A Minimax Approach for the Joint Design of Acoustic Crosstalk Cancellation Filters

Harsha I. K. Rao, *Student Member, IEEE*, V. John Mathews, *Fellow, IEEE*, and Young-Cheol Park, *Member, IEEE*

***Abstract*—This paper presents a method for jointly designing immersive audio rendering filters for a single listener using loudspeakers. The filters for crosstalk cancellation are assumed to have finite impulse responses and are designed using the minimax criterion. In addition to the traditional Atal–Schroeder crosstalk canceler structure, this paper explores an alternate topology that requires the approximation of a single filter. In general, the minimax approach provides improved low-frequency performance leading to a better overall separation of the direct-path and cross-path transfer functions than least-squares designs. The performance of the single-filter structure is better than that of the traditional crosstalk cancellation structure.**

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

I. Introduction

A 3-D AUDIO system can be used to position sounds around a listener and make the listener perceive that the sounds come from arbitrary points in space [1]–[4]. Unlike a conventional stereo system which can only create a virtual image of a loudspeaker in between the two loudspeakers, thereby creating a stereophonic environment, a 3-D audio system can position sounds anywhere around a listener. Consequently, such systems have applications in a variety of areas including multimedia desktop computers, video games, audio, video conferencing, etc.

This paper addresses the problem of crosstalk cancellation in 3-D audio. It is important to maintain sufficient separation between the left and right channels by ensuring that the left ear signal goes to the listener's left ear only and similarly the right ear signal goes to the listener's right ear only. We consider the case of two loudspeakers and one listener for simplicity of analysis and presentation even though generalizations to higher number of filters are possible. Fig. 1 shows the traditional Atal–Schroeder crosstalk canceler [5], [6] 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, l_1 and l_2 are the loudspeaker signals, and $\mathbf{h}_i$ (the ipsilateral term) and $\mathbf{h}_c$ (the contralateral

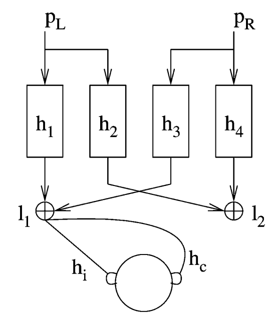


Fig. 1. Atal–Schroeder crosstalk canceler.

term) are the head-related impulse responses (HRIRs) to the same-side and opposite-side ears, respectively. The objective is to design the filters $\mathbf{h}_1$, $\mathbf{h}_2$, $\mathbf{h}_3$, and $\mathbf{h}_4$ such that the crosstalk signals are canceled, i.e., none of the $\mathbf{p}_L$ signal is received at the right ear and similarly none of the $\mathbf{p}_R$ signal is received at the left ear. Assuming that the system is acoustically symmetric so that the transfer functions from the left loudspeaker to the ears are the same as those for the right loudspeaker, it can be shown that the required filter responses in terms of the head-related transfer functions (HRTFs), $H_i(\omega)$ and $H_c(\omega)$ are given by

$$\begin{aligned} H_1(\omega) = H_4(\omega) &= \frac{H_i(\omega)}{H_i^2(\omega) - H_c^2(\omega)} \\ H_2(\omega) = H_3(\omega) &= \frac{-H_c(\omega)}{H_i^2(\omega) - H_c^2(\omega)}. \end{aligned} \tag{1}$$

In the above expressions, the HRTFs $H_i(\omega)$ and $H_c(\omega)$ are the Fourier transforms of the the HRIRs $\mathbf{h}_i$ and $\mathbf{h}_c$, respectively. For the derivations in this paper, we will assume that the HRIRs are time-invariant. Equation (1) can be modified appropriately if the listener is in an asymmetric position with respect to the loudspeakers.

An alternative topology for implementing the crosstalk cancellation filters was suggested in [7]. The frequency responses of the four crosstalk canceler filters in (1) can be expressed in matrix form as

$$\mathbf{H}_d(\omega) = \frac{1}{H_i^2(\omega) - H_c^2(\omega)} \begin{bmatrix} H_i(\omega) & -H_c(\omega) \\ -H_c(\omega) & H_i(\omega) \end{bmatrix}. \tag{2}$$

The above representation results in the realization of the crosstalk cancellation filters shown in Fig. 2. This realization requires the design of only one filter for its implementation. The desired frequency response of the single filter is given by

$$H(\omega) = \frac{1}{H_i^2(\omega) - H_c^2(\omega)}. \tag{3}$$

Manuscript received February 12, 2007; revised June 13, 2007. This paper was presented in part at the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Honolulu, HI, April 2007. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Hiroshi Sawada.

H. I. K. Rao and V. J. Mathews are with the Department of Electrical and Computer Engineering, University of Utah, Salt Lake City UT 84112 USA (e-mail: hrao@eng.utah.edu; mathews@ece.utah.edu).

Y.-C. Park is with the Computer and Telecommunication Engineering Division, Yonsei University, Wonju, Korea (e-mail: young00@yonsei.ac.kr).

Digital Object Identifier 10.1109/TASL.2007.905149

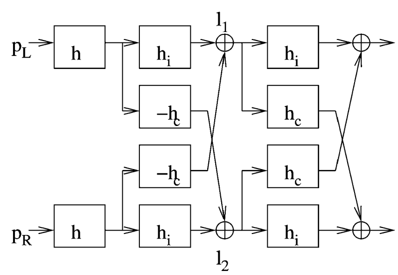


Fig. 2. Single filter structure for crosstalk cancellation.

It is straightforward to see that if the HRTFs are known, choosing $H(\omega)$ as in (3) allows complete cancellation of the crosstalk as well as preservation of the left and right signals without distortion at their destinations.

The general problem of inverse filtering of room acoustics when multiple loudspeakers are employed is addressed in [8]. Numerical issues associated with this problem was investigated in [9]. Several adaptive algorithms to implement the crosstalk cancellation filters are available in the literature [10]–[16]. They typically employ some variations of least-mean-squares (LMS) or recursive least-squares (RLS) adaptation algorithms. Time-domain and frequency-domain algorithms based on regularization principles are also available [17]–[19]. The most common way to design the crosstalk cancellation filters employs the least-squares (LS) optimization technique [6]. However, minimizing the squared difference between the desired and actual impulse responses may not be ideal, and such designs may even result in a cross path magnitude response that is larger than the direct path magnitude response in some frequency range [6]. To counter this problem, we consider the use of a minimax design criterion in this paper.

In our recent work, we presented two approaches for designing individual crosstalk cancellation filters separately using minimax techniques [20]. While the design of the crosstalk cancellation filters using the complex Chebyshev algorithm [21] and second-order cone programming (SOCP) techniques [22] ensures significant channel separation, designing the component filters independently of the others does not result in an optimal design of the overall crosstalk cancellation system. Furthermore, the algorithm required a weighting function for the design of the filters and the choice of this weighting function played a critical role in its performance. The ad hoc and trial-and-error nature of the selection of the weighting function made the design process cumbersome.

In this paper, we present a simple and elegant minimax solution to the joint design of the crosstalk cancellation filters. The filter design for the traditional crosstalk cancellation structure can be formulated in the time-domain as an optimization problem in which the filter coefficients are the minimax solution to an over-determined set of linear equations. With suitable modification, we can also obtain the minimax solution to the single-filter design problem. Our approach utilizes the SOCP [22] techniques and it provides an excellent approximation to the minimax solution. The design problem is formulated as a convex optimization problem and solved using efficient interior point solvers [23]. We have compared the capabilities of the minimax design to those of the LS design for both the Atal–Schroeder structure and the single-filter structure given in [7]. We will present the results of a large number of designs that demonstrate that the minimax design offers higher channel separation, particularly at low frequencies, than the LS method. Our designs also indicate that the minimax formulation of the single-filter structure of [7] provides a superior solution to the crosstalk cancellation problem when compared to the minimax solution for the conventional structure.

The rest of this paper is organized as follows. In Section II, we formulate the minimax problem for designing the Atal–Schroeder crosstalk canceler and the single filter in the alternate topology. We discuss the SOCP algorithm used to design the crosstalk cancellation filters in Section III. Simulation results using experimentally measured room acoustics data comparing the design techniques and the two structures are given in Section IV. Finally, our conclusions are provided in Section V. Throughout the paper, we have denoted vectors and matrices using boldfaced characters.

