Improving Consonant Identification in Noise and Reverberation by Steady-state Suppression as a Preprocessing Approach

Nao Hodoshima, Wataru Yoshida and Takayuki Arai

Department of Information and Communication Sciences, Sophia University, Tokyo, Japan
n-hodosh@sophia.ac.jp

Abstract

Noise (N), reverberation (R), and a combination of N and R (NR) differently degrade speech intelligibility. The current study aims to improve speech intelligibility in public spaces by processing speech signals through public address systems (a preprocessing approach). As a preprocessing approach, we proposed steady-state suppression, and it has improved consonant identification in R. The current study tests the effect of steady-state suppression in N, R, and NR at three signal to noise ratios and at reverberation time of 0.9 s. Results showed that steady-state suppression significantly improved consonant identification by 21 young people in NR and in R. Furthermore, steady-state suppression improved consonant identification more in NR than in R. The results indicate that steady-state suppression may be applicable to public spaces having N and R. The results also indicate that an integration of N and R improves the performance of a preprocessing approach in a certain range of N and R.

Index Terms: speech intelligibility, preprocessing, steady-state suppression, reverberation, noise

1. Introduction

We sometimes experience difficulties in understanding speech sounds in large public spaces. This is partly because noise and/or reverberation mask the speech sounds we are listening to. For an appropriate speech transmission in public spaces used by various types of people (e.g., an emergency alarm given by a spoken message in stations and airports), less noise and reverberation are therefore preferred.

Several approaches have been proposed for improving speech intelligibility in noise and/or reverberation without changing room acoustic characteristics. One approach is preprocessing that processes a speech signal before sending it from loudspeakers. An example of preprocessing in noise is a consonant enhancement [1]. The consonant enhancement significantly improved consonant identification for young people in noise which had the same long-term average spectrum as the speech signal at a signal to noise ratio (SNR) of 0 dB [1].

An example of preprocessing technique in reverberation is steady-state suppression [2, 3]. This technique decreases overlap-masking (i.e., reverberant masking) [5, 6] by suppressing steady-state portions of speech that are relatively unimportant for syllable perception and have much energy compared to spectral transitions [4]. Several listening tests showed that steady-state suppression statistically improved consonant identification for young people with normal hearing [7, 8] and for elderly people [9] at reverberation times (RTs) of 0.7 to 1.3 s.

Noise and reverberation differently degrade speech intelligibility, and the interaction of noise and reverberation adversely affects speech perception to a greater extent than the sum of both effects taken independently [e.g., 10, 11]. This means that speech intelligibility in a combination of noise and reverberation is not simply obtained from that in noise or in reverberation alone. This also indicates that the effect of a signal processing approach in a combination of noise and reverberation will be different from that in noise or in reverberation alone. In order to apply a signal processing approach in public spaces, it must reduce noise and reverberation, not noise alone or reverberation alone, since most public spaces that need higher speech intelligibility have both noise and reverberation. As far as we know, no preprocessing approach has been tested in a combination of noise and reverberation.

The goal of our study is to improve speech intelligibility in public spaces. We think preprocessing will be a suitable approach for improving speech intelligibility in public spaces because the listeners do not need to attach special listening devices. The first purpose of the current study is to test whether steady-state suppression improves consonant identification in a combination of noise and reverberation. If steady-state suppression improves consonant identification in a combination of noise and reverberation as well as in reverberation, we have a possibility to apply steady-state suppression in a wider range of public spaces. The second purpose of the current study is to compare the effect of steady-state suppression in a combination of noise and reverberation with that in noise alone or in reverberation alone. This will find how an integration of noise and reverberation affects the performance of steady-state suppression. We think these discussions are important because the results in the current study will be relevant when we evaluate a signal processing approach used in public spaces. In order to test the purposes, the current study carried out a consonant identification test under two processing conditions (with/without steady-state suppression) and four listening conditions: reverberant, noisy, noisy and reverberant, and non-reverberant.

2. Listening test

2.1. Participants

Participants were 21 native speakers of Japanese (11 males and 10 females, 19 to 23 years old). They had self-reported normal hearing.

