Perceptual Enhancement of Sound Field Reproduction in a Nearly Monaural Sensing System

Chan Jun Chun[1], Hong Kook Kim[1], Seung Ho Choi[2], Sei-Jin Jang[3], and Seok-Pil Lee[4]

[1] School of Information and Communications
Gwangju Institute of Science and Technology, Gwangju 500-712, Korea
{cjchun,hongkook}@gist.ac.kr
[2] Department of Electronic and Information Engineering
Seoul National University of Science and Technology, Seoul 139-743, Korea
shchoi@snut.ac.kr
[3] NSSC Center, Korea Electronics Technology Institute, Goyang 410-380, Korea
sjjang@keti.re.kr
[4] Digital Media Research Center, Korea Electronics Technology Institute,
Seoul 137-070, Korea
lspbio@keti.re.kr

Abstract. In this paper, we propose a method for enhancing the sound field in a nearly monaural sensing system. The proposed method controls the inter-channel coherence (ICC) of a stereo sound source by incorporating early reflections into the sound source. The degree of ICC is controlled using both a decorrelator and a principal component analysis (PCA) technique, while a cascaded all-pass filter is used to induce early reflections. To demonstrate the effectiveness of the proposed method on the perception of the sound field, a subjective test is carried out. It is shown from the test that the proposed method can reduce the ICC, thus it improves the perceived sound field.

Keywords: Sound field, nearly monaural sensing system, inter-channel coherence (ICC), principal component analysis (PCA), decorrelator, early reflections.

1 Introduction

Spatial quality, which is a subset of sound quality, is an important factor in the audio field [1]. It is attributed to the assessment of an auditory image and indicates whether a listener is satisfied or not with the auditory image. There are several spatial attributes associated with spatial quality. Among them, sound field is mostly related to how a listener perceives the width of an auditory image from the audio source, which is shown in Fig. 1. The sound filed is also called the apparent source width (ASW) [2].

Microphone configuration plays an important role in capturing the spatial quality. There are many techniques [3] for recording the spatial quality. One of the techniques suggests that the distance between auditory sensors should be determined by the

T.-h. Kim et al. (Eds.): FGCN 2011, Part II, CCIS 266, pp. 124–131, 2011.

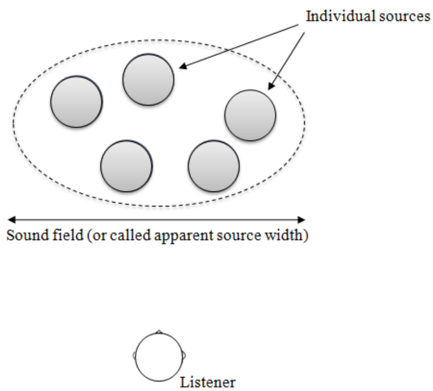

Fig. 1. Concept of the sound field or the apparent source width (ASW)

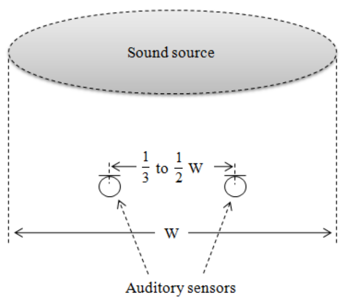

Fig. 2. Configuration of auditory sensors for capturing the spatial quality

width of the sound source. It is usual that the distance is set from one-third to one-half of the sound source width, as depicted in Fig. 2 [4]. In particular, the distance ranging from 20 to 60 cm is recommended for the appropriate distance between auditory sensors [1].

For many portable video or audio devices, however, it is difficult to arrange auditory sensors according to the recommended configuration. There is a mechanical limitation due to the size of the portable device, which results in a nearly monaural sensing system even if the device has stereo microphones. Thus, the audio information obtained from such devices rarely covers the aspects of spatial quality, i.e., the sound field.

In this paper, we propose a sound field enhancement method for a nearly monaural sensing system. First of all, principal component analysis (PCA) is applied [5][6] to decompose the audio signals obtained by stereo auditory sensors into correlated and uncorrelated components. Next, on the basis of the knowledge that inter-channel coherence (ICC) and early reflections affect the perception of the sound field [7][8], the ICC is reduced by using the combination of PCA and a decorrelator, and early reflections are generated by using an all-pass filter.

