Optimal design of loudspeaker arrays for robust cross-talk cancellation using the Taguchi method and the genetic algorithm

Mingsian R. Bai,[a] Chih-Wei Tung, and Chih-Chung Lee
Department of Mechanical Engineering, National Chiao-Tung University, 1001 Ta-Hsueh Road, Hsin-Chu 300, Taiwan, Republic of China

(Received 10 August 2004; revised 13 December 2004; accepted 7 February 2005)

An optimal design technique of loudspeaker arrays for cross-talk cancellation with application in three-dimensional audio is presented. An array focusing scheme is presented on the basis of the inverse propagation that relates the transducers to a set of chosen control points. Tikhonov regularization is employed in designing the inverse cancellation filters. An extensive analysis is conducted to explore the cancellation performance and robustness issues. To best compromise the performance and robustness of the cross-talk cancellation system, optimal configurations are obtained with the aid of the Taguchi method and the genetic algorithm (GA). The proposed systems are further justified by physical as well as subjective experiments. The results reveal that large number of loudspeakers, closely spaced configuration, and optimal control point design all contribute to the robustness of cross-talk cancellation systems (CCS) against head misalignment.
 [DOI: 10.1121/1.1880852]

PACS numbers: 43.38.Vk [AJZ] Pages: 2802–2813

I. INTRODUCTION

Spatial audio or three-dimensional (3D) audio has received much attention in many emerging applications such as computer multimedia, home theater, video games, digital television, etc. Despite the rapid development of the technology, cross talk has been a plaguing problem when loudspeakers are used as the means of rendering. Binaural audio signals containing directional cues are to be reproduced, at the ears of a listener, that he or she would naturally hear. However, excess cross talk can smear these cues and adversely effect the localization of sound images reproduced by loudspeakers. It is thus desirable to preprocess the loudspeaker signals by using the so-called cross-talk cancellation system (CCS) so that the sound from the loudspeakers to contralateral ears is minimized, if not completely eliminated.

Several CCS have been proposed in the past. The idea of CCS was first introduced by Bauer,[1] and later put into practice by Atal and Schroeder,[2] and Damaske and Mellert.[3] The limitation of these early systems is that head movement away from the sweet spot greater than about 75 to 100 mm would significantly degrade the spatial effect. Cooper and Bauck suggested a propagation matrix based on the spherical head model.[4] A similar method by Gardner approximates the effect of the head with a low-pass filter, a delay, and a simple gain.[5] Blumlein,[6] and Cooper and Bauck[7,8] showed that, under the assumption of left–right symmetry, a "shuffler" filter can be used to simplify the implementation of CCS. Note that, if the position of the listener changes over time, then ipsilateral and contralateral transfer function will not be symmetrical, but will vary to reflect the head-related-transfer functions (HRTF) for the listener's new position. A head-tracking CCS was reported in the work of Kyriakakis *et al.* to cope with head movement of the listener.[9,10] Ward and Elko in Bell Labs have conducted a series of less elaborate but insightful analysis of the robustness of the CCS. In their first paper[11] on this topic in 1998, robustness of a simple 2×2 CCS was investigated using weighted cancellation performance measure (at the pass zone and stop zone, respectively). In their second paper[12] in 1999, robustness of a 2×2 CCS was again examined using a different measure that focuses more on numerical stability, as reflected by matrix condition numbers, with respect to data and/or system perturbations during matrix inversion. Both approaches wind up with optimal loudspeaker spacing inversely proportional to frequency. Parallel to the previous work, the present paper explores the robust issue in a more general context. Using multidrive array configurations, more than two loudspeakers are used to provide additional degrees of freedom for control of the sound field. In the optimization procedure, channel separation and beamwidth are employed as a more intuitive robustness measure against head misalignment. The optimization leads to an optimal loudspeaker configuration independent of frequency. An alternative approach was developed by Takeuchi and Nelson to enhance the robustness of CCS against head movement away from the sweet spot. In their system, two loudspeakers are closely spaced to form what they call the "stereo dipole."[13] This idea was further extended by the same researchers to be the optimal source distribution (OSD) system.[14] Their robust analysis of CCS was also based on numerical stability in relation to the errors in matrix inversion. The performance of CCS deteriorates due to these errors resulting possibly from head misalignment and the HRTF modeling variations. Inversion of an ill-conditioned system (with a large matrix condition number) leads to loss of dynamic range and lack of robustness to head misalignment. The authors attempt to pinpoint an optimal configuration of a 2×2 CCS in which loudspeaker spacing is the primary design parameter such that the trade-off among dynamic range, robustness, and control performance

[a]Electronic mail: msbai@mail.nctu.edu.tw

 0001-4966/2005/117(5)/2802/12/$22.50

are best reconciled. Their analysis also yielded optimal loudspeaker spacing inversely proportional to frequency. Since the spacing thus found is frequency dependent, a multidrive configuration of the OSD, comprising pairs of loudspeakers with different spacing, was suggested to deal with cross-talk cancellation in different frequency bands. Apart from the robustness measure and analysis techniques, the present paper differs from their approach in that our approach is a direct multidrive (more than two loudspeakers) array configuration, which requires no crossover circuits that may introduce distortions at the crossover frequencies. In this array configuration, the additional degrees-of-freedom in control of sound field provided by the beamformer can be exploited to the greatest extent.

In this paper, the performance and robustness issues of CCS for various loudspeaker configurations are examined. Traditional stereo CCS systems require that the listener is positioned in the so-called ‘‘sweet spot’’ such that the listener forms an equilateral triangle with respect to the loudspeaker pair. The loudspeakers, therefore, subtend an angle of 60° from the listener.[15] Once the listener moves away from the sweet spot, especially when moving sideway, the conditions for cancellation are no longer met and the spatial sound images are lost. The idea of sweet spot applies with different degrees not only to stereo systems but also to other loudspeaker configurations.

Following the analysis of performance and robustness analysis, this paper is focused on the development of a CCS using a loudspeaker array in an effort to best compromise performance and robustness of the system. An array focusing scheme is also exploited, based on the inverse propagation operator that relates the transducers to a set of chosen control points.[16,17] Optimal design parameters of the array are found using the Taguchi method[18] and the genetic algorithm (GA).[19,20] It has been found in Refs. 11 and 13 that cancellation is least effective because of the narrow sweet spot as the head moves sideway rather than when it moves in the other directions. Hence, only lateral misalignment is investigated in the present paper. As will be detailed later, the optimal configuration is the closely spaced array. Such system is found to be more robust to misalignment of the listener's head. This finding is in agreement with the conclusion of Ref. 21. The 3D audio system resulting from the abovementioned optimization is then implemented on a multimedia Pentium 4 personal computer. The proposed systems are further justified by physical and subjective experiments. Feasibility of the proposed CCS will be discussed in the conclusions.

II. THEORY AND METHODS

A. The propagation matrix

Assume that the process of sound propagation from the loudspeakers to the listener's ears is linear and time invariant. Viewed as a multichannel system, a propagation matrix relates the loudspeaker inputs and a set of chosen ‘‘control points.’’ These control points are allocated along the line linking two ears, as shown in Fig. 1. These control points are crucial to the tailoring of the so-called sweet spot, which is

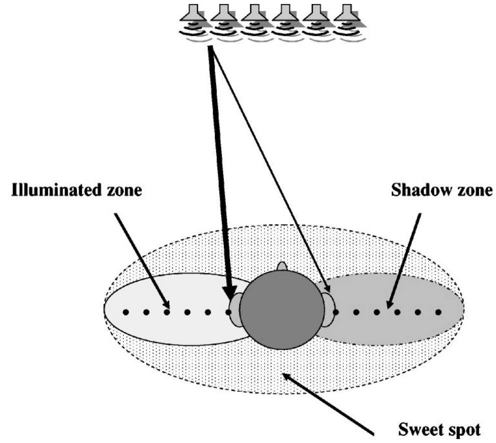


FIG. 1. Allocation of control points for the CCS.

composed of an illuminated zone for the ipsilateral propagation and a shadow zone for the contralateral propagation. The main purpose of a CCS is to minimize, if not completely eliminate, the cross talks associated with the contralateral propagation. To accomplish this, therefore, unity gains are designated to the control points in the illuminated zone, whereas nulls are designated to the control points in the shadow zone. For simplicity, we restrict ourselves to a one-dimensional array and define head-related impulse responses (HRIR), $h_{mj}(n)$, $1 \leqslant m \leqslant M$, $1 \leqslant j \leqslant J$, as the impulse responses corresponding to the mth control point and the jth loudspeaker (n being the discrete-time index). Let $v_j(n)$, $1 \leqslant j \leqslant J$, be the J input signals to the loudspeaker array. The output signals $f_m(n)$, $1 \leqslant m \leqslant M$, received at the control points are given by

$$f_m(n) = \sum_{j=1}^{J} h_{mj}(n) * v_j(n), \quad 1 \leqslant m \leqslant M, \tag{1}$$

where* denotes the convolution operator. Fourier transform of this equation leads to

$$F_m(e^{j\omega}) = \sum_{j=1}^{J} H_{mj}(e^{j\omega}) V_j(e^{j\omega}), \quad 1 \leqslant m \leqslant M. \tag{2}$$

In matrix form

$$\mathbf{f}(e^{j\omega}) = \mathbf{H}(e^{j\omega}) \mathbf{v}(e^{j\omega}), \tag{3}$$

with $\mathbf{v}(e^{j\omega}) = [V_j(e^{j\omega})]_{1 \leqslant j \leqslant J}$ and $\mathbf{f}(e^{j\omega}) = [F_m(e^{j\omega})]_{1 \leqslant m \leqslant M}$ being the column vectors of the Fourier transforms of the loudspeaker input signals and the reproduced signals, respectively. Overall, the transfer matrix $\mathbf{H}(e^{j\omega}) = [H_{mj}(e^{j\omega})]_{1 \leqslant m \leqslant M, 1 \leqslant j \leqslant J}$ represents the frequency-domain multichannel propagation process from the array loudspeakers to the control points at the sweet spot.

