Crosstalk Cancellation for Spatial Sound Reproduction in Portable Devices with Stereo Loudspeakers

Sung Dong Jo[1], Chan Jun Chun[1], Hong Kook Kim[1], Sei-Jin Jang[2], and Seok-Pil Lee[3]

[1] School of Information and Communications
Gwangju Institute of Science and Technology(GIST), Gwangju 500-712, Korea
{sdjo,cjchun,hongkook}@gist.ac.kr
[2] NSSC Center, Korea Electronics Technology Institute, Goyang, Gyeonggi-do 410-380, Korea
sjjang@keti.re.kr
[3] Digital Media Research Center
Korea Electronics Technology Institute, Seoul 137-070, Korea
lspbio@keti.re.kr

Abstract. To reproduce spatial sound through stereo loudspeakers in a portable device environment, it is important to properly design a crosstalk cancellation algorithm that cancels out acoustical crosstalk signals. In other words, the difference between the direct path of head-related transfer functions (HRTFs) and the crosstalk path of HRTFs is very small at certain frequencies, and this causes an excessive boost of frequencies when designing a crosstalk cancellation filter. To mitigate this problem, we propose a crosstalk cancellation filter design method that allows for the selective attenuation of unwanted peaks in the spectrum by constraining the magnitude of the difference between the direct and crosstalk path. The performance of the proposed method is evaluated by subjective source localization and objective tests. It is shown from the tests that the proposed method can provide improved spatial sound effects with very closely spaced stereo loudspeakers.

Keywords: Crosstalk cancellation, inverse filter, HRTFs, fast deconvolution.

1 Introduction

In recent years, numerous portable devices such as mobile phones, laptop computers, MP3 players, portable TV sets, and portable digital imaging devices have become available. Many of these devices are equipped with a pair of small loudspeakers. However, due to the limited size of these devices, the distance between the loudspeakers is very short and this leads to a poor spatial sound effects. Typically, the objective of spatial sound reproduction systems is to synthesize a virtual sound image such that the listener perceives as if the signals reproduced at the listener's ears would have been produced by a specific source located at an intended position relative to the listener [1]. This type of spatial sound reproduction system can provide an immersive sound environment with variable applications in virtual reality, augmented reality, video games, mobile phones, and home entertainment systems, among others. Typically, there are two main environments for rendering a spatial sound. The first tries to generate a spatial

T.-h. Kim et al. (Eds.): FGCN 2011, Part II, CCIS 266, pp. 114–123, 2011.

sound in a headphone-based environment and the second uses two or more loudspeakers to render a spatial sound. In the case of a headphone-based environment, the spatial sound source can be easily reproduced at the listener`s ears, since the headphone separates binaural sound channels to each ear [2]. 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, as well as undesired sound to the opposite-side ear. This problem is known as crosstalk and it degrades the spatial sound reproduction [3]. To overcome this problem, various crosstalk cancellation algorithms have been proposed in order to reduce the crosstalk effect by designing appropriate inverse filters of acoustic transfer functions. Note that the concept of crosstalk cancellation was first introduced in the early 1960s [4]. Since then, a number of sophisticated crosstalk cancellation algorithms have been presented which used two or more loudspeakers to render binaural signals [5].

In practice, a crosstalk cancellation algorithm can be implemented by a two-by-two matrix of digital filters. Unfortunately, a severe inversion problem arises when the difference between the direct path head-related transfer function (HRTF) and the cross-talk path HRTF is very small, and so one ends up having to invert an almost singular two-by-two matrix. This undesirable property causes the optimal solution to amplify certain frequencies by a large amount. This problem, usually referred to as ill-conditioning, is particularly severe at low frequencies when two loudspeakers are positioned close together [6].

In this paper, we propose a crosstalk cancellation method suitable for a portable device environment. In the proposed method, the ill-conditioning at certain frequencies is prevented from constraining the magnitude of difference between the direct and the crosstalk paths of HRTF before designing the inverse filters of the acoustic transfer functions.

