Chinese Journal of Electronics

A Stochastic Approximation Method with Enhanced Robustness for Crosstalk Cancellation*

XU Huaxing[1], WANG Qia[2], XIA Risheng[1], LI Junfeng[1] and YAN Yonghong[1,3]

(1. *Key Laboratory of Speech Acoustics and Content Understanding, Institute of Acoustics, Chinese Academy of Sciences, Beijing 100190, China*)

(2. *China LongYuan Power Group Corporation Limited, Beijing 100034, China*)

(3. *Xinjiang Laboratory of Minority Speech and Language Information Processing, Urumqi 830011, China*)

Abstract — The objective of acoustic crosstalk cancellation is to use loudspeakers to deliver prescribed binaural signals (that reproduce a particular auditory scene) to a listener's ears, which is useful for 3-D audio applications. In practice, the actual transfer function matrix will differ from the design matrix, because of either the listener's head movement or rotation, etc. Crosstalk cancellation system (CCS) is very non-robust to these perturbations. Generally, in order to improve the robustness of CCS, several pairs of loudspeakers are needed whose position varies continuously as frequency varies. In this paper, with the help of assumed stochastic analysis, we propose a stochastic robust approximation crosstalk cancellation method based on random perturbation matrix modeling the variations of the transfer function matrix. Under the free-field condition, simulation results demonstrate the effectiveness of the proposed method.

Key words — 3-D audio, Crosstalk cancellation, Stochastic approximation problem, Robustness.

I. Introduction

3-D audio system reconstructs the acoustic pressures at the listener's eardrums and can deliver an extremely realistic three dimensional virtual acoustic environment to the listener, which could be of great benefit in virtual reality, augmented reality, computer multimedia, home theater, video games, digital television, and so forth[1,2,3]. First, the binaural signals are synthesized by appropriately encoding spatial cues corresponding to the desired target scene, which is suitable for headphone production. In practice, headphone binaural audio production suffers from in-head localization and poor frontal imaging[4,5], while playback over loudspeakers is largely immune to these problems. In addition, compared with headphone reproduction, cues by the involvement of the listener's own head, torso and pinnae in sound diffraction and reflection during playback can enhance the perceived realism of sound reproduction[6]. When the binaural signals are reproduced through loudspeakers, a challenging problem emerges: unwanted crosstalk from each speaker to the opposite ear. For two loudspeaker system, that is, the sound emitted from the right loudspeaker also is heard at the left ear, and vice versa. To overcome this problem, a CCS is needed to pre-filter the binaural audio signals before being delivered through loudspeakers.

First introduced by Bauer[7] and later put into practice by Schoeder and Atal[8], the concept of crosstalk cancellation is designed to equalize and reduce the crosstalk to suppress the distortion of the received signals at the listening position. Since then, various sophisticated and effective methods[6,9,10] based on digital signal processing techniques have been developed by different researchers. Generally, the CCS is optimized to achieve optimum cancellation at a given transfer function matrix corresponding to a nominal listener's position. However, there exist many factors that disturb the transfer function matrix, such as the listener's head tiny movement or rotation, noise, etc. All these disturbances or errors have adverse effects on CCS, especially when the CCS is ill-conditioned. The inverse filter may amplify these small perturbations in the transfer function matrix and result in large distor-

*Manuscript Received Oct. 10, 2015; Accepted Jan. 21, 2016. This work is supported by the National Natural Science Foundation of China (No.11461141004, No.61271426, No.11504406, No.11590770, No.11590771, No.11590772, No.11590773, No.11590774), the Strategic Priority Research Program of the Chinese Academy of Sciences (No.XDA06030100, No.XDA06030500), the National High Technology Research and Development Program of China (863 Program) (No.2015AA016306), the National Basic Research Program of China (973 Program) (No.2013CB329302), and the Key Science and Technology Project of the Xinjiang Uygur Autonomous Region (No.201230118-3).

 DOI:10.1049/cje.2017.0X.0XX

tions in the filter's output. To improve the robustness of CCS, the widely used approach is to analyze the condition number of the transfer function matrix and set different pairs of loudspeakers for reproducing different frequency ranges[11,12].