2.2. Stimuli

The original speech sentences were 14 nonsense consonant-vowel syllables (consonants: /p, t, k, b, d, g, s, ʃ, h, dz, dʒ, tʃ, m, n/ and vowel: /a/) embedded within a Japanese carrier

Table 1. *Conditions used in the current study.*

Reverberation		w/o		w/	
Processing		w/o	w/	w/o	w/
SNR (dB)	∞	nonrev	nonrev_p	rev	rev_p
	10	noisy	noisy_p	noisy_rev	noisy_rev_p
	5				
	0				

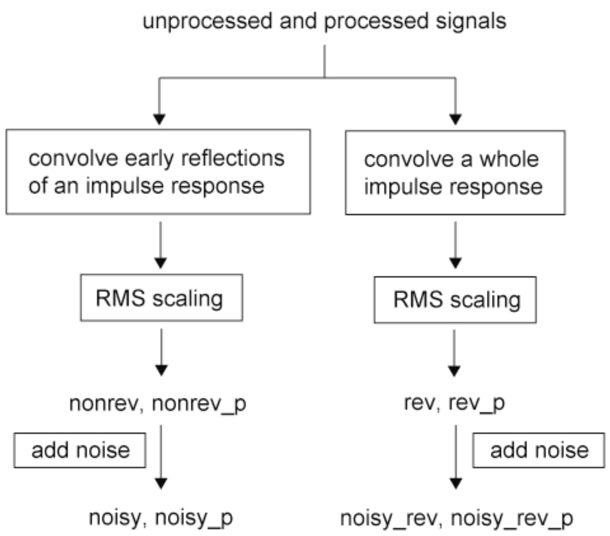

Figure 1. *Flow of preparing stimuli used in the current study.*

phrase obtained from the ATR speech database of Japanese [12], and were the same as those used in the previous studies [7-9]. The talker was a male native speaker of Japanese (40 years old).

Table 1 shows conditions used in the current study. The listening condition is non-reverberant (nonrev), reverberant (rev), noisy, or a combination of noisy and reverberant (noisy_rev). In each listening condition, two processing conditions, unprocessed and processed by steady-state suppression, were applied. As the noisy condition, we used babble noise taken from the NOISEX-92 database [13] in order to simulate noise in public spaces. The intensity of the noise was A-weighted, and was scaled to have the same root mean square (RMS) as that of the non-reverberant and reverberant conditions. Four SNR were prepared: ∞, 10, 5 and 0 dB. SNR of ∞ is equivalent to the condition where no noise is added to speech sentences. As the reverberant condition, we used an impulse response measured in the XEBEC Hall (reverberation time of 0.9 s).

Figure 1 shows the flow of how we prepared stimuli used in the current study. For the non-reverberant and noisy conditions (the left part of Fig. 1), unprocessed and processed speech sentences were convolved with early reflections arriving within 50 ms of the direct sound of the impulse response in order to take account for the precedence effect [14]. Then the intensity of the sentences was A-weighted, and each sentence was scaled to have the same RMS. Then noises were added to the speech sentences.

For the reverberant and a combination of noisy and reverberant conditions (the right part of Fig. 1), unprocessed and processed speech sentences were convolved with the whole impulse response. Then, coefficients used in the RMS scaling were multiplied to the speech sentences so that speech sentences in non-reverberant condition have the same intensities as those in reverberant condition. Then noises were added to the speech sentences.

2.3. Procedure

The listening test was conducted in a sound treated room. The stimuli were presented diotically over headphones (STAX, SR-303) through a digital audio amplifier (Onkyo, MA-500U) that was connected to a computer. Six practice trials were held to familiarize the participants with the procedure beforehand. The playback level was adjusted to 65.0 dBA for all participants beforehand, and this level was maintained throughout the main session. In each trial, a stimulus was presented, after which a computer monitor displayed the 14 syllables used in the listening test, /a/ and "others", all in Kana orthography. The participants were instructed to answer what they heard by clicking the mouse on the computer monitor. When participants chose "others", they were instructed to write down what they heard on an answer sheet. Trials with 196 stimuli except for non-reverberant condition (14 listening conditions x 14 speech materials) were randomly presented first, followed by 28 stimuli for non-reverberant condition (2 listening conditions x 14 speech materials) presented randomly.

3. Results and discussion

3.1. Results

Figure 2 shows mean percent correct responses (correct rates) in each condition. Table 2 shows improvement in intelligibility by steady-state suppression (correct rate in unprocessed condition – that in processed condition) in each listening condition. A 2 x 3 x 2 ANOVA was carried out with reverberation (with/without), noise (SNR of 10, 5 and 0) and processing (with/without) as repeated variables, and the correct rate as the dependent variable. Results showed that processed speech had significantly higher correct rates than unprocessed speech [$p < 0.01$]. Results also showed that the correct rate significantly decreased as SNR decreased [$p < 0.01$]. The significant interaction between reverberation and noise [$p < 0.01$] showed that the correct rate with reverberation decreased significantly more than that without reverberation as SNR decreased.

A Sidak multiple comparison test showed that processed speech had significantly higher correct rates than unprocessed speech at all SNR conditions in reverberation [SNR of 10 dB: $p < 0.05$, SNR of 5 dB: $p < 0.01$, SNR of 0 dB: $p < 0.01$]. A t-test showed that processed speech had significantly higher correct rates than unprocessed speech in reverberation [$p < 0.01$]. T-tests also showed that correct rates of unprocessed and processed speech in reverberation were significantly higher than those of in noise and reverberation at SNR of 10 dB, respectively [$p < 0.01$].

