# Inverse Filter Design for Crosstalk Cancellation in Portable Devices with Stereo Loudspeakers

Sung Dong Jo[1] and Seung Ho Choi[2,*]

*[1] School of Information and Communications*
*Gwangju Institute of Science and Technology, Gwangju 500-712, Korea*
*[2] Dept. of Electronic and IT Media Engineering*
*Seoul National University of Science and Technology*
*Seoul 139-743, Korea*

*E-mail: shchoi@seoultech.ac.kr*
** Corresponding author*

***Abstract***

*To reproduce binaural sound at a listener's ears using stereo loudspeakers in a portable device environment, it is important to design inverse filters that cancel out acoustical crosstalk signals. However, the direct and the crosstalk paths of a head-related transfer function are similar at certain frequencies, resulting in an excessive boost at those frequencies when inverse filters are designed. To mitigate this problem, a cross-talk cancellation filter design method is proposed which allows for the selective attenuation of unwanted peaks in the spectrum by constraining the magnitude of the difference between the direct and crosstalk paths. The performance of the proposed method is then evaluated by means of subjective source localization and objective tests. It is shown from the tests that the proposed method is able to provide binaural sound effects with very closely spaced stereo loudspeakers.*

***Keywords:*** *Crosstalk cancellation, Inverse filter, Head-related transfer function (HRTF), Fast deconvolution, Binaural sound*

## 1. Introduction

The objective of binaural sound reproduction systems is to synthesize a virtual sound image such that the listener perceives a sound as if it had been produced by a specific source located at an intended position relative to the listener [1]. This type of binaural sound reproduction system can provide an immersive sound environment with variable applications in virtual reality [2], augmented reality [3], mobile phones [4-6], and home entertainment systems, among others. Typically, there are two main environments in which a binaural sound is to be rendered. The first attempts to play a binaural sound back in a headphone-based environment, while the second uses two or more loudspeakers to play back a binaural sound [7]. In a headphone-based environment, the binaural sound source is easily reproduced at the listener's ears because the headphone separates the binaural sound channels to each ear [8]. In contrast, in a loudspeaker environment, binaural sounds from each loudspeaker are mixed and simultaneously delivered to both of the listener's ears. That is, each loudspeaker sends sound to the same-side ear, while undesired sound is sent to the opposite-side ear, which is referred to as crosstalk and degrades the binaural sound effect [9].

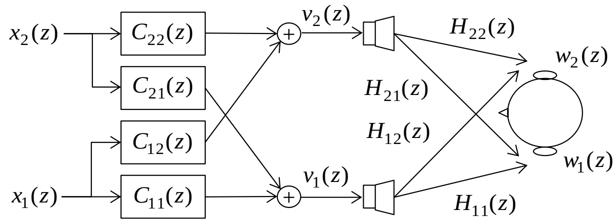



**Figure 1. Block Diagram of Crosstalk Cancellation for Stereo Loudspeakers**

Various crosstalk cancellation methods have been proposed in attempts to reduce the crosstalk effect by designing appropriate inverse filters for acoustic transfer functions [10]. In practice, a crosstalk cancellation method can be implemented with a two-by-two matrix of digital filters. Unfortunately, a severe inversion problem arises when the difference between a head-related transfer function (HRTF) for the direct path and that for the crosstalk path is very small [11]. This circumstance results in having to invert an almost singular two-by-two matrix. Such an undesirable property significantly amplifies certain frequencies of filtered signals. This problem becomes even more severe in a portable device environment because the two loudspeakers in the device are generally in close proximity to one another due to the limited size of the device [12-15].

In this paper, we propose a crosstalk cancellation method suitable for a portable device environment. In the proposed method, an ill-posed problem at particular frequencies is mitigated by constraining the difference between the direct and crosstalk paths prior to designing the inverse filters of the acoustic transfer functions. This is due to the space between speakers in portable devices.

This paper is organized as follows. Following this introduction, a conventional crosstalk cancellation method is briefly reviewed in Section 2. Section 3 describes a fast deconvolution method using frequency-dependent regularization, and Section 4 proposes a new crosstalk cancellation method. In Section 5, the performance of the proposed algorithm is evaluated. Finally, Section 6 concludes the paper.

