SPATIAL COMPRESSION OF MULTI-CHANNEL AUDIO SIGNALS USING INVERSE FILTERS

Toshiyuki KIMURA[1], *Kazuhiko KAKEHI*[1], *Kazuya TAKEDA*[2] *and Fumitada ITAKURA*[2]

[1]Graduate School of Human Informatics, Nagoya University / CIAIR
[2]Graduate School of Engineering, Nagoya University / CIAIR
Furo-cho, Chikusa-ku, Nagoya, 464-8601, Japan
E-mail: kimura@cog.human.nagoya-u.ac.jp

ABSTRACT

A large number of transmission audio channels are necessary for reproduction of a sound field based on Huygens principle. A method is proposed for spatial compression of the multi-channel audio signals. The signals are compressed by convolving them with the inverse filter of the room impulse response to reduce the number of transmission channels to the number of source signals. Then the source signals are transmitted to reproduce the sound field by convolving them with the impulse response. The compression method is evaluated using the signal-to-noise ratio and a subjective assessment. Experimental results show that SNR between a source signal and an extracted signal is more than 40dB and that there is no significant difference of the directional perception due to compression.

1. INTRODUCTION

The sound field reproduction system is important in development of auditory virtual reality techniques. The Huygens principle is used to construct the system, which requires a large number of channels [1] to exhibit a localized, spatial and surrounded perception.

The data size for transmission of the system is proportional to the number of channels. Therefore, it is necessary to compress the channel signals. The audio compression methods such as AC-3 [2] or MPEG2 AAC [3] mainly remove inner-channel redundancy. The ability of the methods to remove inter-channel redundancy [4][5] is not high.

In many applications of the sound field reproduction systems such as piano sonata at a concert hall, the number of sound sources is less than the number of channels. Therefore, transmission of the extracted sound source signals is more efficient than transmission of the channel signals.

For a particular sound source signal, the channel signals are composed of the direct signal and the reflective signal. The direct signal contains the information about the distance between the sound source and the microphone. The reflective signal contains the information about the acoustics characteristics of a room. The distance information and the characteristics information are contained in the room impulse response from a sound source to a receiving point in a room. The sound source signals are extracted by convolving the channel signals with the inverse filter of the room impulse response when the room impulse responses are given. The convolution method is used to reduce the number of channels to be transmitted.

In this paper the proposed compression method is explained first. Then the procedure used for synthesis of multi-channel audio signals is presented. Finally experimental procedures and objective and subjective evaluation of the compression method are shown.

2. SPATIAL COMPRESSION METHOD

2.1. Compression

We consider a sound field reproduction system consisting of N loudspeakers and M microphones. We assume that N is less than M. Let $S_i(\omega)$ be the source signal from the i^{th} loudspeaker (i=1…N) and $X_j(\omega)$ be the channel signal for the j^{th} microphone (j=1…M). The room impulse response from the i^{th} source to the j^{th} channel is denoted by $G_{ij}(\omega)$. Let $H_{ji}(\omega)$ be the inverse filter from the j^{th} channel to the i^{th} source. The extracted signal of the i^{th} source is given by $S'_i(\omega)$. The block diagram of the compression system using the proposed method is given in Figure 1.

The room impulse response matrix $\mathbf{G}$ is an $N\times M$ complex matrix with $G_{ij}(\omega)$ as its element. The inverse filter matrix $\mathbf{H}$ is an $M\times N$ complex matrix with $H_{ji}(\omega)$ as its element. The relationship between $\mathbf{G}$ and $\mathbf{H}$ is as follows:

$$\mathbf{GH} = \mathbf{D} \qquad (1)$$

Here, $\mathbf{D}$ is an $N\times N$ diagonal matrix with complex elements. The i^{th} diagonal component of $\mathbf{D}$ is given by:

