SIMULATION OF AN ACOUSTIC SYSTEM USING POWER ENVELOPE INVERSE FILTERING

Shigekatsu IRIE, Shigeki HIROBAYASHI

UNIVERSITY OF TOYAMA
Department of Intellectual Information Systems Engineering Faculty of Technology
Gofuku 3190, Toyama, Toyama, Japan

ABSTRACT

In the present paper, we examine the development of a high-quality acoustics system called the Sound Field Generation System (SFGS) using a Power Envelope Inverse Filtering (PEIF) proposed by the authors. PEIF is used for pre-processing of the SFGS, and we quantitatively evaluate SFGS using a real signal. When the output signal of SFGS is reproduced in the reproduction field, the effect of the SFGS is diminished due to the impulse response of reproduction field that is added to the created SFGS signal. Therefore, we attempted to reduce the influence of the reproduction field using a PEIF. In the experiment, we used five signals, namely, male speech, female speech, male vocal with music, female vocal with music, and classical music, and these signals were processed by computer simulation. We evaluated these signals by objective evaluation and subjective evaluation.

In the objective evaluation, we obtained an improvement in SFGS for the results ranging from approximately 1 dB to 2 dB under 4 kHz, and up to 5 dB above 4 kHz. In the subjective evaluation, we obtained an improvement of approximately 10% to 60% in an experiment involving 52 subjects. These improvements are significant.

Index Terms— Acoustic fields, Audio systems, Architectural acoustics, Acoustic filters, Signal processing

1. INTRODUCTION

The quality of sound is influenced by echoes generated by the space in which sound is heard. However, repairing or modifying existing concert halls is expensive and it is therefore desirable to control the sound by electric means.

In recent years, higher Active Field Control (AFC) has become possible due to progress in acoustic techniques intended to improve the sound field and expand the functionality of a concert hall [1]. The representative control technique is a Sound Field Generation System (SFGS). The SFGS is widely used due to low cost and ease of use. SFGS is a system that creates reverberant sound by adding impulse responses of optional spaces using a convolution. The SFGS is able to function in real time with practically no signal delay. However, when we use the SFGS in an existing hall, the effect of the SFGS decreases due to the influence of the reproduction field added to the created signal. Specifically, the optional impulse response of the SFGS and the impulse response of the reproduction field are convoluted twice. As a result, the intelligibility deteriorates, echo increases and the sound becomes both difficult to hear and harsh. It is therefore necessary to develop a system that removes the influence of the impulse response of the reproduction field in advance.

In the present paper, we apply a Power Envelope Inverse Filter (PEIF) [2][3], which can reduce reverberation and use pre-processing of dereverberation. The authors then examine the newly designed system, which is capable of improving the sound of existing concert halls without the requirement for major renovations. Quantitative evaluation was performed using a real signal, and the inverse system was constructed to transfer the functions of the reproduction field using a PEIF.

2. DEREVERBERATION

Reverberation is a basic phenomenon in room acoustics. Distortion of the source signal occurs and intelligibility deteriorates due to the source signal being convoluted by the reflection of sound off of the walls of the room. However, good reverberation is required in venues such as concert halls because the sense of spread in spatial sound created by reverberation is added to the source signal. Therefore, the influence of reverberation is a necessary component of the architecture of any concert hall. As shown in the following equation, the audition signal $y(t)$ is described by convoluting the source signal $x(t)$ and room impulse response $h(t)$ in a sound field:

$$\begin{aligned} y(t) &= \int_{-\infty}^{\infty} x(\tau)h(t-\tau)d\tau \\ &= x(t) * h(t) \end{aligned} \tag{1}$$

where $*$ is a convolution operator. We must estimate the transfer system in order to estimate the source signal from the audition signal in a linear system. In addition, we must estimate an adaptive system because $h(t)$ is a time change system. Generally, waveform recovery of a source signal is difficult by inverse filter processing because the transfer system has a non-minimum phase.