## 2. Crosstalk Cancellation for Stereo Speakers

Figure 1 illustrates a block diagram of crosstalk cancellation for two (or stereo) loudspeakers, in which the speakers are positioned in front of a listener. In the figure, a z-transform is used to denote the relevant signals and system responses. Here, $x_1(z)$ and $x_2(z)$ refer to the input binaural signals, $v_1(z)$ and $v_2(z)$ are the inputs to the two loudspeakers, and $w_1(z)$ and $w_2(z)$ are the sound signals approached at the listener's ears. In addition, $H_{ij}(z)(i, j=1,2)$ is an acoustic transfer function from the i-th loudspeaker to the j-th ear, and $C_{ij}(z)\ (i, j=1,2)$ is a crosstalk cancellation function from $x_i(z)$ to the j-th loudspeaker. As shown in Figure 1, we have the following relationships of

$$\begin{bmatrix} v_1(z) \\ v_2(z) \end{bmatrix} = \begin{bmatrix} C_{11}(z) & C_{21}(z) \\ C_{12}(z) & C_{22}(z) \end{bmatrix} \begin{bmatrix} x_1(z) \\ x_2(z) \end{bmatrix} = \mathbf{C}(z) \begin{bmatrix} x_1(z) \\ x_2(z) \end{bmatrix} \tag{1}$$

and

$$\begin{bmatrix} w_1(z) \\ w_2(z) \end{bmatrix} = \begin{bmatrix} H_{11}(z) & H_{21}(z) \\ H_{12}(z) & H_{22}(z) \end{bmatrix} \begin{bmatrix} v_1(z) \\ v_2(z) \end{bmatrix} = \mathbf{H}(z) \begin{bmatrix} v_1(z) \\ v_2(z) \end{bmatrix}. \tag{2}$$

Even if there are four acoustic paths from the loudspeakers to the listener's ears, as in equation (2), the acoustic transfer matrix becomes symmetric for the symmetric listening environment. That is, $H_{11}(z) = H_{22}(z) = H_1(z)$ and $H_{12}(z) = H_{21}(z) = H_2(z)$. Similarly, the crosstalk transfer matrix defined in equation (1) becomes also symmetric such that $C_{11}(z) = C_{22}(z) = C_1(z)$ and $C_{12}(z) = C_{21}(z) = C_2(z)$.

Crosstalk cancellation tries to reproduce the input binaural signals, $x_1(z)$ and $x_2(z)$ at each ear of the listener, *i.e.*, $w_1(z) = x_1(z)$ and $w_2(z) = x_2(z)$. It is quite straightforward to demonstrate that crosstalk cancellation is achieved when the crosstalk cancellation matrix, $\mathbf{C}(z)$, is the inverse of the acoustic transfer matrix, $\mathbf{H}(z)$. Therefore, the main issue in a crosstalk cancellation method lies in how to invert the matrix, $\mathbf{H}(z)$. In practice, achieving an exact inverse is not feasible. In other words, although it is possible to obtain a stable and causal inverse filter with the minimum phase, the acoustic transfer functions are not likely to be the minimum phase. Moreover, $C_1(z)$ and $C_2(z)$ become very large when the difference between the direct path, $H_1(z)$, and the crosstalk path, $H_2(z)$, is very small. This problem becomes particularly severe at very low frequencies since $H_1(z)$ is nearly identical to $H_2(z)$ when the wavelength at low frequencies is very long. Such a phenomenon results in significantly amplifying the spectral magnitude of filtered signals at certain frequencies. Thus, the inverse filter design must carefully consider an environment where stereo loudspeakers are positioned in close proximity to one another.

Crosstalk cancellation filters can be designed in either time domain or frequency domain. Time domain design methods typically guarantee a stable and causal inverse filter with a complex structure and heavy computational complexity. In contrast, frequency domain design methods can be used to easily obtain a sub-optimal filter [16]. In this paper, we aim to design a crosstalk cancellation filter in the frequency domain for a portable device environment. Thus, it is extremely important to mitigate such an ill-posed problem.

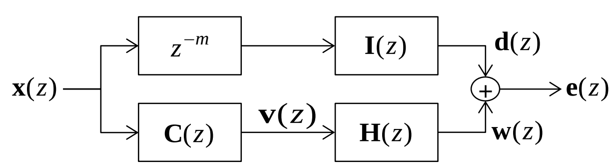



**Figure 2. Block Diagram of a Frequency-Domain Inverse Filter Design in a Least-Square Sense**

## 3. Frequency-Domain Inverse Filter Design with Regularization

In this section, a frequency domain inverse filter design method is described for crosstalk cancellation. In particular, an inverse filter is realized in a least-square sense [16]. It attempts to find the best approximation of the inverse filter in the frequency domain.