The organization of this paper is as follows. Following this introduction, a conventional crosstalk cancellation method is briefly reviewed in Section 2. Section 3 describes the fast deconvolution method using frequency-dependent regularization. After that, we propose a crosstalk cancellation algorithm in Section 4. In Section 5, the performance of the proposed method is analyzed. Finally, this paper is concluded in Section 6.

2 Conventional Crosstalk Cancellation

A block diagram of crosstalk cancellation for two (or stereo) loudspeakers is illustrated in Fig. 1, which is positioned symmetrically in front of a single listener. In the figure, z-transform is used to denote the relevant signals and system responses. $x_1(z)$ and $x_2(z)$ are 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 generated at the listener`s ears. There are two acoustic transfer functions from stereo loudspeakers to the listener`s ears; the direct path $C_1(z)$ and the crosstalk path $C_2(z)$. In addition, $H_1(z)$ and $H_2(z)$ are the crosstalk cancellation functions that transform the binaural

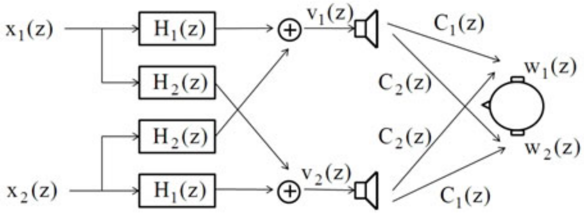


Fig. 1. Block diagram of a conventional crosstalk cancellation system for stereo loudspeakers

signals, $x_1(z)$ and $x_2(z)$, into the loudspeakers input signals, $v_1(z)$ and $v_2(z)$. By inspecting Fig. 1, it is verified that

$$\begin{bmatrix} v_1(z) \\ v_2(z) \end{bmatrix} = \begin{bmatrix} H_1(z) & H_2(z) \\ H_2(z) & H_1(z) \end{bmatrix} \begin{bmatrix} x_1(z) \\ x_2(z) \end{bmatrix} \tag{1}$$

and

$$\begin{bmatrix} w_1(z) \\ w_2(z) \end{bmatrix} = \begin{bmatrix} C_1(z) & C_2(z) \\ C_2(z) & C_1(z) \end{bmatrix} \begin{bmatrix} v_1(z) \\ v_2(z) \end{bmatrix}. \tag{2}$$

An ideal cross-talk cancellation system reproduces the input binaural signals, $x_1(z)$ and $x_2(z)$, at each ear of the listener. Thus, it is straightforward to show that this is achieved when the crosstalk cancellation matrix $\mathbf{H}(z)$ should be the inverse matrix of acoustic transfer functions, $\mathbf{C}(z)$. That is,

$$\mathbf{H}(z) = \begin{bmatrix} H_1(z) & H_2(z) \\ H_2(z) & H_1(z) \end{bmatrix} = \frac{1}{C_1(z)^2 - C_2(z)^2} \begin{bmatrix} C_1(z) & -C_2(z) \\ -C_2(z) & C_1(z) \end{bmatrix}. \tag{3}$$

2.1 Inverse Filter Design

The main issue in a crosstalk cancellation system is to invert a matrix $\mathbf{C}(z)$. In practice, the exact inverse is not possible. It is possible to get a stable and causal inverse system when a system is minimum phase. However, acoustic transfer functions are not likely to be a minimum phase system. Another problem encountered with the exact inversion is in that $H_1(z)$ and $H_2(z)$ become very large when the difference between the direct path $C_1(z)$ and the crosstalk path $C_2(z)$ is very small. This problem particularly becomes severe at very low frequencies, since the direct path, $C_1(z)$, and the crosstalk path, $C_2(z)$, are almost equal when the wavelength is very long. This undesirable property causes the optimal solution to amplify certain frequencies by a large amount. Thus, the design should carefully consider the inverse

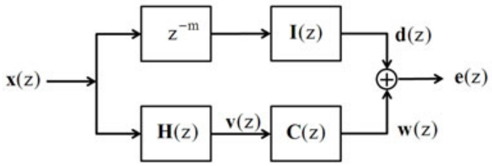


Fig. 2. Block diagram of the fast deconvolution method

functions in an environment in which stereo loudspeakers are to be positioned close together. Consequently, it is not realistic to expect an exact inverse of the acoustic transfer functions. However, by designing an approximated inverse filter, we can get an optimal solution close to the exact solution.

