Multirate Adaptive Filtering for Immersive Audio

Jong-Soong Lim and Chris Kyriakakis

Immersive Audio Laboratory
Integrated Media Systems Center
University of Southern California
3740 McClintock Ave., EEB 432
Los Angeles, California 90089-2564, USA
E-mail: jongsoon@sipi.usc.edu, ckyriak@imsc.usc.edu

ABSTRACT

This paper describes a method for implementing immersive audio rendering filters for single or multiple listeners and loudspeakers. In particular, the paper is focused on the case of single or two listeners with different loudspeaker arrays to determine the weighting vectors for the necessary FIR and IIR filters using the LMS (least-mean-squares) adaptive inverse algorithm. It describes transform-domain LMS adaptive inverse algorithm that is designed for crosstalk cancellation necessary in loudspeaker-based immersive audio rendering. Specifically, each weighting vector of the inverse filter is generated based on psychoacoustic critical band filters and uses the LMS adaptive inverse algorithm to improve performance in the sensitive frequency bands. We also investigate the sensitivity of the listening position under different number of listeners and loudspeakers with various loudspeaker geometries. Performance is measured based on the ipsilateral signal to contralateral signal (crosstalk) ratio that results from the different filter types with and without psychoacoustic critical band filtering.

1. INTRODUCTION

An important issue in the immersive audio rendering is the reproduction of 3-D sound fields that preserve the desired spatial location, frequency response, and dynamic range of the sound. There are two general methods for 3-D audio rendering that can be categorized as headphone reproduction and loudspeaker reproduction [1]. Head-related binaural recording, or dummy-head stereophony methods, attempt to accurately reproduce at each eardrum of the listener the sound pressure generated by a set of sources and their interactions with the acoustic environment [2]. Transaural audio is a method used to deliver binaural signals to the ears of listeners using multiple loudspeakers. The basic idea is to filter the binaural signal such that the subsequent stereo presentation produces binaural signals at the ears of the listener. The technique was first put into practice by Schroeder and Atal [3, 4] and later refined by Cooper and Bauck [5], who first referred to it as "transaural audio". Previous research [5, 7, 9, 12] shows the theoretical and practical methods for generalizing crosstalk cancellation filter design using matrix formulation. In the work of Cooper and Bauck [12] further conceptual ideas were discussed for developing transaural systems for multiple listeners with multiple loudspeakers. Nelson *et al* [7] and J.-S. Lim *et al* [9] showed the equalization of multichannel sound reproduction systems using LMS adaptive algorithm for a listener using two loudspeakers or two listeners using four loudspeakers in a symmetric or non-symmetric environment. As the eigenvalue spread of the input autocorrelation matrix increases, the convergence speed of LMS for multichannel adaptation is too slow. To solve this problem, algorithms such as DFT/LMS and DCT/LMS (discrete Fourier and discrete cosine transform-LMS), which attempt to decorrelate the inputs by preprocessing them with a transformation that is independent of the input signal, have been proposed. In this paper we present results from three crosstalk cancellation filter design approaches based on the LMS adaptive inverse algorithm, the normalized frequency domain adaptive filter (NFDAF) LMS inverse algorithm, and a multirate critical band adaptive inverse algorithm.

2. CROSSTALK CANCELLATION FILTER DESIGN

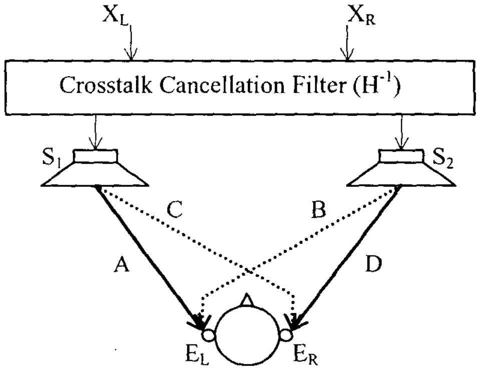

Fig. 1. Geometry and signal paths from input binaural signals to ears that shows the ipsilateral (solid) and contralateral (dashed) signal paths.

