A JOINT MINIMAX APPROACH FOR BINAURAL RENDERING OF AUDIO THROUGH LOUDSPEAKERS

Harsha I.K. Rao*, V. John Mathews

Department of Electrical & Computer Engineering University of Utah Salt Lake City, UT 84112, USA

ABSTRACT

This paper describes a method for jointly designing the crosstalk cancellation filters to facilitate binaural rendering of audio through loudspeakers. The minimax criterion is used to design the immersive audio rendering filters having finite impulse responses for a single listener using loudspeakers. The work presented is applied to the traditional Atal-Schroeder crosstalk canceler structure. The minimax approach provides improved low frequency performance and a better overall separation of the direct path and cross path transfer functions than the conventional least-squares designs.

Index Terms— Acoustic signal processing, crosstalk, loudspeakers, minimax methods.

1. INTRODUCTION

A conventional stereo system is known to create a virtual image of a loudspeaker in between the two loudspeakers, thereby creating a stereophonic environment. However, a 3-D audio system can position sounds anywhere around a listener and it has the ability to make the listener perceive that the sounds come from arbitrary points in space [1]. The applications of 3-D audio systems includes multimedia desktop computers, video games, audio and video conferencing, etc.

In this paper, we address the problem of crosstalk cancellation in 3-D audio. The case of the two loudspeaker case is presented here for simplicity. The ideas can be easily extended to the multiple loudspeaker case. The traditional Atal-Schroeder crosstalk canceler [2], [3] is shown in Fig. 1 in which a set of filters in combination with the acoustic paths cancels the crosstalk signals at the ears. In this figure, the left and right signals are \mathbf{p}_L and \mathbf{p}_R respectively, \mathbf{l}_1 and \mathbf{l}_2 are the loudspeaker signals and \mathbf{h}_i (the ipsilateral term) and \mathbf{h}_c (the contralateral term) are the head-related impulse responses (HRIRs) to the same-side and opposite-side ears, respectively. It is essential to maintain sufficient channel separation by ensuring that the left ear signal \mathbf{p}_L goes to the listener's left ear only and similarly the right ear signal \mathbf{p}_R goes to the listener's right ear only. Hence, our objective is to design the filters \mathbf{h}_1 , \mathbf{h}_2 , \mathbf{h}_3 and \mathbf{h}_4 to cancel the crosstalk signals.

The least-squares (LS) optimization technique is widely used to design the crosstalk cancellation filters [3]. However, this may not be an ideal approach as minimizing the average channel separation between the direct path and the cross path over the frequency range of interest may result in the overlap of the direct path and cross path

Young-Cheol Park

Computer & Telecommunication Engineering Division Yonsei University Wonju, Korea



Fig. 1. The Atal-Schroeder crosstalk canceler.

responses. To counter this problem, we consider the use of minimax design criterion in this paper. In our recent work, we presented two approaches based on minimax techniques for designing individual crosstalk cancellation filters [4]. While the design of the crosstalk cancellation filters in the minimax sense ensures significant channel separation, the individual design of the component filters does not result in an optimal design of the overall crosstalk cancellation system. In addition, the ad hoc nature of the selection of the weighting function made the design process cumbersome.

This paper describes a simple and elegant minimax approach to jointly design the crosstalk cancellation filters. The filter design problem is formulated in the time-domain as a joint optimization problem with the filter coefficients being the minimax solution to an over-determined set of linear equations. We make use of secondorder cone programming (SOCP) [5] techniques to derive excellent approximations to the minimax solutions. The performance of the minimax filters have been compared to those of the LS design for the traditional Atal-Schroeder structure [2]. Our designs indicate that the minimax design provides higher channel separation and significantly improves the low frequency performance.

The rest of this paper is organized as follows. The minimax formulation for jointly designing the crosstalk cancellation filters is presented in Section 2. We discuss the use of the SOCP algorithm to solve the minimax problem in Section 3. Simulation results comparing the design techniques are given in Section 4. Finally, we conclude in Section 5. We make use of the assumption of acoustic symmetry in our derivations. In what follows, we have denoted vectors and matrices using bold-faced characters.

