Acoust. Sci. & Tech. **23**, 4 (2002)

ACOUSTICAL LETTER

Effects of suppressing steady-state portions of speech on intelligibility in reverberant environments

Takayuki Arai*, Keisuke Kinoshita, Nao Hodoshima, Akiko Kusumoto and Tomoko Kitamura

Department of Electrical and Electronics Engineering, Sophia University

(*Received 14 February 2002, Accepted for publication 4 April 2002*)

Keywords: Speech enhancement, Reverberation, Speech intelligibility, Masking, Modulation filtering

PACS number: 43.72.Ew

1. Introduction

When listening to a lecture in a large auditorium it is often difficult to understand the speech. Among other factors, comprehension may be impaired by reverberation, which is sound reflecting from the wall, interfering with direct sound. Based on the modulation transfer function (MTF), the speech transmission index (STI) has been proposed as an objective measure for speech intelligibility in rooms [1].

There are two types of processing for improving speech intelligibility under reverberant conditions. One is called pre-processing, where signal processing is done before a signal is radiated through a loudspeaker to prevent intelligibility degradation. The other is called post-processing, where signal processing is done on reverberant speech. Langhans *et al.* [2] proposed a method for post-processing using a single microphone, and Avendano *et al.* [3] extended their method. Both researchers used so-called ''modulation filtering,'' or filtering on a temporal envelope, after dividing the speech signal into a series of subbands.

Langhans *et al.* [2] also tried modulation filtering for the pre-processing method. In their study, however, a large improvement was not reported. Our group has also tested speech intelligibility based on a similar type of modulation filtering [4–7]. In our previous studies, we emphasized modulation characteristics of speech around 3–6 Hz which corresponds to the syllabic rate of speech, and as a result, we observed some tendencies toward improvement, although they were not significant.

In the present study, we focus on suppressing the steady-state portions of speech to investigate the effect of modulation filtering on speech intelligibility. In light of the fact that modulation filtering essentially suppresses steady-state portions and emphasizes speech dynamics, we hypothesize that steady-state suppressed speech sounds are more robust in the case of reverberation. To test this hypothesis, we compare intelligibilities of speech after adding reverberation to speech signals with and without steady-state suppression.

2. Principle

2.1. Masking effect

One of the major reasons reverberation lowers speech intelligibility is that it smears subsequent speech segments. In other words, the reverberant components of one segment mask the segments that follow. As a result, the segments following reverberating segments are harder to hear. This is especially true when the reverberating segment has more power, such as a vowel, and the subsequent, less, such as a consonant.

Figures 1(a) and (b) show this effect conceptually. Figure 1(a) is the original waveform of a nonsense word /aka/, and Fig. 1(b) is a simulated waveform obtained by taking the convolution of the impulse response of a hall and only the first vowel part of the original waveform. In Fig. 1(b), one can see that due to reverberation the tail of the first vowel /a/ has relatively high power compared to that of the original waveform of the following /ka/ (shown with a lighter color in this figure).

Thus, in reverberation, a segment is always overlaid by the reverberant components of previous segments. To reduce this effect, one may suppose that suppressing the power of previous segments will increase intelligibility. However, because this approach results in the suppression of all segments, overall intelligibility actually declines. Instead, suppressing the power of only steady-state portions is a promising venture, because the information in steady-state portions of the speech signal is relatively redundant with that in transient segments [8].

2.2. Steady-state suppression

We used the *D* parameter by Furui (1986) to measure the degree of steadiness of speech [9]. In Furui's study, *D* is defined as the mean square of the regression coefficients for each time trajectory of cepstrum. In this study, we used the mean square of the regression coefficients for each time trajectory of the logarithmic envelope of a subband, where a subband was obtained by 1/3-octave analysis to simulate the critical bands. The regression coefficients were calculated from the five adjacent values of the time trajectory, after downsampling from the original sampling rate of 16 kHz, to 100 Hz.

We define a speech portion as steady-state when *D* is less than a certain threshold. Once a portion is considered steady-state, the amplitude of the portion is multiplied by a factor less than 1.0 (in this study, we used 0.4).

Figure 2(a) is an original speech signal, and Fig. 2(b) is the reverberant signal simulated by taking the convolution of the impulse response of a hall and the original signal (a). Figure 2(c) shows the suppressed waveform of the original (a). Figure 2(d) is the reverberant signal of the suppressed signal (c).