2.1. POWER ENEVELOPE INVERSE FILTER

In a previous study [2], the authors showed that the power of the audition signal $y(t)$ can be expressed in a convolution of the power envelope of a source-signal $e_x(t)$ and the impulse response $e_h(t)$ based on the theory of the Modulation Transfer Function (MTF) suggested by T. Houtgast [4].

Generally, an impulse response is not a minimum phase, but $e_h(t)^2$ is near a minimum phase because it dampens in a manner similar to an exponential function. Therefore, as shown in the following equation, the power envelope property of the estimation signal $\hat{P}_x(\omega)$ is given by multiplying the power envelope property of the reverberation signal $P_y(\omega)$ and the minimum phase inverse property $P_{h\min}(\omega)^{-1}$ of the power envelope transfer function $P_h(\omega)$:

$$\hat{P}_x(\omega) \approx P_y(\omega) P_{h\min}(\omega)^{-1} \tag{2}$$

where $\hat{P}_x(\omega)$ and $P_y(\omega)$ are the frequency properties of $\hat{e}_x(t)^2$ and $e_y(t)^2$, respectively. The recovery signal $\hat{x}(t)$ is given by the following equation in terms an amplitude envelope of $\hat{e}_x(t)^2$ and $e_y(t)^2$:

$$\hat{x}(t) = \hat{e}_x(t) e_y^{-1}(t) y(t) \tag{3}$$

where the right side $\hat{e}_x(t)$ is the recovery envelope calculated by inverse Fourier transform in equation (2). In addition, $e_y^{-1}(t)y(t)$ is a fine structure signal of a reverberant signal and is a flattened-amplitude envelope of an output signal. Equation (3) is referred to as the Power Envelope Inverse Filter (PEIF) and is obtained by adding the amplitude envelope estimated in equation (2) to the output signal, which is a flattened amplitude envelope. In addition, the minimum phase property of the transfer function of equation (2) can be calculated by cepstrum processing [5].

2.2. REBERBERATION CORRECT SYSTEM

SFGS can generate the signal, $X(\omega)$, of a purpose field by measuring the transfer function, $H_{\mathrm{imag}}(\omega)$, of the reproduction field (purpose field), and multiplying its source signal, $S(\omega)$, shown in the following equation:

$$X(\omega) = S(\omega) H_{\mathrm{imag}}(\omega) \tag{4}$$

However, when $X(\omega)$ is reproduced in the reproduction field, SFGS is not able to correctly reproduce the sound of the purpose field due to the influence of the transfer function, $H_{\mathrm{real}}(\omega)$, of the reproduction field is added. In other words, the listener of the reproduction field hears signal $Y(\omega)$, which is given by the following equation:

$$\begin{aligned} Y(\omega) &= X(\omega) H_{\mathrm{real}}(\omega) \\ &= S(\omega) H_{\mathrm{imag}}(\omega) H_{\mathrm{real}}(\omega) \end{aligned} \tag{5}$$

First, we have to reduce the influence of the impulse response of the reproduction field from the purpose signal, $X(\omega)$, to match $Y(\omega)$ to $X(\omega)$.

A summary of the system is shown in Figure 1. We can obtain a correct signal $\hat{X}(\omega)$ as shown in equation (6) by building a system that is able to reduce the effect of the impulse response $H_{\mathrm{real}}(\omega)$ of the reproduction field beforehand, for the $X(\omega)$, which added $H_{\mathrm{imag}}(\omega)$ to $S(\omega)$ by using a SFGS.

$$\hat{X}(\omega) \quad = \quad X(\omega)/H_{\mathrm{real}}(\omega) \tag{6}$$

As with equation (6), the listener can hear a virtual sound of the purpose field by outputting a signal that reduces the influence of the reproduction field on the purpose signal beforehand. Since the audience can hear the sound of the purpose field due to $H_{\mathrm{real}}(\omega)$ being added to $\hat{X}(\omega)$. In this paper, we realize equation (6) by using a PEIF, and quantitatively evaluate the effect of the system.