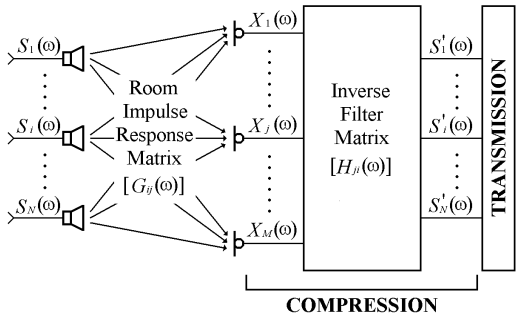

Figure 1: Compression systems

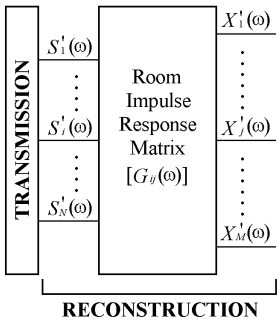

Figure 2: Reconstruction systems

$$D_i(\omega) = S'_i(\omega)/S_i(\omega) \qquad (2)$$

When $N<M$, there are several solutions for the Equation (1). We consider a solution using the Moore-Penrose pseudo inverse matrix [6]. From the singular value decomposition of **G**, we have the following:

$$\mathbf{G} = \mathbf{U}\mathbf{S}\mathbf{V}^H \qquad (3)$$

Here, **U** is an $N\times N$ unitary matrix. The columns of **U** are the eigenvectors of $\mathbf{G}\mathbf{G}^H$, where $\mathbf{G}^H$ is a conjugate transpose of **G**. The columns of the $M\times M$ unitary matrix **V** are the eigenvectors of $\mathbf{G}^H\mathbf{G}$. If the $N\times M$ diagonal matrix **S** has a rank d, then the d diagonal entries of **S** are non-zero. The matrix $\mathbf{S}^+$ is defined by transposing the matrix **S** and then replacing its non-zero entries with their reciprocals. The inverse filter matrix **H** is now obtained as:

$$\mathbf{H} = \mathbf{V}\mathbf{S}^+\mathbf{U}^H\mathbf{D} \qquad (4)$$

The M channel signals are compressed by convolving them with the inverse filter of the room impulse response to obtain the N extracted source signals.

2.2. Reconstruction

Reconstruction of the channel signals is achieved by convolving the extracted source signals with the room impulse response, as shown in Figure 2.

3. SYNTHESIS OF MULTI-CHANNEL AUDIO SIGNALS

Multi-channel audio signals are synthesized by convolving "dry source signals," or the direct signal from sound source, with the room impulse response.

3.1. Measurement of Room Impulse Response

The variable reverberation room at Nagoya University is used in the experiments. The reverberation time of the room can be changed from 151ms to 303ms by adjusting the absorption rate of the wall. The shape of a room, the positions of a loudspeaker and the microphones in the arrangement to be used in an experiment are shown in Figure 3.

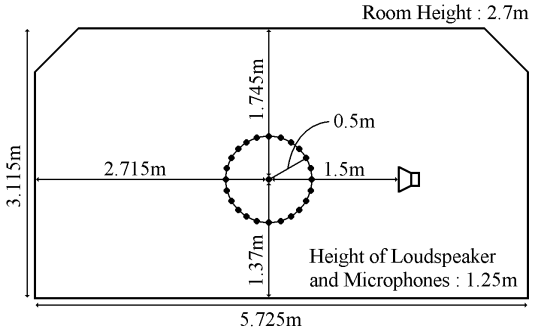

Figure 3: Position of a loudspeaker and microphones

24 microphones are placed on a circle of 0.5-meter radius, with an interval of 15 degrees. A microphone is also placed at the center of the circle. The measurement conditions are given in Table 1. The room impulse response is measured for the reverberation time condition of 151ms and 303ms, respectively.

Table 1: Measurement conditions

Room temperature	21ºC
Noise level	19.6dB(A)
Sampling frequency	48kHz
Reference signal	TSP (1.37sec)
The number of repetitions	8
FIR filter order	16384 (0.34sec)

3.2. Synthesis of Multi-channel Audio Signals

Two types of dry sources considered in our experiments are: a white noise and a female speech. The duration of the source signals is 3 seconds. The bandwidth of the source signals is from 50Hz to 5kHz. The signals are sampled at the rate of 12kHz in case of reproduction. From Table 1, it may be noted that the sampling frequency used in measuring the room impulse response is 48kHz. Therefore, the reverberation time will be multiplied by a factor of 4 and the reverberation time of the source signals is 0.6sec and 1.2sec.