*e-mail: arai@sophia.ac.jp

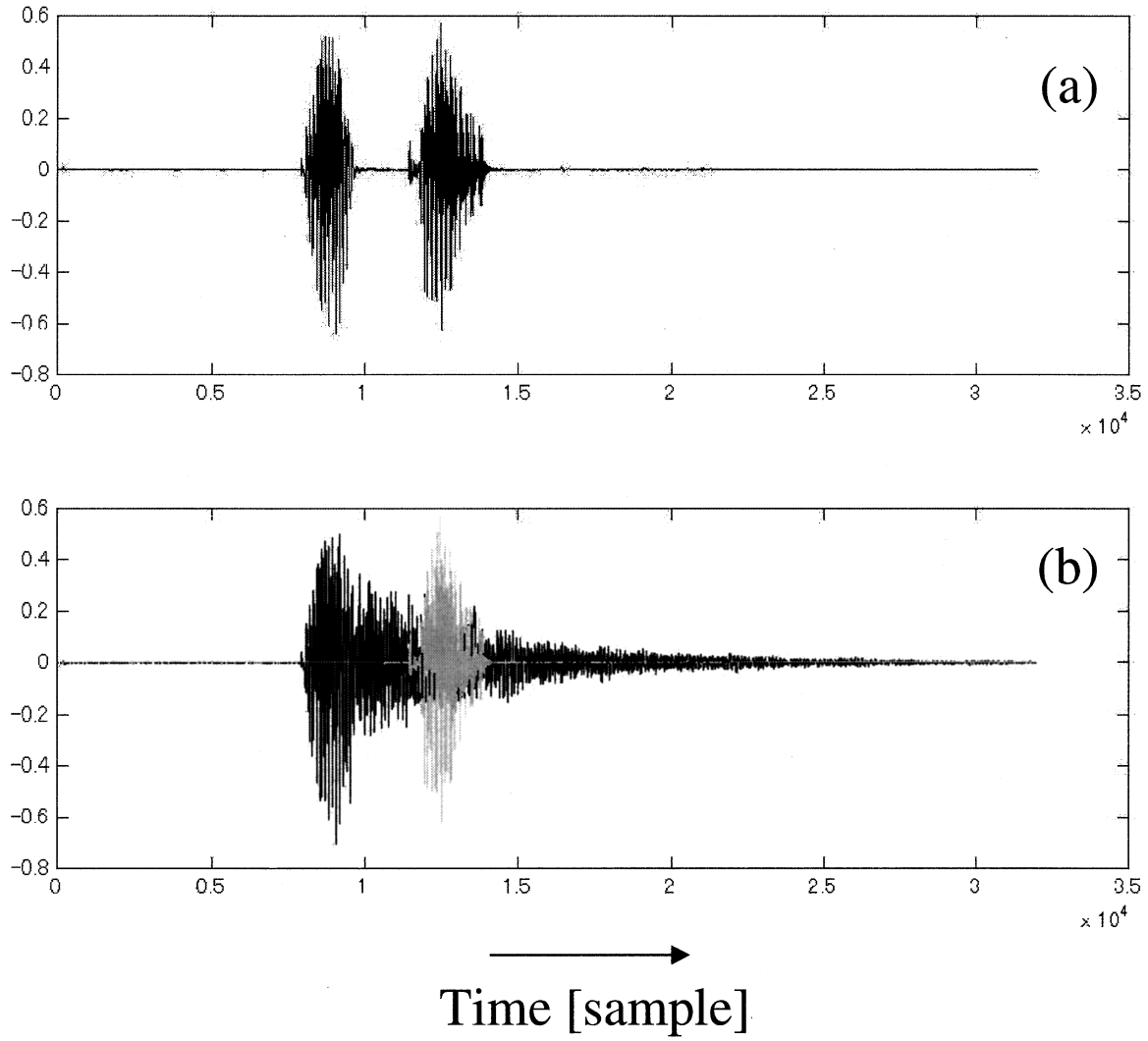

Fig. 1 (a) Original waveform of a nonsense word /aka/, (b) simulated waveform obtained by taking the convolution of the impulse response of a hall and only the first vowel part of the original waveform. The /ka/ portion of the original waveform is shown with a lighter color in (b).