Figure 2 shows the block diagram of a frequency-domain inverse filter design method [17]. In the figure, crosstalk cancellation is assumed to be performed in the z-transform domain. Here, $\mathbf{x}(z)$ is a vector for the input binaural signals and $\mathbf{v}(z)$ is a vector for the loudspeaker input signals. Also, $\mathbf{w}(z)$ is a vector for the sound signals approached at the listener's ears, $\mathbf{d}(z)$ is a vector for the desired signals, and $\mathbf{e}(z)$ is a vector for the error signals. In addition, $\mathbf{C}(z)$ is a matrix composed of the crosstalk cancellation filters, $\mathbf{H}(z)$ is a matrix for the acoustical transfer functions, and $\mathbf{I}(z)$ refers to an identity matrix. Notice that $z^{-m}$ implements a modeling delay of m samples, induced by processing both crosstalk cancellation and acoustical transfer functions. As shown in the figure, several relationships among the signals are denoted as

$$\mathbf{v}(z) = \mathbf{C}(z)\mathbf{x}(z), \tag{3}$$

$$\mathbf{w}(z) = \mathbf{H}(z)\mathbf{v}(z), \tag{4}$$

$$\mathbf{d}(z) = z^{-m}\mathbf{I}(z)\mathbf{x}(z), \tag{5}$$

and

$$\mathbf{e}(z) = \mathbf{d}(z) - \mathbf{w}(z). \tag{6}$$

Employing the Tikhonov regularization [18], the filter design is attained by minimizing a cost function denoted as

$$\mathbf{J}_1(z) = \mathbf{e}^H(z)\mathbf{e}(z) + \beta_1 \mathbf{v}^H(z)\mathbf{v}(z) \tag{7}$$

where the first term at the right side, $\mathbf{e}^H(z)\mathbf{e}(z)$, is the performance error and the right term, $\beta_1 \mathbf{v}^H(z)\mathbf{v}(z)$, corresponds to the effort penalty. In addition, $H$ represents the Hermitian operator, which transposes and conjugates its argument. In equation (7), $\beta_1(\geq 0)$ refers to a regularization parameter to control the contribution of the effort penalty [19]. Instead of using a real number for $\beta_1$, the cost function $\mathbf{J}_1(z)$ is modified as

$$\mathbf{J}_2(z) = \mathbf{e}^H(z)\mathbf{e}(z) + \beta_2(\mathbf{B}(z)\mathbf{v}(z))^H(\mathbf{B}(z)\mathbf{v}(z)). \tag{8}$$

Comparing $\mathbf{J}_1(z)$ and $\mathbf{J}_2(z)$, the $\beta_1$ for $\mathbf{J}_1(z)$ is a regularization factor and corresponds to $\beta_2 \mathbf{B}^H(z)\mathbf{B}(z)$ for $\mathbf{J}_2(z)$, where $\mathbf{B}^H(z)\mathbf{B}(z)$ is a regularization matrix to give an emphasis on $\mathbf{v}(z)$. $\mathbf{J}_2(z)$ in equation (8) is minimized in a least-square sense, resulting in $\mathbf{C}(z)$ as

$$\mathbf{C}(z) = [\mathbf{H}^{\mathrm{H}}(z)\mathbf{H}(z) + \beta_2 \mathbf{B}^H(z)\mathbf{B}(z)\mathbf{I}]^{-1}\mathbf{H}^{\mathrm{H}}(z)z^{-m} \tag{9}$$

where * denotes the complex conjugate operator. In this paper, $\mathbf{B}(z)$ is a diagonal matrix whose element is the z-transform of a filter that amplifies the undesired frequencies boosted by crosstalk cancellation. That is, $B_{11}(z) = B_{22}(z)$, and $B_{12}(z) = B_{21}(z) = 0$.