4. EXPERIMENTAL STUDIES

4.1. Experiment of Compression

In our experiments, the channel signals from 25 microphones shown in Figure 1 are compressed to obtain a single extracted source signal. The room impulse response measured from each channel signal is transformed with the frequency domain using a 16384-point FFT. $D_1(\omega)$ is obtained by computing the 16384-point FFT on the FIR band-pass filter with a delay of 8192 points and a bandwidth from 50Hz to 5kHz. The inverse filter matrix **H** is then computed using Equation (3) and (4). The inverse filter in the time domain is obtained by computing the 16384-point IFFT on the matrix **H**.

4.2. Evaluation of Extraction

The method for extraction of the sound source is evaluated by computing the signal-to-noise ratio (*SNR*) between the source signal and the extracted source signal. *SNR* between a source signal $s_1(n)$ and an extracted signal $s'_1(n)$ is as follows:

$$SNR = 10\log_{10}\frac{\sum_n \{s_1(n)\}^2}{\sum_n \{s_1(n) - s'_1(n+8192)\}^2} \qquad (5)$$

The *SNR* obtained for the two types of dry source signals (white noise and female speech) and for two reverberation times (0.6 seconds and 1.2 seconds), is shown in Figure 4.

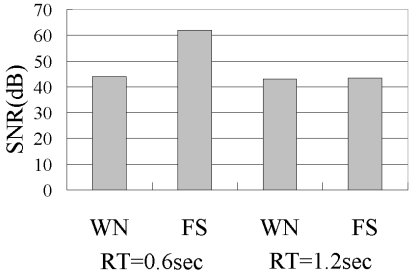

Figure 4: *SNR* of extraction

It may be noted that the *SNR* is more than 40dB for all the four cases. The high *SNR* indicates that compression is 'strict'.

4.3. Subjective Assessment

The compression method is evaluated by a subjective assessment because an objective evaluation method is not yet established. We study the effect of compression on the directional perception.

The experiment for subjective assessment is carried out in a low-reverberation room. 24 loudspeakers are placed on a circle with 2-meters radius. The height of loudspeakers is about 1.2m from the floor. The ear of the subject is at the same height. The loudspeakers are placed at an angle interval of 15 degrees, as shown in Figure 5. The sound signals are played from the 12 loudspeakers shown in gray color in Figure 5.

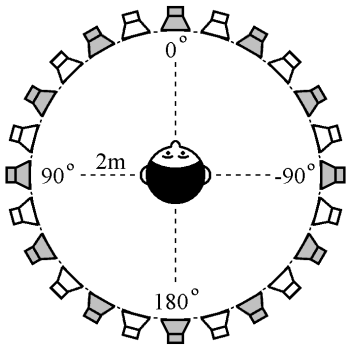

Figure 5: Position of loudspeakers

The experimental conditions are given in Table 2.

Table 2 Experimented conditions

Subjects	13 (8 males and 5 females)
Sound pressure level	About 70dB(A) at center of circle
Sessions	4 = 2(dry sources) ×2(reverberation times)
Trials	96 = 12(presented directions) ×2(compression or not) ×4(repetitions)

The order of sessions is random for each subject. The order of trials is also random in each session.

The experimental procedure is as follows: The direction of the head of the subject is set at 0 degrees at start of a trial. When the sound signal is played, the subject is asked to locate the direction of the sound image by reporting the index indicated on a loudspeaker. The subject is allowed to rotate the head during locating the direction of the sound image.

The number of the front-back errors is very small. So, the front-back errors are corrected in the calculation of the directional perception error. A histogram of the perception errors, the mean and the standard deviation are computed for each presented direction of the sound image in each session. The results for the four sessions are shown in Figure 6 to 9, where a positive perception error indicates that the perception error is counterclockwise. In the figures, the gray circles indicate the means for the non-compression condition. The white circles indicate the means for the compression condition. The error bars indicate the standard deviation values.