The proposition system separates the purpose signal $X(\omega)$ into the envelope and the carrier and carries out inverse filter processing using a power envelope of the transfer function $H_{\mathrm{real}}(\omega)$ of the purpose field for the power envelope of the purpose signal $X(\omega)$. In addition, the proposition system composes an output-signal and a carrier of the purpose signal. The method of extracting an envelope involves the calculation of an absolute signal after the Hilbert transform and uses the low-path filter (cut-off frequency : f_c).

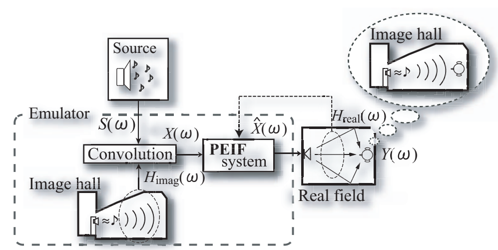

Fig. 1. Power Envelope Inverse Filter System

3. EXPERIMENT

In the experiment, we assumed the purpose field to be Aubade Hall (Fig. 2), located in Toyama City, Toyama, Japan, which seats 2,200 people and has a reverberation time of 1.3 (s). We assumed the reproduction field to be Shimin Plaza Hall (Fig. 3), which seats 308 people and has a reverberation time of 1.1(s), and is also located in Toyama City. Aubade Hall is used mainly for concerts, and Shimin Plaza Hall is used for both lectures and concerts. In this experiment, we measured the impulse response of both halls and carried out processing by computer simulation. We set up an audition point in the center of each hall and accurately measured the impulse responses of each hall using a swept-sine signal [6] at a sampling frequency of 32 kHz. The waveform of each impulse response is shown in Figure 4. We calculated the cross-correlation between Aubade Hall and Shimin Plaza Hall by dividing the frequency range by 64. The average cross-correlation of each frequency range was calculated to be approximately 0.05. In addition, we only used an envelope that was divided in the frequency range of the impulse response of the reproduction field. We did not use a carrier of the impulse response in improvement processing of the experiment. The source of the speech signal was the "database of speech of 20 languages" from NTT Advance Technology, and we used the first track of Japanese speech as the male speech and the fourth track of Japanese speech as the female speech. The source of the music signal was J-POP music. We used "*Winterfall*" by L'Arc-en-Ciel as a male vocal with music, "*Deattakoronoyouni*" by Every Little Thing as a female vocal with music, and "Polonaise No. 6," by Chopin as classical music. In the objective experiment, we used a speech signal that is approximately 10 seconds in length without processing, and a music signal that is approximately 10 seconds in length. In the subjective experiment, we used signals that were processed for 3 and 10 seconds that were used in the objective experiment. The sampling frequency of each signal was 32 kHz.

In the objective experiment, we evaluated the processed signal for a frequency range of 250 Hz (division number N = 64) and a cut-off frequency of 10 Hz using a PEIF. In addition, for the evaluation of the processed signal, we used the evaluation index of equation (7) because, generally, the intelligibility of the reverberation signal is closely related to changes of the envelope [7].

$$I_p = 10\log_{10}\frac{\int_0^T \left\{e_x(t) - e_y(t)\right\}^2 dt}{\int_0^T \left\{e_x(t) - \hat{e}_x(t)\right\}^2 dt}(\mathrm{dB}) \tag{7}$$

where T is the analysis time, $e_x(t)$ represents the envelope of the purpose signal, $e_y(t)$ represents the envelope signal that is the convoluted impulse response of Shimin Plaza Hall and Aubade Hall by SFGS, and $\hat{e}_x(t)$ represents the envelope signal of the recovery signal obtained using the proposed system. In the present paper, the value calculated in equation (7) is a recovery value. This index is not generally a standard index, but, until now, this index has been used as an effective way to easily verify the recovery effect of reverberation or a similar signal [8].