B. Inverse filtering with Tikhonov regularization

The CCS aims to cancel the cross talks in stereo loudspeaker rendering so that the binaural signals are reproduced at two ears like those from a headphone. This can be viewed as a model-matching problem, shown in Fig. 2. In the block diagram, $\mathbf{x}(z)$ is a vector of H program input signals (z being the z-transform variable), $\mathbf{u}(z)$ is a vector of $I=2$ binaural

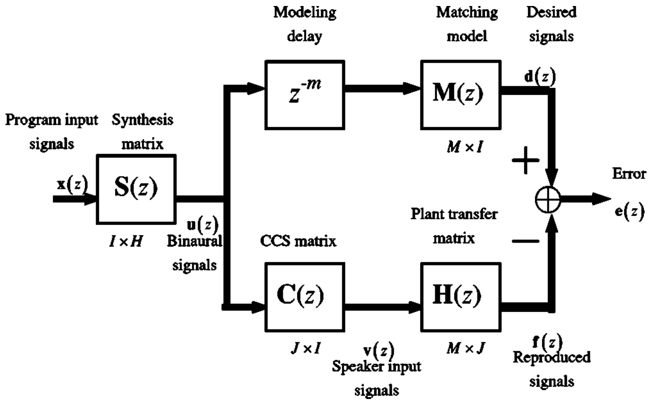


FIG. 2. The block diagram of a multichannel model-matching problem in the CCS design.

signals, $\mathbf{v}(z)$ is a vector of J loudspeaker input signals, $\mathbf{f}(z)$ is a vector of M reproduced signals, $\mathbf{d}(z)$ is a vector of M desired signals, and $\mathbf{e}(z)$ is a vector of M error signals. $\mathbf{M}(z)$ is an $M\times I$ matrix of matching model, $\mathbf{H}(z)$ is an $M\times J$ plant transfer matrix, and $\mathbf{C}(z)$ is a $J\times I$ matrix of the CCS filters. The term z^{-m} accounts for the modeling delay to ensure causality of the CCS filters. It is straightforward to establish the following relationships:

$$\mathbf{v}(z)=\mathbf{C}(z)\mathbf{u}(z), \tag{4}$$

$$\mathbf{f}(z)=\mathbf{H}(z)\mathbf{v}(z), \tag{5}$$

$$\mathbf{d}(z)=z^{-m}\mathbf{M}(z)\mathbf{u}(z), \tag{6}$$

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

Ideal model matching requires that $\mathbf{H}(z)\mathbf{C}(z)=z^{-m}\mathbf{M}(z)$. In general, $\mathbf{H}(z)$ is noninvertible because it is usually ill-conditioned and even nonsquare. To overcome this difficulty, we employ the Tikhonov regularization procedure in the matrix inversion process.[22] In the method, one seeks to minimize a frequency-domain objective function $O(e^{j\omega})$ defined as

$$O(e^{j\omega})=\mathbf{e}^H(e^{j\omega})\mathbf{e}(e^{j\omega})+\beta^2\mathbf{v}^H(e^{j\omega})\mathbf{v}(e^{j\omega}). \tag{8}$$

The regularization parameter β weighs the input power $\mathbf{v}^H\mathbf{v}$ against the performance error $\mathbf{e}^H\mathbf{e}$. The optimal solution $\mathbf{v}_{\mathrm{opt}}(e^{j\omega})$ of Eq. (8) is

$$\mathbf{v}_{\mathrm{opt}}(e^{j\omega})=[\mathbf{H}^H(e^{j\omega})\mathbf{H}(e^{j\omega})+\beta^2\mathbf{I}]^{-1}\mathbf{H}^H(e^{j\omega})\mathbf{M}(e^{j\omega})\mathbf{u}(e^{j\omega}). \tag{9}$$

Consequently, the CCS matrix can be readily identified as

$$\mathbf{C}(e^{j\omega})=[\mathbf{H}^H(e^{j\omega})\mathbf{H}(e^{j\omega})+\beta^2\mathbf{I}]^{-1}\mathbf{H}^H(e^{j\omega})\mathbf{M}(e^{j\omega}). \tag{10}$$

In our approach, the parameter β is frequency dependent and constrained by a gain threshold applied to $\mathbf{C}(e^{j\omega})$, e.g., 12 dB. This is in contrast to the approach in Ref. 16, where a constant β applied to all frequencies.

Traditionally, the desired signals $\mathbf{d}(z)$ are just the binaural signals $\mathbf{u}(z)$. The matrix $\mathbf{M}(z)$ is an identity matrix of order 2, i.e., $\mathbf{M}=\mathbf{I}$, and the frequency responses of the corresponding optimal filters are given by

$$\mathbf{C}(e^{j\omega})=[\mathbf{H}^H(e^{j\omega})\mathbf{H}(e^{j\omega})+\beta^2\mathbf{I}]^{-1}\mathbf{H}^H(e^{j\omega}). \tag{11}$$

The frequency response matrix $\mathbf{C}(e^{j\omega})$ is then sampled at N_c equally spaced frequencies with discrete-frequency index k

$$\mathbf{C}(k)=[\mathbf{H}^H(k)\mathbf{H}(k)+\beta^2\mathbf{I}]^{-1}\mathbf{H}^H(k), \quad k=1,2,...,N_c. \tag{12}$$

The impulse responses of the inverse filters can be calculated using the inverse fast Fourier transform (IFFT) of the frequency samples of Eq. (12) with appropriate windowing. Circular shifts may be necessary to guarantee the causality of CCS filters; hence, the modeling delay z^{-m} in Fig. 1.

The present method differs from the foregoing conventional approach in that, instead of "single-point" matching, a number of control points are distributed in the illuminated zone and the shadow zone so that the sweet spot can be widened. This is accomplished by choosing a more complex matching model akin to the window design in the time-domain digital signal processing. An example of choosing control points is illustrated as follows. Suppose we wish to choose three points in the illuminated zone and six points in the shadow zone for each ear. These control points can only be located at six discrete locations on each side of the head, as shown in Fig. 1. In this scenario, we may choose a 9×1 matching model for the left ear with the following pattern:

$$\mathbf{M}_L=[1\ \ 1\ \ 1\ \ 0\ \ 0\ \ 0\ \ 0\ \ 0\ \ 0]^T, \tag{13}$$

where the subscript L stands for the left ear, and the ones and zeros correspond to the designated control points in the illuminated zone and the shadow zone, respectively. Hence, the desired signal for the left ear is

$$\mathbf{d}_L=z^{-m}\mathbf{M}_L\mathbf{u}=z^{-m}[u_L\ \ u_L\ \ u_L\ \ 0\ \ 0\ \ 0\ \ 0\ \ 0\ \ 0]^T. \tag{14}$$

After the matching model is selected, the optimal CCS filters can be calculated according to Eq. (10). The same procedure applies to the ear on the right side. In general, more points in the shadow zone are needed than in the illuminated zone, since the performance in the former region is more critical to cancellation of cross talks. It should be noted that the increased complexity of the sweet spot widening technique lies purely in the off-line design procedure. The number of channels $(J\times I)$ of the resulting CCS filter remains the same.

III. OPTIMIZATION OF ARRAY CONFIGURATION FOR ROBUST CROSS-TALK CANCELLATION

Many design factors are involved in the CCS, e.g., array configurations, spacing and positions, number of control points in filter design, and so forth. Different configurations have effects with varying degree on the performance as well as robustness of the CCS. To minimize the effort of trial and error, a systematic design procedure of CCS based on the Taguchi method and the genetic algorithm (GA) is presented as follows.

A. Taguchi method

The Taguchi method is an experiment design procedure well suited to multivariable optimization. The method is intended for three engineering applications: system design, parameter design, and tolerance design. For our optimization

TABLE I. Parameter design using the orthogonal array of the Taguchi method. Nine observations and three factors for three levels are required.