The requirements for designing a crosstalk cancellation filter can be broadly classified into either time domain or frequency domain design method. Time domain design methods guarantee a stable and causal solution by approximating the inverse filter in the time domain, though they have a complex structure and heavy computational complexity. In contrast, frequency domain design methods can be used easily obtain an optimal solution and to theoretically analyze the solution [5]. In this paper, we aim to design a crosstalk cancellation filter for a portable device environment where stereo loudspeakers are positioned close together. Thus, it is very important to mitigate the ill-conditioning problem.

3 Inverse FIR Filter Design Using Fast Deconvolution with Frequency-Dependent Regularization

This section outlines the theory upon which the fast deconvolution algorithm is based. Fast deconvolution is a method based on a fast Fourier transform (FFT) in combination with the least-square approximation method, which is commonly used for designing crosstalk cancellation filters [7]. This method does not try to find the exact solution, but rather the best approximation, which results in minimum errors in a least-squares sense.

3.1 Least-Squares Approximations

Fig. 2 shows a block diagram of the fast deconvolution method. In the figure, it is assumed that the crosstalk cancellation system works in the z-transform domain. Here, $\mathbf{x}(z)$ is a vector of input binaural signals, $\mathbf{H}(z)$ is a matrix composed of the crosstalk cancellation filters, $\mathbf{v}(z)$ is a vector of the loudspeaker input signals, $\mathbf{C}(z)$ is a vector of acoustical transfer functions, $\mathbf{w}(z)$ is a vector of the sound signals reproduced at the listener`s ears, $\mathbf{I}(z)$ is the identity matrix, $\mathbf{d}(z)$ is a vector of the desired signals, and $\mathbf{e}(z)$ is a vector of the performance error signals. In addition, z^{-m} implements a modeling delay of m samples to ensure that the crosstalk cancellation system is causal and performs well not only in terms of amplitude but also in terms of phase [8]. The relationships among the signals are denoted as follows

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

$$\mathbf{w}(z) = \mathbf{C}(z)\mathbf{v}(z) \tag{5}$$

$$\mathbf{d}(z) = z^{-m}\mathbf{I}(z)\mathbf{x}(z) \tag{6}$$

$$\mathbf{e}(z) = \mathbf{d}(z) - \mathbf{w}(z)\,. \tag{7}$$

Then, by employing the Tikhonov regularization [9], the filter design is performed by minimizing the cost function of

$$\mathbf{J}(z) = \mathbf{e}^H(z)\mathbf{e}(z) + \beta\mathbf{v}^H(z)\mathbf{v}(z) \tag{8}$$

where the first term $\mathbf{e}^H(z)\mathbf{e}(z)$ is the performance error term and $\beta\mathbf{v}^H(z)\mathbf{v}(z)$ is the effort penalty term. In addition, H represents the Hermitian operator, which transposes and conjugates its argument, and the positive real number β is a regularization parameter that determines the weight given to the effort term.

3.2 Fast Deconvolution Method Based on Frequency-Dependent Regularization

The cost function $\mathbf{J}(z)$ is a minimum in the least-squares sense when $\mathbf{H}(z)$ is given as

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

where * denotes the complex conjugate operator. In this case, the regularization parameter is the product of two components: a gain factor β and a shape factor $B(z)$, where $B(z)$ is the z-transform of a digital filter which amplifies the undesired frequencies boosted by crosstalk cancellation [10]. Thus, we can suppress the signal value boosted at certain frequencies by adjusting the regularization term. In other words, it is important to find the optimal frequency-dependent regularization parameters that compromise crosstalk cancellation in order to minimize the ill-conditioning problem. To this end, Eq. (9) gives an expression as a continuous function of frequency.

