한국소음진동공학회 2002년도 춘계학술대회논문집, pp. 559~564.

직접 역필터 설계법을 이용한 스테레오 재생시스템의 Cross-talk 제거

김상명, 한도근, 왕세명

A DIRECT INVERSE-BASED CROSS-TALK CANCELLATION METHOD FOR STEREO AUDIO SYSTEMS

Sang-Myeong Kim, Dogeun Han, and Semyung Wang

ABSTRACT

Cross-talk cancellation, inverse filter design or deconvolution in a generic term, is a vital process for a virtual sound realization in the stereo sound reproduction system. Most, if not all, of the design algorithms available for the inverse filter are based on a linearized model of the real physical plant. The result of such a plant-based design method, which may be referred to here as the indirect method, is biased due to both modelling and inversion errors. This paper presents a novel direct cross-talk cancellation method that may be free from the inversion error. The direct method can *directly* models the inverse filter by a suitable rearrangement of the input and output ports of the original plant so that no inversion is required here. Advantages are discussed with various experiments in an anechoic chamber using a PC soundcard. Binaural reproduction tests conducted showed that the conventional indirect method yields about 8 % reproduction performance error on both ear positions, whereas the direct method offers about 3 %.

1. INTRODUCTION

Binaural synthesis and *cross-talk cancellation (binaural reproduction)* may constitute binaural technology for virtual sound realisation in the stereo reproduction system using two loudspeakers. Binaural synthesis can be defined as a technique to synthesize a pair of binaural sound signals from an original monaural source signal with provision of the 3-dimensional head-related acoustic information: distance and direction of the source, space of the reproduction field[1-5].

On the other hand, cross-talk cancellation (binaural reproduction), or *inverse filtering* or *deconvolution* in a generic term, is a signal processing technique to design the inverse filter of an acoustic plant in order to cancel out the influence of the plant in which the binaurally recorded sound signals are reproduced[6-11].

정회원, 광주과학기술원 기전공학과
email: ksm@kjist.ac.kr

This paper presents a novel direct cross-talk cancellation method that may be free from the inversion error. The direct method can *directly* models the inverse filter by a suitable rearrangement of the input and output ports of the original plant so that no inversion is required ere. Advantages are discussed with various experiments in an anechoic chamber using a PC soundcard.

2. CROSS-TALK CANCELLATION ALGORITHMS

2.1 Overview of binaural technology for perfect binaural reproduction

A complete binaural technology implementation procedure for perfect binaural sound reproduction at the listener's eardrums is illustrated in Figure 1 for a stereo reproduction system that uses two loudspeakers[1,2]. A given monaural source signal $s(t)$ is first processed using the two filters, $A_1(j\omega)$ and $A_2(j\omega)$, to yield a pair of the binaurally synthesised sound signals, $u_1(t)$ and $u_2(t)$. It is assumed in this paper that the binaurally synthesised signals, $\mathbf{u}(t) = \{u_1(t) \quad u_2(t)\}^T$ in vector form, are already given. Thus we will now only consider the problem of cross-talk cancellation (binaural reproduction), or inverse or deconvolution in generic term, that is to perfectly reproduce the binaural signals at the listener's eardrums $\mathbf{d}(t) = \{d_1(t) \quad d_2(t)\}^T$ via cancellation of the electro-acoustic plant $G(j\omega)$ using the inverse filter $H(j\omega)$. The eardrum responses in the frequency domain can be written as

$$\mathbf{d} = \mathbf{GHu} \tag{1}$$

where frequency dependency of the variables is suppressed for brevity. In theory, the physically realisable perfect inverse filter for a given linear plant **G** can be designed by introducing a time delay term as given by [12]

$$\mathbf{H} = \mathbf{G}^{-1} e^{-j\omega\tau} \tag{2}$$

where the time delay τ should be set large enough for the impulse responses of the filters to be causal stable. Substituting equation (2) to (1) offers perfect reproduction of the binaural signals, but with some time delay, as

$$\mathbf{d} = \mathbf{u}e^{-j\omega\tau} \text{ or } \mathbf{d}(t) = \mathbf{u}(t-\tau) \tag{3}$$

where $\mathbf{d}(t)$ is the time domain representation of **d**.