Run	Factor A	Factor B	Factor C	Fitness
1	1	1	1	317.35
2	1	2	2	74.653
3	1	3	3	173.62
4	2	1	2	141.81
5	2	2	3	35.169
6	2	3	1	80.04
7	3	1	3	43.49
8	3	2	1	90.706
9	3	3	2	206.65

Factor	Level 1	Level 2	Level 3
A	Closely spaced	Apart and facing ears	Apart and facing front
B	5 control points (2 points on the illuminated zone, 3 points on the shadow zone)	5 control points (2 points on the illuminated zone, 1 point at the head center, 2 points on the shadow zone)	2 control points (one at each ears)
C	2×6 CCS	2×3 CCS	2×2 CCS

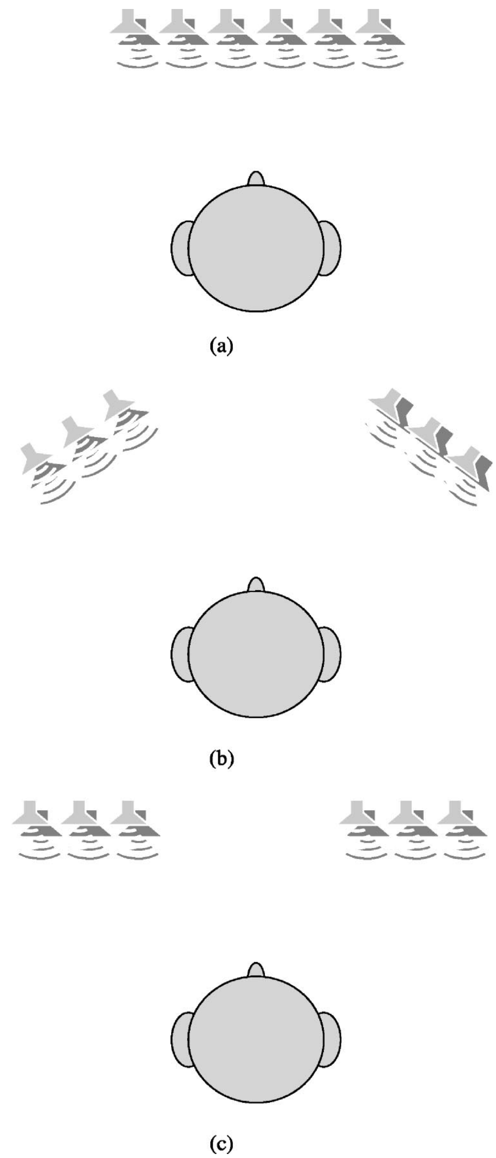


FIG. 3. Three configurations of loudspeaker arrays. (a) closely spaced loudspeakers (six loudspeakers in a row); (b) two wide-apart three-element loudspeaker arrays (subtending 60°), facing the ears; (c) two wide-apart three-element loudspeaker arrays (subtending 60°), facing the front.

problem at hand, we focus primarily on the parameter design with application in determining array configuration.

The greatest benefit of using the Taguchi method is that, instead of an exhaustive search, much fewer experiments are required in search of the optimal combination of design parameters. This is accomplished by means of orthogonal arrays which are based on statistical experimental design theory. To illustrate, we consider three factors and three levels problem. Assume that no interactions exist and the variation is very small in each observation. The orthogonal array, denoted as $L_9(3^3)$, is shown in Table I, where the numbers 1–3 correspond to three discrete levels of the design factors. According to the table, only nine runs of experiment are required, which is fewer than original 27 searches. The $L_9(3^3)$ orthogonal array is applied to the design of a robust CCS. The parameters to optimize include (A) the configurations of loudspeaker arrays; (B) the distribution of the control points; and (C) the dimension of the CCS matrix. As shown in Fig. 3, the factor (A) is categorized into three levels: (A1) represents the case in which the loudspeakers are closely spaced (six loudspeakers in a row); (A2) represents the case in which two three-element loudspeaker arrays are wide apart (subtending 60°) and facing the ears; (A3) represents the case in which two three-element loudspeaker arrays are wide apart (subtending 60°) and facing the front. The factor (B) is categorized into three levels: (B1) represents the case of five control points in which two points are placed in the illuminated zone and three points are placed in the shadow zone; (B2) represents the case of five control points in which two points are placed in the illuminated zone, one point is placed at the center of the head with 0.5 weighting, and two points are placed in the shadow zone; (B3) represents the case of two control points in which one point is placed at the ipsilateral ear and one point is placed at the contralateral ear. As shown in Fig. 4, the factor (C) is categorized into three levels. (C1) represents the case of a 2×6 CCS in which six loudspeakers are driven with different signals to reproduce the binaural signals (12 filters are needed). (C2) represents the case of a 2×3 CCS in which only a three-element array is considered in the CCS design to focus on the ipsilateral ear and nullify the beam at the contralateral ear loudspeakers. The 2×3 CCS design procedure is applied to each side of the ear to reproduce the binaural signals (six filters are needed). (C3) represents the case of a 2×2 CCS in which only two stereo loudspeakers are driven with different signals to reproduce the binaural signals (four filters are needed).

Both performance and robustness are considered with appropriate weighting W in the objective function

$$f = \text{performance} + W \times \text{robustness}. \qquad (15)$$

To assess the performance and robustness, the channel sepa-

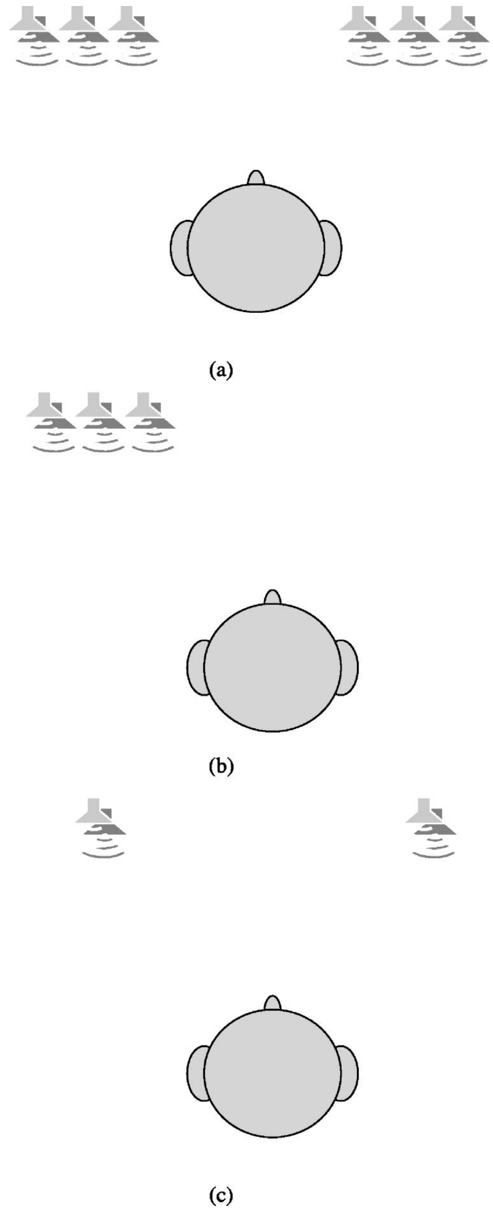


FIG. 4. The dimension of the CCS matrix. (a) a 2×6 CCS in which six loudspeakers are driven with different signals to reproduce the binaural signals. (b) a 2×3 CCS in which only a three-element array is considered in the CCS design to focus on the ipsilateral ear and nullify the beam at the contralateral ear loudspeakers. The 2×3 CCS design procedure is applied to each side of the ear to reproduce the binaural signals. (c) a 2×2 CCS in which only two stereo loudspeakers are driven with different signals to reproduce the binaural signals (four filters are needed).

ration is calculated using the interaural transfer functions (ITFs)

$$ITF_L = \frac{H_{LR}}{H_{LL}}, \quad ITF_R = \frac{H_{RL}}{H_{RR}}, \tag{16}$$

where H_{LR} and H_{RL} are the contralateral frequency responses; H_{LL} and H_{RR} are the ipsilateral frequency responses. The performance function is defined as the channel separation at the nominal position, and the robustness function is defined as the lateral beamwidth when the channel separation drops below −20 dB. The lower the channel separation, the better is the performance of cross-talk cancellation. The larger the beamwidth, the more robust is the CCS

TABLE II. The results of optimal parameters obtained using the Taguchi method. The numbers in the second column are obtained by summing the fitness functions of the corresponding parameter levels. The optimal combinations of parameters of the robust CCS are closely spaced loudspeakers, five control points (with two points on the illuminated zone, three points on the shadow zone), and with an 2×6 CCS matrix.

Levels of parameters	Average of objective function	Chosen level
A1	317.35+74.653+173.62=565.623	Closely spaced
A2	141.81+35.169+80.04=257.019	
A3	43.49+90.706+206.65=340.846	
B1	317.35+141.81+43.49=502.65	5 control points (2 points on the illuminated zone, 3 points on the shadowzone)
B2	74.653+35.169+90.706=200.528	
B3	173.62+80.04+206.65=460.31	
C1	317.35+80.04+90.706=488.096	2×6 CCS
C2	74.653+141.81+206.65=423.113	
C3	173.62+35.169+43.49=252.279	

against lateral misalignment of the listener's head. The results with the weighting $W=10$ are summarized in Table II. From Table II, the optimal parameters (with maximum values of objective function) of the robust CCS are found to be closely spaced arrays, five control points (two points on the illuminated zone, three points on the shadow zone), and a 2×6 CCS matrix.

