50 and 100 ms.

3. Reverberant conditions

We conducted the experiments in XEBEC hall (Kobe, Japan). This hall has an electrical reverberator so that the reverberation parameters, such as the delay of reflected sound, and the ratio of direct and reflected sounds, can be changed electrically (see Figs. 1 and 2). In Experiment 1, seven reverberant conditions were set and two of them were also used in Experiment 2. Table 1 shows the values of "Deutlichkeit" (D) and reverberation time (RT) for each reverberant condition.

We used Early Decay Time (EDT), which is the time it takes for the reverberation to decay 10 dB, and we multiplied that decay time by 6 to extrapolate the reverberation time. To calculate the values of D and RT, an impulse response was first split into suboctave bands. Then, the mean Ds and RTs were calculated for each subband having a center frequency of 500, 1000, 2000 Hz, respectively..

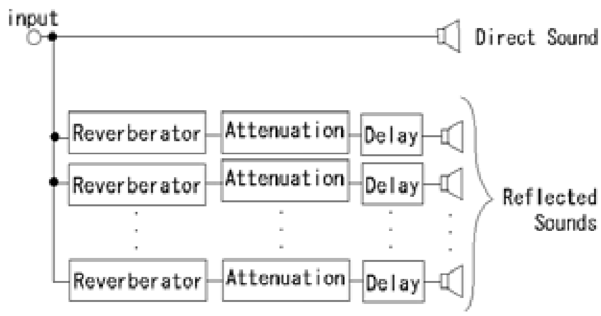

Figure 1. Electrical reverberator system in XEBEC hall (Kobe, Japan)

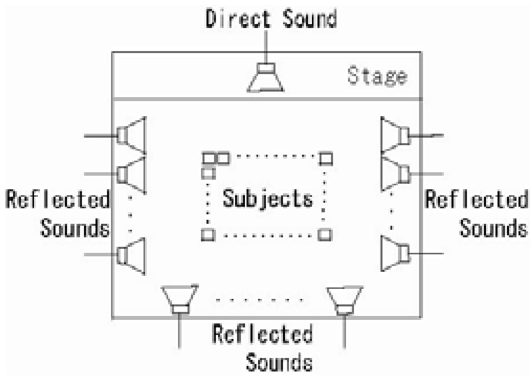

Figure 2. Schematic layout of XEBEC hall (Kobe, Japan)

Table 1. Reverberant conditions

Condition#	1	2	3	4	5	6	7
D [%]	24.1	25.7	28.0	29.6	31.3	32.7	33.1
RT [s]	3.3	3.2	2.7	2.9	2.7	2.6	2.7

4. Experiments

The original speech samples consisted of nonsense consonant-vowel syllables (14 CV syllables) embedded in a Japanese carrier phrase. Thirty-one normal-hearing subjects (native speakers of Japanese) participated in all experiments. The subjects were asked to listen to the CVs and write them down on answer sheets. Each reverberant condition was changed after all stimuli in the reverberant condition were presented in random order.

4.1 Experiment 1

Stimuli were the speech samples with and without steady-state suppression; we call the former "Proc" and the latter "Org". The intelligibility of those two sets of speech samples was compared under the 7 reverberant conditions showed in Table 1.

4.2 Experiment 2

We compared the intelligibility of speech with and without the proposed zero-padding method. As the length of the padded zero sequence T_z was variable in this method, we tested two values: 50 and 100 ms (Proc50 and Proc100). In this experiment, we used two different. reverberant conditions; Conditions 1 and 4 in Table 1. There were four experimental setups in total, that is, two T_z under the two reverberant conditions.

5. Experimental results

5.1 Experiment 1

The results of Experiment 1 are shown in Fig. 3. The mean percent of correct answers with Proc was lower than that of Org in 5 out of the 7 conditions. However, no significant differences were seen between Proc and Org across all conditions ($p > 0.05$). Therefore, steady-state suppression cannot be said to decrease speech intelligibility. Rather, the fact that each mean percent correct was less than 50% supports the idea that perception was much more difficult in the present study than in the previous studies [4-6] using the same CV pairs and carrier phrase.

5.2 Experiment 2

Steady-state zero-padding improved speech intelligibility under the two reverberant environments ($p < 0.05$), even when the reverberation time was relatively long (more than 2.0 s) (see Fig. 4). The results show that a longer T_z prevents the degradation of speech intelligibility more than a shorter one, and that the longer T_z is needed particularly for a longer reverberation time.