3. Experiment

We conducted a perceptual experiment to see whether our proposed method prevents degradation of speech intelligibility in the presence of reverberation. In the experiment, 20 Japanese syllables were used for an articulation test. Each syllable (target) was embedded within a carrier sentence. We used three processing conditions: ''no_proc,'' ''half_proc,'' and ''whole_proc.'' The ''no_proc'' condition was the control, that is, no processing was applied. In the case of ''half_proc,'' the steady-state suppression was applied but only for the first half of the sentence (up to just before the target). Because we wanted to see the sole effect of masking due to reverberation, this ''half_proc'' condition was included in the experiment. In the case of ''whole_proc,'' the steady-state suppression was applied for the whole sentence.

Two impulse responses from two halls were used: Hall A (reverberation time was 1.1 s) and Hall B (reverberation time was 1.8 s). For each sentence, there were six conditions (3 types of processing × 2 halls). Twelve subjects participated in the experiment.

No subject reported having any previous hearing problem. Each subject listened to 60 stimuli and was asked to choose one of 20 syllables based on what they heard for each trial.

4. Results and discussions

The experimental results are listed in Table 1. The subjects performed better when processing both the partial and whole speech signals for Hall A, although the differences were not statistically significant ($p > 0.05$). For Hall B, however, the subjects performed better only when half of the signal was processed, but not when the whole signal was processed. It seems that because the reverberation was so strong in Hall B, the subjects had difficulty perceiving

Table 1 % of correct responses.