B. The genetic algorithm

The above-mentioned Taguchi method is more suited to design parameters with finite number of discrete levels. In

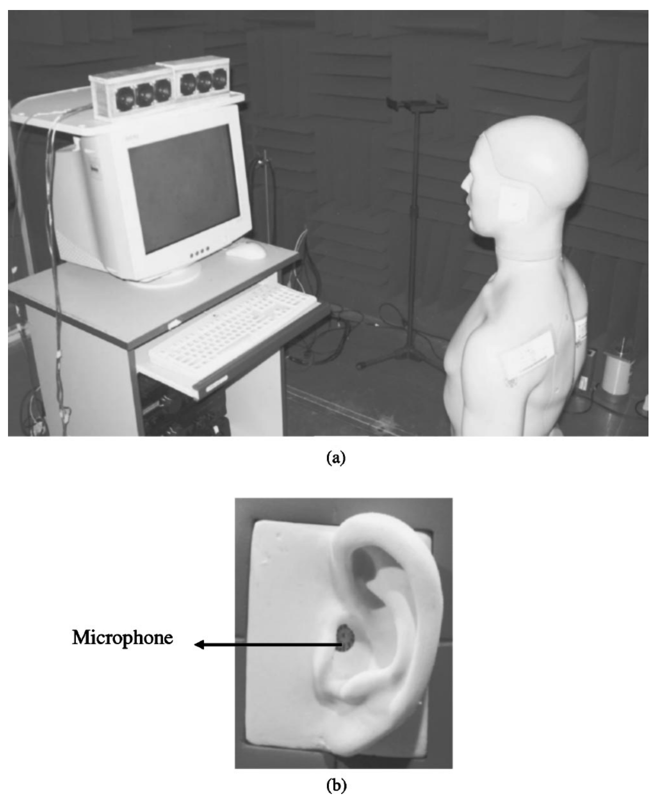


FIG. 5. The photo of the experimental arrangement. (a) The robust CCS with an loudspeaker array. (b) The 1/2-in. condenser microphone embedded in the manikin's ear.

the sequel, an alternative approach that is useful for optimization of continuous parameters is exploited to find the best configuration of the CCS.

1. Encoding and decoding

In the method of GA, all parameters are encoded into binary strings called the *chromosomes*. The resolution of a parameter is dependent on the amount of bits per string and search domain. For instance, we wish to find the optimal spacing $x \in [U_{\min}, U_{\max}]$ ($U_{\min}$ and $U_{\max}$ being the lower limit and the upper limit of the search space) of the loudspeaker array. This parameter is then mapped to an unsigned integer in $[0, 2^l]$, where l is the number of bits. Thus, the resolution of this coding scheme is

$$\Gamma = \frac{U_{\max} - U_{\min}}{2^l - 1}. \tag{17}$$

2. Fitness evaluation

In the GA optimization, the objective one seeks to achieve is termed the fitness function. A chromosome with high fitness has higher probability to survive the natural selection and reproduce offspring in the next generation. The fitness function is the performance function (channel separation) and the robustness function (beam width) with appropriate weighting W

$$f = \text{performance} + W \times \text{robustness}. \tag{18}$$

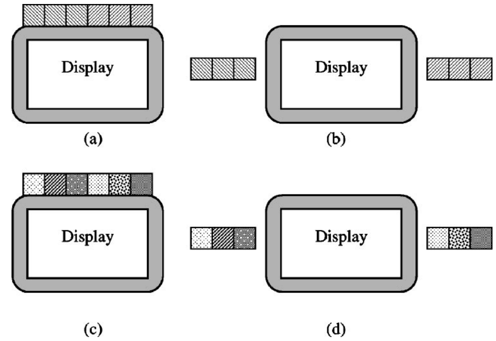


FIG. 6. Illustrations of four loudspeaker array configurations. (a) Configuration 1: closely spaced 2×2 CCS. (b) Configuration 2: wide apart (subtending 60°) 2×2 CCS. (c) Configuration 3: closely spaced 2×6 CCS. (d) Configuration 4: wide-apart (subtending 60°) 2×6 CCS.

3. Reproduction, crossover, and mutation

Reproduction directs the search of GA towards the best individuals. During the process, the reproduction probability of the chromosome is determined by the fitness function. First, the chromosome of the present population is reproduced in the next generation according to the reproduction probability S_i

$$S_i = \frac{f_i}{\sum_{k=1}^{P_l} f}, \tag{19}$$

where P_l is the population size.

Crossover exchanges the contents of chromosomes via probabilistic decision in the mating pool. It is done in three steps. First, the crossover ratio C_r is defined (in general,

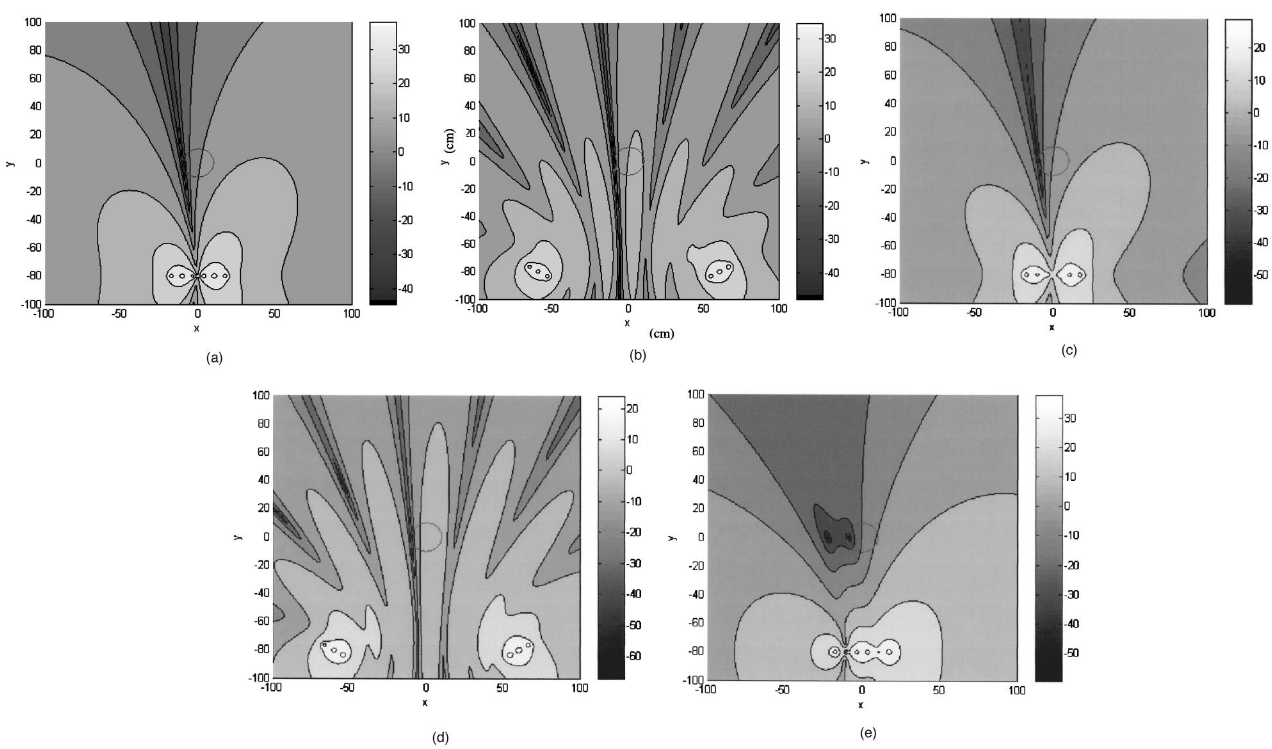


FIG. 7. The contour plots of beam patterns at 1 kHz of various CCS configurations. (a) Configuration 1. (b) Configuration 2. (c) Configuration 3. (d) Configuration 4. (e) Configuration 3 with the optimal 2×6 CCS obtained in the GA procedure.