II. Problem Formulation

A. Conventional Crosstalk Canceler

The traditional Atal–Schroeder crosstalk canceler is shown in Fig. 1. To simplify our analysis, we make use of the assumption of acoustic symmetry and also consider only the reproduction of the left signal $\mathbf{p}_L$ at the left ear. The goal is to obtain the filter coefficients $h_1(k)$, $k = 0, 1, \ldots, K-1$ and $h_2(k)$, $k = 0, 1, \ldots, K-1$ 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

$$\hat{a}_1(k) = h_i(k) * h_1(k) + h_c(k) * h_2(k) \tag{4a}$$

$$\hat{a}_2(k) = h_c(k) * h_1(k) + h_i(k) * h_2(k) \tag{4b}$$

where * denotes convolution. These equations can be put together in matrix form as

$$\begin{bmatrix} \hat{\mathbf{a}}_1 \\ \hat{\mathbf{a}}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{C}_1 & \mathbf{C}_2 \\ \mathbf{C}_2 & \mathbf{C}_1 \end{bmatrix} \begin{bmatrix} \mathbf{h}_1 \\ \mathbf{h}_2 \end{bmatrix} \tag{5}$$

or equivalently

$$\hat{\mathbf{a}} = \mathbf{C}\mathbf{h} \tag{6}$$

where $\hat{\mathbf{a}}_1 = [\hat{a}_1(0), \ldots, \hat{a}_1(M+K-2)]^T$ is an $(M+K-1) \times 1$ element vector, and $(\cdot)^T$ denotes matrix transpose. The vector $\hat{\mathbf{a}}_2$ is similarly defined. The matrix $\mathbf{C}_1$ is an $(M+K-1) \times K$ element matrix defined as

$$\mathbf{C}_1 = \begin{bmatrix} h_i(0) & & \mathbf{0} \\ \vdots & \ddots & \\ h_i(M-1) & & h_i(0) \\ & \ddots & \vdots \\ \mathbf{0} & & h_i(M-1) \end{bmatrix}. \tag{7}$$

We can define the matrix $\mathbf{C}_2$ in a similar manner. The vectors $\mathbf{h}_1$, $\mathbf{h}_2$ are the K-coefficient filters to be designed and are defined as $\mathbf{h}_n = [h_n(0), \ldots, h_n(K-1)]^T$, $n = 1, 2$.

We employ a minimax design criterion to obtain the filter coefficients $\mathbf{h}_1$ and $\mathbf{h}_2$. The approach involves minimizing the maximum weighted error between the desired impulse responses and the actual impulse responses. The desired impulse response vector $\mathbf{a} = \left[\mathbf{a}_1^T, \mathbf{a}_2^T\right]^T$, where $\mathbf{a}_1$ is a pure delay and $\mathbf{a}_2$ is a zero vector. On extensive experimentation, satisfactory results were obtained when the delay was chosen to be $K/2$. This particular value of the delay is also commonly used in the literature [11], [24]. While implementing practical crosstalk cancellation systems, it might be necessary to use an attainable desired direct path response that takes into account the actual loudspeaker response [6]. The overall impulse response vector $\hat{\mathbf{a}}$ is obtained upon convolution of the crosstalk cancellation filters with the acoustic path as described by (4a) and (4b). The motivation behind our approach is that by minimizing the joint maximum deviation of the direct path and the cross path impulse responses from their ideal values, we may be able to maximize the minimum channel separation between the direct and cross paths.

The cost function to be minimized to obtain the filter coefficients $\mathbf{h} = \left[\mathbf{h}_1^T, \mathbf{h}_2^T\right]^T$ is

$$J_c(\mathbf{h}) = \|\mathbf{W}(\mathbf{a} - \mathbf{C}\mathbf{h})\|_\infty \tag{8}$$

where $\|(\cdot)\|_\infty$ denotes the L_∞ norm defined as $\|\mathbf{x}\|_\infty = \max_i | x_i |$, where x_i is the ith element of the vector $\mathbf{x}$. We also suggest the 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

$$\mathbf{W} = \begin{bmatrix} \alpha\mathbf{I} & \mathbf{0} \\ \mathbf{0} & \beta\mathbf{I} \end{bmatrix} \tag{9}$$

where $\mathbf{I}$ is an identity matrix of dimension $(M+K-1) \times (M+K-1)$. The parameters α and β are positive numbers used to weight the direct path and the cross path impulse responses, respectively. Different choices of these parameters allow the designer to emphasize the cross path or the direct path behavior. In all the results presented later, we chose α to be one and $\beta > \alpha$ so as to reduce the cross path errors and increase the minimum channel separation. However, a large value of β can distort the direct path response significantly and adversely affect the minimum channel separation. In this paper, β was chosen such that the channel separation was maximum subject to the constraint that after the crosstalk cancellation was applied, the deviation of the direct path frequency response from its ideal value of one (0 dB) was less than a specified maximum value. A similar weighting matrix can also be applied to the LS approach to control the direct path and the cross path errors.

B. Single-Filter Crosstalk Canceler

The filter design problem formulated in Section II-A for the traditional crosstalk canceler can be suitably modified for the single-filter case. On comparing the single-filter structure described by (2) and shown in Fig. 2 with the conventional realization of Fig. 1, it is apparent that we can represent the two crosstalk cancellation filters $\mathbf{h}_1$ and $\mathbf{h}_2$ as follows:

$$h_1(k) = h(k) * h_i(k) \tag{10a}$$

$$h_2(k) = -h(k) * h_c(k) \tag{10b}$$

where $\mathbf{h} = [h(0), \ldots, h(K-1)]^T$ represent the coefficients of the single crosstalk filter. Using these equations, we can simplify (4a) and (4b) and put them together in a matrix as follows:

$$\begin{bmatrix} \hat{\mathbf{a}}_1 \\ \hat{\mathbf{a}}_2 \end{bmatrix} = \begin{bmatrix} \tilde{\mathbf{C}}_1 \\ \tilde{\mathbf{C}}_2 \end{bmatrix} [\mathbf{h}]. \tag{11}$$

We can see that (11) is of the form $\hat{\mathbf{a}} = \tilde{\mathbf{C}}\,\mathbf{h}$, where $\hat{\mathbf{a}}_1 = [\hat{a}_1(0), \ldots, \hat{a}_1(2M+K-3)]^T$ is an $(2M+K-2) \times 1$ element vector, and $\hat{\mathbf{a}}_2$ is similarly defined

$$\tilde{\mathbf{C}}_1 = \begin{bmatrix} h_d(0) & & \mathbf{0} \\ \vdots & \ddots & \\ h_d(2M-2) & & h_d(0) \\ & \ddots & \vdots \\ \mathbf{0} & & h_d(2M-2) \end{bmatrix} \tag{12}$$

is a $(2M+K-2) \times K$ element matrix with $\mathbf{h}_d = \mathbf{h}_i^*\mathbf{h}_i - \mathbf{h}_c^*\mathbf{h}_c$. $\tilde{\mathbf{C}}_2$ is a zero matrix of dimension $(2M+K-2) \times K$, and $\mathbf{h}$ is the K-coefficient single filter to be designed. Similar to the design of conventional filters, we minimize the maximum error between the desired impulse responses and the actual impulse responses. The desired impulse response vector is given by $\mathbf{a} = \left[\mathbf{a}_1^T, \mathbf{a}_2^T\right]^T$, with $\mathbf{a}_1$ being a pure delay and $\mathbf{a}_2$ being a zero vector. Since all the elements in $\mathbf{a}_2$ and $\tilde{\mathbf{C}}_2$ are zero, the optimization process boils down to minimizing the maximum direct path error to obtain the single-filter coefficients $\mathbf{h}$. As a result of the complete absence of the cross path impulse response, we shall also not be introducing any weighting matrix in the problem formulation. The cost function for the single-filter design is defined as

$$J_s(\mathbf{h}) = \|(\mathbf{a} - \tilde{\mathbf{C}}h)\|_\infty = \|(\mathbf{a}_1 - \tilde{\mathbf{C}}_1 h)\|_\infty. \tag{13}$$

III. Minimax Optimization

We use the SOCP approach to jointly design the crosstalk cancellation filters in the minimax sense. The SOCP technique has been used to design minimax infinite-impulse response (IIR) filters [25] and recently, Yan *et al.* [22] have provided a unified framework for the use of the SOCP technique to design FIR filters using various norms. The filter coefficients can be obtained as the minimax solution to the over-determined system of linear equations described in (6) for the two-filter realization and in (11) for the single-filter topology. SOCP provides us with a flexible structure to formulate the FIR filter design problem, and the optimization problem can be solved using interior point solvers such as those included in the Self-Dual-Minimization (SeDuMi) toolbox of MATLAB [23].

The procedure for both topologies are identical. To understand how the SOCP approach works, we reformulate (8) in terms of the filter coefficients $\mathbf{h}$ so as to minimize the L_∞ norm of the error given by

$$r(\mathbf{h}) = \max_k W(k,k)|a(k) - \mathbf{C}_k\mathbf{h}|; \quad k = 1, \ldots, L \tag{14}$$

where $W(k,k)$ is the kth diagonal element of the weighting matrix $\mathbf{W}$, $a(k)$ is the kth element of the desired impulse response vector $\mathbf{a}$, and $\mathbf{C}_k$ is the kth row of $\mathbf{C}$. In (14), L represents the length of $\mathbf{a}$.