In the subjective experiment, we evaluated the signal in the audition experiment with the collaboration of 52 students from Engineering Department of the University of Toyama. The subjects heard the purpose signal, which was processed by computer, and heard the unprocessed signal and processed signal. Finally, subjects reported which signal was nearer to the purpose signal. Stimulation was one set for three signals. The substance of the stimulation set evaluated by a participant was two sets. The first set was, in turn, the purpose signal, the unprocessed signal, and the processed signal, and the second set was, in turn, the purpose signal, the processed signal, and the unprocessed signal. The subjects heard the stimulation set using headphones. In addition, the number of total answers was 102 because there is a stimulus set of two evaluation objects.

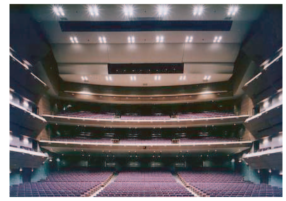

Fig. 2. Aubedo hall

Fig. 3. Shimin Plaza hall

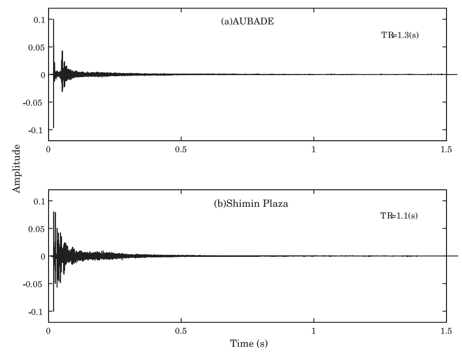

Fig. 4. Impulse response of Aubado hall and Shimin Plaza hall

3.1. EXPERIMENTAL RESULTS

In the objective experiment, we calculated the average of the recovery value of each frequency range by dividing the frequency range between below and above 4 kHz. In the results for the male speech signal, PEIF gives an average improvement over SFGS of approximately 2.1 dB for frequencies below 4 kHz, and an improvement of up to 5.6 dB for frequencies above 4 kHz. In the results for the female speech signal, PEIF gives an average improvement over SFGS of approximately 1.4 dB for frequencies below 4 kHz, and an average improvement of up to 4.7 dB for frequencies above 4 kHz. In the results for the male vocal with music signal, PEIF gives an average improvement over SFGS of approximately 0.8 dB for frequencies below 4 kHz, and an average improvement of up to 5.8 dB for frequencies above 4 kHz. In the results for the female vocal with music signal, PEIF gives an average improvement over SFGS of approximately 1.4 dB for frequencies below 4 kHz, and an average improvement of up to 4.6 dB for frequencies above 4 kHz. In the results for the classical music signal, PEIF gives an average improvement over SFGS of approximately 1.3 dB for frequencies below 4 kHz, and an average improvement of up to 5.7 dB for frequencies above 4 kHz. Figure 5 shows the recovery value of the processed signal of each frequency range by equation (7).

In the subjective experiment, approximately 68% of subjects answered that the processed signal of male speech was closest to the purpose signal. Approximately 52% of subjects answered that the processed signal of the female speech was closest to the purpose signal. Approximately 83% of subjects answered that processed signal of male vocal with music was closest to the purpose signal. Approximately 45% of subjects answered that processed signal of the female vocal with music was closest to purpose signal. Approximately 59% of subjects answered that processed signal of classical music was closest to the purpose signal. We examined the significant difference between the processed signal and unprocessed signal by examining the ratio. In the results, the male speech was judged to be significantly different at a 1% level. The female speech had no significant difference. The male vocal with music was judged to be significantly different at a 1% level. The female vocal with music had no significant difference. The classical music was judged to be significantly different at a 10% level. Table 1 shows the results of the subjective experiment.