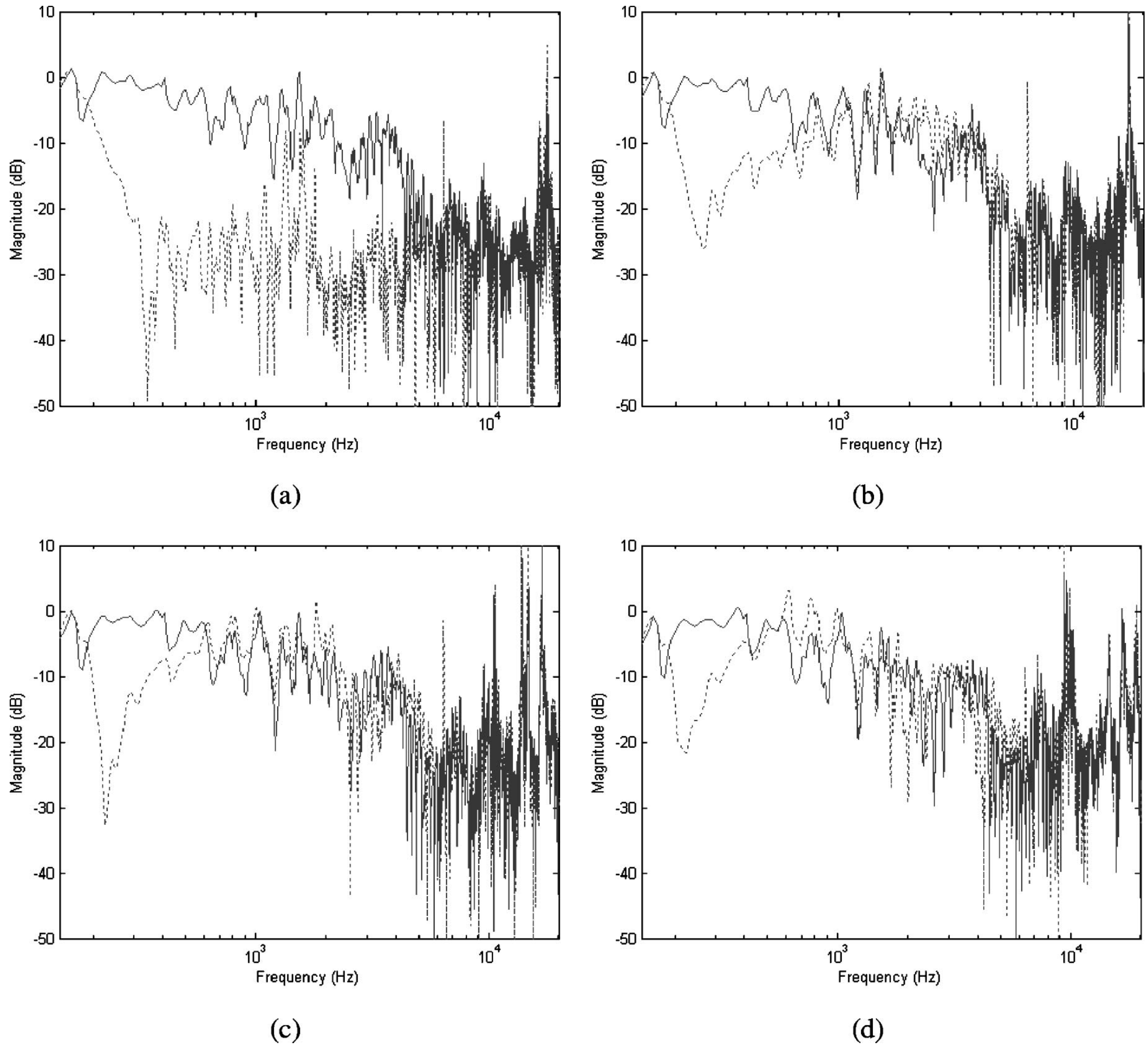


FIG. 8. Channel separations of the left ear obtained using the 2×2 CCS for the wide-apart configurations. The solid lines represent the natural separations and the dotted lines represent the separations with cross-talk cancellation. (a) The channel separation with no displacement. (b) The channel separation with 5-cm displacement to the left. (c) The channel separation with 10-cm displacement to the left. (d) The channel separation with 15-cm displacement to the left.

$0.8 \leq C_r \leq 1$ and we choose $C_r = 0.85$) and two chromosomes in the present population are selected randomly. Second, a splice point at the chromosomes is selected randomly. Third, the chromosomes codes after the splice point are interchanged.

Normally, the chromosomes become increasingly homogeneous as one particular gene begins to dominate after several generations and eventually leads to premature convergence. To obviate this problem, *mutation* is introduced into the GA procedure. Let the mutation ratio be M_r (in general, $0 \leq M_r \leq 0.01$ and we choose $M_r = 0.008$). The mutation point is determined randomly and carried out by alternating the gene from zero to 1, or vice versa. Note, however, that mutation should be used sparingly. The GA would behave like a random search if the mutation rate were too high.

The aforementioned GA procedure was applied to optimize the design of the robust CCS. The design parameters we wish to optimize are similar to those in the Taguchi method, i.e., the spacing between loudspeaker arrays, the distribution of the control points, and the dimension of CCS matrix. When the robustness weighting of fitness function is set to be 1, the optimal design parameters of the robust CCS obtained with the aid of the GA procedure are 0 cm spacing (closely spaced arrays), six control points (one point in the illuminated zone and five points in the shadow zone), and a 2×6 CCS matrix. This result is consistent with the optimal configuration obtained previously using the Taguchi method.

IV. NUMERICAL AND EXPERIMENTAL INVESTIGATIONS

In the paper, the performance of CCS and the associated robustness against head misalignment is examined via numerical and experimental investigations. Only lateral misalignment is considered because it affects the performance of the CCS more significantly than the other types of misalignment.[23,24] The objective performance index is channel separation as defined previously. The experimental arrangement is shown in Fig. 5. A loudspeaker array is mounted on a computer monitor. The distance between the array and the manikin is 80 cm. The loudspeaker array is 10 cm higher than the ears of the manikin. A 1/2-in. condenser

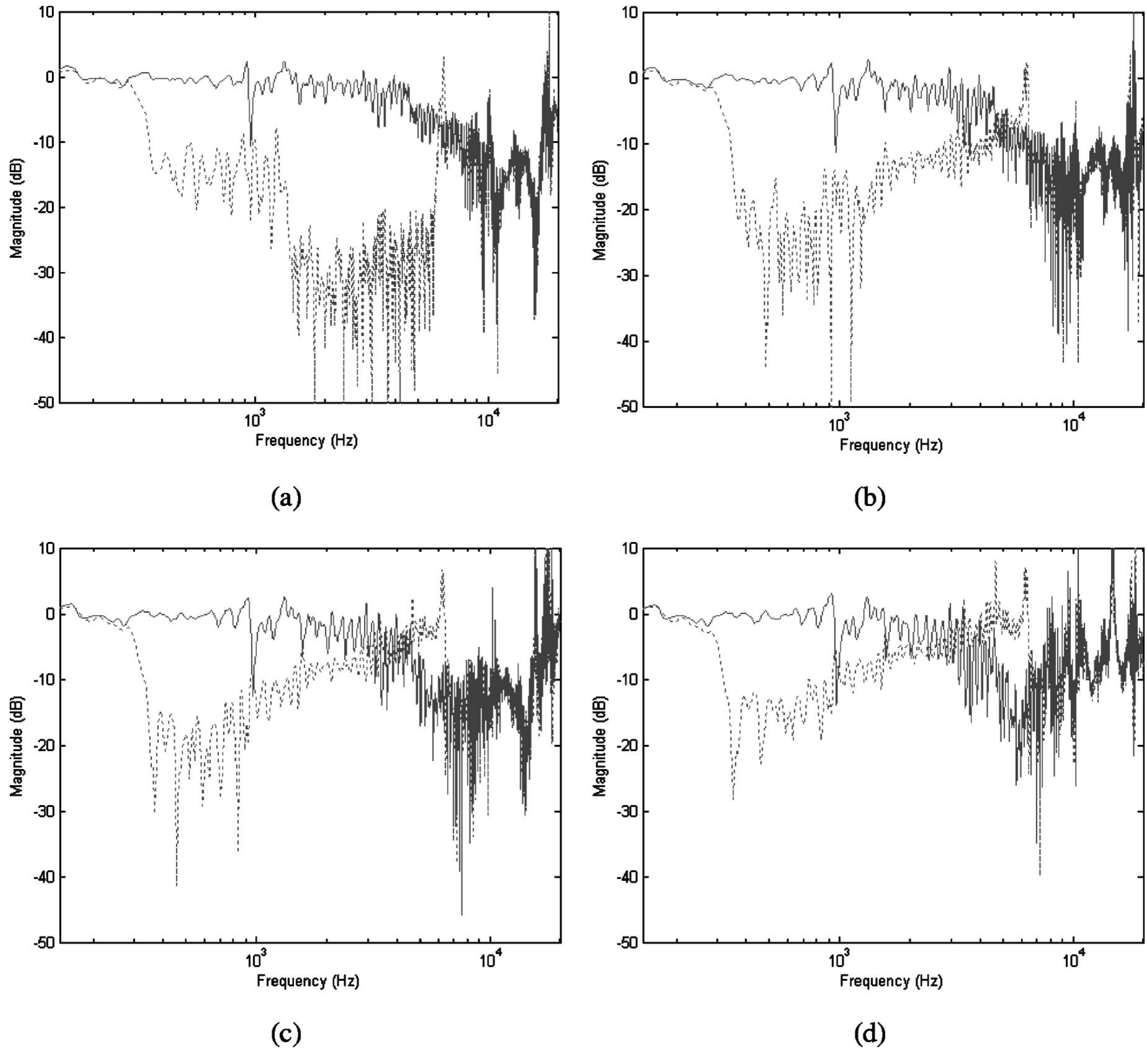


FIG. 9. Channel separations of the left ear obtained using the 2×2 CCS for the closely spaced configuration. The solid lines represent the natural separations and the dotted lines represent the separations with cross-talk cancellation. (a) The channel separation with no displacement. (b) The channel separation with 5-cm displacement to the left. (c) The channel separation with 10-cm displacement to the left. (d) The channel separation with 15-cm displacement to the left.

microphone is fitted inside the ear of the manikin. The sampling rate is 51.2 kHz. The CCS matrix of inverse filtering is calculated by using Eq. (10). The length of each filter is 512 samples and the modeling delay m is 256 samples. The *overlap-add* method is employed to perform block convolution efficiently.[25]

A. Numerical simulations

Before embarking on the experimental investigations, a numerical simulation is carried out to gain more insights into the loudspeaker array configurations in relation to the robustness issue of the CCS. The simulation is conducted for the configurations shown in Fig. 6. In configurations 1 and 2, the 2×2 CCS is simulated, where only one control point is placed in the illuminating zone and another in the shadow zone. There are six loudspeakers in each configuration, where three out of the six loudspeakers form a cluster. The loudspeakers in the same cluster are driven by the same input signal, as indicated by the same pattern of shading. The two clusters are placed side by side in configuration 1, while the two clusters are placed apart (subtending 60°) in configuration 2. In configurations 3 and 4, the 2×6 CCS is simulated. The six loudspeakers are driven by independent signals. Similar to configurations 1 and 2, the only difference between configurations 3 and 4 is whether the loudspeaker clusters are placed side by side or apart. For simplicity, the loudspeakers are assumed to be point sources and the head diffraction as well as room reflection is neglected.