In addition, even with such multiple loudspeakers reproduction, it will still be necessary to provide some inherent robustness during designing the crosstalk cancellation filters. This raises the need for dealing with the robustness of designing crosstalk cancellation filters against slight disturbances or errors. In this paper we try to achieve this goal on the basis of statistical modeling. From the statistics of view, changes in the transfer function caused by all the uncertain disturbances or errors can be modeled as random variables subject to a certain distribution. Different from the aforementioned method[6,9,10], in this paper, a random variable matrix is introduced to characterize the variations of the transfer function matrix between the loudspeakers and the listener. Then the traditional crosstalk cancellation problem turns into a stochastic robust approximation problem[13]. Simulation results demonstrate that this method can improve the crosstalk robustness, especially when the nominal transfer matrix is ill-conditioned.

The rest of this paper is organized as follows: Section II describes the brief overview of generalized crosstalk cancellation problem. Section III illustrates the novel, stochastic robust approximation crosstalk cancellation method. Simulations are performed to test the validity of proposed method in Section IV, and finally, some conclusions are drawn on aspects of the stochastic robust crosstalk cancellation method in Section V.

II. Crosstalk Cancellation Problem Formulation

A classic Atal-Schroeder CCS is illustrated in Fig.1, in which p_L and p_R are the left and right input audio signals, respectively, and $h_n^L(m)$, $n = 1, 2$; $m = 0, \ldots, M-1$, represent the Impulse response(IR) from the nth loudspeaker to the left ear (a similar pair of IRs for the right ear, denoted by $h_n^R(m)$, for concision, are not shown). For simplification, only considering the reproduction of the left audio signal, *i.e.*, $p_R = 0$, and the IR between p_L and the left and right ears, respectively, is

$$\hat{b}_L(n) = h_1^L(m) \star c_1(k) + h_2^L(m) \star c_2(k) \tag{1}$$

$$\hat{b}_R(n) = h_1^R(m) \star c_1(k) + h_2^R(m) \star c_2(k) \tag{2}$$

where $\star$ denotes the linear convolution operator and the symbol $\hat{b}$ is used to distinguish the actual IR, $\hat{b}(n)$, from the desired IR, $b(n)$. In matrix form, it can be mathematically expressed as follows

$$\begin{bmatrix} \hat{\boldsymbol{b}}_{\boldsymbol{L}} \\ \hat{\boldsymbol{b}}_{\boldsymbol{R}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{A}_{\boldsymbol{1}}^{\boldsymbol{L}} & \boldsymbol{A}_{\boldsymbol{2}}^{\boldsymbol{L}} \\ \boldsymbol{A}_{\boldsymbol{1}}^{\boldsymbol{R}} & \boldsymbol{A}_{\boldsymbol{2}}^{\boldsymbol{R}} \end{bmatrix} \begin{bmatrix} \boldsymbol{c_1} \\ \boldsymbol{c_2} \end{bmatrix} \tag{3}$$

where $\hat{\boldsymbol{b}}_{\boldsymbol{L}} = \left[\hat{b}_L(0), \ldots, \hat{b}_L(M+K-2)\right]^{\mathrm{T}}$ is an $(M+K-1) \times 1$ vector ($\hat{\boldsymbol{b}}_{\boldsymbol{R}}$ is similarly defined). $\boldsymbol{A}_{\boldsymbol{1}}^{\boldsymbol{L}} = \mathrm{convmtx}(\left[h_1^L(0), \ldots, h_1^L(M-1)\right]^{\mathrm{T}}, K)$ is an $(M+K-1) \times K$ convolution matrix (similarly for $\boldsymbol{A}_{\boldsymbol{2}}^{\boldsymbol{L}}$, $\boldsymbol{A}_{\boldsymbol{1}}^{\boldsymbol{R}}$ and $\boldsymbol{A}_{\boldsymbol{2}}^{\boldsymbol{R}}$), which is expressed as

$$\boldsymbol{A}_{\boldsymbol{1}}^{\boldsymbol{L}} = \begin{bmatrix} h_1^L(0) & 0 & \cdots & 0 \\ h_1^L(1) & h_1^L(0) & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & h_1^L(M-1) \end{bmatrix} \tag{4}$$

$\boldsymbol{c_1} = [c_1(0), \ldots, c_1(K-1)]^{\mathrm{T}}$ is a $K \times 1$ vector (similarly for $\boldsymbol{c_2}$). A more simplified form in matrix can be expressed as

$$\boldsymbol{A}\boldsymbol{c} = \boldsymbol{b} \tag{5}$$

where the transfer function matrix $\boldsymbol{A}$ is made up by $\boldsymbol{A}_{\boldsymbol{1}}^{\boldsymbol{L}}$, $\boldsymbol{A}_{\boldsymbol{2}}^{\boldsymbol{L}}$; $\boldsymbol{A}_{\boldsymbol{1}}^{\boldsymbol{R}}$ and $\boldsymbol{A}_{\boldsymbol{2}}^{\boldsymbol{R}}$. $\boldsymbol{b} = [\boldsymbol{b_L}, \boldsymbol{b_R}]^{\mathrm{T}}$ is the desired IR vector and $\boldsymbol{c} = [\boldsymbol{c_1}, \boldsymbol{c_2}]^{\mathrm{T}}$ is the corresponding crosstalk cancellation filter coefficient vector.