^{*}Email: hrao@eng.utah.edu

2. MINIMAX FORMULATION

Consider the traditional Atal-Schroeder crosstalk canceler shown in Fig. 1. The objective is to obtain the filter coefficients $h_1(k)$ and $h_2(k)$, $k = 0, 1, \dots, K - 1$ such the left signal arrives at the left ear with unit gain and it is not reproduced at the right ear. The impulse response between \mathbf{p}_L and the left and right ears, respectively are

$$\hat{a}_1(l) = h_i(m) * h_1(k) + h_c(m) * h_2(k)$$
 (1a)

$$\hat{a}_2(l) = h_c(m) * h_1(k) + h_i(m) * h_2(k)$$
 (1b)

where '*' denotes convolution, $m = 0, 1, \dots, M - 1$ and $l = 0, 1, \dots, M + K - 2$. These equations can be represented in matrix form as follows:

$$\begin{bmatrix} \mathbf{\hat{a}}_1 \\ \mathbf{\hat{a}}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{C}_1 & \mathbf{C}_2 \\ \mathbf{C}_2 & \mathbf{C}_1 \end{bmatrix} \begin{bmatrix} \mathbf{h}_1 \\ \mathbf{h}_2 \end{bmatrix}$$
(2)

or equivalently

$$\hat{\mathbf{a}} = \mathbf{C}\mathbf{h}$$
 (3)

where $\hat{\mathbf{a}}_1 = [\hat{a}_1(0), \dots, \hat{a}_1(M + K - 2)]^T$ is an $(M + K - 1) \ge 1$ element vector and $(\cdot)^T$ denotes matrix transpose. The vector $\hat{\mathbf{a}}_2$ is similarly defined. The crosstalk cancellation filters are represented by the $K \ge 1$ element vectors $\mathbf{h}_n = [h_n(0), h_n(1), \dots, h_n(K-1)], n = 1, 2$. \mathbf{C}_1 is an $(M + K - 1) \ge K$ element matrix defined as

$$\mathbf{C}_{1} = \begin{bmatrix} h_{i}(0) & \mathbf{0} \\ \vdots & \ddots & \\ h_{i}(M-1) & h_{i}(0) \\ & \ddots & \vdots \\ \mathbf{0} & h_{i}(M-1) \end{bmatrix}.$$
 (4)

The matrix C_2 is defined in a similar manner.

We employ a minimax design criterion to solve the over-determined set of linear equations defined in (3). The motivation behind our approach is that by minimizing the maximum deviation of the direct path and the cross path impulse responses from their ideal values, we will be able to maximize the minimum channel separation between the direct and cross paths. We define the cost function to be minimized to obtain the filter coefficients $\mathbf{h} = [\mathbf{h}_1^T, \mathbf{h}_2^T]^T$ as follows:

$$J(\mathbf{h}) = ||\mathbf{W}(\mathbf{a} - \mathbf{C}\mathbf{h})||_{\infty}$$
(5)

The desired impulse response vector $\mathbf{a} = [\mathbf{a}_1^T, \mathbf{a}_2^T]^T$, where \mathbf{a}_1 is a pure delay and \mathbf{a}_2 is a zero vector. Satisfactory results were obtained when the delay was chosen to be K/2 [6]. In (5), $\|(\cdot)\|_{\infty}$ denotes the L_{∞} norm defined as $\|\mathbf{x}\|_{\infty} = \max_i |x_i|$, where x_i is the *i*th element of the vector \mathbf{x} . We also make use of a weighting matrix \mathbf{W} , with the weights located along the diagonal of the matrix. The $2(M + K - 1) \times 2(M + K - 1)$ weighting matrix \mathbf{W} is defined as

$$\mathbf{W} = \begin{bmatrix} \alpha & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 0 & \alpha & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & & \vdots & \vdots & & & \vdots \\ 0 & 0 & \cdots & \alpha & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & \beta & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & \beta & \cdots & 0 \\ \vdots & & & \vdots & \vdots & & & \vdots \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & \beta \end{bmatrix}.$$
(6)

 α and β are positive numbers used to weight the direct path and the cross path impulse responses respectively. To ensure that we are

able to reduce the cross path errors without overly affecting the direct path performance, we chose α to be one and β took values between 10 and 100.