The following contour plots in $x-y$ coordinates compare the beam patterns for the right-ear signals resulting from the foregoing loudspeaker configurations. Only the results of the right-side control are shown. The head and the six loudspeakers are indicated in the figures. The results of configurations 1 and 2 are shown in Figs. 7(a) and (b), respectively. The configuration when all loudspeakers are closely placed results in a wider beam. In contrast, many grating lobes with narrow beamwidth can be seen in the pattern produced by the wide-apart configuration. This shows that the closely spaced configuration is more robust than the wide-apart configuration in cross-talk cancellation, albeit the two CCS perform equally well.

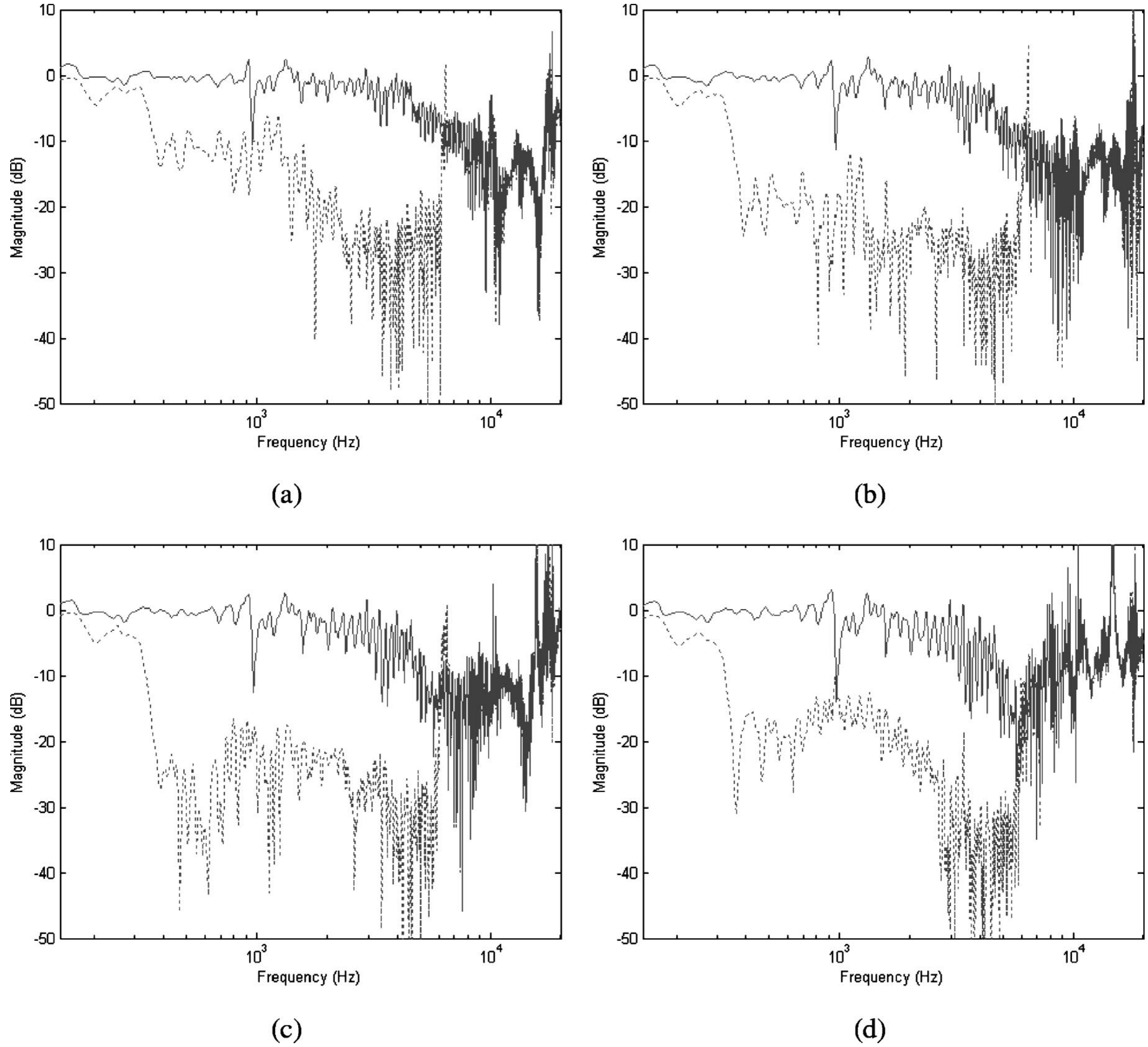


FIG. 10. Channel separations of the left ear obtained using the optimal closely spaced 2×6 configuration designed using six control points (one at the ipsilateral ear and five at the contralateral ear). The solid lines represent the natural separations and the dotted lines represent the separations with cross-talk cancellation. (a) The channel separation with no displacement. (b) The channel separation with 5-cm displacement to the left. (c) The channel separation with 10-cm displacement to the left. (d) The channel separation with 15-cm displacement to the left.

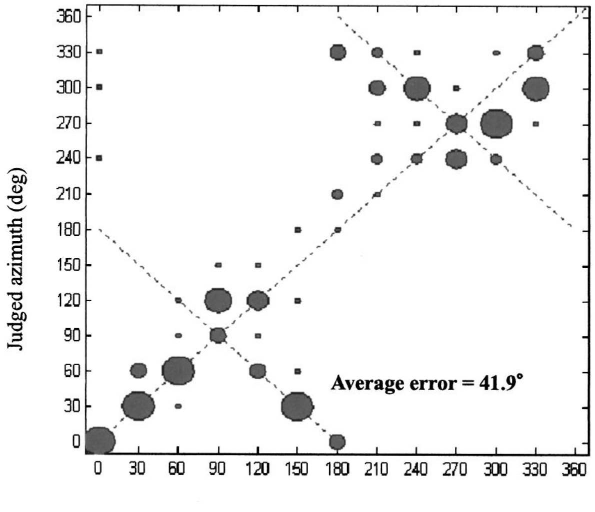
The results of configurations 3 and 4 are shown in Figs. 7(c) and (d), respectively. Inspection of these figures reveals that performance of the two CCS is better than configurations 1 and 2. The wide-apart configuration performs better than the closely spaced configuration, especially at low frequency. However, the closely spaced configuration appears to be more robust than the wide-apart configuration in cross-talk cancellation.

The last four beam patterns are based on the CCS design with only one control point at each ear. Figure 7(e) shows the beam pattern of configuration 3 for the optimal 2×6 CCS obtained in the aforementioned GA procedure. Six control points are used in the design: one at the ipsilateral ear and five at the contralateral ear. As compared to the previous configurations, the sweet spot of the CCS has been effectively widened using the control point technique without significant compromise of cancellation performance.

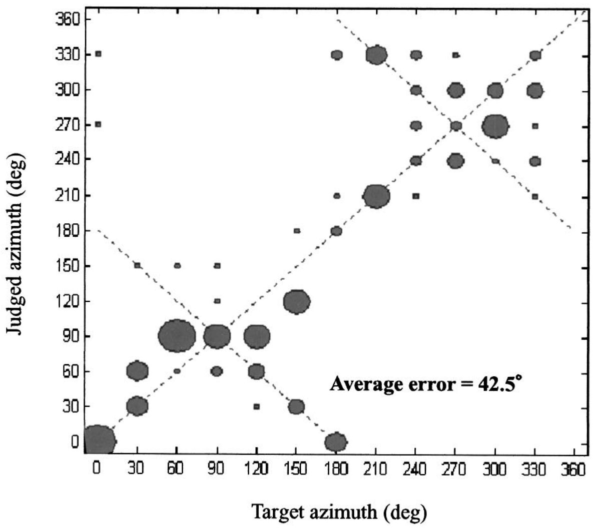
B. Physical tests

In this section, experiments were conducted to examine how channel separation degrades when the listener's head is laterally displaced from the nominal location in the ideal listening scenario. The experiment was performed in an anechoic room, where a CCS bandlimited to 6.4 kHz was tested.