Such a realisation of perfect reproduction is only possible in a noise-free ideal situation with assumptions that the plant **G** is LTI(linear time invariant) and the filter **H** is causal and stable and is the perfect inversion of **G**. In practice, however, the plant is slightly non-linear and time-variant so that **G** in Figure 1 is only a linearlized model that is unavoidably subject to some *modelling error*[13]. Even with a perfect model of the LTI plant, furthermore, calculation of **H** in equation (2) can induce *inverse error* due to the constraints of causality and robustness on **H**.

Thus, we should change the mission to authentic reproduction instead of perfect reproduction. The performance error of the binaural reproduction system for each channel may be defined as

$$J_i(\%) = \frac{E[e_i^2(t)]}{E[u_i^2(t)]} \times 100 \tag{4}$$

where $E[\circ^2]$ denotes the variance, and the error signal is defined as $e_i(t) = d_i(t) - u_i(t-\tau)$ in which $i = 1, 2$ denoting the left and right channels.

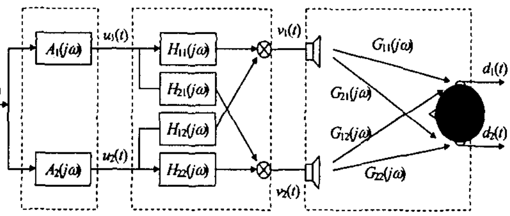


Figure 1. Binaural technology for perfect reproduction where $A_1(j\omega)$ and $A_2(j\omega)$ *conduct binaural synthesis and* $\mathbf{H}(j\omega)$ *conducts cross-talk cancellation of the electro-acoustic plant* $\mathbf{G}(j\omega)$. *The signal* $s(t)$ *is a monaural source signal, the signal vector* $\mathbf{u}(t) = \{u_1(t) \quad u_2(t)\}^T$ *denotes the binaurally synthesised sound signals,* $\mathbf{v}(t) = \{v_1(t) \quad v_2(t)\}^T$ *is the loudspeaker output signals, and* $\mathbf{d}(t) = \{d_1(t) \quad d_2(t)\}^T$ *is the sound pressures at the listener's eardrums.*

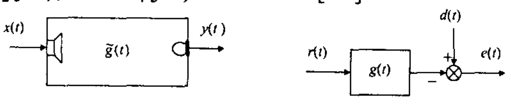
2.2 Conventional indirect methods

A. Time domain design methods

Figure 2(a) illustrates the input-output relationship of the real physical plant $\tilde{g}(t)$, where $x(t)$ and $y(t)$ are the test input signal and it corresponding output signal, respectively. Note $\tilde{g}(t)$ denotes the real physical plant, while $g(t)$ denotes its linearlised model. To identify the impulse response $g(t)$, the Wiener filter can be most preferable used, which can be explained in block diagram form as shown in Figure 2(b). The Wiener filter $g(t)$ is the optimal filter that minimises the error $e(t)$ in mean square sense when the desired $d(t)$ and received $r(t)$ signals are ergodic and stationary random[14]. Thus, by setting $r(t) = x(t)$, $d(t) = y(t)$, and $e(t) = m(t)$, we get

$$y(t) = g(t) * x(t) + m(t) \qquad (5)$$

where the impulse response $g(t)$ is the optimally linearlised model obtained by Wiener filtering, and $m(t)$ denotes the plant modelling error. There are also much faster frequency domain modelling methods, such as, $H_1(j\omega)$, $H_2(j\omega)$, and $H_v(j\omega)$ estimators[13].



(a) Impulse response test *(b) Wiener filter*

Figure 2. Impulse response identification in a SISO(single-input-single-output) system; (a) Impulse response test (b) Wiener's problem of finding h(t) in order to minimise the error signal in mean square sense when the received and desired signals are ergodic and stationary random.

