MODULATION ENHANCEMENT OF SPEECH AS A PREPROCESSING FOR REVERBERANT CHAMBERS WITH THE HEARING-IMPAIRED

A. Kusumoto, T. Arai, T. Kitamura, M. Takahashi, Y. Murahara

Dept. of Electrical and Electronics Eng., Sophia University
7–1 Kioi-cho, Chiyoda-ku, Tokyo, JAPAN

ABSTRACT

In this paper we report on a method for reducing the degradation of speech intelligibility in public halls caused severe reverberation. Hall reverberation makes speech more difficult to understand, particularly for the hearing-impaired. Our method involves processing the speech audio signal between a microphone and a loudspeaker that radiates the speech into the room. As there is a strong correlation between the modulation spectrum and the intelligibility of speech, we filtered the speech in the modulation frequency domain. Using several modulation filters, we conducted perceptual experiments with hearing-impaired subjects and asked their preference in a church. The experiments indicate that enhancing the modulation frequencies between 2 and 8 Hz improves intelligibility in reverberant environments. The four hearing-impaired subjects rated the processed speech easier to hear than the unprocessed speech.

1. INTRODUCTION

Many public halls are designed for multiple purposes such as musical concerts and lectures. Reverberation is usually preferable for musical performances. Unfortunately, reverberation also degrades speech intelligibility for lectures held in the same hall [1, 2].

It has been reported that there is a strong correlation between the modulation transfer function (MTF) and the intelligibility of speech [3, 4]. RASTI (RApid Speech Transfer Index) based on the MTF is widely used for measuring speech intelligibility in auditorium [5, 6].

The peak of the modulation spectrum of speech is located about 4 Hz in the modulation frequency region; when the speech is reverberant, its peak shifts to lower modulation frequency and its modulation index declines [5]. Thus, the MTF with a reverberant condition has low-pass characteristics in the modulation frequency domain.

In a large auditorium, speech is usually converted into an electric signal by a microphone, and then it is amplified by an amplifier. Finally, it is radiated from a loudspeaker as an acoustic wave that reaches the ears of the audience. In the present study, we developed an algorithm for a pre-processor within the conversion process between the microphone and loudspeaker to prevent severe degradation of the speech intelligibility due to the reverberation. In Section 2, we describe the algorithm based on several modulation filters that are applied to the temporal contours of subband envelopes of speech. The perceptual experiment using the processed speech is described in Section 3.

2. MODULATION FILTERING

As there is a strong correlation between the MTF and the speech intelligibility, a potential method to mitigate degradation of the intelligibility consists of enhancing a specific modulation frequency components of speech in the modulation spectral domain prior to radiating it from a loudspeaker. Figure 1 shows a block diagram of the signal processing. This processing is based on the RASTA processing used in automatic speech recognition [8].

In Fig. 1, an input signal was divided into 16 frequency bands by constant-Q band-pass filters (BPFs) with 1/3-octave bandwidths, which approximates critical bands of human auditory system. For each band-passed signal, the envelope was extracted using Hilbert analysis. After downsampling (by the factor of M), we applied several modulation filters on the temporal contour of the envelope. To convert the filtered envelope back to the original sampling rate, upsampling was done with the same factor of M. (In the present study, the sampling rate was 16 kHz and the factor M was 160.) After the upsampling, we applied half-wave rectification to remove any negative value artifacts introduced by the modulation filtering. The modified envelope was multiplied by the excitation of the original band-passed signal. By applying BPF to the resultant signal, we removed frequency components outside the range of the band that is produced by the process. Finally, the output signal (i.e., the processed speech signal) is obtained by summing up all the processed signals from each band.