A typical two-loudspeaker listening situation is shown in Figure 1, where X_L and X_R are the binaural signals sent to the speakers (S_1 and S_2), and E_L and E_R are the signals at the listener's ears. The system can be fully described by the matrix equation

$$\begin{bmatrix} E_L \\ E_R \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \cdot \begin{bmatrix} S_1 \\ S_2 \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \cdot \left[H^{-1} \right] \cdot \begin{bmatrix} X_L \\ X_R \end{bmatrix} \qquad (1)$$

0-7803-7041-4/01/$10.00 ©2001 IEEE

where **A**, **B**, **C**, and **D** are transfer functions (vectors) from speaker (**S**) to ear (**E**). If **X** is the binaural signal, we need to deliver the left channel binaural signal X_L to E_L, the right channel binaural signal X_R to E_R, and eliminate the unwanted crosstalk terms. If the above definition is satisfied, matrix equation (1) is formulated as follows

$$\begin{bmatrix} E_L \\ E_R \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \cdot \left[H^{-1} \right] \cdot \begin{bmatrix} X_L \\ X_R \end{bmatrix} = \begin{bmatrix} X_L \\ X_R \end{bmatrix}. \qquad (2)$$

For optimum performance the matrix product of the transfer function and the inverse matrix (crosstalk cancellation filter) should be the identity matrix

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} \cdot \left[H^{-1} \right] = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}. \qquad (3)$$

From matrix equation (3), the matrix H^{-1} is

$$\left[H^{-1} \right] = \begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} \cdot \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \frac{1}{AD - BC} \begin{bmatrix} D & -B \\ -C & A \end{bmatrix} \overset{let}{=} \begin{bmatrix} W_1 & W_3 \\ W_2 & W_4 \end{bmatrix} \qquad (4)$$

where, **W** is the weight vector.

3. ADAPTIVE INVERSE CONTROL FILTER

The first crosstalk canceller described is implemented using the LMS adaptive inverse algorithm [6]. We need to rearrange matrix equations (2) and (4) based on the adaptive inverse algorithm for multiple channels [7, 8, 9] as follows:

$$\begin{bmatrix} E_L \\ E_R \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \cdot \left[H^{-1} \right] \cdot \begin{bmatrix} X_L \\ X_R \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \cdot \begin{bmatrix} W_1 & W_3 \\ W_2 & W_4 \end{bmatrix} \cdot \begin{bmatrix} X_L \\ X_R \end{bmatrix} = \begin{bmatrix} X_L \\ X_R \end{bmatrix}. \qquad (5)$$

If the weight vector **W** is the optimum crosstalk canceller, the signal **E** delivered to each ear is exactly the same signal as the input binaural signal **X**. Equation (5) is modified as follows

$$\begin{bmatrix} E_L \\ E_R \end{bmatrix} = \begin{bmatrix} AW_1 + BW_2 & AW_3 + BW_4 \\ CW_1 + DW_2 & CW_3 + DW_4 \end{bmatrix} \cdot \begin{bmatrix} X_L \\ X_R \end{bmatrix} = \begin{bmatrix} AW_1X_L + BW_2X_L + AW_3X_R + BW_4X_R \\ CW_1X_L + DW_2X_L + CW_3X_R + DW_4X_R \end{bmatrix}. \qquad (6)$$

where, AW_1+BW_2 and CW_3+DW_4 are the ipsilateral transfer functions, and AW_3+BW_4 and CW_1+DW_2 are the contralateral transfer functions (crosstalk). Equation (6) can be separated into matrix product of crosstalk canceller and other vectors

$$\begin{bmatrix} E_L \\ E_R \end{bmatrix} = \begin{bmatrix} AX_L & BX_L & AX_R & BX_R \\ CX_L & DX_L & CX_R & DX_R \end{bmatrix} \cdot \begin{bmatrix} W_1 \\ W_2 \\ W_3 \\ W_4 \end{bmatrix} = \begin{bmatrix} X_L \\ X_R \end{bmatrix}. \qquad (7)$$

Equation (7) can be formulated as shown in Figure 2 and the weight vectors are updated as follows

$$W_1(n+1) = W_1(n) + \mu(-\hat{\nabla}_1(n)),$$
$$W_2(n+1) = W_2(n) + \mu(-\hat{\nabla}_2(n)),$$
$$W_3(n+1) = W_3(n) + \mu(-\hat{\nabla}_3(n)), \textit{ and}$$
$$W_4(n+1) = W_4(n) + \mu(-\hat{\nabla}_4(n)). \qquad (8)$$

Where, $\hat{\nabla}_1(n) = -2[e_1(n) \times (A * X_L) + e_2(n) \times (C * X_L)]$,

$$\hat{\nabla}_2(n) = -2[e_1(n) \times (B * X_L) + e_2(n) \times (D * X_L)],$$
$$\hat{\nabla}_3(n) = -2[e_1(n) \times (A * X_R) + e_2(n) \times (C * X_R)], \text{ and}$$
$$\hat{\nabla}_4(n) = -2[e_1(n) \times (B * X_R) + e_2(n) \times (D * X_R)]. \qquad (9)$$