## 4. Proposed Inverse Filter Design Method

This section proposes a crosstalk cancellation design method for providing binaural sound reproduction through a pair of loudspeakers in very close proximity to one another. As mentioned in Section 3, there is a greater potential for the ill-posed problem to arise when the two loudspeakers are positioned close together, particularly in a portable device environment, due to the similar nature of the direct and crosstalk paths from the two loudspeakers to the listener's ears [17].

In order to mitigate such a problem, the regularization methods described in Section 3 can be applied. However, each regularization method is highly dependent upon the selection of a regularization factor or a regularization matrix. However, it may degrade due to excessive suppression of the spectrum at frequencies that have no ill-posed problem. Therefore, a method must be devised to prevent such a problem at certain frequencies by constraining the magnitude difference between the direct and crosstalk paths before designing the inverse filters of the acoustic transfer functions.

Figure 3 shows a configuration of the stereo loudspeakers used in this paper. Each speaker has a diameter of 2 cm long and is placed 30 cm away from the listener. The acoustic path from the speakers to a listener's ears is modeled by an HRTF measured on a KEMAR dummy head.

Figure 4 illustrates the magnitude response of the determinant of $\mathbf{H}^{\mathrm{H}}(z)\mathbf{H}(z)$ in equation (9), which is used to compute the matrix inversion. As shown in the figure, the magnitude is extremely low at three frequencies below approximately 1 kHz, at approximately 8 kHz, and over 20 kHz. The low magnitudes are due to steep notches in HRTFs, caused by a pinna reflection and anti-aliasing filters [19]. Therefore, if a

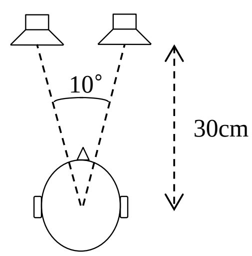



**Figure 3. Configuration of Two Loudspeakers to Simulate a Portable Device Environment**

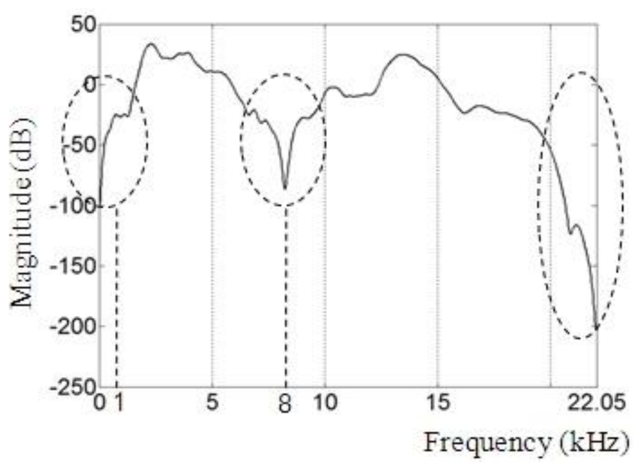



**Figure 4. Magnitude response of the determinant of $\mathbf{H}^{\mathrm{H}}(z)\mathbf{H}(z)$**

crosstalk cancellation filter could be designed by inverting the transfer functions, $H_1(z)$ and $H_2(z)$, the spectrum of the output signal processed by crosstalk cancellation would contain large peaks near the same frequencies.

This phenomenon is easily observed in Figure 5 that shows the magnitude responses for the crosstalk cancellation filters, $C_1(z)$ and $C_2(z)$, without any regularization. Here, the gain regularization factor, $\beta_2$, in equation (9) is zero. As mentioned earlier, there are sharp peaks at frequencies below approximately 1 kHz, at approximately 8 kHz, and over 20 kHz. Consequently, the crosstalk cancellation filters should be carefully designed so as to prevent sharp peaks at such frequencies.

In the proposed method, the magnitude of the determinant of $\mathbf{H}^{\mathrm{H}}(z)\mathbf{H}(z)$ is constrained. First, a discrete version of the determinant, $D(k)$, is defined as

$$D(k) = \det(\mathbf{H}^{\mathrm{H}}(z)\mathbf{H}(z))\Big|_{z=\exp(j\frac{2\pi k}{N})} \tag{10}$$