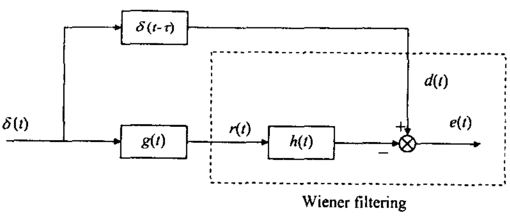


Figure 3. Deconvolution of g(t) using the inverse filter h(t).

The identified linear model $g(t)$ can now be used for design of its inverse $h(t)$. As illustrated in Figure 3, this design task is no more than a deconvolution problem that is to deconvolute $g(t)$ using the inverse filter $h(t)$[6]. For the general non-minimum phase plant $g(t)$, a perfect deconvolution is only possible by introducing an appropriate time delay τ in $\delta(t-\tau)$ [12]. Note here the deterministic approach using an impulse $\delta(t)$ instead of white noise $w(t)$ is used as the input in Figure 3 since they are equivalent as far as the Wiener filter is concerned[15]. Advantages of this approach in the discrete time implementation are its fast calculation time and high accuracy. From the desired signal can be written as

$$\delta(t-\tau) = g(t) * h(t) * \delta(t) + n(t) \qquad (6)$$

where $n(t)$ can be defined as the inverse error that could be induced by, for example, a short time delay τ. Rewriting equation (6) for the inverse filter with suppressing the inverse error term gives

$$h(t) \approx g^{-1}(t) * \delta(t-\tau) \qquad (7)$$

The filter can be obtained using the deterministic approach to the Wiener's problem with setting $r(t) = g(t)$ and $d(t) = \delta(t-\tau)$ [6, 14, 15].

The case of binaural reproduction for a pair of binaural sound signals can be similarly described by extending that for the monaural signal described above. A corresponding description for equation (6), but excluding the inverse error term $n(t)$, can be written as

$$\mathbf{I}\delta(t-\tau) = \mathbf{H}(t) * \mathbf{G}(t) * \mathbf{I}\delta(t) \qquad (8)$$

where $\mathbf{I}$ is the (2×2) identity matrix, and the impulse response matrices of the inverse filter and plant are $\mathbf{H}(t)$ and $\mathbf{G}(t)$, respectively. The inverse filter is given by[16]

$$\mathbf{H}(t) = \Delta^{-1}(t) * \begin{bmatrix} g_{22}(t) & -g_{12}(t) \\ -g_{21}(t) & g_{11}(t) \end{bmatrix} * \delta(t-\tau) \quad (9)$$

where $\Delta^{-1}(t) = (g_{11}(t) * g_{22}(t) - g_{12}(t) * g_{21}(t))^{-1}$. The elements in the matrix are guaranteed to be causal stable, and the rest term $\Delta^{-1}(t) * \delta(t-\tau)$ indicates again a deconvolution problem for the SISO system as given in equation (7).

B. Frequency domain design methods

A robust algorithm, as presented by Kirkeby *et. al.*[10] is required, which is given by

$$\mathbf{H} = \left(\mathbf{G}^{\mathbf{H}}\mathbf{G} + \beta\mathbf{I}\right)^{-1}\mathbf{G}^{\mathbf{H}}e^{-j\omega\tau} \qquad (10)$$

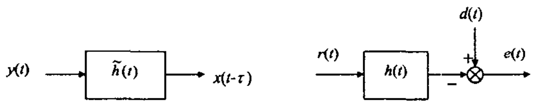
where the positive number β is a regularisation parameter that determines the trade-off between performance and robustness.

2.3 New direct method

Let a non-linear filter $\tilde{h}(t)$ denote the exact inverse filter of the real physical plant $\tilde{g}(t)$ shown in Figure 2(a). Then the input-output relationship can be represented as shown in Figure 4, where a time delay is intentionally convoluted into the output signal $x(t-\tau)$ in order to have a causal filter $\tilde{h}(t)$. Identifying $\tilde{h}(t)$ is again the Wiener filter problem, thus the output can be written as

$$x(t-\tau) = h(t) * y(t) + l(t) \tag{11}$$

where the optimal inverse filter $h(t)$ can be calculated using the Wiener filter, and $l(t)$ is the inverse filter modelling error. With the direct inverse method, it is strikingly that no explicit inversion process so that there is no inverse error $n(t)$ is involved here.