Each error is:

$$e_1(n) = d_1(n) - y_1(n) = X_L(n - M)$$
$$-[(W_1 * A + W_2 * B) * X_L + (W_3 * A + W_4 * B) * X_R],$$
$$e_2(n) = d_2(n) - y_2(n) = X_R(n - M)$$
$$-[(W_1 * C + W_2 * D) * X_L + (W_3 * C + W_4 * D) * X_R]. \qquad (10)$$

In Figure 2, **d(n)** could simply be a pure delay, say of M samples, which will assist in the equalization of the minimum phase components of the transfer function matrix in equation (7). The inclusion of an appropriate modeling delay significantly reduces the mean square error produced by the equalization process. The filter length, as well as the delay M, can be selected based on the minimization of the mean squared error. This method can be used either off-line or in real time according to virtual sound position and movement of the listener's head. The weight vector (crosstalk cancellation filter) can be chosen to be either an FIR or an IIR impulse sequence.

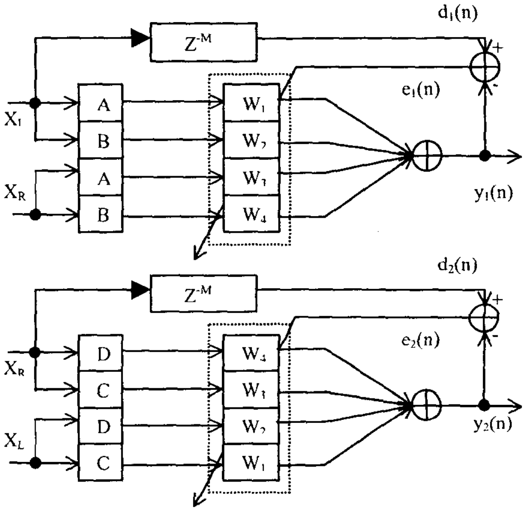

Fig. 2. LMS block diagrams for the estimation of the crosstalk cancellation filter.

4. TRANSFORM DOMAIN ADAPTIVE INVERSE CONTROL FILTER

Frequency domain implementations of the LMS adaptive filter have several advantages over time domain implementations that include improved convergence and reduced computational complexity. In practical implementations of frequency domain LMS adaptive filters, the input power varies dramatically over the different frequency bins. To overcome this, the NFDAF LMS inverse algorithm can be used estimating the input power in each frequency bin. The power estimate was included directly in the frequency domain LMS algorithm [10]. Our adaptive inverse filter algorithm in Figure 2 is modified in the frequency domain using the overlap-save method NFDAF LMS inverse algorithm [11], which is shown in Figure 3. The general form of FDAF algorithms can be expressed as follows:

$$W(k+1) = W(k) + 2G\mu(k)X^{H}(k)E(k) \qquad (11)$$

in which the superscript H denotes complex conjugate transpose. The time-varying matrix $\mu(k)$ is diagonal and it contains the step sizes $\mu_m(k)$. Generally, each step size is varied according to the signal power in that frequency bin. In our crosstalk cancellation filter implementation:

$$\begin{aligned}
W_1(k+1) &= W_1(k) + \mu \times \mathit{fft}([\mathit{ifft}(\frac{S_1^{H}(k)}{P_1(k)} \cdot E_1(k) + \frac{S_5^{H}(k)}{P_5(k)} \cdot E_2(k)]),\\
W_2(k+1) &= W_2(k) + \mu \times \mathit{fft}([\mathit{ifft}(\frac{S_2^{H}(k)}{P_2(k)} \cdot E_1(k) + \frac{S_6^{H}(k)}{P_6(k)} \cdot E_2(k)]),\\
W_3(k+1) &= W_3(k) + \mu \times \mathit{fft}([\mathit{ifft}(\frac{S_3^{H}(k)}{P_3(k)} \cdot E_1(k) + \frac{S_7^{H}(k)}{P_7(k)} \cdot E_2(k)]),\\
W_4(k+1) &= W_4(k) + \mu \times \mathit{fft}([\mathit{ifft}(\frac{S_4^{H}(k)}{P_4(k)} \cdot E_1(k) + \frac{S_8^{H}(k)}{P_8(k)} \cdot E_2(k)]) \qquad (12)
\end{aligned}$$