4 Proposed Crosstalk Cancellation

In this section, we propose a crosstalk cancellation design method for providing spatial sound reproduction through a pair of very closely spaced loudspeakers. As mentioned in Section 3, there is a greater potential for the ill-conditioning problem to arise when the two loudspeakers are positioned close together, such as the case in a portable device environment. This is because both the direct and crosstalk paths from

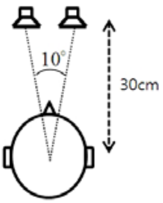


Fig. 3. Environment of the two close small loudspeakers arrangement

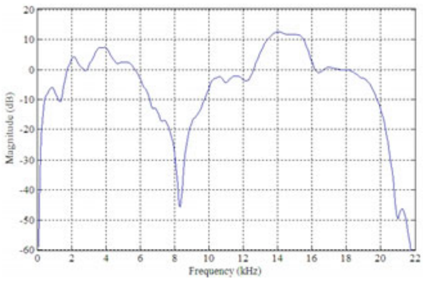
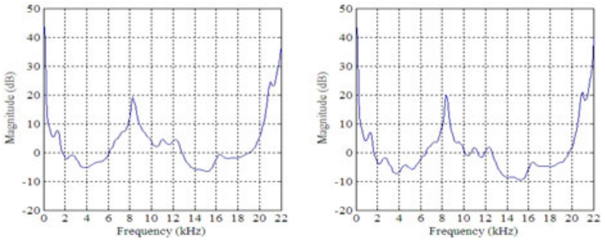


Fig. 4. Magnitude responses of difference between $C_1(k)$ and $C_2(k)$

the two loudspeakers to the listener's ears are similar. In order to mitigate this problem, a frequency-dependent regularization method based on fast deconvolution was previously proposed in [7]. However, the purpose of the regularization technique is to impose a subjective constraint on the solution. The constraint may also degrade the performance of crosstalk cancellation because of some losses in the inversion accuracy if the regularization value affects the output signals by suppressing the signal value at frequencies that have no ill-conditioning problem. Thus, we propose a technique to prevent ill-conditioning at certain frequencies by constraining the magnitude difference between the direct and crosstalk paths of HRTF before designing the inverse filters of the acoustic transfer functions.

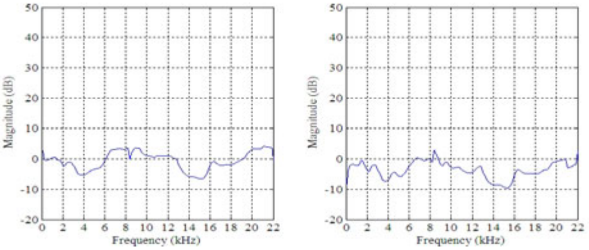
4.1 Design of Crosstalk Cancellation Filter

Fig. 3 shows the configuration of stereo loudspeakers in this study. Each speaker had 2 cm diameter. It were placed at 30 cm apart from the listener at a listening angle of 10° or -10°. In this paper, we used the HRTFs data, which were measured on a KEMAR dummy-head [11]. Fig. 4 shows the magnitude responses of the difference between the direct path, $C_1(k)$, and the crosstalk path, $C_2(k)$, according to the configuration of Fig. 3. As shown in Fig. 4, the magnitude responses of the differences between $C_1(k)$ and $C_2(k)$ were very small at low frequencies. In other words, $C_1(k)$ and $C_2(k)$ were quite similar because stereo loudspeakers were placed very close together. In addition, the magnitudes were quite small at frequencies around 8 kHz since the HRTFs had steep notches at around 8 kHz. This



(a) (b)

Fig. 5. Magnitude responses of (a) $H_1(k)$ and (b) $H_2(k)$ calculated using fast deconvolution without any regularization