where $N$ is set to 1024 in this paper. Then, $D(k)$ is modified as

$$\hat{D}(k) = \max(D(k), T_D) \tag{11}$$

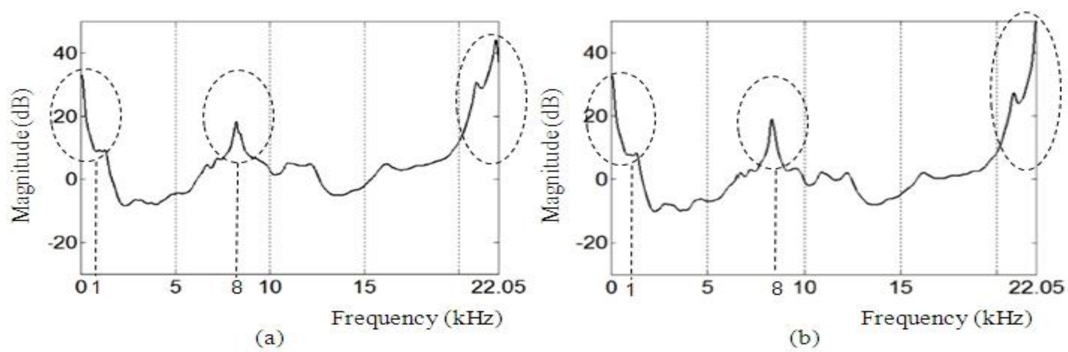



**Figure 5. Magnitude Responses of (a) $C_1(z)$ and (b) $C_2(z)$ Obtained without any Regularization**

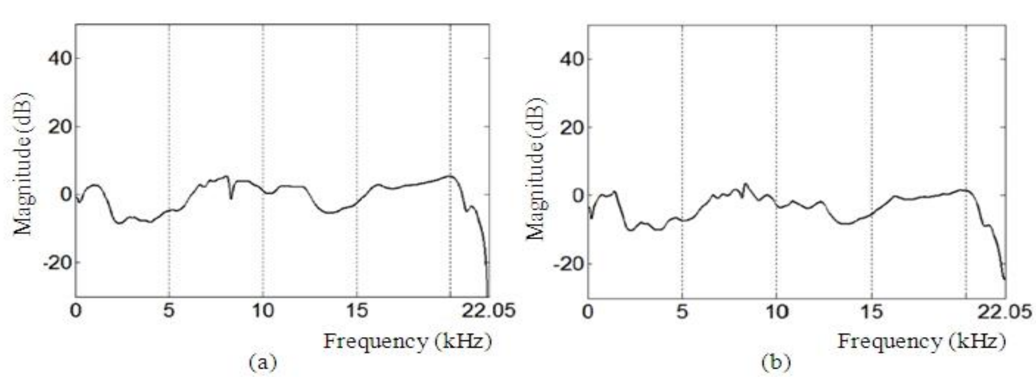



**Figure 6. Magnitude Responses of (a) $C_1(z)$ and (b) $C_2(z)$ Obtained by the Proposed Method**

where $T_D$ is a threshold and set to $10^{-5}$. Representing $\hat{D}(k)$ into the z-transform domain, $\hat{\mathbf{D}}(k)$, the inverse filter is designed as

$$\mathbf{C}(z) = \hat{\mathbf{D}}^{-1}(z)\mathbf{D}(z)[\mathbf{H}^{\mathrm{H}}(z)\mathbf{H}(z)]^{-1}\mathbf{H}^{\mathrm{H}}(z)z^{-m} \ . \tag{12}$$

Figure 6 shows the magnitude responses of $C_1(z)$ and $C_2(z)$ obtained by using equation (12). As shown in the figure, the sharp peaks of the magnitude responses disappeared at these frequencies, as compared to Figure 5.

## 5. Performance Evaluation

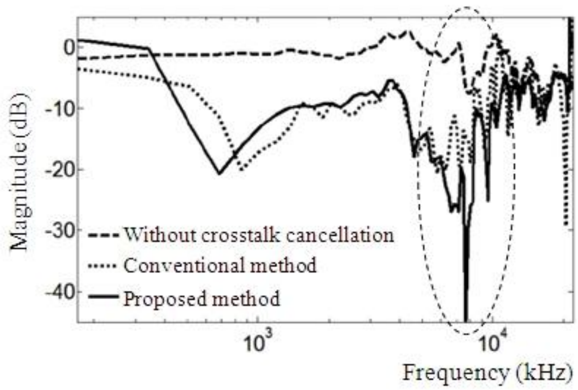



**Figure 7. Performance Comparison of Channel Separation**

In this section, the performance of the proposed method was first evaluated from the perspective of a channel separation that was defined as the ratio between the contralateral magnitude response and the ipsilateral one [20]. Figure 7 compares the performance of channel separation for three different methods, including no crosstalk cancellation, a conventional method, and the proposed method. Note here that the conventional method designed a crosstalk cancellation filter using equation (8) with $\beta_2 = 0.05$. As shown in the figure, the proposed method outperformed the conventional method, especially at around 8 kHz.