The minimax optimization problem can be equivalently stated as

$$\begin{aligned} \min_{\mathbf{h}} \quad & \delta \\ \text{subject to} \quad & W(k,k)|a(k) - \mathbf{C}_k\mathbf{h}| \le \delta; \\ & k = 1, \ldots, L. \end{aligned} \tag{15}$$

This particular formulation solves for the coefficient vector $\mathbf{h}$ by ensuring that none of the weighted error values are greater than δ. We employed the SeDuMi toolbox to solve this problem. The key step involved in solving the L second-order cone constraints for $\mathbf{h}$ using SeDuMi lies in rewriting the convex optimization problem (15) in the standard form of a dual SOCP (given by (8) in [23]). To do so, we redefine the conic problem as

$$\begin{aligned} \max \quad & \mathbf{b}^T\mathbf{y}, \\ \text{subject to} \quad & \mathbf{c} - \mathbf{A}^T\mathbf{y} \in Q_{\text{cone}_1^q} \times Q_{\text{cone}_2^q} \\ & \times \cdots \times Q_{\text{cone}_L^q} \end{aligned} \tag{16}$$

where $\mathbf{y} = [\delta, \mathbf{h}^T]^T$, $\mathbf{b} = [-1, 0_{1\times N}]^T$, N represents the length of $\mathbf{h}$, such that $\delta = -\mathbf{b}^T y$

$$\mathbf{c} = \begin{bmatrix} 0 \\ W(k,k)a(k) \end{bmatrix} \qquad \mathbf{A}^T = \begin{bmatrix} -1 & \mathbf{0}_{1\times N} \\ 0 & W(k,k)\mathbf{C}_k \end{bmatrix} \tag{17}$$

and $Q_{\text{cone}_k^q}$ is the kth symmetric cone of dimension $q = 2$, that is defined as

$$Q_{\text{cone}_k^q} \equiv \left(\begin{bmatrix} x_1 \\ \mathbf{x_2} \end{bmatrix} : x_1 \in \Re, \mathbf{x_2} \in C^{q-1}, x_1 \ge \|\mathbf{x_2}\| \right). \tag{18}$$

In the above equations, $\Re$ refers to the set of real numbers, and C^q refers to a second-order cone of dimension q.

SeDuMi is an add-on for MATLAB, which can solve optimization problems with linear, quadratic, and semidefinite constraints. A primal-dual interior point algorithm is implemented in SeDuMi [23]. The algorithm has a proven $O(\sqrt{n}\log(1/\epsilon))$ worst case iteration bound, where n is the size of the problem, and ϵ is the stopping criterion. The algorithm treats initialization issues by means of the self-dual embedding technique over a homogeneous model [26]. The idea is to embed the optimization problem into a slightly larger problem which always has a solution [27]. An appropriate solution to the embedded problem either provides a certificate of infeasibility or an optimal solution to the original problem. A detailed exposition about the implementation of the algorithm in SeDuMi can be found in [28].

IV. Performance Evaluation

Head-related transfer functions provide a measure of the transformation of sound from a point in space to the ear canal [1], [3]. Each HRTF contains information about the time delay, amplitude, and tonal transformation of the signal from a specific location of the source to a specific ear. Three-dimensional audio systems usually use nonindividualized HRTFs, i.e., HRTFs measured from manikins. This is because the measurement of HRTFs is usually a complicated procedure and it is not practical to obtain the HRTFs for each and every individual. The HRTFs used in our simulations were taken from the extensive set of HRTFs measured at the University of California, CIPIC Interface Laboratory [29], [30]. The HRTFs are available for 45 different subjects including two Knowles Electronic Manikins for Acoustic Research (KEMARs) (one with large pinna and another with small pinna). The HRIRs are 200 taps long, sampled at 44.1 kHz and are available for 25 different azimuths and 50 different elevations for each subject. The HRIRs have been compensated for the response of the loudspeakers, the details of this procedure are explained in [29].

A. *Performance Metrics*

We use measures of equalization and cancellation as the metrics for performance evaluation in this work. Recall that, assuming symmetry, the frequency response of the direct path is

$$T(\omega) = H_1(\omega)H_i(\omega) + H_2(\omega)H_c(\omega) \tag{19}$$

and the frequency response of the cross path is given by

$$S(\omega) = H_1(\omega)H_c(\omega) + H_2(\omega)H_i(\omega). \tag{20}$$

Ideally, $T(\omega) = 1$ and $S(\omega) = 0$.

For the conventional design, the parameters that were used to measure the performance were the minimum channel separation, average channel separation, maximum cross path values, and maximum deviation of the direct path from 0 dB. Channel separation is defined as

$$R(\omega) = 20\log_{10}|T(\omega)| - 20\log_{10}|S(\omega)|. \tag{21}$$

The maximum cross path value is given by the maximum of $20\log_{10}|S(\omega)|$ and the maximum deviation of the direct path from 0 dB, given by the maximum of $|20\log_{10}|T(\omega)||$. All the parameters were measured in the frequency range from 50 Hz to 20 kHz. Since our design criterion involves minimizing the maximum error of the direct path and the cross path from their respective desired impulse responses, we intuitively expect the minimax method to maximize the minimum channel separation. We also look at the average performance of the system over the entire frequency range of interest. Hence, we use the average channel separation to be one of the performance metrics in our evaluation.

For the single-filter case, the channel separation as defined in (21) will be a constant, irrespective of the choice of the single filter. This can be verified by substituting $H_1(\omega) = H(\omega)\,\hat{H}_i(\omega)$ and $H_2(\omega) = -H(\omega)\,\hat{H}_c(\omega)$ in (19) and (20). $H(\omega)$ is the desired frequency response of the single filter, and $\hat{H}_i(\omega)$ and $\hat{H}_c(\omega)$ are the estimated ipsilateral and

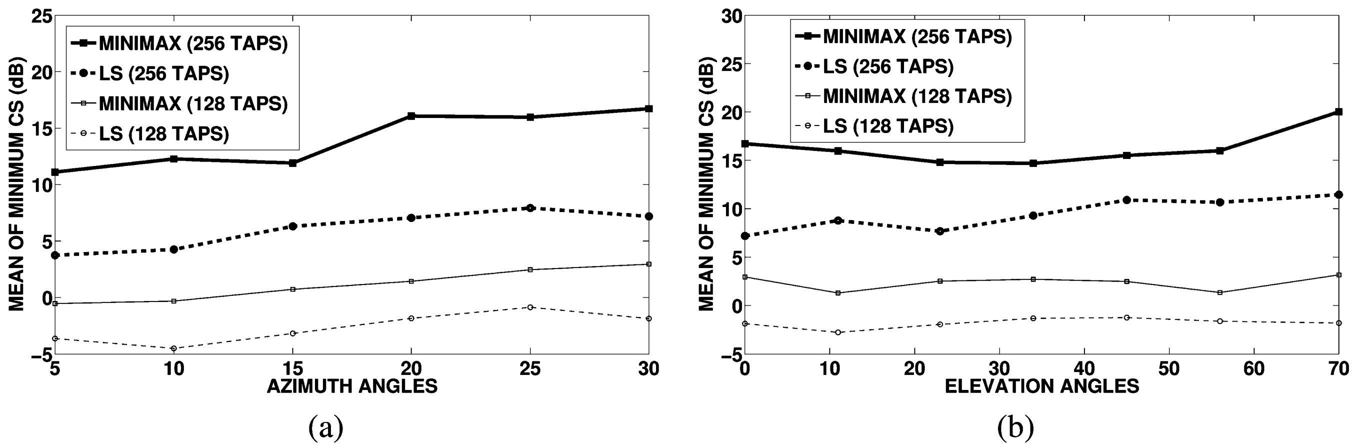


Fig. 3. (a) Variation of the minimum channel separation for elevation angle = 0°. (b) Variation of the minimum channel separation for azimuth angle = 30°. (Thick lines—256 taps, thin lines—128 taps, solid lines—minimax, dashed lines—LS).

contralateral HRTFs, respectively. Given that the cost function primarily results in the minimization of the direct path error, the best measure for comparing the performance of the minimax and LS algorithms for the single-filter topology appears to the maximum deviation of the direct path from 0 dB.

The design of all the crosstalk canceling filters attempted to constrain the maximum direct path deviation to 25 dB. For this purpose, we designed each set of crosstalk cancellation systems using ten different values of β uniformly distributed in the range [10,100]. The design corresponding to the β value that resulted in the maximum channel separation and met the constraint on direct path deviation was selected as the solution in this paper. If the constraint was not met by any of the designs, the constraint was relaxed in steps of 10 dB, until it was achieved, and the resulting system was deemed the solution to the design problem. The β values were independently selected for the minimax algorithm as well as the LS algorithm according to the procedure outlined above. This ensures that we will be comparing the best possible minimax design with the best possible LS solution.