(a) (b)

Fig. 6. Magnitude responses of (a) $H_1(k)$ and (b) $H_2(k)$ calculated using fast deconvolution with a threshold magnitude of T_{dB} = -5 dB

was caused by a pinna reflection and anti-aliasing filters [8]. Therefore, if a crosstalk cancellation filter was designed by inverting the transfer functions, $C_1(k)$ and $C_2(k)$, the solution would contain large peaks around the frequencies where $C_1(k)$ and $C_2(k)$ were similar. This undesirable property could lead to strong boosts at those frequencies.

Fig. 5 shows the magnitude responses of the crosstalk cancellation filters, $H_1(k)$ and $H_2(k)$, which were calculated using the fast deconvolution method with no regularization. Here, the gain regularization factor β was zero in Eq. (9). As shown in the figure, the magnitude responses of the crosstalk cancellation filters had sharp peaks at the frequencies below around 1 kHz, at around 8 kHz, and over 20 kHz. Importantly, there was a very sharp peak at around 1 kHz, which could cause an excessive boost at the low frequencies. Consequently, a crosstalk cancellation filter should be carefully implemented in order to avoid overloading the loudspeakers. To prevent the sharp peaks at such frequencies, we constrained the magnitude difference to between $C_1(k)$ and $C_2(k)$ before designing the crosstalk cancellation filters. The magnitude difference between $C_1(k)$ and $C_2(k)$ was defined as

$$m(k) = 20\log_{10}(|C_1(k) - C_2(k)|)\,. \tag{10}$$

Then, the magnitude difference between $C_1(k)$ and $C_2(k)$ was redefined as $\hat{m}(k)$ by using a threshold, T_{dB}, as

$$\hat{m}(k) = \max\left(m(k), T_{dB}\right). \tag{11}$$

Fig. 6 shows the magnitude responses of crosstalk cancellation filters, $H_1(k)$ and $H_2(k)$, which were calculated using the fast deconvolution method with a threshold of T_{dB} = -5 dB. As shown in the figure, the sharp peaks of magnitude responses disappeared at the frequencies, as compared to Fig. 5.

5 Performance Evaluation

In this section, we evaluated the performance of the proposed method. First, we conducted a subjective localization test in the environment illustrated in Fig. 3. To this end, six subjects participated in each test and the subjects were instructed to sit at a position in front of stereo loudspeakers. The test stimulus was a pink noise and binaural signals were rendered by filtering the pink noise with HRTFs. The directions were on the front horizontal plane from -90° to 90° at a step of 30° on the azimuth. Each stimulus was played 5 times at durations of 25 ms with 50 ms silent intervals. In this test, we measured the perceived source direction for the two kinds of pink noise. One was a pink noise rendered by HRTFs only, and the other was processed via the crosstalk cancellation from the pink noise rendered by HRTFs. The results of the source localization test were shown in terms of target azimuth versus the judged azimuth.

Figs. 7(a) and 7(b) show the results of the subjective localization test of the azimuth without and with crosstalk cancellation, respectively. When crosstalk cancellation was not applied, the judged azimuths were measured to within $\pm 30^{\circ}$ of all target azimuths. However, the judged azimuths were closer to the target azimuth after crosstalk cancellation was applied. Next, we measured the channel separation at the listener's ears, which was defined as the ration between the contralateral magnitude response, $C_2(k)$, and the ipsilateral magnitude response, $C_1(k)$ [12]. Fig. 8 shows the channel separation for the proposed crosstalk cancellation method. It was shown from the figure that crosstalk cancellation was indeed effective.