3. SOCP OPTIMIZATION

Second-order cone programming (SOCP) approach is used to obtain a minimax solution to the filter design problem defined in (5). The joint optimization problem can be solved using efficient interior point solvers such as the SeDuMi toolbox of MATLAB [5]. SeDuMi is an add-on for MATLAB and it uses a primal-dual interior point algorithm to solve optimization problems with linear, quadratic and semidefinite constraints [5]. A detailed description about the implementation of the algorithm in SeDuMi can be found in [7]. We reformulate (5) in terms of the required filter coefficients **h** so as to minimize the L_{∞} norm of the error **r** given by

$$r(\mathbf{h}) = \max_{k} W(k,k) |a(k) - \mathbf{C}_k \mathbf{h}|; \ k = 1, \cdots, L$$
(7)

where **h** is the filter coefficient vector of length *N*, C_k is the *k*th row of the HRIR matrix **C**, a(k) is the *k*th element of the desired impulse response vector **a** of length *L* and W(k, k) is the *k*th diagonal element of the weighting matrix **W**.

The minimax optimization problem can be equivalently stated as

 $\min_{\mathbf{h}} \delta$ subject to $W(k,k)|a(k) - \mathbf{C}_k \mathbf{h}| \le \delta; \ k = 1, \cdots, L.$ (8)

The optimization process minimizes the *L* error values by ensuring that none of these *L* values are greater than δ . The *L* second-order cone constraints can be solved to obtain **h** using SeDuMi [5] by rewriting the convex optimization problem (8) in the standard form of dual SOCP, *i.e.*,

max
$$\mathbf{b}^T \mathbf{y}$$
 subject to $\mathbf{c} - \mathbf{A}^T \mathbf{y} \in Q_{cone_1^q} \mathbf{x} Q_{cone_2^q} \mathbf{x} \cdots \mathbf{x} Q_{cone_L^q}$
(9)
where $\mathbf{y} = [\delta, \mathbf{h}^T]^T$, $\mathbf{b} = [-1, \mathbf{0}_{1 \times N}]^T$, such that $\delta = -\mathbf{b}^T \mathbf{y}$,

$$\mathbf{c} = \begin{bmatrix} 0 \\ W(k,k)a(k) \end{bmatrix}, \mathbf{A} = \begin{bmatrix} -1 & \mathbf{0}_{1 \times N} \\ 0 & W(k,k)\mathbf{C}_k \end{bmatrix}$$
(10)

and $Q_{cone_i^q}$ is the *i*th symmetric cone of dimension q = 2, that can be defined as

$$Q_{cone_i^q} \equiv \left(\begin{bmatrix} x_1 \\ \mathbf{x_2} \end{bmatrix} : x_1 \in \Re, \mathbf{x_2} \in C^{q-1}, x_1 \ge \|\mathbf{x_2}\| \right) .$$
(11)

In (11), \Re refers to the set of real numbers and C^q refers to a secondorder cone of dimension q. Thus, we have outlined a procedure to efficiently obtain a minimax solution to an over-determined system of linear equations of the form **W C h** = **W a**, with **h** being the unknown parameter to be solved.

4. RESULTS AND DISCUSSION

The HRIRs used in our simulations were taken from the extensive set of HRIRs measured at the University of California, CIPIC interface laboratory [8]. The HRIRs are 200 taps long, sampled at 44.1 kHz and are available for 25 different azimuths and 50 different elevations for 45 different subjects (inclusive of 2 KEMARs). The performance of the crosstalk canceler was evaluated using measures of channel separation, cross path values and the deviation of the direct path from 0 dB. The parameters were measured in the audio frequency range 20 Hz to 20 kHz. Channel separation is defined as

$$C(\omega) = 20\log_{10}|T(\omega)| - 20\log_{10}|S(\omega)|$$
(12)

where $T(\omega)$ is the direct path transfer function and $S(\omega)$ is the cross path transfer function. We will present results for elevation and azimuth angles corresponding to the stereo dipole $(0^{\circ}, 5^{\circ})$ and the ITU setup $(0^{\circ}, 30^{\circ})$ [9]. The performance of the crosstalk canceler was evaluated using 256-tap and 128-tap FIR filters.