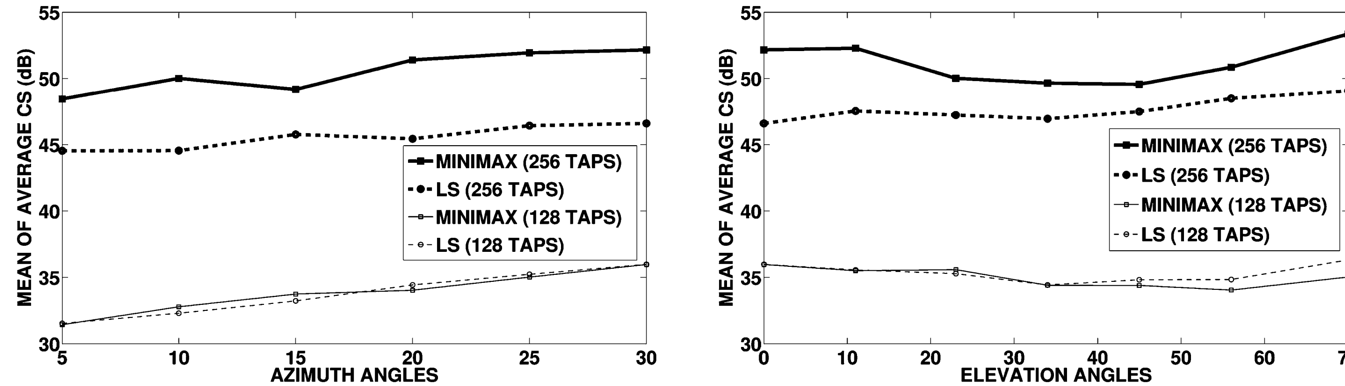
B. Results and Discussion

We have done extensive performance evaluation for both the conventional crosstalk canceler and the single-filter structure using the HRIRs from the CIPIC database [30]. The HRIRs available for seven different angles of elevation uniformly spaced between 0° and 67.5° and six different angles of azimuth uniformly spaced between 5° and 30° were used to design the crosstalk cancellation filters for each of the 45 subjects. As a result, filters were designed for a total of 1890 crosstalk cancellation systems. For each case, 256-tap and 128-tap filters were designed for the analysis. For a fixed angle of elevation, the mean values of the performance metrics over all the subjects were calculated for various angles of azimuth. Similarly, for a fixed angle of azimuth, the mean values of the performance metrics over all the subjects were calculated for various angles of elevations.

Several authors [1] have suggested that the speakers should be mounted on opposite sides of the head at high elevations in a crosstalk cancellation system, so that the HRTFs are relatively flat. The HRTFs at lower elevations contain notches which can make inverse responses ill-conditioned. However, placing the loudspeakers at high elevations may not be a practical solution. The results presented in this paper include the magnitude responses of the direct path and cross path for the commonly used configuration of a stereo dipole whose elevation and azimuth angles are 0° and 5°, respectively [31]. In addition to the above results, we will also present histograms of the performance metrics over all the 1890 crosstalk cancellation systems comparing the two algorithms for each type of design.

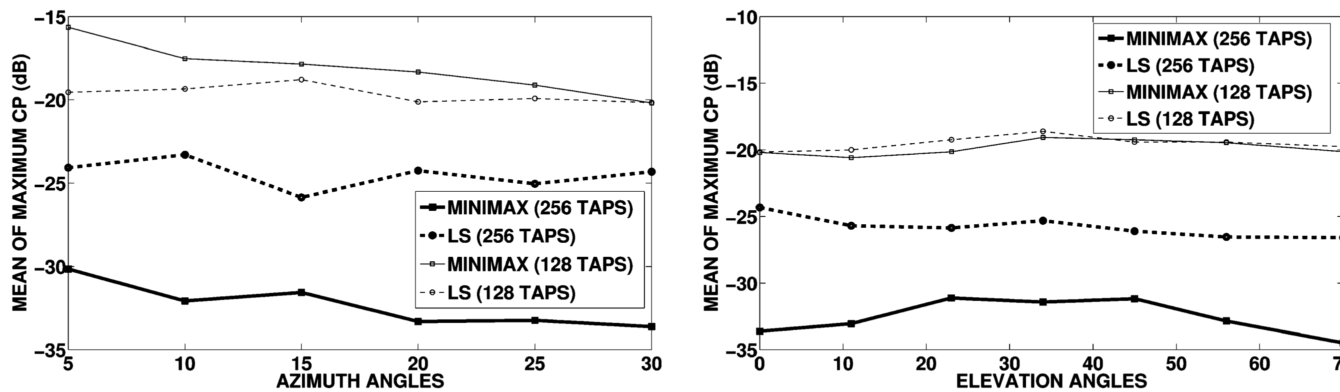
Fig. 3(a) shows the mean values of the minimum channel separation over all the 45 subjects calculated for various angles of azimuth and for a fixed angle of elevation. Similarly, for a fixed angle of azimuth, the mean values of the minimum channel separation over all the 45 subjects were calculated for various angles of elevation, and the plot is depicted in Fig. 3(b). We have chosen the fixed angle of elevation to be 0° and the fixed angle of azimuth to be 30° for our plots. The overall behavior of the metrics for other elevation angles and azimuth angles were similar to those shown in this paper. The results indicate that on average, the minimax method offers 5–10 dB or more minimum channel separation than the LS technique for the 256-coefficient design. For the 128-coefficient design, the performance improvement of the minimax design technique is slightly lower, but still significant. Furthermore, there is significant overlap of the direct path with the cross path responses using the LS method for this case. There are cases when the direct path and the cross path responses slightly overlap for the 128-coefficient minimax filters, but such overlaps are of smaller magnitudes than the overlaps that occurs for the 128-coefficient LS filters.

We have compared the results of the mean values of the average channel separation in Fig. 4. Despite the fact that the minimax cost function only minimizes the maximum errors, the crosstalk cancellation systems designed using 256-tap minimax filters provide higher average channel separation when compared with a corresponding LS system. However, the 128-tap design resulted in comparable performance for the two design methods. That the 256-coefficient, minimax design resulted in significantly better performance than the corresponding least-squares design may be somewhat surprising to some. However, we note here that the joint least-squares design, as formulated in



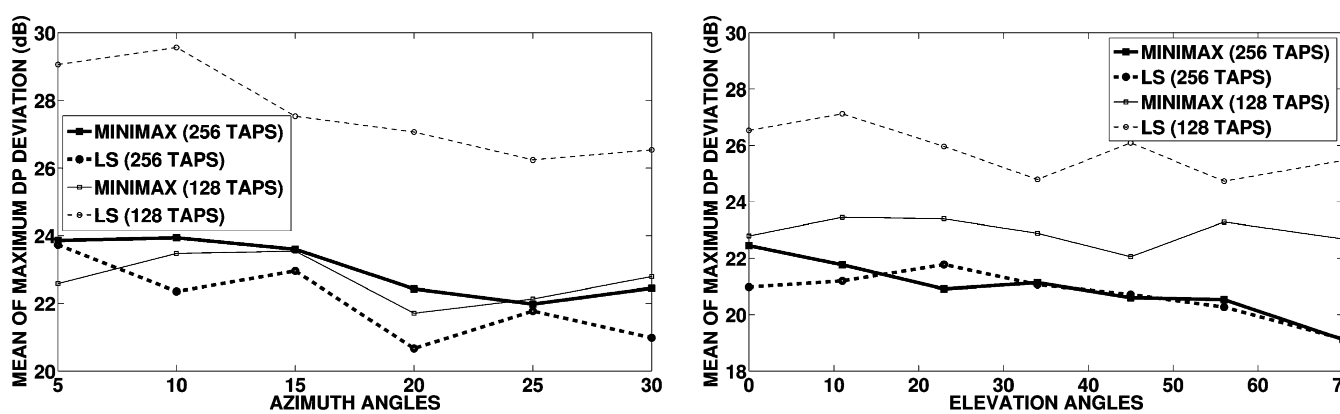
(a) (b)

Fig. 4. (a) Variation of the average channel separation for elevation angle $= 0^\circ$. (b) Variation of the average channel separation for azimuth angle $= 30^\circ$. (Thick lines—256 taps, thin lines—128 taps, solid lines—minimax, dashed lines—LS).



(a) (b)

Fig. 5. (a) Variation of the maximum cross path values for elevation angle $= 0^\circ$. (b) Variation of the maximum cross path values for azimuth angle $= 30^\circ$. (Thick lines—256 taps, thin lines—128 taps, solid lines—minimax, dashed lines—LS).



(a) (b)

Fig. 6. (a) Variation of the maximum direct path deviation for elevation angle $= 0^\circ$. (b) Variation of the maximum direct path deviation for azimuth angle $= 30^\circ$. (Thick lines—256 taps, thin lines—128 taps, solid lines—minimax, dashed lines—LS).

this paper, does not guarantee the maximization of the average channel separation in the frequency domain.