3.2. Non-reverberant and reverberant conditions

As expected, correct rates in the non-reverberant condition were very high (94.9% for unprocessed and 97.6% for processed), and decreased to 59.2% for unprocessed and 66.3% for processed condition in reverberation. Consistent previous studies [7-9], steady-state suppression significantly

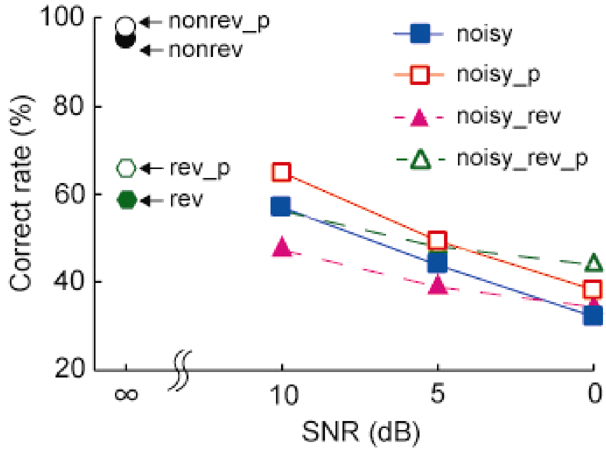

Figure 2. *Mean percent correct responses in each condition. Filled symbols show unprocessed condition and open symbols ('_p' in captions) show processed condition.*

improved speech intelligibility in reverberation.

3.3. Noisy conditions

Consistent with the previous study (e.g., [15]), correct rates significantly decreased as SNR decreased. Processed speech had higher correct rates than unprocessed speech in noise (see Table 2), but no significant differences were observed between unprocessed and processed conditions in all SNRs. This result showed that steady-state suppression was not effectively able to reduce babble noise alone because steady-state suppression aimed to reduce overlap-masking by decreasing amplitudes of steady-state portions. This result indicates that reverberation and noise differently affect the performance of a signal processing approach.

Steady-state suppression improved consonant identification by a small amount in noise, possibly because consonants and spectral transitions were relatively enhanced by steady-state suppression [16]. This possible reason was derived from the results of the listening test where steady-state suppression significantly improved consonant identification by elderly people in the non-reverberant condition [16]. The results indicate that steady-state suppression might be also effective for elderly people in noise.

3.4. Combination of noisy and reverberant conditions

As was shown in Fig. 2, overall correct rates significantly decreased as SNR decreased in reverberation. At SNR of 10 dB, stimuli presented in noise had significantly higher correct rates than stimuli presented in a combination of noise and reverberation [$p < 0.05$]. At SNR of 0 dB, on the other hand, stimuli presented in a combination of noise and reverberation had higher correct rates than stimuli presented in noise. This interesting result was consistent with the finding [17] that reverberation increased speech intelligibility under a high noisy condition (e.g., SNR of -5 and -10 dB) when reverberation time is increased from 2 to 4 s, while reverberation usually decreases speech intelligibility under such a long reverberation time. Our results indicate that an integration of noise and reverberation differently affects speech intelligibility with different combinations of noise and reverberation levels. Further research would find the integration mechanism with different combinations of SNR and reverberation time from the conditions used in the current study.

Table 2. *Improvement in intelligibility by steady-state suppression in each listening condition. * indicates a statistically significant improvement at $p < 0.05$.*

Condition	SNR (dB)	Improvement in intelligibility by steady-state suppression (%)
nonrev	∞	2.7
rev	∞	7.1*
noisy	10	6.2
	5	5.4
	0	5.8
noisy_rev	10	8.5*
	5	8.5*
	0	9.5*

Contrary to no improvement in intelligibility in noise alone, steady-state suppression significantly improved speech intelligibility in a combination of noise and reverberation as well as in reverberation. This result showed that we have a possibility to apply steady-state suppression in public spaces that have both noise and reverberation or have reverberation where the listening condition is similar to the current study.

When we compared the effect of steady-state suppression in each listening condition, the improvement in intelligibility was slightly larger in a combination of noise and reverberation than that in reverberation alone (see Table 2). This may be because of the benefit of noise and reverberation integration. That is, reverberation increased speech intelligibility when noise is added to speech signals as was observed at SNR of 0 dB. Further research would find if the noise and reverberation integration would help steady-state suppression at SNR of lower than 0 dB, and would find the range of listening condition of noise and reverberation where steady-state suppression is effective.