It has been reported in several studies that the important modulation frequency range for speech intelligibility is between 1 and 16 Hz [7, 9, 10] and centered at around 4 Hz,

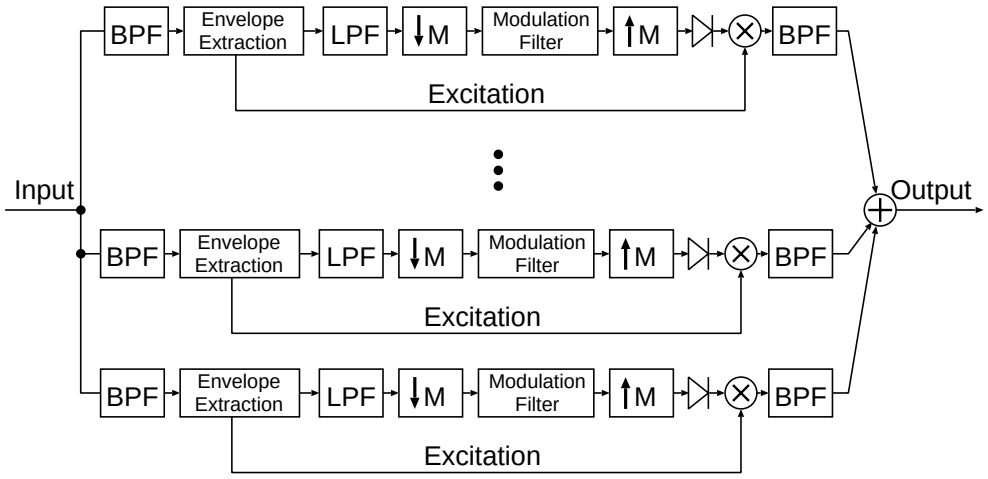


Figure 1: Block diagram for modulation filtering

which corresponds to the syllabic rate of speech [11]. RASTA processing also enhances this frequency region, and successful automatic speech recognition results attest to the importance of this modulation frequency range [8, 12].

Figure 2 shows the frequency responses of the modulation filters that were used in the experiment. Among these modulation filters, filters (a) through (c) enhance the 4 Hz components and filter (d) enhances about 6 Hz in the modulation frequency domain. They differ in the peaks and shapes of their frequency responses. Figure 3 shows the waveform of the original and the processed speech signal using the modulation filter (d).

By comparing the processed speech signal (b) to the original one (a) in Fig. 3, we can see that the envelope corresponding to each syllable is enhanced. Figure 4 shows the modulation spectra before and after processing within a band using modulation filter (d). From this figure, we can see that the components between 2 and 8 Hz are enhanced.

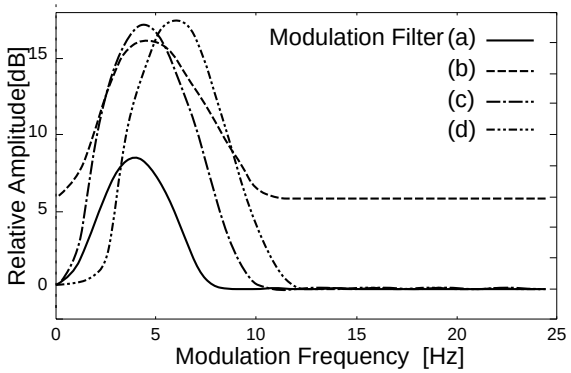


Figure 2 : Frequency response of the modulation filters of (a) through (d).

3. PERCEPTUAL EXPERIMENT

3.1 Stimuli

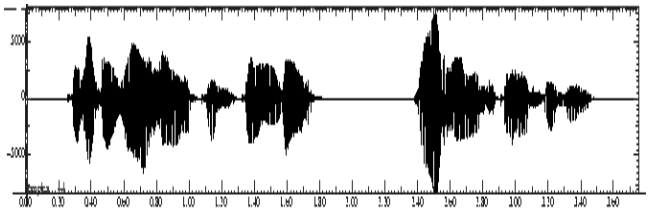
We used the sentence "Terebi gēmu ya pasokon de gēmu o shite asobu (playing with video games and PCs)" uttered by a Japanese male speaker in the ATR speech corpus (Set C) with 16 kHz sampling and 16-bit quantization. We processed this utterance by the signal processing based on the system in Fig. 1 using four different modulation filters. We used the same modulation filter among the 16 bands for each stimuli.

3.2 Subjects

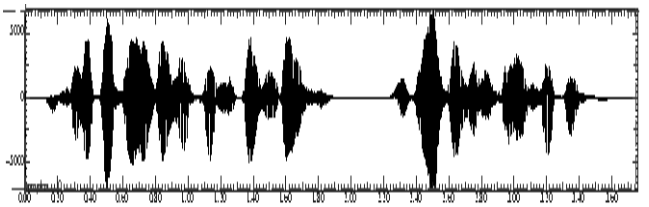
Table 1: Hearing-level of each subject.