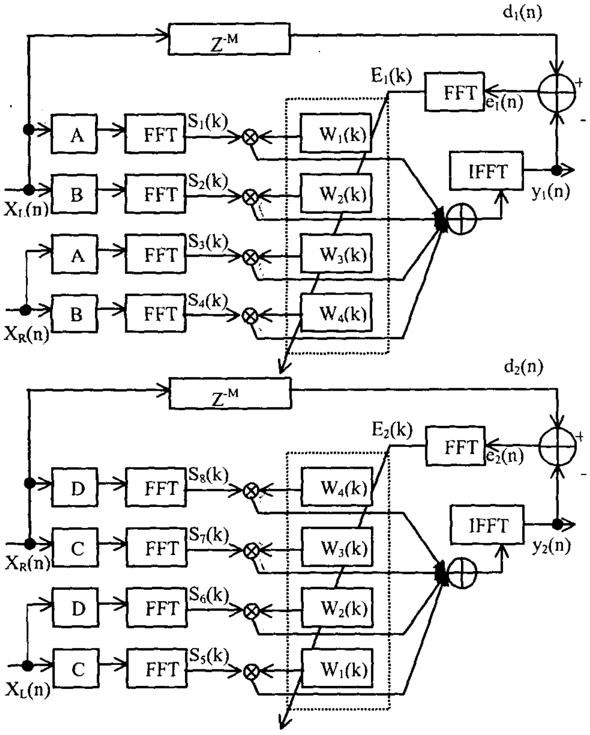

Fig. 3. NFDAF LMS inverse algorithm block diagrams for the estimation of the crosstalk cancellation filter.

where $P_n(k)$ is an estimation of the signal power in the nth input signal.

$$P_n(k) = \lambda P_n(k-1) + \alpha |C_n(k)|^2 . \qquad (13)$$

$P_n(k)$ and $C_n(k)$ are vectors composed of N (vector size) different bins. $\lambda = 1 - \alpha$ is a forgetting factor.

6. RESULTS

We implemented crosstalk cancellation filters using the algorithms described above. The value of the delay M and tap size of each FIR filter in the adaptive algorithm were chosen so as to minimize adaptation error and make the FIR filter causal. The training data used for each adaptive algorithm were random noise signals with zero mean and unity variance that include frequency range between 200 Hz and 10 kHz. The performance of crosstalk canceller was measured based on the equation (6). We have found that the desired characteristics in the frequency domain for the ipsilateral and contralateral signal transfer functions require that the magnitude response of ipsilateral signal transfer functions in the frequency domain should have $|A(\omega)W_1(\omega) + B(\omega)W_2(\omega)| = 1$, and $|C(\omega)W_3(\omega) + D(\omega)W_4(\omega)| = 1$ for lossless signal transfer in the expected frequency band. The ipsilateral signal transfer function should be linear phase: $\angle[A(\omega)W_1(\omega) + B(\omega)W_2(\omega)] = \exp(-jn\omega)$, and $\angle[C(\omega)W_3(\omega) + D(\omega)W_4(\omega)] = \exp(-jn\omega)$. The magnitude response of contralaterall signal transfer functions should be $|A(\omega)W_3(\omega) + B(\omega)W_4(\omega)| = 0$, and $|C(\omega)W_1(\omega) + D(\omega)W_2(\omega)| = 0$ for perfect crosstalk cancellation. All of the requirements described above apply to the frequency range between 200Hz and 10KHz.

Figure 4 presents typical results for the LMS adaptive inverse algorithm. The magnitude response of the ipsilateral signal is about 0dB in the frequency range between 200Hz and 10KHz with linear phase. Therefore the desired signal can be transferred from loudspeaker to ear without distortion. The magnitude response of the contralateral signal is at least 20dB below the ipsilateral signal in the same range. Figure 5 presents the result of the normalized frequency domain adaptive filter inverse algorithm. The magnitude response of ipsilateral signal

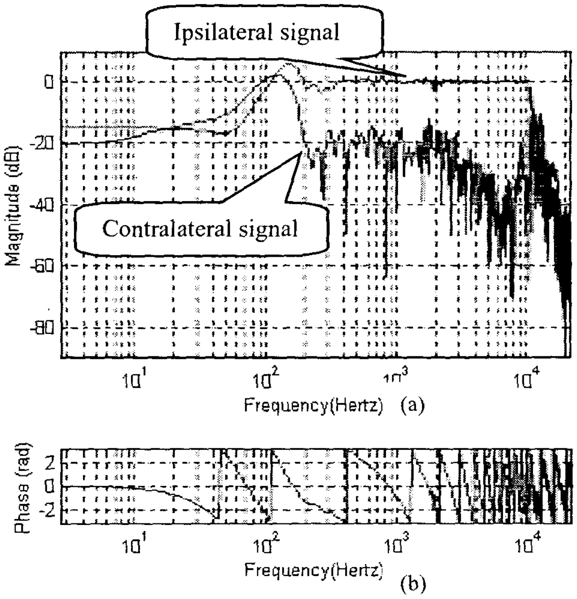

Fig. 4. Frequency response of LMS adaptive inverse algorithm. (a) Magnitude response, (b) Phase response.