The mean of the maximum cross path values as shown in Fig. 5 are significantly lower for the 256-coefficient minimax filters implying that these systems offer better crosstalk cancellation. The performance of the 128-tap designs are comparable in this case. Fig. 6 shows that the maximum deviations of the direct path from 0 dB are comparable for the 256-tap designs. We should also keep in mind that the crosstalk cancellation systems have been designed subject to a constraint which limits the maximum direct path deviations to be below a prescribed threshold. On the other hand, the direct path deviations are much lower

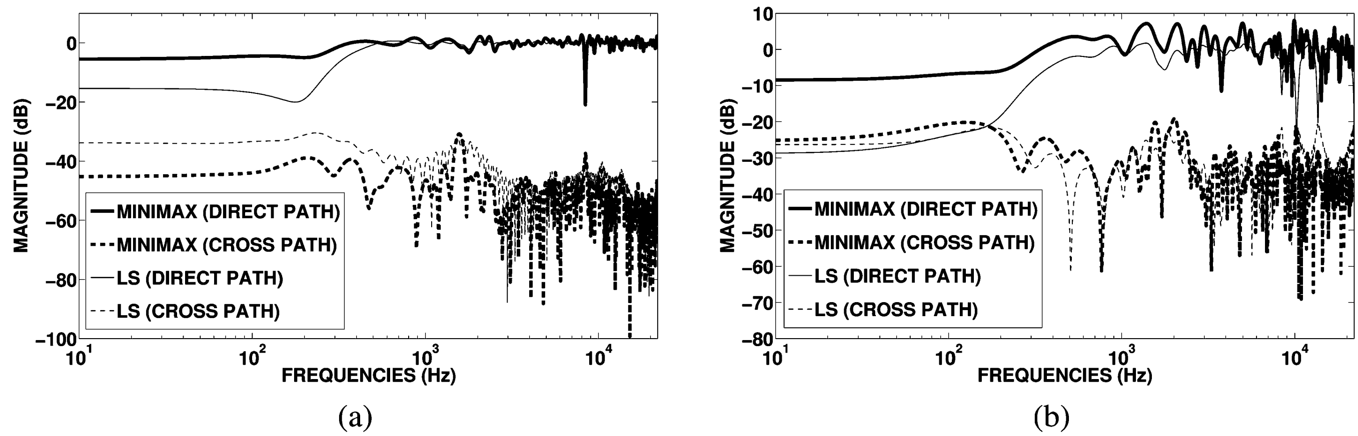


Fig. 7. Direct path and cross path magnitude responses for elevation $= 0°$, azimuth $= 5°$ using a conventional crosstalk canceler having (a) 256 coefficients and (b) 128 coefficients. (Thick lines—minimax and thin lines—LS).

for the 128-coefficient minimax filters when compared to the 128-coefficient LS filters. Thus, we can see from Figs. 5 and 6 that a combination of the low cross path characteristics along with the superior direct path behavior of the minimax design explains the minimum channel separation and the average channel separation as depicted in Figs. 3 and 4, respectively.

A representative plot of the magnitude responses of the direct and cross paths for a 256-tap crosstalk cancellation system for elevation angle 0° and azimuth angle 5° illustrating the superior performance of the minimax method is shown in Fig. 7(a). We can see from the figure that the minimax systems offers excellent channel separation with no overlap of the direct path and the cross path magnitude responses across the entire frequency range of interest. In the low-frequency region, as a consequence of the minimax filters providing lower crosstalk, the channel separation is much higher than what we can obtain using the LS filters. We also observe that minimax design exhibits superior performance even at frequencies less than 50 Hz. This may not be perceptually significant to a listener, nevertheless, we feel that it is important to highlight the consistent behavior of the minimax algorithm over a wide range of frequencies. Fig. 7(b) displays the magnitude responses associated with the two designs for elevation angle 0° and azimuth angle 5° for the 128-coefficient filters. From these results also, we can see the improvement in channel separation provided by the minimax filters. In contrast to the minimax design, the direct path and cross path responses overlap significantly in the LS design. The results also indicate that the performance of the LS designs are unacceptably poor at some frequencies, justifying our preference for the minimax design. Fig. 8 compares the direct path and the cross path impulse responses of the minimax solution with the impulse responses obtained using the LS solution. While these plots are not a true indicator of the performance improvement obtained using the minimax algorithm, it does serve to illustrate the achievement of a near-ideal response for the direct path as well as the cross path.

The histograms of the performance metrics of the 256-tap crosstalk cancellation systems for all the design cases are shown in Figs. 9 and 10. In general, the minimax crosstalk cancellation systems exhibit higher channel separation and lower maximum

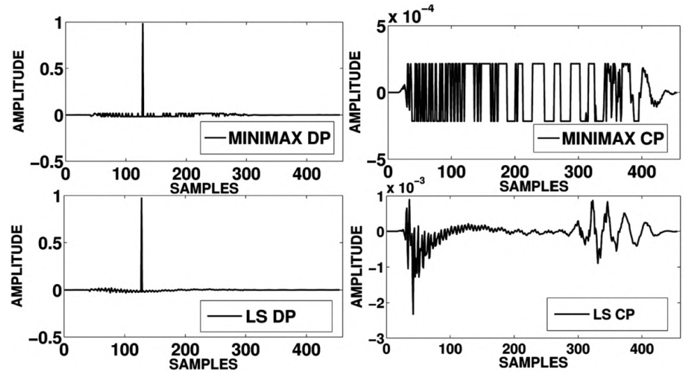
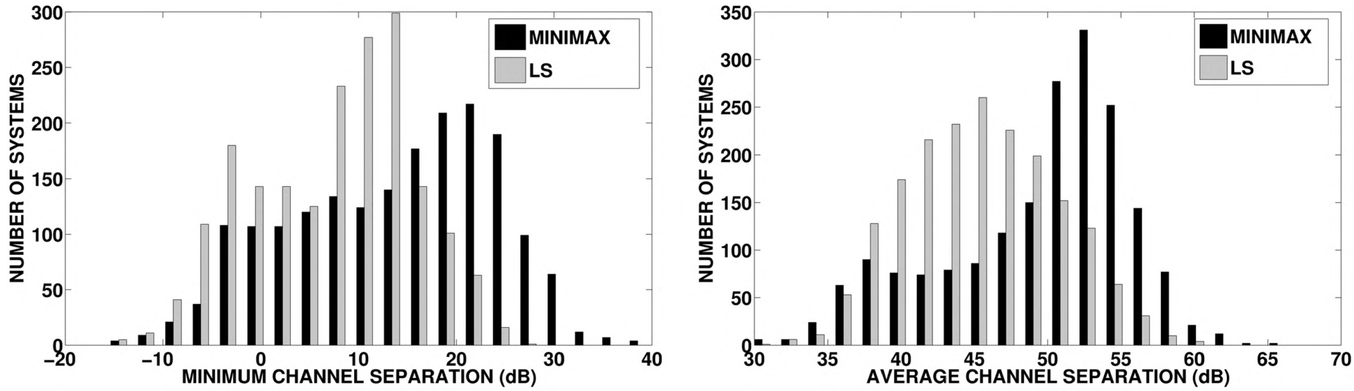


Fig. 8. Direct path (DP) and cross path (CP) impulse responses for elevation $= 0°$, azimuth $= 5°$ using a conventional crosstalk canceler having 256 coefficients.

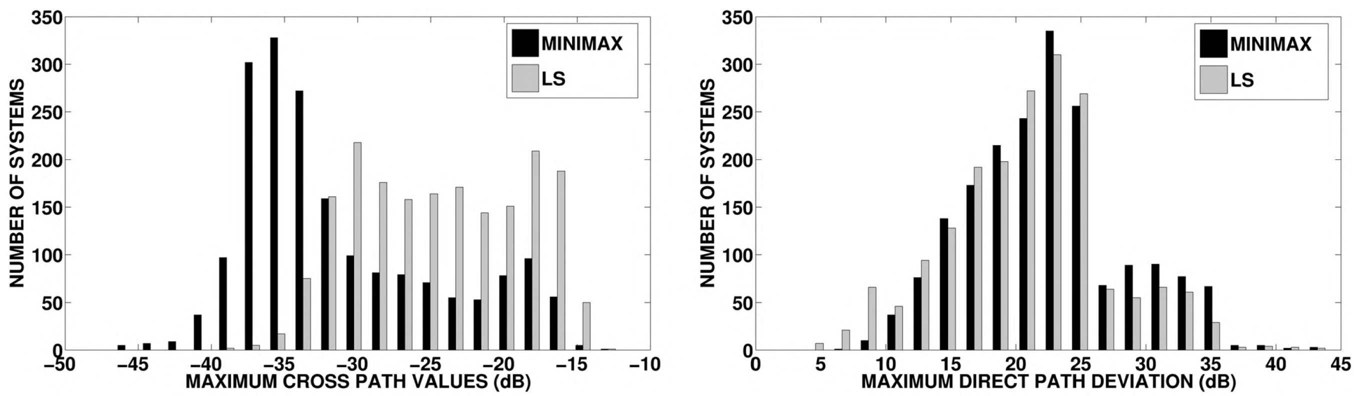
cross path values. The maximum direct path deviation is comparable for both designs, even though the minimax designs may have a slight advantage in this area also. Recall that both the designs constrain the maximum value of the direct path deviation. The comparable performance of the two designs may be due to this constraint.