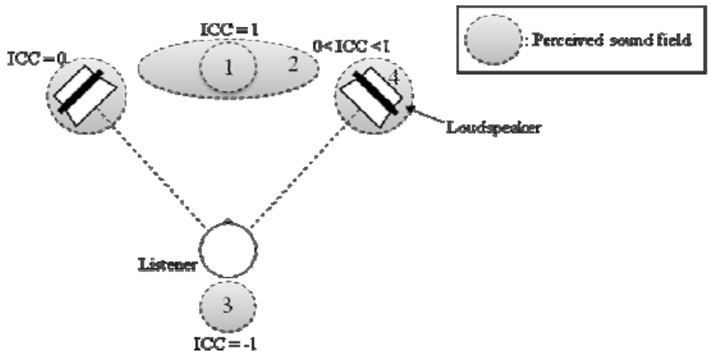

Fig. 3. Perception of the sound field depending on ICC

The remainder of this paper is organized as follows. Following the introduction, we describe the basics of spatial hearing with regard to the perception of the sound field depending on the ICC and the intensity of the early reflections in Section 2. In Section 3, we propose a sound field enhancement method that incorporates PCA, a decorrelator, and early reflections in order to control the ICC. Next, we evaluate the performance of the proposed method by calculating the ICC and by measuring the perceived sound field in Section 4. Finally, we conclude our findings in Section 5.

2 Spatial Hearing

In this section, we discuss the basics of spatial hearing with regard to the sound field. Fig. 3 shows the perception of the sound field depending on the ICC between a pair of loudspeakers. The ICC indicates the coherence between two channels and it is expressed as

$$ICC = \frac{\sum_{n=-\infty}^{\infty} x_L(n) x_R(n)}{\sqrt{\sum_{n=-\infty}^{\infty} x_L^2(n) \sum_{n=-\infty}^{\infty} x_R^2(n)}} \quad (1)$$

where $x_L(n)$ and $x_R(n)$ are the left and right audio samples at time *n*, respectively.

Fig 3 shows how the ICC affects the auditory perception [7]. When the ICC is almost 1.0, the auditory image is in the middle of two loudspeakers. Thus, the perception of the sound field is actually narrow, as shown in the shadowed area labeled as ‘1’ in Fig. 3. On the other hand, as the ICC is reduced to 0, the perception of the sound field is increased, as illustrated in the shaded area labeled as ‘4’ in Fig. 3. Thus, in order to adjust the ICC, both PCA and a decorrelator are applied in this work.

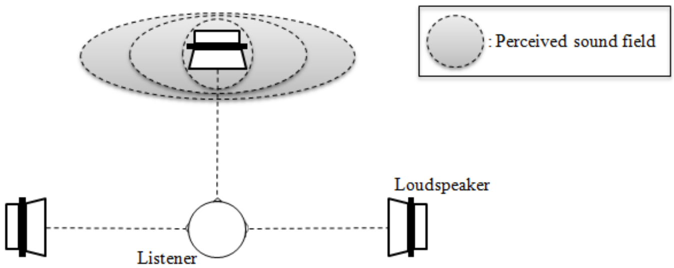

Fig. 4. Perception of the sound field depending on the intensity of early reflections

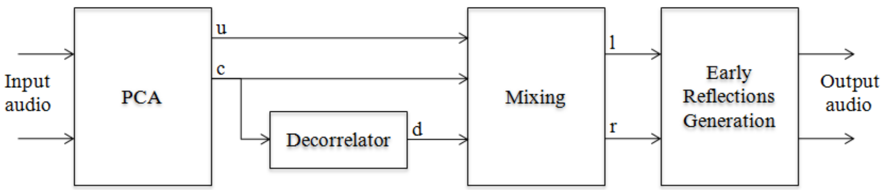

Fig. 5. Overall structure of the proposed sound field enhancement method

In addition to the ICC, early reflections are also related to the perception of the sound [8]. Fig. 4 illustrates the relationship between the early reflections and the sound field. It is found that the perception of the sound field increases with the intensity of the early reflections. Thus, in this work, we generate early reflections by using an all-pass filter.

3 Proposed Sound Field Enhancement Method

In this section, we propose a sound field enhancement method in order to enhance the sound field in a nearly monaural sensing system. Fig. 5 shows an overall structure of the proposed method. As shown in the figure, the proposed method incorporates the PCA technique and the generation of early reflections. In addition, it also uses a decorrelator.