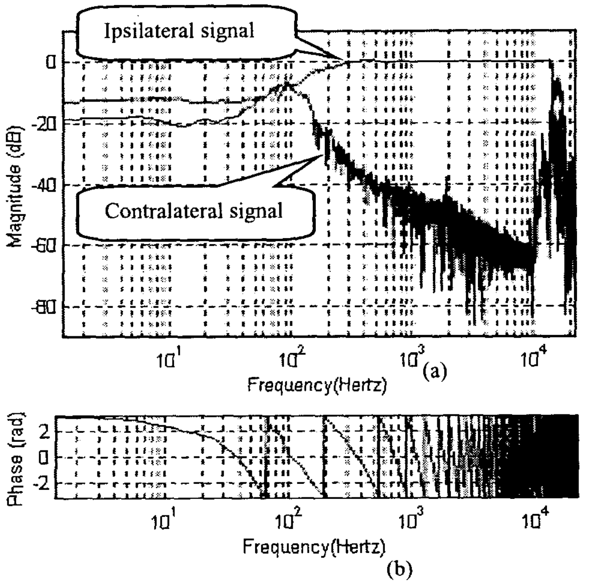

Fig. 5. Frequency response of NFDAF LMS inverse algorithm. (a) Magnitude response, (b) Phase response.

is about 0dB in the frequency range between 200Hz and 10KHz with linear phase. It has almost the same magnitude response as Figure 4. However, the magnitude response of contralateral signal is more than 40 dB below the ipsilateral signal. From above results it is clear that the NFDAF LMS inverse algorithm performs better and faster than the LMS adaptive inverse algorithm.

7. CONCLUSION

In this paper we designed a generalized transaural crosstalk cancelling filter to produce virtual sound source spatialization through multiple loudspeakers. Each filter's coefficients were calculated using LMS inverse algorithms in time and frequency domain. To increase the convergence speed and improve the performance of the 3-D rendering effect, the frequency domain can be divided into multi-bands based on the critical band structure dictated by psychoacoustics. Crosstalk cancellation filters in each critical band can be obtained using NFDAF. In future work we will apply above algorithms to multiple listeners and more than two loudspeakers.

8. REFERENCES

[1] H. W. Gierlich, "The application of binaural technology," *Appl. Acoust.*, vol. 36, pp. 219-243, 1992.

[2] H. Moller, "Fundamentals of binaural technology," *Appl. Acoust.*, vol. 36, pp. 171-218, 1992

[3] M. R. Schroeder and B. S. Atal, "Computer simulation of sound transmission in rooms," *IEEE Conv. Record,* 7:150-155, 1963

[4] M. R. Schroeder, D. Gottlob, and K. F. Siebrasse, "Comparative Stude of European Concert Halls," *J. Acoust. Soc. Am.*, 56:1195-1201, 1974.

[5] D. H. Cooper and J. Bauck, "Prospects for Transaural Recording," *J. Audio Eng. Soc.*, 37(1/2):3-19, 1989.

[6] B. Widrow and E. Walach, *Adaptive Inverse Control.* Prentice Hall, 1995.

[7] P. A. Nelson, H. Hamada, and S. J. Elliott, "Adaptive Inverse Filters for Stereophonic Sound Reproduction," *IEEE Trans. Signal Process.*, vol. 40, pp. 1621-1632 (1992 July).

[8] A. Mouchtaris, J.-S. Lim, T. Holman, and C. Kyriakakis, "Head-Related Transfer Function Synthesis for Immersive Audio," *Proceedings of the IEEE Multimedia Signal Processing Workshop (MMSP '98)*, Redondo Beach, California, December, 1998.

[9] J.-S. Lim and C. Kyriakakis, " Virtual Loudspeaker Rendering for Multiple Listeners," *The 109th AES Convention*, Preprint 5183 (C-1), Los Angeles, California, September, 2000.

[10] N. J. Bershad and P. L. Feintuch, "A normalized frequency domain LMS adaptive algorithm," *IEEE Trans. Acoust., Speech Signal Processing,* vol. ASSP-34, pp. 452-461, June 1986.

[11] E. R. Ferrara, Jr., " Fast implementation of LMS adaptive filters," *IEEE Trans. Acoust., Speech Signal Processing*, vol. ASSP-28, no. 4, pp. 474-475, Aug. 1980.

[12] J. Bauck and D. H. Cooper, "Generalized transaural stereo and applications," *J. Audio Eng. Soc.* Vol. 44, 683-705, 1996.