Fig. 11 shows the effect of the weighting matrix $\mathbf{W}$ on the minimax designs. The minimax systems implemented by selecting α to be one and β as chosen by the algorithm were compared with the minimax filters designed using unity values for α and β. Fig. 11(a) plots the magnitude responses of the direct and cross paths for elevation angle 0° and azimuth angle 5° for the cases when the minimax algorithm was implemented using ($\alpha = 1, \beta = 1$) and ($\alpha = 1, \beta = 30$). Similarly, the crosstalk cancellation system for elevation angle 0° and azimuth angle 30° was implemented using ($\alpha = 1, \beta = 1$) and ($\alpha = 1$, $\beta = 100$) and the magnitude responses are plotted in Fig. 11(b). We once again recognize the fact that these values for β were selected by the algorithm according to the procedure described in Section IV-A. Though the use of nonuniform weights increases the distortion of the direct path, there is a tremendous improvement in the cross path behavior leading to a significant increase in the channel separation.



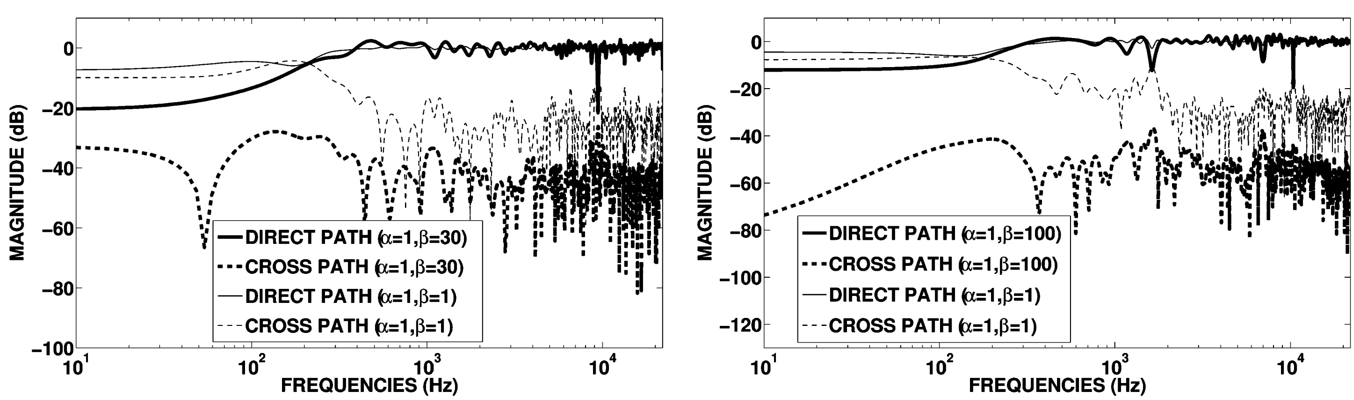
(a) (b)

Fig. 9. Histograms comparing (a) the minimum channel separation and (b) the average channel separation for the conventional crosstalk canceler. (Dark boxes—minimax, light boxes—LS).



(a) (b)

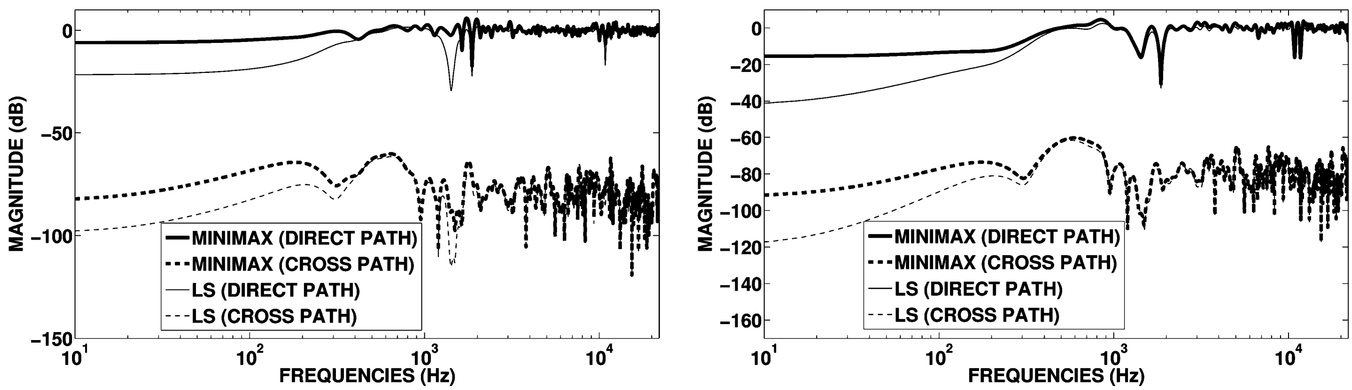
Fig. 10. Histograms comparing (a) the maximum cross path values and (b) the maximum direct path deviations for the conventional crosstalk canceler. (Dark boxes—minimax, light boxes—LS).



(a) (b)

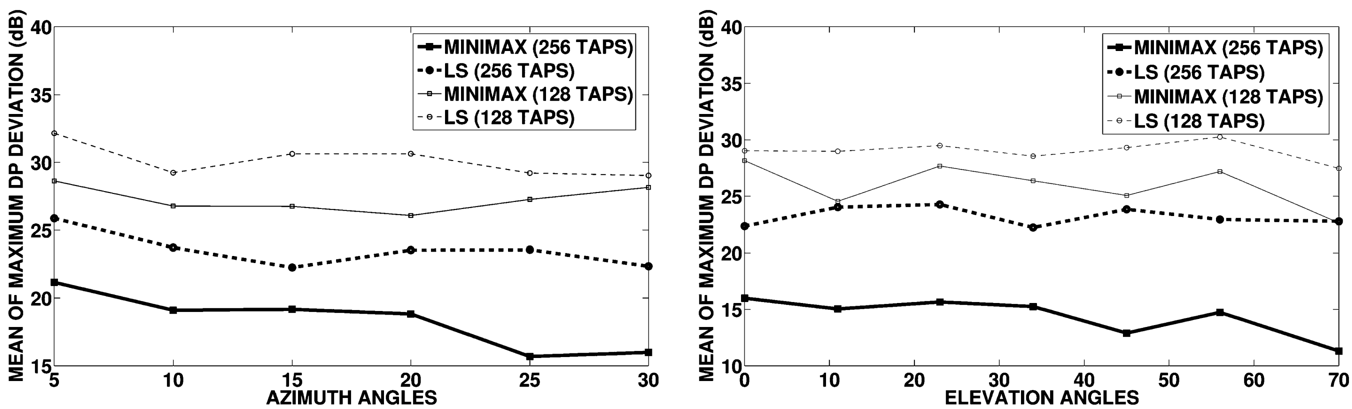
Fig. 11. Direct path and cross path magnitude responses for a 256-tap minimax crosstalk canceler for (a) elevation $= 0^\circ$, azimuth $= 5^\circ$. [Thick lines—($\alpha = 1$, $\beta = 30$), thin lines—($\alpha = 1$, $\beta = 1$)] (b) elevation $= 0^\circ$, azimuth $= 30^\circ$. [Thick lines—($\alpha = 1$, $\beta = 100$), thin lines—($\alpha = 1$, $\beta = 1$)].

In the next set of results, we present a similar statistical analysis to demonstrate the effectiveness of the minimax algorithm for designing crosstalk cancelers realized using the single-filter topology. We can obtain reasonably accurate estimates of the HRIRs using an LMS algorithm or another appropriate technique. As a result, the single-filter structure will ensure that the cross path response given by (20) is almost zero, irrespective of the method used. Consequently, we consider only the maximum direct path deviation as the performance measure for the single-filter case. Representative magnitude response plots for the systems designed using 256 coefficients and 128 coefficients are shown in Fig. 12(a) and (b), respectively, for elevation



(a) (b)

Fig. 12. Direct path and cross path magnitude responses for elevation $= 0^\circ$, azimuth $= 5^\circ$ using a single-filter crosstalk canceler having (a) 256 coefficients and (b) 128 coefficients. (Thick lines—minimax and thin lines—LS).



(a) (b)

Fig. 13. Statistics for the single-filter structure—(a) Variation of the maximum direct path deviation for elevation angle $= 0^\circ$. (b) Variation of the maximum direct path deviation for azimuth angle $= 30^\circ$. (Thick lines—256 taps, thin lines—128 taps, solid lines—minimax, dashed lines—LS).

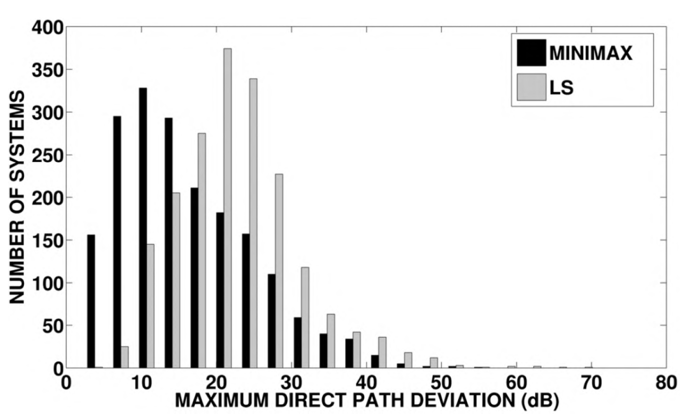


Fig. 14. Histogram comparing the maximum deviation of the direct path for the single-filter structure. (Dark boxes—minimax, light boxes—LS).