Subject	Left	Right
1	75 dB	75 dB
2	83 dB	82 dB
3	105 dB	95 dB
4	105 dB	95 dB

There were four hearing-impaired persons (four females). All of them have sensorineural hearing-loss. Table 1 shows their hearing-level. They wore their hearing-aids during the experiment.



(a)



(b)

Figure 3: Speech waveforms. (a) Original speech signal, and (b) processed signal using the modulation filter (d).

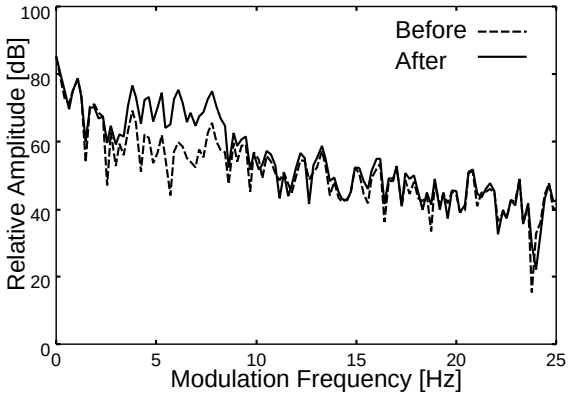


Figure 4: Modulation-spectral change before and after modulation filtering.

3.3 Procedure

We conducted a perceptual experiment at the St. Ignatius Church in Tokyo, which is located next to Sophia University. There were two halls in the church, and we will call them Hall 1 and 2. The floor area is 1,362 m^2 in Hall 1, that takes oval form, and 140 m^2 in Hall 2, that takes rectangular one. In Hall 1, the reverberation time is 3.1–3.2 seconds. It could be a little longer when we used the built-in audio system [13]. In Hall 1 we used the built-in audio system and loudspeakers and played digitally recorded stimuli on a compact disk (the sampling rate: 44.1 kHz). In Hall 2, on the other hand, we played the same digital signals from a loudspeaker through the audio system that we provided.

We used five stimuli, four processed speech signal by using modulation filters (a) to (d) and the original one. The combinations of pairs out of five stimuli were played with a random order, and the subjects were forced to choose one of two stimuli on the basis of which one was easier to hear. We presented a set of stimuli twice at each time. For reference, we asked impression of these stimuli for normal hearing audience simultaneously.

3.4 Experimental results

Table 2 shows the results in Hall 1. The values in this table show the ratios of subjects who heard the processed speech signal easier than the original one. In case of stimuli using modulation filter (a), all of the subjects answered that the processed speech signal was easier to hear. In case of filter (d) three of them, and in case of filter (b) and (c) half of them answered the processed speech signals were easier to hear.

In Hall 2 one of four subjects chose this processed speech signal using modulation filter (a) that had good results in Hall 1. In case of modulation filter (b), two of them answered that the processed speech is easier to understand than original one.

When the persons who have normal hearing heard these processed speech signal, there is no impact to understand the speech both in Hall 1 and Hall 2.

Table 2: Ratio of subjects who heard the processed speech signal easier than the original one in Hall 1.

Modulation Filter	(a)	(b)	(c)	(d)
Ratio	4/4	2/4	2/4	3/4

3.5 Discussion

In Hall 1 all of the subjects answered that the processed speech using modulation filter (a) was easier to hear than the original one. On the other hand, Subjects 3 and 4, who have relatively severe hearing-loss, preferred the prosecced speech in case of modulation filter (b). By comparing modulation filter (a) with (b), modulation filter (b) contains higher modulation frequency components. This appears to indicate that a slightly wider modulation frequency emphasis is desirable for the severely hearing-impaired.

In Hall 2, the loudspeaker was located very close to the subjects when we conducted the experiment making the direct sound dominant. Only Subject 1 who has slight hearing-loss, preferred the processed speech using modulation filter (a) that were preferred by all subjects in Hall 1. The results were different between Hall 1 and Hall 2 using the same

modulation filter (a).