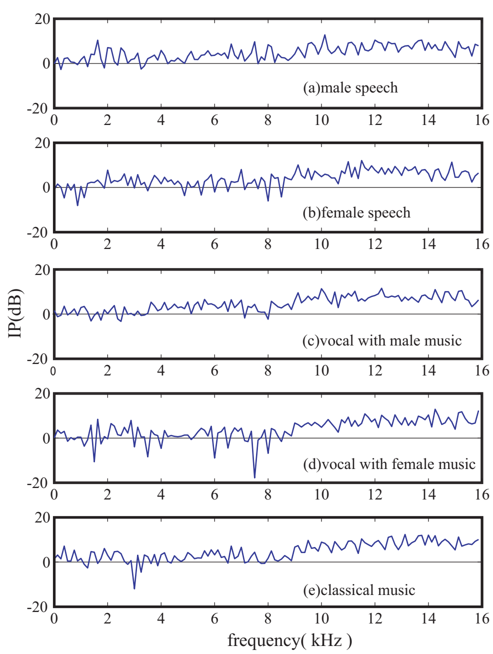

Fig. 5. Recovery value for each frequency range

Table 1. Result of subject experiment

	Processed	Unprocessed	Significant difference
Male speech	68%	32%	YES
Female speech	52%	48%	NO
Male vocal with music	83%	17%	YES
Female vocal with music	45%	55%	NO
Classical music	59%	41%	YES

4. CONCLUSIONS

In the present paper, for an improved SFGS, we proposed a system that applied a Power Envelope Inverse Filter, and carried out quantitative evaluation of the proposed system. Using the previous SFGS it is difficult to reproduce the purpose signal correctly because the impulse response of the purpose field and the impulse response of the reproduction field are doubly convoluted. Therefore, we attempted to reduce the influence of reverberation before reproduction by using a PEIF that has the possibility of being able to unify the controlling space. We evaluated the system by computer simulation using signals of male speech, female speech, male vocal with music, female vocal with music, and classical music.

The result of the objective experiment confirmed the recovery value of the processed signal of approximately 1 dB to 2 dB at 4 kHz and below, and the recovery value of the processed signal of approximately 5 dB to 6 dB at 4 kHz and above. The ratio improved from approximately 18% to 66% for male speech, male vocal with music, and classical music, all of which are significantly different. However female speech and female vocal with music do not have a significant difference.

Henceforth, we intend to examine control in a concert hall in which the reverberation time or control range is very different in a real field.

5. REFERENCES

[1] K. Takahashi, and T. Yamakawa, "Architectural Acoustics for Hall Design-Pursuit of Rich Sound," SICE, vol. 43, no. 4, pp. 348-352, 2004.

[2] S. Hirobayashi, and H. Nomura, "Speech Waveform Recovery from a Reverberant Speech Signal Using Inverse Filtering of the Power Envelope Transfer Function," IEICE, vol. J81-A, no. 10, pp. 1323-1330, 1998.

[3] S. Hirobayashi, and T. Yamabuchi, "Validation of Blind Dereverberation Using Power Envelope Inverse Filtering and Filter Banks," IEICE, vol. J83-A, no. 8, pp. 1029-1033, 2000.

[4] T. Houtgast, H. J. M. Steeneken, and R. Plomp. " Predicting speech intelligibility in room acoustics," Acustica, vol. 46, pp. 60-72, 1980.

[5] S. Hirobayashi, H. Kimura, M. Tohyama, and T. Oyabu. "Gas-Generating Events Estimation Using Remote Sensing and Cepstrum Processing in an Indoor space," IEICE, vol. J81-1, no. 11, pp. 1600-1610, 1998.

[6] Y. Suzuki, F. Asano and T. Sone. "On the simulation of a transfer function of an acoustic system (Part 2)," vol. 45, no. 1, pp. 44-50, 1989.

[7] R. Frullman, J.M. Festen, and R. Plomp. " Effect of temporal envelope smearing on speech reception," J. Acoust, Soc AM, vol. 95, pp. 1053-1064, 1989.

[8] S. Hirobayashi, H. Terashima, and T. Yamabuchi. "Evaluation of Envelope Estimation Method of Acoustic Signal in Reverberant Field," JSST, vol. 22, no. 3, pp. 208-215, 2003.