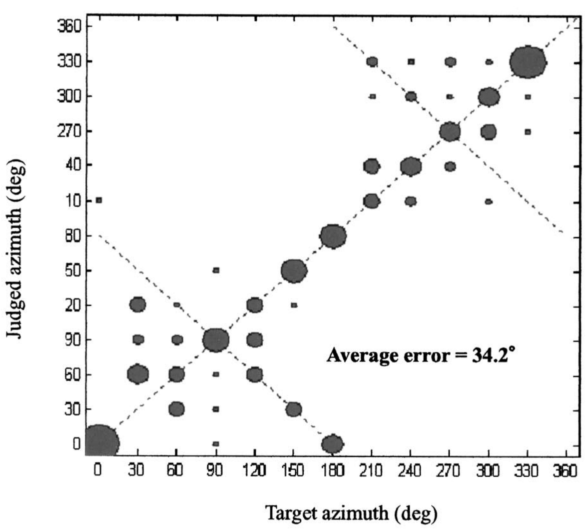
The channel separations of the left ear obtained using the 2×2 CCS are shown in Fig. 8 and Fig. 9 for the wide-apart and closely spaced configurations, respectively. The solid lines represent the natural separations and the dotted lines represent the separations with cross-talk cancellation. In low frequencies, due to diffraction effect, there is almost no natural separation below 400 Hz in the wide-apart configuration (Fig. 8) and below 900 Hz in the closely spaced configuration (Fig. 9). Head shadowing effect becomes visible in high frequencies, where the wide-apart configuration offers better natural separation than the closely spaced configuration. The peaks at higher frequencies result from the inversion of the notches in the ipsilateral responses. Inspection of the results indicates that the 2×2 CCS is not very robust. The performance degrades by 20 dB above 1.5 kHz as the head is displaced leftward by more than 5 cm irrespective of which configuration is used. Nevertheless, the closely spaced con-



(a)

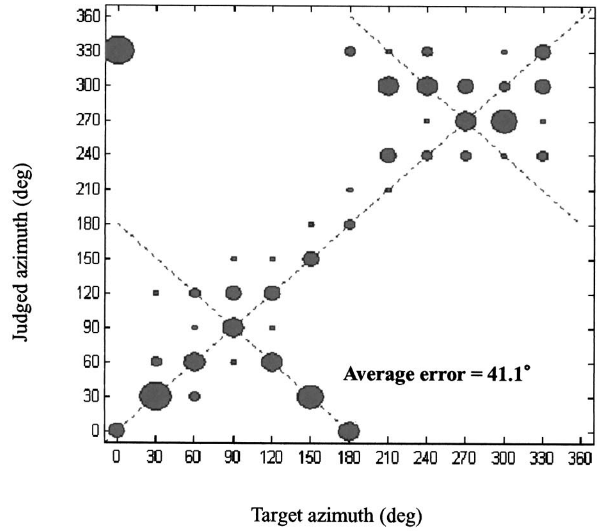


(b)

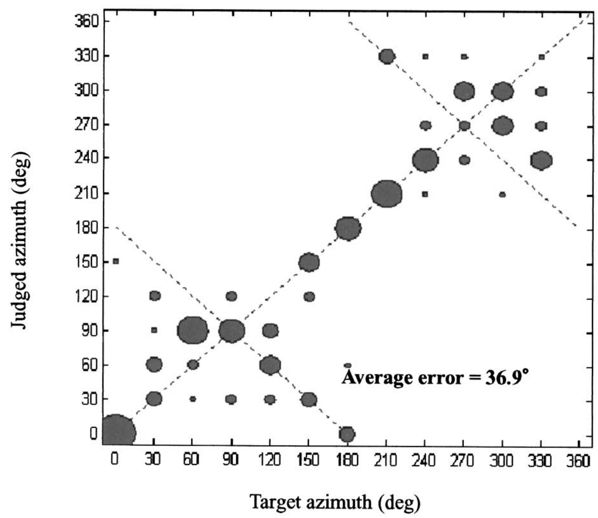


(c)

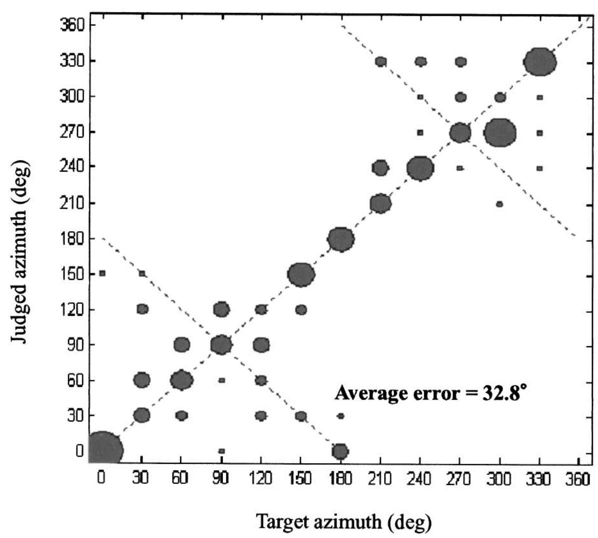
FIG. 11. Azimuth localization results of the subjective test with no head displacement. (a) wide-apart 2×2 CCS; (b) closely spaced 2×2 CCS; (c) The optimal closely spaced 2×6 configuration designed using six control points (one at the ipsilateral ear and five at the contralateral ear).



(a)



(b)



(c)

FIG. 12. Azimuth localization results of the subjective test with 5-cm head displacement to the left. (a) wide-apart 2×2 CCS; (b) closely spaced 2×2 CCS (c). The optimal closely spaced 2×6 configuration designed using six control points (one at the ipsilateral ear and five at the contralateral ear).

figuration in Fig. 9 appears to be slightly more robust than the wide-apart configuration. In Figs. 8(c) and (d), the wide-apart CCS almost lost entire performance above 1 kHz, and the channel separations are nearly the same as the natural channel separations.

In order to improve the robustness of CCS, the optimal closely spaced 2×6 configuration designed using six control points (one at the ipsilateral ear and five at the contralateral ear) was utilized in the next experiment. Figure 10 shows the channel separations obtained using this CCS. It is evident from these plots that the robustness of the 2×6 CCS has been significantly improved over the previous 2×2 CCS. Figure 10(d) shows that the channel separation of the optimal CCS remains as low as −30 dB above 2.5 kHz. The regularization parameter β is frequency dependent and constrained by a 12-dB gain threshold. Because of thus applied regularization, some peaks in Figs. 8–10 can be seen due to imperfect cancellation. From the observation of these results, it is fair to say that large number of loudspeakers, closely spaced configuration, and optimal control point design all contribute to the robustness of CCS against head misalignment.

C. Subjective listening tests

In order to compare various configurations of CCS, a subjective localization experiment was performed in the anechoic room. The test stimulus was a random noise band-limited to 20 kHz. Each stimulus was played for 5-s in duration and switched off for 2 s before the next stimulus was switched on. Virtual sound images at 12 directions on the horizontal plane with increment of 30° azimuth were generated through the filtering of head-related transfer functions (HRTFs). The CCS configurations used in the experiment were the wide-apart 2×2 CCS, the closely spaced 2×2 CCS, and the optimal closely spaced 2×6 CCS. Nine human subjects with normal hearing participated in the experiment.

The experimental results of the judged angles versus the target angles in the localization tests are shown in Figs. 11–13, corresponding to the cases of no misalignment, 5-cm misalignment, and 10-cm misalignment. In each case, all three CCS configurations were tests. The area of each circle is proportional to the number of the listeners who localized the same perceived angle. The 45-deg line represents the perfect localization. The average errors of localization are shown in the figures. As can be seen from the results, the optimal closely spaced 2×6 CCS exhibited remarkable performance and robustness among all configurations. The average localization error using this configuration is only 32.8° (approximately 1 increment of angle) for 5-cm misalignment.