Next, a subjective test for sound localization was carried out in the environment illustrated in Figure 3. Six people participated in this test and were instructed to sit at a position in front of the stereo loudspeakers. The test stimulus was a pink noise rendered by filtering with KEMAR HRTFs, where the pink noise was 25 msec long. The azimuth was changed from -90° to 90° at a step of 30°. Each participant was asked to measure the perceived source direction in degree after listening to each stimulus. The stimuli were composed of three different pink noises, of which one was the noise rendered by HRTFs only, and the others were the noises processed by the conventional method and by the proposed method, respectively.

Figure 8 shows the perceived directions for the pink noises processed by the three different methods. When no crosstalk cancellation was applied, the azimuth was measured to be within ±5° of all desired azimuths. However, the azimuths were measured as close to the target azimuth after crosstalk cancellation methods were applied. Moreover, the proposed method provided better performance than the conventional method.

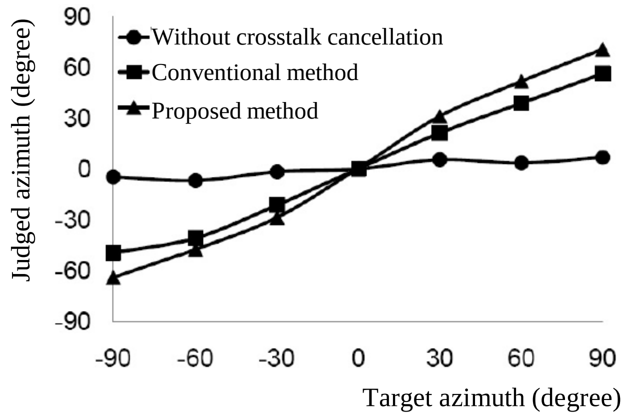



**Figure 8. Results of the Subjective Test for Sound Localization**

## 6. Conclusion

In this paper, a crosstalk cancellation design method was proposed to provide better binaural sound reproduction through a pair of very closely spaced loudspeakers. The proposed method was able to selectively attenuate unwanted peaks in the spectrum. The performance of the proposed method was evaluated by both a subjective test for sound localization and an objective test for channel separation. It was shown from the tests that the proposed method could indeed provide binaural sound effects using just a pair of very closely spaced loudspeakers, compared to a conventional crosstalk cancellation filter design method.

## References

[1] J. L. Bauck and D. H. Cooper, Generalized transaural stereo and applications. Journal of the Audio Engineering Society, vol. 44, pp. 683-705. **(1996)**.

[2] H. Lee, Y. Tateyama and T. Ogi, “Motion templates based user interface for immersive virtual reality environment”, International Journal of Control and Automation, vol. 5, no. 4, **(2012)** December, pp. 113-116.

[3] O. A. Folorunso, M. A. Ikotun and O. J. Oyelami, “Visualizing pipeline sensors datasets using augmented reality based prototype”, International Journal of Advanced Science and Technology, vol. 32, **(2011)** July, pp. 97-108.

[4] J. A. Kang and H. K. Kim, “Adaptive redundant speech transmission over wireless multimedia sensor networks based on estimation of perceived speech quality”, Sensors, vol. 11, no. 9, **(2011)** September, pp. 8469-8484.

[5] J. A. Kang and H. K. Kim, “An adaptive packet loss recovery method based on real-time speech quality assessment and redundant speech transmission”, International Journal of Innovative Computing, Information and Control, vol. 7, no. 12, **(2011)** December, pp. 6773-6783.

[6] N. I. Park and H. K. Kim, “Artificial bandwidth extension of narrowband speech applied to CELP-type speech coding”, Information: an International Interdisciplinary Journal, vol. 16, no. 3(B), **(2013)** March, pp. 3153-3164.