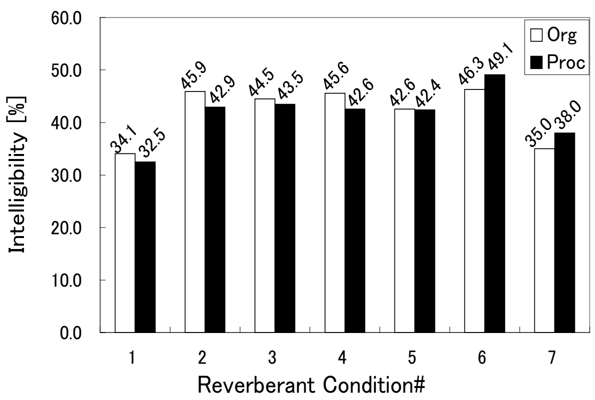

Figure 3. Rresults of Experiment 1

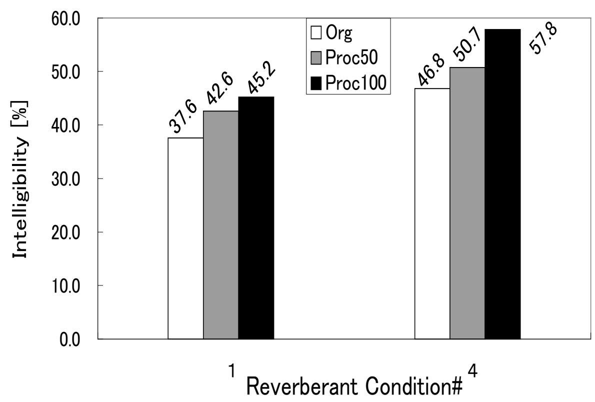

Figure 4. Results of Experiment 2

6. Discussion

Steady-state suppression under the reverberant conditions in Experiment 1 produced no statistically significant improvement in intelligibility. This is thought to result from a floor effect which is caused when the acoustical environment is so poor that speech intelligibility is low. Under such conditions, as in the present study, all stimuli (both Org and Proc) were poorly recognized, and as a result there was no difference in correct rates between Org and Proc. Therefore, it is reasonable to assume that Experiment 1 was conducted under reverberant conditions where the floor effect occurred, because the reverberation times of this experiment are much longer than ones used in the previous experiments [4-6]; 2.6 - 3.3s in the present study and 0.7 –1.3s in the previous studies.

As shown in Fig 3, the intelligibility of Proc in Conditions 6 and 7 is greater than that of Org. Conditions 6 and 7 are the two conditions in which the reverberation time and D value are highest among all the conditions in the experiment. Thus, it can be expected that steady-state suppression improves speech intelligibility in an environment where the RT is shorter and the D value is higher than the ones in the present experiment. Therefore, we suspect that steady-state suppression is effective within a certain range of reverberant conditions. When the reverberant conditions are better or worse than those in the effective range, processing becomes less effective. Further study is necessary to clarify the relationship between the effects of steady-state suppression and the reverberation time and D value.

In Experiment 2, it was shown that steady-state zero-padding is effective as a pre-processing approach to improve speech intelligibility in reverberant environments. The remarkable point is that steady-state zero-padding showed statistically significant improvements whereas steady-state suppression did not show any improvements under the same reverberant conditions. Thus, it can be said that steady-state zero-padding is effective in a room with long reverberation time whereas steady-state suppression is not effective when the reverberation time is long.

7. Conclusions

In these experiments, we confirmed that steady-state zero-padding is effective for long reverberation times, and that steady-state suppression has a lower limit with respect to the reverberant condition under which it improves speech intelligibility. Furthermore, an electrical reverberator in a hall is very useful, because it is possible to conduct multiple experiments under many different reverberant conditions with the same set of subjects. However, accurate set up of the acoustical environment is necessary.

In future experiments, we would like to clarify the effect of steady-state suppression in a hall under other conditions such as shorter reverberation time and higher D value to investigate the correlation of steady-state suppression and reverberant environments.

The length of a speech signal with steady-state zero-padding is longer than the original speech sample. Therefore, it is preferable that the speaker not be situated in the same room where the processed speech signal is broadcast, otherwise listeners may experience a visual-auditory discontinuity. This phenomenon is particularly noticeable under conditions such as an emergency broadcast in a traffic tunnel. Tunnels are spaces in which the reverberation time is long, so steady-state zero-padding might be more useful than steady-state suppression in these applications. Further tests in such reverberant environments are necessary.

Acknowledgements

This research was supported in part by a Grant-in-Aid for Scientific Research (A-2, 16203041) from the Japan Society for the Promotion of Science. The authors would like to thank Messrs. Tsuyoshi Inoue, Takahito Goto, Fumihiro Tadokoro from the Arai Lab., Sophia University, for their assistance in carrying out the experiments, and XEBEC Corporation for providing the hall used in the experiments.

References

[1] R. H. Bolt and A. D. Macdonald, “Theory of speech masking by reverberation,” J. Acoust. Soc. Am., 21(6), 577-580,1949.

[2] 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(4), 229-232, 2002.

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

[4] 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), 58-60, 2004.

[5] 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, 1365-1368, 2003.

[6] 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.

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