(a) Impulse response test *(b) Wiener filter*

Figure 4. Impulse response identification of the inverse filter for (a) a mono reproduction system and (b) its identification using the Wiener filter

Now consider the stereo TITO system as shown in Figure 5(a). A test input signal $x_1(t)$ is acting at the speaker 1 only, and the corresponding outputs at the microphones are $y_{11}(t)$ and $y_{21}(t)$. Likewise, the corresponding output signals for the source input at the speaker 2 can be written as $y_{12}(t)$ and $y_{22}(t)$. Thus, we get

$$\mathbf{x}_1(t-\tau) = \mathbf{H}(t) * \mathbf{y}_1(t) + \mathbf{l}_1(t),$$
$$\mathbf{x}_2(t-\tau) = \mathbf{H}(t) * \mathbf{y}_2(t) + \mathbf{l}_2(t) \tag{12a,b}$$

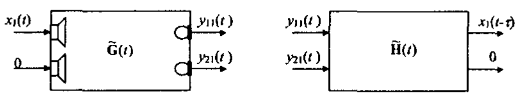
where $\mathbf{x}_1(t-\tau) = \begin{Bmatrix} x_1(t-\tau) \\ 0 \end{Bmatrix}$, $\mathbf{x}_2(t-\tau) = \begin{Bmatrix} 0 \\ x_2(t-\tau) \end{Bmatrix}$, $\mathbf{y}_1(t) = \begin{Bmatrix} y_{11}(t) \\ y_{21}(t) \end{Bmatrix}$, $\mathbf{y}_2(t) = \begin{Bmatrix} y_{12}(t) \\ y_{22}(t) \end{Bmatrix}$, $\mathbf{H}(t) = \begin{bmatrix} h_{11}(t) & h_{12}(t) \\ h_{21}(t) & h_{22}(t) \end{bmatrix}$, $\mathbf{l}_1(t) = \begin{Bmatrix} l_{11}(t) \\ l_{21}(t) \end{Bmatrix}$, and $\mathbf{l}_2(t) = \begin{Bmatrix} l_{12}(t) \\ l_{22}(t) \end{Bmatrix}$. Assume now for simplicity that the same white noise $w(t)$ is used for both the test input signals so that $x_1(t) = x_2(t) = w(t)$, then we get

$$\mathbf{I}w(t-\tau) = \mathbf{H}(t) * \mathbf{Y}(t) + \mathbf{L}(t) \tag{13}$$

By rearranging them, we finally have four separate Wiener filtering problems as given by

$$\begin{bmatrix} y_{22}(t) & -y_{12}(t) \\ -y_{21}(t) & y_{11}(t) \end{bmatrix} * w(t-\tau) \approx \begin{bmatrix} h_{11}(t) & h_{12}(t) \\ h_{21}(t) & h_{22}(t) \end{bmatrix} * \Delta(t) \tag{14}$$

where $\Delta(t) = (y_{11}(t) * y_{22}(t) - y_{12}(t) * y_{21}(t))$. For example, the Wiener filter $h_{11}(t)$ can be calculated by setting the received and desired signals as $r(t) = \Delta(t)$ and $d(t) = y_{22}(t) * w(t-\tau)$. The other three filter responses can be similarly obtained via Wiener filtering.



(a) Impulse response test (b) Inverse filter

Figure 5. Input-output relationships for (a) the physical plant $\tilde{\mathbf{G}}(t)$ *and (b) its corresponding inverse filter* $\tilde{\mathbf{H}}(t)$.