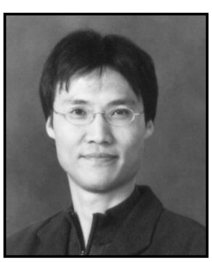
[7] C. J. Chun, H. K. Kim and S. R. Kim, “Enhanced reproduction of auditory image recorded using closely spaced stereo microphones”, Information: an International Interdisciplinary Journal, vol. 16, no. 3(B), **(2013)** March, pp. 3275-3284.

[8] D. R. Begault, “Challenges to the successful implementation of 3-D sound”, Journal of the Audio Engineering Society, vol. 39, **(1990)**, pp. 868-870.

[9] D. H. Cooper and J. L. Bauck, "Prospects for Transaural recording", Journal of the Audio Engineering Society, vol. 37, **(1989)**, pp. 3-19.
[10] P. A. Nelson, "Active control of acoustic fields and the reproduction of sound", Journal of Sound and Vibration, vol. 177, **(1994)**, pp. 447-477.
[11] S. T. Neely and J. B. Allen, "Invertibility of a room impulse response", Journal of the Acoustical Society of America, vol. 66, **(1979)**, pp. 165-169.
[12] J. H. Park, H. K. Kim, M. B. Kim and S. R. Kim, "A user voice reduction algorithm based on binaural signal separation for portable digital imaging devices", IEEE Transactions on Consumer Electronics, vol. 58, no. 2, **(2012)** May, pp. 679-684.
[13] F. Baig, S. Beg and M. F. Khan, "Zigbee based home appliances controlling through spoken commands using handheld devices", International Journal of Smart Home, vol. 7, no. 1, **(2013)** January, pp. 19-26.
[14] J. A. Kang, C. J. Chun, H. K. Kim, M. B. Kim and S. R. Kim, "A smart background music mixing algorithm for portable digital imaging devices", IEEE Transactions on Consumer Electronics, vol. 57, no. 3, **(2011)** August, pp. 1258-1263.
[15] Y. R. Oh, J. S. Yoon, H. K. Kim, M. B. Kim and S. R. Kim, "A voice-driven scene-mode recommendation service for portable digital imaging devices", IEEE Transactions on Consumer Electronics, vol. 55, no. 4, **(2009)** November, pp. 1739-1747.
[16] O. Kirkeby, P. A. Nelson, H. Hamada and F. Orduna-Bustamante, "Fast deconvolution of multichannel systems using regularization", IEEE Transactions on Speech and Audio Processing, vol. 6, **(1998)**, pp. 189-194.
[17] Y. L. Parodi and P. J. Rubak, "Objective evaluation of the sweet spot size in spatial sound reproduction using elevated loudspeakers", Journal of the Acoustical Society of America, vol. 128, **(2010)**, pp. 1045-1055.
[18] A. N. Tikhonov, "Solution of incorrectly formulated problems and the regularization method", Soviet Mathematics Doklady, vol. 4, **(1963)**, pp. 1035-1038.
[19] O. Kirkeby and P. A. Nelson, "Digital filter design for inversion problems in sound reproduction", Journal of the Audio Engineering Society, vol. 47, **(1999)**, pp. 583-595.
[20] W. G. Gardner, "3-D Audio Using Loudspeakers", Ph.D. Dissertation, MIT Media Lab., Cambridge, MA, **(1997)**.

## Authors

**Sung Dong Jo** received a B.S. degree in Electronics Engineering from Inha University, Korea in 2010, and an M.S. degree in Information and Communications from Gwangju Institute of Science and Technology (GIST), Korea in 2012. Since February 2012, he has been working in Info-Communications development team at Hyundai motor company as a research engineer. His current research interests include 3D audio signal processing and speech recognition.



**Seung Ho Choi** received a B.S. degree in Electronic Engineering from Hanyang University, Korea in 1991. He then received both M.S. and Ph.D. degrees in Electrical Engineering from Korea Advanced Institute of Science and Technology (KAIST), Korea in 1993 and 1999, respectively. He was a senior researcher at Samsung Advanced Institute of Technology, Korea, from 1996 to 2002. He was a visiting professor at University of Florida, USA, from 2008 to 2009. Since August 2002, he has been with the Department of Electronic and IT Media Engineering at Seoul National University of Science and Technology as a professor. His current research interests include speech and audio signal processing, human-computer interaction. E-mail: shchoi@seoultech.ac.kr.