We consider the performance of the 256-tap FIR filters first. Figures 2 and 3 show the magnitude responses of the direct path and cross path for elevation and azimuth angles of $(0^{\circ}, 5^{\circ})$ and $(0^{\circ}, 30^{\circ})$, respectively. The performance metrics are tabulated in Ta-



Fig. 2. Direct path and cross path magnitude responses for a 256-tap crosstalk canceler, elevation = 0° and azimuth = 5° (thick lines - minimax, thin lines - LS).



Fig. 3. Direct path and cross path magnitude responses for a 256-tap crosstalk canceler, elevation = 0° and azimuth = 30° (thick lines - minimax, thin lines - LS).

ble 1. The 256-tap minimax crosstalk cancellation system offers higher minimum channel separation than an equivalent LS system. It also exhibits lower levels of maximum crosstalk values and maximum value of the deviation of the direct path response from 0 dB. The minimax system provides higher channel separation with absolutely no overlap of the direct path and cross path responses over the entire frequency range of interest. In contrast, the direct and cross path frequency responses can overlap in the LS design. In the low frequency region, as a consequence of the minimax filters offering lower crosstalk, the channel separation is much higher than what we can obtain using the LS filters. Despite the fact that the minimax cost function only minimizes the maximum errors, the crosstalk cancellation systems designed using 256-tap minimax FIR filters provide higher average channel separation and also higher average crosstalk cancellation over the frequency range of interest when compared with a corresponding LS system.

For real-time applications of the crosstalk cancelers, it might be desirable to use lower-order FIR filters. However, the use of lower-order crosstalk cancellation filters will result in poorer performance. The minimax crosstalk cancellation system was designed using 128-tap FIR filters and its performance was compared with an equivalent LS system. The magnitude responses of the direct path and the cross path for $(0^{\circ}, 5^{\circ})$ and $(0^{\circ}, 30^{\circ})$ shown in Fig. 4 and Fig. 5, respectively indicate that the 128-tap minimax crosstalk cancellation system provides comparatively higher minimum channel separation. The performance metrics are tabulated in Table 2. Simulation results have shown that the direct path and the cross



Fig. 4. Direct path and cross path magnitude responses for a 128-tap crosstalk canceler, elevation = 0° and azimuth = 5° (thick lines - minimax, thin lines - LS).



Fig. 5. Direct path and cross path magnitude responses for a 128-tap crosstalk canceler, elevation = 0° and azimuth = 30° (thick lines - minimax, thin lines - LS).

path responses can slightly overlap for the 128-coefficient minimax filters. But such overlaps are of lower magnitudes and occur over smaller frequency ranges than the overlaps that occurs for the LS filters. Though the maximum cross path values obtained using the two algorithms are comparable, the minimax filters exhibits considerably lower values of the maximum deviation of the direct path response

(Elevation,	Minimum		Maximum		Maximum		Average		Average		Average	
Azimuth)	channel		cross	cross path direct path		t path	channel		cross path		direct path	
	separation (dB)		values (dB)		deviation (dB)		separation (dB)		values (dB)		deviation (dB)	
	Minimax	LS	Minimax	LS	Minimax	LS	Minimax	LS	Minimax	LS	Minimax	LS
$(0^{o},5^{o})$	13.138	10.037	-32.491	-27.524	20.729	26.839	53.382	49.464	-53.824	-50.070	1.527	0.875
$(0^{o}, 30^{o})$	22.844	-3.276	-34.707	-28.708	18.881	38.687	51.710	49.282	-51.963	-49.764	1.214	0.686

Table 1. Performance metrics for a 256-tap crosstalk canceler

	Fable 2	2. Performance	metrics for a	a 128-tap	crosstalk	cancele
--	---------	----------------	---------------	-----------	-----------	---------