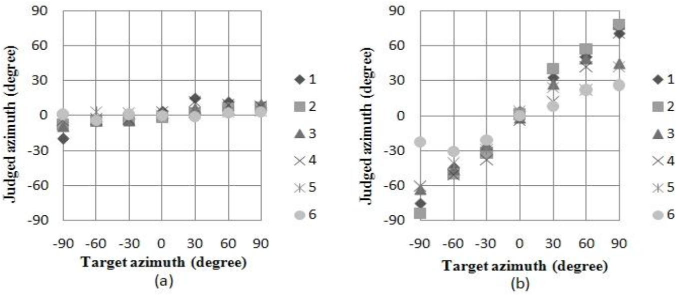


Fig. 7. Results of the subjective localization test of the azimuth (a) without and (b) with crosstalk cancellation

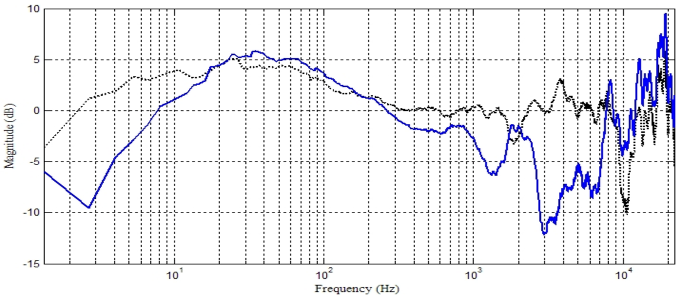


Fig. 8. Comparison of channel separation: the dashed line is a natural head shadowing channel separation and the solid line shows the channel separation after crosstalk cancellation

6 Conclusion

In this paper, we proposed a crosstalk cancellation design method that provided spatial sound reproduction through a pair of very closely spaced loudspeakers, such as those found in portable devices. The proposed method allowed the selective attenuation of unwanted peaks in the spectrum by constraining the magnitude of the difference between the direct and crosstalk paths. Then, the performance of the proposed method was evaluated by a subjective source localization test and an objective test. It was shown from the tests that the proposed method could provide spatial sound effects using only a pair of very closely spaced loudspeakers.

References

1. Bauck, J.L., Cooper, D.H.: Generalized transaural stereo and applications. Journal of the Audio Engineering Society 44(9), 683–705 (1996)
2. Begault, D.R.: Challenges to the successful implementation of 3-D sound. Journal of the Audio Engineering Society 39(11), 864–870 (1990)
3. Cooper, D.H., Bauck, J.L.: Prospects for transaural recording. Journal of the Audio Engineering Society 37(1/2), 3–19 (1989)
4. Atal, B.S., Schroeder, M.R.: Apparent sound source translator. U.S. Patent (3), 236, 949 (1966)
5. Nelson, P.A.: Active control of acoustic fields and the reproduction of sound. Journal of Sound and Vibration 177(4), 447–477 (1994)
6. Kirkeby, O., Nelson, P.A., Hamada, H.: The “stereo dipole” – a virtual source imaging system using two closely spaced loudspeakers. Journal of the Audio Engineering Society 46(5), 387–395 (1998)
7. Kirkeby, O., Nelson, P.A., Hamada, H., Orduna-Bustamante, F.: Fast deconvolution of multichannel systems using regularization. IEEE Trans. Speech and Audio Processing 6(2), 189–194 (1998)
8. Parodi, Y.L., Rubak, P.: Objective evaluation of the sweet spot size in spatial sound reproduction using elevated loudspeakers. Journal of the Acoustical Society of America 128(3), 1045–1055 (2010)

9. Tikhonov, A.N.: Solution of incorrectly formulated problems and the regularization method. Soviet Mathematics Doklady 4(2), 1035–1038 (1963)
10. Kirkeby, O., Nelson, P.A.: Digital filter design for inversion problems in sound reproduction. J. Audio Eng. Soc. 47(7/8), 583–595 (1999)
11. Gardner, B., Martin, K.: HRTF Measurements of a KEMAR dummy-head microphone. MIT Media Lab., http://sound.media.mit.edu/KEMAR.html
12. Gardner, W.G.: 3-D audio using loudspeakers. Ph.D. Dissertation. MIT Media Lab, Cambridge (1997)