3. BINAURAL SOUND REPRODUCTION TEST

3.1 Experimental set-up

Binaural as well as monaural reproduction tests were performed in an anechoic chamber, and the geometric setting for the binaural case is shown in Figure 6(a), where the distance between the centres of the loudspeakers and microphones l was 0.5m, and the distances between loudspeakers and microphones were 0.6m and 0.2m, respectively. For the monaural case, the distance between the loudspeaker and microphone was 0.5m. The whole experiments including impulse response identification and binaural reproduction were conducted in a PC using a soundcard(LINX TWO®) as a front-end device for data acquisition and transmission. The loudspeakers used were ordinary small two-way PC loudspeakers, and the microphones were B&K 1/2-inch condenser microphones(Type : 4189). Matlab® signal processing and data acquisition toolboxes were used for signal processing and data acquisition and transmission.

Each of the four impulse responses of the (2×2) plant matrix was first identified with a wire connection setting as illustrated in Figure 6(b). A MLBS(maximum length binary sequence) of length (2^{18}-1) was used for the input signal, and the lengths of FIR filters for the plant and the inverse filter were both 1024. The sampling frequency was 32000 Hz. The mean coherence values of each plant were all above 0.94, which indicates that the electro-acoustic plant is quite linear. Finally, Figure 6(c) illustrates the wire connection for binaural reproduction tests that will be described in the following section. For monaural reproduction, a single channel was used in both the input and output ports.

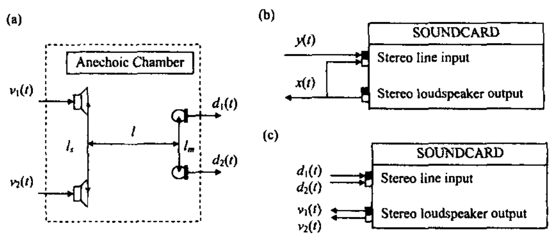
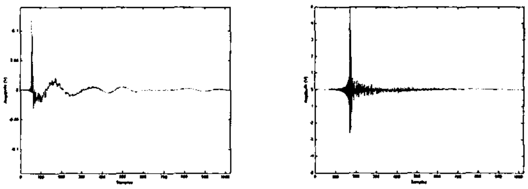


Figure 6. Experimental set-up: (a) geometric set-up, (b) Wire connection for impulse response identification, and (c) Wire connection for binaural reproduction

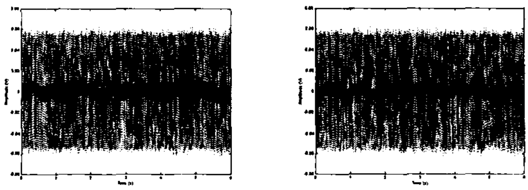
3.2 Monaural reproduction test

The electro-acoustic plant impulse response was calculated by using the Wiener filter, and it is shown in Figure 7(a). The indirect and direct inverse filters for authentic reproduction looked very similar so that only the indirect inverse filter is shown in Figure 7(b). Each of them was implemented separately as the inverse filter, and the reproduction results are shown in Figure 8 where dotted lines denote the recorded data while the solid lines denote the reproduction error $e_i(t)$. The reproduction performance errors for the indirect and direct methods are 1.4% and 0.9%, respectively.



(a) Plant response *(b) Inverse filter*

Figure 7. Identification of the electro-acoustic plant and its inverse filter using the indirect inverse method

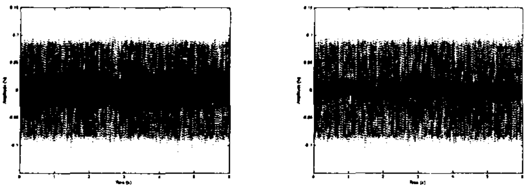


(a) Indirect method *(b) Direct method*

Figure 8. Monaural reproduction performance errors: (a) 1.4% and (b) 0.9% where the dotted lines denote the recorded signal and the solid lines denotes the reproduction error

3.3 Binaural reproduction test

Both indirect and direct inverse filter matrixes were implemented separately, and the reproduction results are shown in Figure 9(a, b) as the same form of Figure 8. Only the response on the left microphone is shown since that on the right was similar. For the indirect inverse filter, the performance errors of the left and right channels were 8.2% and 8.1%, respectively. On the other hand, when the direct inverse filter matrixes were implemented, they were 2.7 % on both left and right channels. The reproduction performance errors for both monaural and binaural sound signals are tabulated in Table 1 for easy reference.