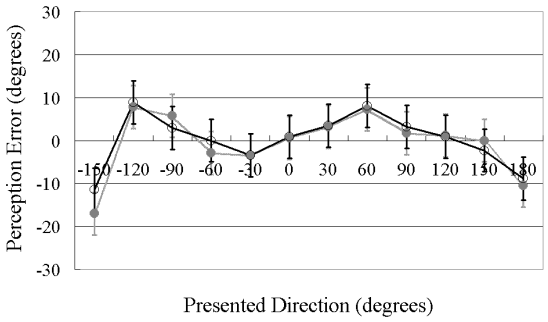

Figure 6: White noise, Reverberation time 0.6sec

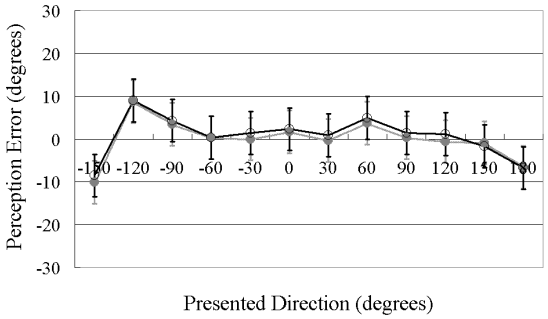

Figure 7: Female speech, Reverberation time 0.6sec

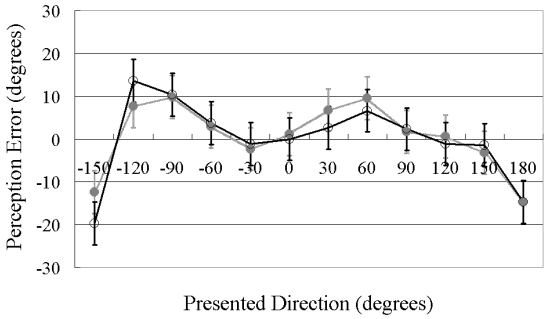

Figure 8: White noise, Reverberation time 1.2sec

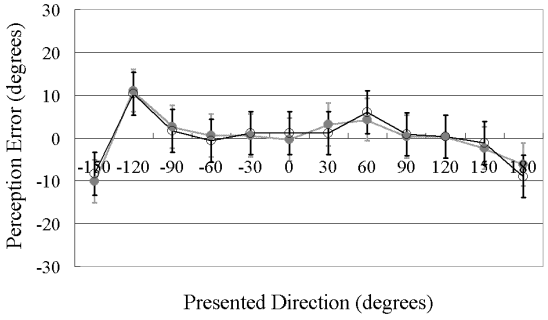

Figure 9: Female speech, Reverberation time 1.2sec

The means for the compression condition are almost the same as the means for the non-compression condition. An ANOVA of 3 factors is applied for each reverberation time. The three factors are the compression condition, the presented angle and the type of dry source. The result indicates that there was no significant difference of the main effect and the interaction about the compression condition. Therefore, the directional perception is not affected by the compression method.

5. CONCLUSION

In this paper, a new method for compression of multi-channel audio signals using the spatial information is proposed. The compression method is based on measuring the room impulse response and then convolving the channel signals with the inverse filter of the room impulse response. The signal-to-noise ratio of the extracted source signals indicates that the waveform is not distorted by the compression. The subjective assessment of the compression method indicates that the directional perception is preserved by the compression.

The future work involves investigation of the other type of spatial impression such as the distance, reverberation, spaciousness, and surroundedness. It is also necessary to study the case of multiple sources and moving sources.

6. REFERENCES

[1] M. Camras, “Approach to recreating a sound field,” J. Acoust. Soc. Am. **43**, 1425-1431, 1968.

[2] C. C. Todd, G. A. Davidson, M. F. Davis, L. D. Fielder, B. D. Link and S. Vernon, “AC-3: Flexible perceptual coding for audio transmission and storage,” AES preprint 3796, presented at the 96th Convention, February 1994, Amsterdam.

[3] ISO/IEC 13818-7, “Information technology – Generic coding of moving pictures and associated audio information – Part 7 Advanced Audio Coding”.

[4] D. Yang, H. Ai, C. Kyriakakis and C.-C. J.Kuo, “An inter-channel redundancy removal approach for high-quality multichannel audio compression,” AES preprint 5238, presented at the 109th Convention, September 2000, Los Angeles.

[5] Y. Wang, M. Vilermo, M. Väänänen and L. Yaroslavsky, “A multichannel audio coding algorithm for inter-channel redundancy removal,” AES preprint 5295, presented at the 110th Convention, May 2001, Amsterdam.

[6] J. Bauck and D. H. Cooper, “Generalized transaural stereo and applications,” J. Audio Eng. Soc., **44**, no.9, pp.683-705, 1996.