4. Conclusions

The current study 1) examined whether steady-state suppression improved consonant identification in a combination of noise and reverberation and 2) compared the effect of steady-state suppression in a combination of noise and reverberation with that in noise or in reverberation alone. The results showed that steady-state suppression improved consonant identification in a combination of noise and reverberation as well as in reverberation alone under SNR of 10, 5 and 0 and under a reverberation time of 0.9 s. This indicates a possibility for applying steady-state suppression to a wider range of public spaces. The results also showed that steady-state suppression improved speech intelligibility more in a combination of noise and reverberation than in reverberation alone. This indicates the benefit of noise and reverberation integration where reverberation increases speech intelligibility under a certain range of high noise condition [17]. Further research would find the range of listening condition of noise and reverberation where the noise and reverberation integration increases speech intelligibility and where steady-state suppression is effective. For an application of steady-state suppression to public spaces, we would need more realistic listening environments such as a binaural listening test or a sentence comprehension test. It is also interesting to test a hybrid technique of steady-state suppression and some other approach such as a noise cancelation so that we may improve speech intelligibility in public spaces where noise is too dominant compared to reverberation.

5. Acknowledgements

This research was supported by Sophia University Open Research Center from MEXT.

6. References

[1] Hazan, V. and Simpson, A., "The effect of cue-enhancement on consonant intelligibility in noise: Speaker and listener effects", Language and Speech, 43(3):273-294, 2000.
[2] Arai, T., Kinoshita, K., Hodoshima, N., Kusumoto, A. and Kitamura, T., "Effects of suppressing steady-state portions of speech on intelligibility in reverberant environments", Proc. Autumn Meet. Acoust. Soc. Jpn., 1:449-450, 2001.
[3] Arai, T., Kinoshita, K., Hodoshima, N., Kusumoto, A. and Kitamura, T., "Effects on suppressing steady-state portions of speech on intelligibility in reverberant environments", Acoust. Sci. Tech., 23:229-232, 2002.
[4] Furui, S., "On the role of spectral transition for speech perception", J. Acoust. Soc. Am., 80:1016-1025, 1986.
[5] Bolt, R. H. and MacDonald, A. D., "Theory of speech masking by reverberation", J. Acoust. Soc. Am., 21:577-580, 1949.
[6] Nábělek, A. K., Letowski, T. R. and Tucker, F. M., "Reverberant overlap- and self-masking in consonant identification", J. Acoust. Soc. Am., 86:1259-1265, 1989.
[7] Hodoshima, N., Arai, T., Kusumoto, A. and Kinoshita, K., "Improving syllable identification by a preprocessing method reducing overlap-masking in reverberant environments", J. Acoust. Soc. Am., 119(6):4055-4064, 2006.
[8] Hodoshima, N., Goto, T., Ohata, N., Inoue, T. and Arai, T., "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):212-214, 2005.
[9] Miyauchi, Y., Hodoshima, N., Yasu, K., Hayashi, N., Arai, T. and Shindo, M., "A preprocessing technique for improving speech intelligibility in reverberant environments: The effect of steady-state suppression on elderly people", Proc. Interspeech, 2769-2772, 2005.
[10] Nábělek, A. K. and Pickett, J. M., "Monaural and binaural speech perception through hearing aids under noise and reverberation", J. Speech Hear. Res., 17:724-739, 1974.
[11] Cranedll, C. C. and Smaldino, J. J., "Classroom acoustics for children with normal hearing and with hearing impairment", Language, Speech and Hearing Services in Schools, 31:362-370, 2002.
[12] Takeda, K., Sagisaka, Y., Katagiri, S. and Kuwabara, H., "A Japanese speech database for various kinds of research purposes", J. Acoust. Soc. Jp., 44(10):747-754. 1988.
[13] Varga, A. and Steeneken, H. J. M., "Assessment for automatic speech recognition II: NOISEX-92: a database and an experiment to study the effect of additive noise on speech recognition systems", Speech Communication, 12(3):247–251, 1993.
[14] Watkins, A. J., "The influence of early reflections on the identification and lateralization of vowels", J. Acoust. Soc. Am., 106(5): 2933-2944, 1999.
[15] Bradley, J. S., Reich, R. D. and Norcross, S. D., "On the combined effects of signal-to-noise ratio and room acoustics on speech intelligibility", J. Acoust. Soc. Am., 106(4):1820-1828, 1999.
[16] Kobayashi, K., Yasu, K., Hodoshima, N., Arai, T. and Shindo, M., "A study of syllable enhancement for elderly listeners by suppressing energy of steady-state portions of vowels", J. Acoust. Soc. Jpn., 2008 (to be published).
[17] Sato, H., Nagatomo, M., Yoshino, H. and Yajima, Y., "Experimental study on evaluation of effect of reverberation and noise on speech intelligibility in rooms", J. Archit. Plan. Environ. Eng., 484:1-8, 1996.