Hall	A	B
no_proc	81.7	70.0
half_proc	87.5	72.5
whole_proc	85.8	60.8

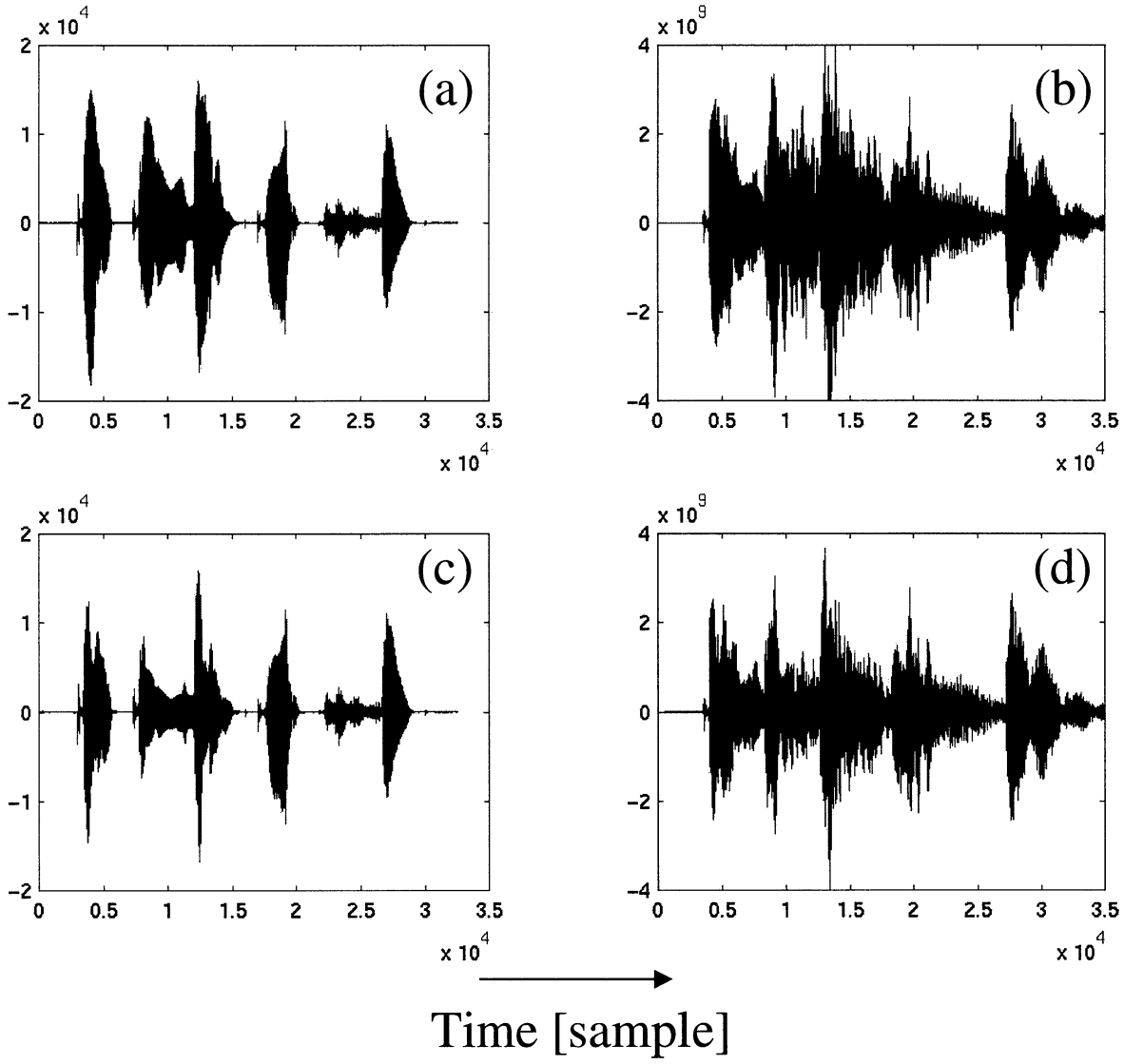

Fig. 2 (a) Original waveform of a speech signal, (b) reverberant signal simulated by taking the convolution of the impulse response of a hall and the original signal, (c) suppressed signal of the original, and (d) reverberant signal of (c).

syllables correctly when the whole sentence was processed. Even with this strong reverberation, we observed some tendency toward improvement (from 70.0% to 72.5%) by reducing amount of masking (this difference was also not statistically significant).

5. Conclusions

We compared intelligibilities of speech in reverberant environments with and without steady-state suppression. As a result, we confirm that 1) reducing masking of the reverberant components of the previous segments prevents the degradation of speech intelligibility, and 2) suppressing steady-state portions of speech is useful as a pre-processing technique in a certain reverberation condition. Because modulation filtering essentially suppresses steady-state portions and the steady-state suppression reduces amount of masking, we can conclude that the modulation filtering also reduces this masking effect.

Acknowledgement

We would like to thank to Prof. Tachibana of the Univ. of Tokyo, and the members of his lab., especially to Kanako Ueno and Sakae Yokoyama, who let us use the impulse responses of the halls. We would also like to thank to Prof. Furui of Tokyo Institute of Technology, who provided the speech data for the articulation test.

References

[1] T. Houtgast and H. J. Steeneken, "A review of the MTF concept in room acoustics and its use of estimating speech intelligibility in auditoria," *J. Acoust. Soc. Am.*, **77**, 1069–1077 (1985).

[2] T. Langhans and H. W. Strube, "Speech enhancement by nonlinear multiband envelope filtering," *Proc. IEEE ICASSP*, pp. 156–159 (1982).

[3] C. Avendano and H. Hermansky, "Study on the dereverberation of speech based on temporal envelope filtering," *Proc. ICSLP*, pp. 889–892 (1996).

[4] A. Kusumoto, T. Arai, T. Kitamura, M. Takahashi and Y. Murahara, "Speech processing on the room acoustics for the

hearing-impaired,'' *Proc. Autumn Meet. Acoust. Soc. Jpn.*, Vol. 1, pp. 389–390 (1999).

[5] A. Kusumoto, T. Arai, T. Kitamura, M. Takahashi and Y. Murahara, ''Modulation enhancement of speech as a preprocessing for reverberant chambers with the hearing-impaired,'' *Proc. IEEE ICASSP*, Vol. 2, pp. 853–856 (2000).

[6] T. Kitamura, T. Arai, A. Kusumoto and Y. Murahara, ''Modulation filtering in a robust speech processing against reverberation for the hearing-impaired,'' *Proc. Spring Meet. Acoust. Soc. Jpn.*, Vol. 1, pp. 333–334 (2000).

[7] T. Kitamura, K. Kinoshita, T. Arai, A. Kusumoto and Y. Murahara, ''Designing modulation filters for improving speech intelligibility in reverberant environments,'' *Proc. ICSLP*, Vol. 3, pp. 586–589 (2000).

[8] H. Hermansky and N. Morgan, ''RASTA processing of speech,'' *IEEE Trans. Speech Audio Process.*, **2**, 578–589 (1994).

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