First of all, the audio signals obtained by stereo auditory sensors are decomposed into correlated and uncorrelated components using the PCA technique. Here, the decomposition into the two signal components is achieved by using a 2×2 covariance matrix, **A**, which is defined as

$$\mathbf{A} = \begin{bmatrix} \mathrm{cov}(\mathbf{x}_L, \mathbf{x}_L) & \mathrm{cov}(\mathbf{x}_L, \mathbf{x}_R) \\ \mathrm{cov}(\mathbf{x}_R, \mathbf{x}_L) & \mathrm{cov}(\mathbf{x}_R, \mathbf{x}_R) \end{bmatrix} \quad (2)$$

where $\mathrm{cov}(\mathbf{x}_p, \mathbf{x}_q)$ is the covariance of $\mathbf{x}_p$ and $\mathbf{x}_q$, and p and q represent the left channel, L, and the right channel, R, respectively. The covariance matrix in Eq. (2)

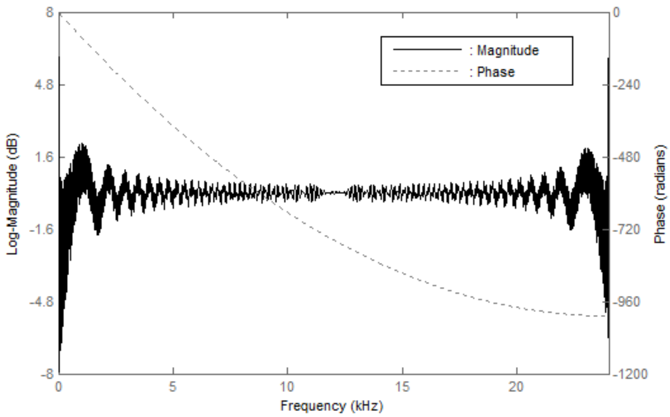

Fig. 6. Magnitude and phase response of the decorrelator

has two eigenvectors, which become the basis vectors for a new coordinate system. These eigenvectors are then used as the weight vectors corresponding to the left and right channels to generate the correlated and uncorrelated signal components.

Next, the correlated signal component is fed into the input of a decorrelator via the PCA. Fig. 6 illustrates the spectrum of the 640-tapped finite-duration impulse response (FIR) decorrelator. In addition, the correlated component, uncorrelated component, and the decorrelator output are mixed into stereo audio signals, such as

$$l(n) = \frac{1}{\sqrt{3}}\left(c(n) + u(n) + d(n)\right) \tag{3}$$

$$r(n) = \frac{1}{\sqrt{3}}\left(c(n) - u(n) - d(n)\right) \tag{4}$$

where $c(n)$ and $u(n)$ are the correlated and uncorrelated components obtained via the PCA, respectively. In addition, $d(n)$ denotes the decorrelator output. It should be noted here that the stereo audio outputs are divided by $\sqrt{3}$ for conserving a constant acoustic energy between the input and the processed audio signals.

An all-pass filter is used to generate early reflections, as shown in Fig. 7 [9]. The transfer function of the all-pass filter is expressed as

$$H(z) = \frac{z^{-M_1} - g_1}{1 - g_1 z^{-M_1}} \cdot \frac{z^{-M_2} - g_2}{1 - g_2 z^{-M_2}} = \frac{z^{-(M_1+M_2)} - g_1 z^{-M_2} - g_2 z^{-M_1} + g_1 g_2}{1 - g_1 z^{-M_1} - g_2 z^{-M_1} + g_1 g_2 z^{-(M_1+M_2)}} \tag{5}$$

where M_1 and M_2 denote delay lengths of 240 and 840, respectively. The gain coefficients, g_1 and g_2, are set to 0.67 and 0.37, respectively.

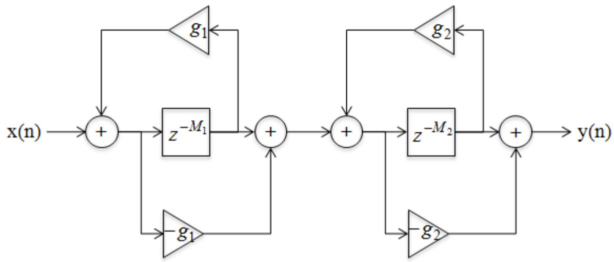

Fig. 7. An all-pass filter for generating early reflections

4 Performance Evaluation

In this section, we evaluated the performance of the proposed sound field enhancement method. First, we measured the ICC to demonstrate how the PCA and the decorrelator worked together to decompose the audio signals into correlated and uncorrelated components. Fig. 8 shows the ICCs according to frame index, where audio signals are segmented into a sequence of frames whose length is 1024. It was shown from the figure that the ICC was reduced by applying the proposed method.