V. CONCLUSIONS

Performance and robustness issues are examined through extensive numerical and experimental investigations. An array beamforming technique using control points is exploited in the design of the CCS filters. Various configurations are compared in the numerical simulations. In terms of the cancellation performance, the wide-apart configura-

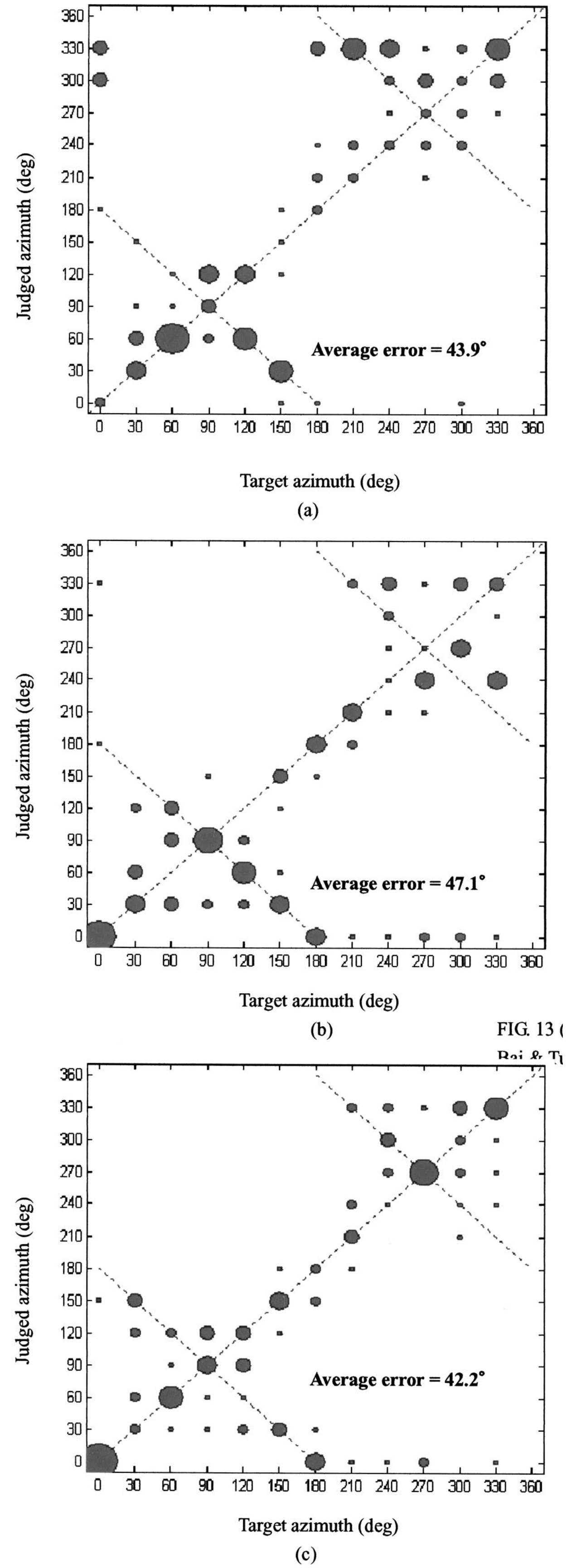


FIG. 13. Azimuth localization results of the subjective test with 10-cm head displacement to the left. (a) wide-apart 2×2 CCS; (b) closely spaced 2×2 CCS. (c) The optimal closely spaced 2×6 configuration designed using six control points (one at the ipsilateral ear and five at the contralateral ear).

TABLE III. Comparison of the CCS configurations. Performance is the average channel separation throughout 20 kHz. Robustness is the lateral displacement (in 5-cm increments) of the head that allows for decay of channel separation within 5 dB. The numbers corresponding to the rows of performance and robustness are experimental data.

	2×2 CCS close	2×2 CCS apart	2×6 CCS close	2×6 CCS apart
Performance	−20 dB	−23 dB	−15 dB	−15 dB
Robustness	±5 cm	±0 cm	±10 cm	±5 cm
Subjective localization test	good	good	excellent	fair
Number of CCS filters	4 filters	4 filters	12 filters	12 filters
Ranking	2	3	1	4

tions could achieve higher channel separation than the closely spaced configurations. However, the closely spaced configurations appear to be more robust than the wide-apart configurations against the lateral misalignment of the head. There is a trade-off that we have to reconcile between the performance and robustness. To facilitate this trade-off, a procedure based on the Taguchi method and the GA has been developed to find optimal configurations of CCS and loudspeaker arrays that attain the best compromise between the performance (channel separation) and robustness (beamwidth). Four configurations are compared by means of objective and subjective experiments. The results are summarized in Table III.

The experimental results indicate that the optimal closely spaced 2×6 CCS is the best choice in terms of performance and robustness. It is fair to say that large number of loudspeakers, closely spaced configuration, and optimal control point design all contribute to the robustness of CCS against head misalignment. Such array design is well suited to equipment that must be spatially compact, e.g., laptop computer, portable audio, mobile phone, etc. A limitation of the 2×6 design of loudspeaker array is that it is more computationally intensive than the 2×2 system. The 2×6 CCS requires 12 filters versus 4 filters in the 2×2 CCS. If computation loading is an issue, however, the closely spaced 2×2 CCS is perhaps the second best choice. Some limitations of the employed optimization methods should also be mentioned. Although the Taguohi method is well suited to problems with discrete levels, the choices must be prespecified. The number of combinations becomes exceedingly large when too many factors to investigate are involved. The same situation happens to the GA; the search requires a very long time to converge for problems with long-coded chromosomes. However, this is not a problem for the CCS in the paper since only loudspeaker spacing is the major design variable. It should be borne in mind that the configuration of the CCS suggested may not be the ultimate optimal, but is the best of the configurations considered.

The horizontally placed loudspeaker array suggested in the paper could have potential impact on the way people implement 3D sound in practical applications. For example, conventional wide-apart stereo loudspeakers are commonplace in PC multimedia and TV applications, but are not effective configurations in the context of 3D audiovisual reproduction. The new loudspeaker configuration proposed in this paper provides a useful alternative.

ACKNOWLEDGMENT

The work was supported by the National Science Council in Taiwan, Republic of China, under the Project Number NSC 92-2212-E009-030.

[1] B. B. Bauer, ''Stereophonic earphones and binaural loudspeakers,'' J. Audio Eng. Soc. **9(2)**, 148–151 (1961).

[2] M. R. Schroeder and B. S. Atal, ''Computer simulation of sound transmission in rooms,'' IEEE Conv. Record. **7**, 150–155 (1963).

[3] P. Damaske and V. Mellert, ''A procedure for generating directionally accurate sound images in the upper- half space using two loudspeakers,'' Acustica **22**, 154–162 (1969).

[4] D. H. Cooper, ''Calculator program for head-related transfer functions,'' J. Audio Eng. Soc. **30**, 34–38 (1982).

[5] W. G. Gardner, ''Transaural 3D audio,'' MIT Media Laboratory Tech. Report, 342, (1995).

[6] A. D. Blumlein, ''Improvements in and relating to sound-transmission, sound-recording and sound-reproducing systems,'' J. Audio Eng. Soc. **6(2)**, 91–99 (1958).

[7] D. H. Cooper and J. L. Bauck, ''Prospects for transaural recording,'' J. Audio Eng. Soc. **37**, 3–19 (1989).

[8] J. L. Bauck and D. H. Cooper, ''Generalized transaural stereo and applications,'' J. Audio Eng. Soc. **44**, 683–705 (1996).

[9] C. Kyriakakis, T. Holman, J. S. Lim, H. Homg, and H. Neven, ''Signal processing, acoustics, and psychoacoustics for high-quality desktop audio,'' J. Visual Commun. Image Represent **9**, 51–61 (1997).

[10] C. Kyriakakis, ''Fundamental and technological limitations of immersive audio systems,'' Proc. IEEE **86**, 941–951 (1998).

[11] D. B. Ward and G. W. Elko, ''Optimal Loudspeaker Spacing for Robust Crosstalk Cancellation,'' Proc. ICASSP 98, IEEE, 3541–3544 (1998).

[12] D. B. Ward and G. W. Elko, ''Effect of loudspeaker position on the robustness of acoustic crosstalk cancellation,'' IEEE Signal Process. Lett. **6(5)**, 106–108 (1999).

[13] T. Takeuchi and P. A. Nelson, ''Robustness to head misalignment of virtual sound imaging systems,'' J. Audio Eng. Soc. **109**, 958–971 (2001).

[14] T. Takeuchi and P. A. Nelson, ''Optimal source distribution for binaural synthesis over loudspeakers,'' J. Audio Eng. Soc. **112**, 2786–2797 (2002).

[15] A. Sibbald, ''Transaural acoustic crosstalk cancellation,'' Sensaura White Papers (1999) (http://www.sensaura.co.uk)

[16] M. Tanter, J.-L. Thomas, and M. Fink, ''Time reversal and the inverse filter,'' J. Acoust. Soc. Am. **108**, 223–234 (2000).

[17] O. Kirkeby, P. A. Nelson, and H. Hamada, ''Fast deconvolution of multichannel systems using regularization,'' IEEE Trans. Speech Audio Process. **6**, 189–194 (1998).

[18] H. C. Robert, Jr., *Fundamental Concepts in the Design of Experiments* (Saunders College Publishing, Philadelphia, 1982).

[19] J. H. Holland, ''Outline for a logical theory of adaptive system,'' J. ACM **3**, 297–314 (1962).

[20] C. T. Lin and C. S. G. Lee, *Neural Fuzzy Systems* (Prentice-Hall, Englewood Cliffs, NJ, 1996).

[21] T. Takeuchi, P. A. Nelson, and H. Hamada, ''Robustness to head misalignment of virtual sound imaging systems,'' J. Acoust. Soc. Am. **109**, 958–971 (2001).

[22] A. Schuhmacher and J. Hald, ''Sound source reconstruction using inverse boundary element calculations,'' J. Acoust. Soc. Am. **113**, 114–127 (2003).

[23] B. S. Atal, M. Hill, and M. R. Schroeder. ''Apparent sound source translator.'' U.S. Patent No. 3236949. 22 Feb., 1966.

[24] W. G. Gardner, *3D Audio using Loudspeakers* (Kluwer Academic, Dordrecht, 1998).

[25] A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing*, 2nd ed. (Prentice-Hall, Englewood Cliffs, NJ, 1999).