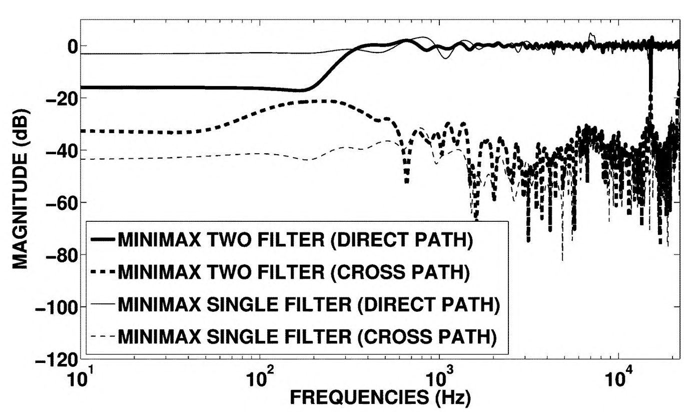
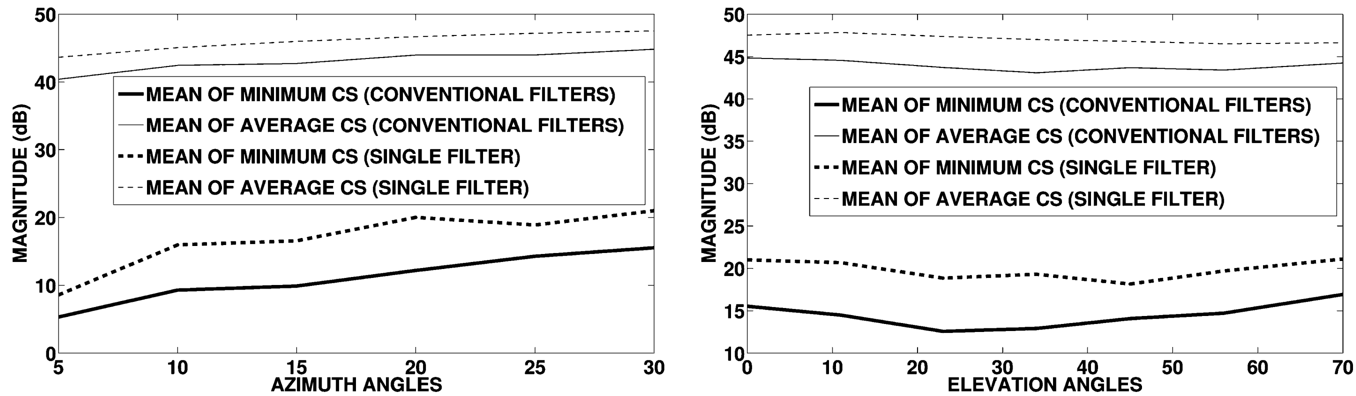


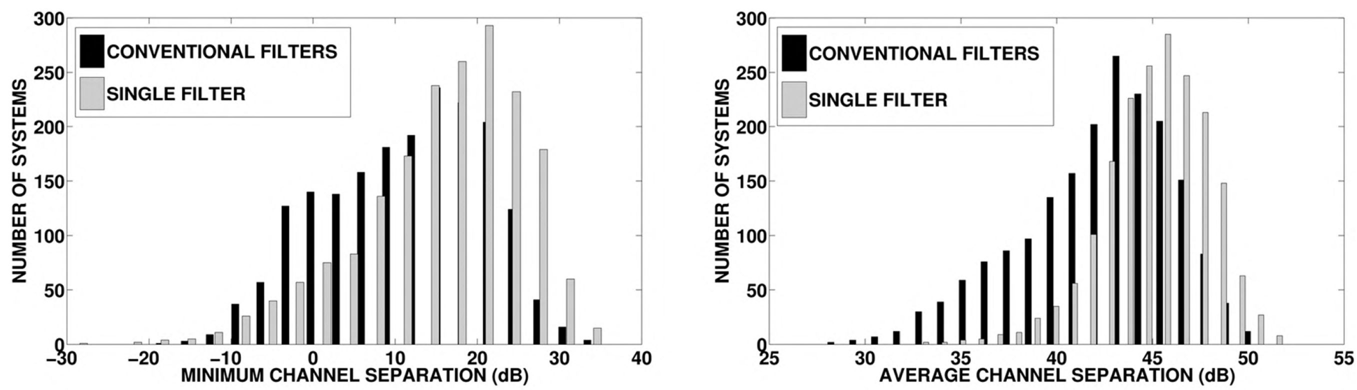
Fig. 15. Comparison of the direct path and cross path magnitude responses of a conventional system with those of a single-filter system for elevation $= 0^\circ$ and azimuth $= 5^\circ$. (Thick lines—conventional design, thin lines—single-filter design).

angle 0° and azimuth angle 5°. Fig. 13 shows the statistics for the single-filter structure. These results indicate that when compared with the LS method, the maximum deviation of the direct path from 0 dB is significantly lower for the minimax method for both the 256-tap and 128-tap filters. Besides facilitating a lower value of the maximum direct path deviation than the LS approach, the minimax method also results in direct path gains that are closer to one in the low frequency region. It might appear that while the direct path performance of the LS method is poorer, it offers lower crosstalk over the same frequency range. Since both the minimax and the LS methods offer more than



(a) (b)

Fig. 16. a) Variation of the minimum and average channel separations for elevation angle $= 0^\circ$. (b) Variation of the minimum and average channel separations for azimuth angle $= 30^\circ$. (Thick lines—minimum values, thin lines—average values, solid lines—conventional design, dashed lines—single-filter design).



(a) (b)

Fig. 17. Histograms comparing (a) the minimum channel separation and (b) the average channel separation for the conventional and single-filter systems. (Dark boxes—conventional design, light boxes—single-filter design).

60-dB crosstalk cancellation, this is not a perceptible advantage. Fig. 14 displays the histogram plot comparing the 256-tap minimax and LS crosstalk cancelers. We can see that the minimax design results in a higher distribution of crosstalk cancellation systems that exhibit lower values of maximum direct path deviation than the LS designs.

Finally, we compare the performance of the traditional Atal–Schroeder crosstalk cancellation system with the single-filter topology designed using the minimax approach. Let M be the length of the acoustic HRIRs, K_c the length of the conventional crosstalk cancellation filters, and K_s be the length of the single filter. The total number of coefficients in a conventional system is $4K_c$ and a single-filter system has $4M + 2K_s$ coefficients. In general, if $K_c = K_s = K$ and $K \leq 2M$, the complexity of the conventional system is less than that of a single-filter system.

Fig. 15 compares the performance of the two topologies designed using the minimax criterion when their implementation complexities were the same. In this case, the single-filter system required the design of a 112-coefficient filter, and the conventional system was based on the design of the two filters with 256 coefficients each. To make the simulations realistic, we estimated the HRIRs using an LMS algorithm from a simulated acoustic environment. The simulations employed the CIPIC database to generate the acoustic signals used in the estimation process. The signal-to-noise ratio (SNR) of the measured signals was 30 dB in all cases. The performance comparison indicate that the single-filter structure offers excellent crosstalk cancellation, and it exhibits satisfactory direct path performance. Consequently, the overall channel separation is comparatively higher than the conventional crosstalk cancellation system.

Finally, we present the statistics of the minimum and average channel separations for the two topologies. Fig. 16 displays the mean values of the minimum and average channel separations over all the 45 subjects. The single-filter design provides 5 dB or more minimum channel separation than the conventional design. The average channel separation is also higher for the single-filter design. The superior performance of the single-filter structure is further confirmed by the histograms shown in Fig. 17. The distributions indicate that the single-filter design results in more number of crosstalk cancellation systems having higher minimum and average channel separations than the conventional crosstalk canceler.

V. Conclusion

In this paper, we presented a method to jointly design the crosstalk cancellation filters for a conventional crosstalk cancellation system as well for a single-filter topology using minimax techniques. The algorithm is easy to formulate, can be implemented efficiently, and achieves excellent channel separation, especially at low frequencies. Simulation results indicate that the single-filter structure is more robust to errors in estimation of the HRIRs. The single-filter design also offers better cancellation and channel separation than the conventional crosstalk canceler structure. These characteristics of the minimax crosstalk cancellation filters makes the authors believe that the approach presented in this paper is an attractive alternate to techniques currently available in the literature. Loudspeaker inputs can play an important role in practical crosstalk cancellation systems, and regularization techniques are commonly used to avoid overloading the loudspeaker at low frequencies. It may be worthwhile to devise a regularization scheme applicable for the minimax techniques described in this paper.

Acknowledgment

The authors would like to thank B. Farhang-Boroujeny for his review of the paper and useful suggestions.

References

[1] W. G. Gardner. *3-D Audio Using Loudspeakers*. Norwell, MA: Kluwer, 1997.