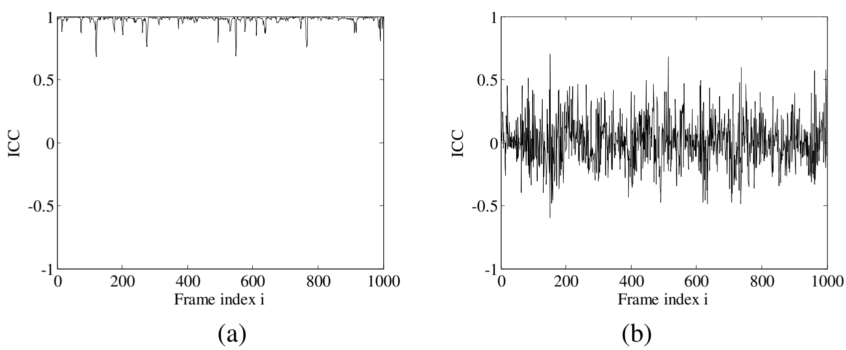

Fig. 8. ICC according to the frame index; (a) original audio signals and (b) processed audio signals

A subjective test was also conducted to evaluate the degree of the perception of the sound field perception. Six subjects with no auditory diseases participated in the test. We prepared audio files of different genres including four music genres such as ballad, dance, hip-hop, and rock. Fig. 9 illustrates the loudspeaker configuration used in the test. In this test, we measured the perceived sound field. Two types of audio contents were prepared for the test. One was the original audio content, which was obtained in a nearly monaural sensing system. The other was the audio content processed from the original file using the proposed method. For the test, randomly selected audio files were reproduced, and each subject was asked to estimate the length of the perceived sound field by ear.

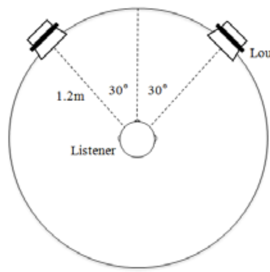

Fig. 9. Loudspeaker configuration for the subjective test

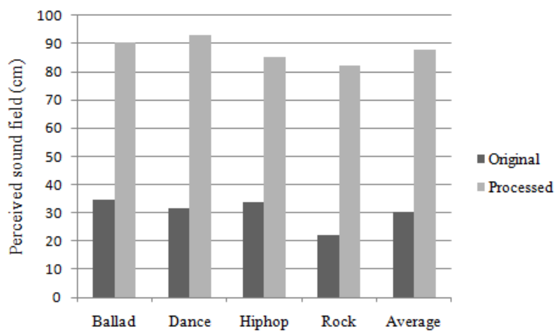

Fig. 10. Comparison of the average perceived sound field measured in cm for different genres of audio signals

Fig. 10 illustrates the result of the subjective test. As shown in the figure, the sound filed was improved after applying the proposed method. By averaging the perceived sound field over all the audio signals, it was measured as 30.50 cm for the input audio signals. However, it was increased as 87.63 cm after applying the proposed method.

5 Conclusion

In this paper, we proposed a sound field enhancement method in a nearly monaural sensing system. To this end, the ICC was controlled, and early reflections were incorporated. In other words, to decrease the ICC between the audio signals reproduced by a pair of loudspeakers, a PCA technique and a decorrelator were used. Moreover, an all-pass filter was designed to generate early reflections. The performance of the proposed method was evaluated by measuring the ICC and by performing a subjective test. It was shown from the ICC measurement and the subject test that the proposed method could reduce the ICC and improve the perception of the sound field.

References

1. Rumsey, F.: Spatial quality evaluation for reproduced sound: terminology, meaning, and a scene-based paradigm. Journal of the Audio Engineering Society 50(9), 651–666 (2002)
2. Rumsey, F.: Spatial Audio. Focal Press, Oxford (2001)
3. Gayford, M.: Microphone Engineering Handbook. Focal Press, Oxford (1994)
4. Dooley, W.L., Streicher, R.D.: MS stereo: a powerful technique for working in stereo. Journal of the Audio Engineering Society 30(10), 707–718 (1982)
5. Jolliffe, I.T.: Principal Component Analysis. Springer, Heidelberg (2002)
6. Chun, C.J., Kim, Y.G., Yang, J.Y., Kim, H.K.: Real-time conversion of stereo audio to 5.1 channel audio for providing realistic sounds. International Journal of Signal Processing, Image Processing and Pattern Recognition 2(4), 85–94 (2009)
7. Blauert, J.: Spatial Hearing: The Psychophysics of Human Sound Localization. MIT Press, Cambridge (1997)
8. Barron, M., Marshall, H.: Spatial impression due to early lateral reflections in concert halls: the derivation of a physical measure. Journal of Sound and Vibration 77(2), 211–232 (1981)
9. Schroeder, M.R., Logan, B.F.: Colorless artificial reverberation. Journal of the Audio Engineering Society 9(3), 192–197 (1961)