(a)Indirect method *(b) Direct method*

Figure 9. Binaural reproduction performance errors on the left microphone

Table 1. Binaural reproduction performance errorsof the indirect and direct inverse methods

Reproduction system		Indirect inverse method	Direct inverse method
Mono		1.4 %	0.9 %
Stereo	Left	8.2 %	2.7 %
	Right	8.1 %	2.7 %

4. CONCLUSIONS

An efficient method for binaural reproduction of a pair of binaural sound signals has been considered. Unlike the conventional methods, the so-called direct inverse method considered in this research does not refer to a linearlised model of the real plant, but *directly* models the inverse filter by a suitable rearrangement of the input and output ports of the original plant so that no inversion is required here. Advantages are discussed with various experiments in an anechoic chamber using a PC soundcard. Binaural reproduction tests conducted showed that the conventional indirect method yields about 8 % reproduction performance errors on both ear positions, whereas the direct method offers about 3 %.

ACKNOWLEDGEMENT

This work was supported by grant No. 2000-1-30400-006-3 from the Korea Science and Engineering Foundation.

REFECRENCES

1. J.P. Blauert, 1996, *Spatial Hearing*, rev. ed. Cambridge, MA: MIT Press
2. D.R. Begault, 1994, *3-D Sound for Virtual Reality and Multimedia*, New York: Academic,
3. B. Gardner and K. Martin, *KEMAR HRTF Measurements*, MIT's Media Lab, http://sound.media.mit.edu/KEMAR.html
4. C.P. Brown and R.O. Duna, 1998, *IEEE Trans. on Speech and Audio Processing*, 6(5), 476-488, A structural model for binaural sound synthesis
5. P.A. Nelson and Y. Kahana, 2001, *Journal of Sound and Vibration*, 239(4), 607-637, Spherical harmonics, singular-value decomposition and the head-related transfer function
6. S. Haykin, 1996, Adaptive Filter Theory, 3rd. ed., Prentice Hall
7. B. Widrow and E. Walach, 1996, *Adaptive Inverse Control*, Prentice Hall International, Inc.
8. B.B. Bauer, 1961, *Journal of Audio Engineering Society*, vol. 9, 148-151, Stereophonic earphones and binaural loudspeakers
9. P.A. Nelson, H. Hamada, and S.J. Elliott, 1992, *IEEE Trans. on Signal Processing*, 40, 1621-1632, Adaptive inverse filters for stereophonic sound reproduction
10. O. Kirkeby, P.A. Nelson, H. Hamada, and F. Orduna-Bustamante, 1998, *IEEE Trans. on Speech and Audio Processing*, 189-194, Fast deconvolution of multichannel systems using regularisation
11. W.G. Gardner, 1997, *Proceedings of WASPAA*, Head-tracked 3-D audio using loudspeakers, New Paltz, NY
12. A.V. Oppenheim and R.W. Schafer, 1989, Discrete-Time Signal Processing, Prentice Hall, Inc.
13. J.S. Bendat and A.G. Piersol, 2000, Random Data: Analysis and Measurement Procedures, 3rd. ed., John Wiley & Sons, Inc.
14. A. papoulis, 1984, *Signal Analysis*, New York: McGraw-Hill International Ed.
15. M.J. Brennan and S.M. Kim, 2001, *Journal of Sound and Vibration*, 246(2), 281-296, Feedforward and feedback control of sound and vibration – a Wiener filter approach
16. H.S. Kim, P.M. Kim, and H.B. Kim, 2000, *ETRI Journal*, 22(2), 11-19, Cross-talk cancellation algorithm for 3D sound reproduction