(Elevation,	Minimum		Maximum		Maximum		Average		Average		Average		
Azimuth)	channel		cross path		direct path		channel		cross path		direct path		
	separation (dB)		values (dB) d		deviatio	deviation (dB)		separation (dB)		values (dB)		deviation (dB)	
	Minimax	LS	Minimax	LS	Minimax	LS	Minimax	LS	Minimax	LS	Minimax	LS	
$(0^{o},5^{o})$	9.552	-9.803	-21.122	-21.873	17.324	36.834	33.953	35.670	-34.118	-37.235	2.213	1.956	
$(0^{o}, 30^{o})$	4.963	-24.020	-23.439	-23.524	18.476	51.321	39.655	36.132	-39.932	-37.408	2.544	1.719	

from 0 dB. Consequently, unlike the LS systems, the minimax systems provide higher channel separation, especially at low frequencies. We also observe that the average performance of the minimax filters over the frequency range of interest is comparable with that of the LS filters. Since the LS designs attempts to minimize the average squared deviation from the ideal solution, one should expect the performance metrics that measures average characteristics over the frequency range to be comparable or slightly better for the LS design than the minimax design, as is the case here. However, the results also indicate that the performance of the LS designs are unacceptably poor at some frequencies, justifying our preference for the minimax design.

To conclude our analysis of the performance of the minimax filters, we would like to illustrate the importance of the weighting matrix **W**. The weighting function provides us with the flexibility to improve the cross path performance without significantly distorting the direct path behavior. The minimax algorithm was implemented using β values of 1 and 70 and the magnitude responses of the direct and cross paths are shown in Fig. 6. Though the use of weights



Fig. 6. Direct path and cross path magnitude responses for a 256-tap crosstalk canceler, elevation = 0° , azimuth = 30° (thick lines - minimax with weights, thin lines - minimax without weights).

lowers the direct path performance, there is a tremendous improvement in the cross path behavior leading to a significant increase in the channel separation.

5. CONCLUSIONS

We presented a novel minimax method to jointly design the crosstalk cancellation filters for the Atal-Schroeder crosstalk canceler. The algorithm is easy to formulate, efficient in implementation and it can achieve excellent channel separation, especially at low frequencies. These features makes the joint minimax design an attractive solution to the crosstalk cancellation problem.

6. REFERENCES

- W. G. Gardner, *3-D Audio Using Loudspeakers*, Kluwer Academic, Norwell, MA, 1997.
- [2] B.S. Atal and M. R. Schroeder, "Apparent sound source translator," U.S. Patent 3 236 949, Feb. 1966.
- [3] D. B. Ward, "Joint least squares optimization for robust acoustic crosstalk cancellation," *IEEE Trans. on Speech and Audio Processing*, vol. 8, no. 2, pp. 211–215, Feb. 2000.
- [4] H. I. K. Rao, V. J. Mathews, and Y.-C. Park, "Inverse filter design using minimax approximation techniques for 3-D audio," in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, Toulouse, France, May 2006, vol. 5, pp. 353–356.
- [5] J. F. Strum, "Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones," *Optimization Methods Software*, vol. 11-12, pp. 625–653, Aug. 1999.
- [6] P. A. Nelson, H. Hamada, and S. J. Elliot, "Adaptive inverse filters for stereophonic sound reproduction," *IEEE Trans. on Signal Processing*, vol. 40, no. 7, pp. 1621–1632, July 1992.
- [7] J. F. Sturm, "Implementation of interior point methods for mixed semidefinite and second order cone optimization problems," *Optimization Methods and Software*, vol. 17, no. 6, pp. 1105–1154, Dec. 2002.
- [8] V. R Algazi, R. O Duda, D. M Thompson, and C. Avendano, "The CIPIC HRTF database," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, NY, Oct. 2001, pp. 99–102.
- [9] O. Kirkeby, P. A. Nelson, and H. Hamada, "The stereo dipole -A virtual source imaging system using two closely spaced loudspeakers," *Journal of Audio Engineering Society*, vol. 46, no. 5, pp. 387–395, May 1998.