Since the "best" modulation filter can differ between halls and subjects, we may be required to tailor them to each room or subject. Future work will investigate the following modulation filter modifications:

- Consider the dynamic range of typical hearing-loss at every frequency
- Use different modulation filters among frequency bands instead of using the same modulation filter as in the present study.

4. CONCLUSIONS

When speech is radiated in halls that have long reverberation time, the audience often has difficulties understanding speech. In this paper we examined a method of mitigate this problem. Our method enhances specific frequencies of the modulation spectrum of the speech signal. We conducted a perceptual experiment and confirmed that this technique is useful for the hearing-impaired. In this study, the preference was asked to the subjects. However, these pilot experiments should be supplemented with more intelligibility tests at word and sentence levels.

Our results indicate that the proposed method is helpful to the hearing-impaired and those of advanced years; the method is expected to contribute to the coming aging society in a noteworthy way.

ACKNOWLEDGMENTS

We gratefully acknowledge the cooperation of the people who participated in the perceptual experiment. We also acknowledge the person concerned who made us to use St. Ignatius Church.

We would like to express our grateful thanks to Hynek Hermansky of Oregon Graduate Institute of Science and Technology, and also to Michael L. Shire of International Computer Science Institute.

5. REFERENCES

[1] Y. Ando and M. Imamura, "Subjective Preference Tests for Sound Fields in Concert Halls Simulated by the Aid of a Computer," *J. Sound and Vib.*, vol. 65, pp. 229–239, 1979.

[2] Y. Ando, M. Okura and K. Yuasa, "On the Preferred Reverberation Time in Auditorium," *Acustica*, vol. 50, pp. 134–141, 1982.

[3] M. R. Schroeder, "Modulation Transfer Functions: Definition and Measurement," *IEEE Trans. ASSP*, vol. 26, pp. 179–182, 1978.

[4] T. Houtgast, H. J. M. Steeneken and R. Plomp, "Predicting Speech Intelligibility in Rooms from the Modulation Transfer Function I. General Room Acoustics," *Acoustica*, vol. 46, pp. 60–71, 1980.

[5] T. Houtgast and H. J. M. Steeneken, "A Review of the MTF Concept in Room Acoustics and its Use for Estimating Speech Intelligibility in Auditoria," *J. Acoust. Soc. Am.*, vol. 77, pp. 1069–1077, 1985.

[6] T. Nakajima, "Measurements of the Modulation Transfer Function (MTF) and Speech Transmission Index (STI) in Room Acoustics," *J. Acoust. Soc. Jpn.*, vol. 49, pp. 103–110, 1993 (in Japanese).

[7] T. Arai, M. Pavel, H. Hermansky and C. Avendano, "Syllable Intelligibility for Temporally Filtered LPC Cepstral Trajectories," *J. Acoust. Soc. Am.*, vol. 105, pp. 2783–2791, 1999.

[8] H. Hermansky and N. Morgan, "RASTA Processing of Speech," *IEEE Trans. Speech and Audio*, vol. 2, pp. 578–589, 1994.

[9] R. Drullman, J. M. Festen and R. Plomp, "Effect of Temporal Envelope Smearing on Speech Reception," *J. Acoust. Soc. Am.*, vol. 95, pp. 1053–1064, 1994.

[10] R. Drullman, J. M. Festen and R. Plomp, "Effect of Reducing Slow Temporal Modulations on Speech Reception," *J. Acoust. Soc. Am.*, vol. 95, pp. 2670–2680, 1994.

[11] S. Greenberg, "On the Origins of Speech Intelligibility in the Real World," *Proceedings of the ESCA Workshop on Robust Speech Recognition for Unknown Communication Channels*, pp. 23-32, 1997.

[12] N. Kanedera, T. Arai, H. Hermansky and H. M. Pavel, "On the Importance of Various Modulation Frequencies for Speech Recognition," *Proc. EUROSPEECH*, vol. 3, pp. 1079–1082, 1997.

[13] H. Tachibana and K. Iida, "Acoustic Design of the St. Ignatius Church," *Architectural Design*, Oct., 1999 (in Japanese).