[2] W. G. Gardner, "Head tracked 3-D audio using loudspeakers," in *Proc. IEEE ASSP Workshop Applicat. Signal Process. Audio and Acoust.*, New Paltz, NY, Oct. 1997, pp. 898–901.

[3] W. G. Gardner, "3-D audio and acoustic environmental modeling," 1999 [Online]. Available: http://www.headwize.com/tech/gardner_tech.htm

[4] C. Kyriakakis, "Fundamental and technological limitations of immersive audio systems," *Proc. IEEE*, vol. 86, no. 5, pp. 941–951, May 1998.

[5] B. S. Atal and M. R. Schroeder, "Apparent sound source translator," U.S. patent 3,236,949, Feb. 1966.

[6] D. B. Ward, "Joint least squares optimization for robust acoustic crosstalk cancellation," *IEEE Trans. Speech Audio Process.*, vol. 8, no. 2, pp. 211–215, Feb. 2000.

[7] C. Bourget and T. Aboulnasr, "Inverse filtering of room impulse response for binaural recording playback through loudspeakers," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Adelaide, Australia, Apr. 1994, vol. III, pp. 301–304.

[8] M. Miyoshi and Y. Kaneda, "Inverse filtering of room acoustics," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 36, no. 2, pp. 145–152, Feb. 1988.

[9] W. Putnam, D. Rocchesso, and J. Smith, "A numerical investigation of the invertibility of room transfer functions," in *Proc. IEEE ASSP Workshop Applicat. Signal Process. Audio and Acoust.*, New Paltz, NY, Oct. 1995, pp. 249–252.

[10] M. Mouchtaris, P. Reveliotis, and C. Kyriakakis, "Inverse filter design for immersive audio rendering over loudspeakers," *IEEE Trans. Mutimedia*, vol. 2, no. 2, pp. 77–87, Jun. 2000.

[11] P. A. Nelson, H. Hamada, and S. J. Elliot, "Adaptive inverse filters for stereophonic sound reproduction," *IEEE Trans. Signal Process.*, vol. 40, no. 7, pp. 1621–1632, Jul. 1992.

[12] J.-S. Lim and C. Kyriakakis, "Multirate adaptive filtering for immersive audio," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Salt Lake City, UT, May 2001, vol. 5, pp. 3357–3360.

[13] A. Gonzalez and J. J. Lopez, "Time domain recursive deconvolution in sound reproduction," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Istanbul, Turkey, Jun. 2000, vol. 2, pp. 833–836.

[14] S. M. Kuo and G. H. Canfield, "Dual-channel audio equalization and cross-talk cancellation for 3-D sound reproduction," *IEEE Trans. Consumer Electron.*, vol. 43, no. 4, pp. 1189–1196, Nov. 1997.

[15] J. Garas and P. Sommem, "Adaptive filtering approach to multichannel audio reproduction." [Online]. Available: http://www.dspalgorithms.com/download/pdf/nerg.pdf

[16] J. Garas. *Adaptive 3-D Sound Systems*. Norwell, MA: Kluwer, 2000.

[17] S.-M. Kim and S. Wang, "A Weiner filter approach to the binaural reproduction of stereo sound," *J. Acoust. Soc. Amer.*, vol. 114, no. 6, pp. 3179–3188, Dec. 2003.

[18] O. Kirkeby, P. A. Nelson, H. Hamada, and F. Orduna-Bustamante, "Fast deconvolution of multichannel systems using deregularization," *IEEE Trans. Speech Audio Process.*, vol. 6, no. 2, pp. 189–195, Mar. 1998.

[19] O. Kirkeby and P. A. Nelson, "Digital filter design for inversion problems in sound reproduction," *J. Audio Eng. Soc.*, vol. 47, no. 7, pp. 583–595, Jul. 1999.

[20] 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. Acoust., Speech, Signal Process.*, Toulouse, France, May 2006, vol. 5, pp. 353–356.

[21] L. J. Karam and J. H. McClellan, "Complex Chebyshev approximation for FIR filter design," *IEEE Trans. Circuits Syst.-II*, vol. 42, no. 3, pp. 207–216, Mar. 1995.

[22] S. F. Yan and Y. Ma, "A unified framework for designing FIR filters with arbitrary magnitude and phase response," *Digital Signal Process.*, vol. 14, no. 6, pp. 510–522, Nov. 2004.

[23] J. F. Sturm, "Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones," *Optim. Meth. Softw.*, vol. 11–12, pp. 625–653, Aug. 1999.

[24] H.-S. Kim, P.-M. Kim, and B. K. Kim, "Cross-talk cancellation algorithm for 3D sound reproduction," *ETRI J.*, vol. 22, no. 2, pp. 11–19, Jun. 2000.

[25] W.-S. Lu and T. Hinamoto, "Optimal design of IIR digital filters with robust stability using conic-quadratic-programming updates," *IEEE Trans. Signal Process.*, vol. 51, no. 6, pp. 1581–1592, Jun. 2003.

[26] Y. Ye, M. J. Todd, and S. Mizuno, "An $O(\sqrt{n}L)$-iteration homogeneous and self-dual linear programming algorithm," *Math. Oper. Res.*, vol. 19, no. 1, pp. 53–67, Feb. 1994.

[27] E. D. Anderson, C. Roos, and T. Terlaky, "On implementing a primal-dual interior point method for conic quadratic optimization," *Math. Programm.*, vol. 95, no. 2, pp. 249–277, Feb. 2003.

[28] J. F. Sturm, "Implementation of interior point methods for mixed semidefinite and second order cone optimization problems," *Optimization Methods Softw.*, vol. 17, no. 6, pp. 1105–1154, Dec. 2002.

[29] V. R. Algazi, R. O. Duda, D. M. Thompson, and C. Avendano, "The CIPIC HRTF database," in *IEEE Workshop Applicat. Signal Process. Audio Acoust.*, New Paltz, NY, Oct. 2001, pp. 99–102.

[30] V. R. Algazi *et al.*, "The CIPIC HRTF Database," 2001 [Online]. Available: http://interface.cipic.ucdavis.edu/CIL_html/CIL_HRTF_database.htm

[31] O. Kirkeby, P. A. Nelson, and H. Hamada, "The stereo dipole—A virtual source imaging system using two closely spaced loudspeakers," *J. Audio Eng. Soc.*, vol. 46, no. 5, pp. 387–395, May 1998.

Harsha I. K. Rao (S'06) received the B.E. degree in electronics and communication engineering with highest honors from the National Institute of Technology, Tiruchirappalli, India, in 2003. He is currently working towards the Ph.D. degree at the University of Utah, Salt Lake City.

From 2003 to 2004, he was a Design Engineer at ABB, Bangalore, India. His current research is in acoustic signal processing and adaptive filtering.

V. John Mathews (S'82–M'84'–SM'90–F'02) received the B.E. (honors) degree in electronics and communication engineering from the University of Madras, Madras, India, in 1980, and the M.S. and Ph.D. degrees in electrical and computer engineering from the University of Iowa, Iowa City, in 1981 and 1984, respectively.

At the University of Iowa, he was a Teaching/Research Fellow from 1980 to 1984 and a Visiting Assistant Professor in the Department of Electrical and Computer Engineering during the 1984–1985 academic year. He joined the University of Utah, Salt Lake City, in 1985, where he is a Professor in the Department of Electrical and Computer Engineering. He served as the Chairman of the Department from 1999 to 2003. His research interests include adaptive and nonlinear filtering and applications of signal processing techniques in communication, audio, aerospace, and biomedical engineering. He is the author of the book *Polynomial Signal Processing* (Wiley, 2000), coauthored with Prof. G. L. Sicuranza, University of Trieste, Italy. He has published more than 100 technical papers.

Dr. Mathews served as the Vice President-Finance of the IEEE Signal Processing Society from 2003 to 2005, and has served on the publications board and the conference board of the society. He is a past Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING and the IEEE SIGNAL PROCESSING LETTERS. He serves on the editorial boards of the *IEEE Signal Processing Magazine* and the IEEE JOURNAL ON SPECIAL TOPICS IN SIGNAL PROCESSING. He was a member of the Signal Processing Theory and Methods and Education Technical Committees and served as the General Chairman of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) 2001.

Young-Cheol Park (S'88–M'93) was born in Korea in 1964. He received the B.S., M.S., and Ph.D. degrees in electronic engineering from Yonsei University, Seoul, Korea, in 1986, 1988, and 1993, respectively.

From September 1993 to October 1995, he was a Postdoctoral Research Scholar at The Pennsylvania State University, University Park, where he worked on active noise control. From March 1996 to August 1998, he was with the Samsung Electronics Company, Ltd., Kiheung, Korea, and from April 1999 to February 2002, he was with InTime Corporation, Seoul, where he worked on digital hearing aids, digital audio processors, and VoIP systems. In 2002, he joined the Computer and Telecommunications Engineering Division, Yonsei University, where he is currently an Associate Processor. His research interests include adaptive filtering algorithms, three-dimensional audio signal processing, high-quality audio coding, digital hearing aids, and their implementations.