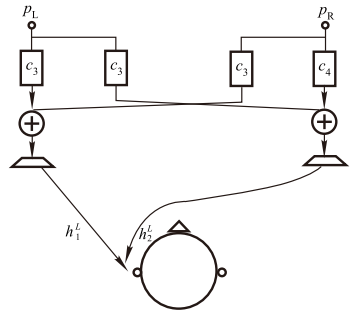


Fig. 1. The typical Atal-Schroder crosstalk cancellation system

A common design criterion for crosstalk cancellation is Least mean squares(LMS), which minimizes the squared distance between a set of desired IRs and the actual IRs obtained with the head in a prescribed position. Ideally, $\boldsymbol{b_L}$ is a pure delay, and $\boldsymbol{b_R}$ is a zero vector. At a given head position, CCS filter coefficients can be achieved by minimizing the following function

$$J_0(\boldsymbol{c}) = \|\boldsymbol{b} - \boldsymbol{A}\boldsymbol{c}\|_2^2 \tag{6}$$

The optimum filter coefficients are obtained as follows

$$\boldsymbol{c_{opt}} = \arg\min J_0(\boldsymbol{c}) = \boldsymbol{A}^{\dagger}\boldsymbol{b} \tag{7}$$

where $\boldsymbol{A}^{\dagger} = (\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}$ is the Moore-Penrose generalized inverse of real-valued $\boldsymbol{A}$. Obviously, such CCS is only effective when the listener is in the prescribed position and the so-called "sweet spot" is too small.

III. Proposed Stochastic Robust Crosstalk Cancellation Method

In mathematics, according to equation $\boldsymbol{Ac} = \boldsymbol{b}$, the crosstalk cancellation can be considered as an approximation problem with basic objective in the different forms (for example, Minmax[14], LMS[9]). In practical applications, the transfer function matrix $\boldsymbol{A}$ is unavoidably influenced by some perturbations and errors due to misalignments, tiny head movement, *etc.* In this section, we consider some statistical characteristics of the variations in $\boldsymbol{A}$ from the statistics of view.

1. Stochastic robust approximation problem

Assuming that $\boldsymbol{A}$ is a random variable taking values in $\mathbb{R}^{m\times n}$ with mean $\bar{\boldsymbol{A}}$, $\boldsymbol{A}$ can be described as $\boldsymbol{A} = \bar{\boldsymbol{A}} + \boldsymbol{U}$, where $\boldsymbol{U}$ is a random matrix with zero mean. Here, the constant matrix $\bar{\boldsymbol{A}}$ represents the average value of $\boldsymbol{A}$, and $\boldsymbol{U}$ describes its statistical variation. Naturally, using the expected value as the objective function, we can get

$$\min E\|\boldsymbol{Ac} - \boldsymbol{b}\| \tag{8}$$

where E represents the mathematical expectation. This problem is referred as the stochastic robust approximation problem[15]. As a simple case, when $\boldsymbol{A}$ is a discrete random variable with only a finite number of values, *i.e*,

$$\text{prob}(\boldsymbol{A} = \boldsymbol{A_i}) = p_i, i = 1, \ldots, k \tag{9}$$

where prob means the probability of different $\boldsymbol{A_i} \in \mathbb{R}^{m\times n}$, $\mathbf{1}^{\mathrm{T}}\boldsymbol{p} = 1$, $\boldsymbol{p} \geq 0$. In this case, the problem has the form

$$\min\left(p_1\|\boldsymbol{A_1c} - \boldsymbol{b}\| + \ldots + p_k\|\boldsymbol{A_kc} - \boldsymbol{b}\|\right) \tag{10}$$

So both the joint multi-position optimization[16] and multi-position weighted optimization[17] for crosstalk cancellation can be seen as a special case of the stochastic robust approximation, given by Eq.(10). Considering the LMS criterion, the statistical robust LMS method for crosstalk cancellation can be described as

$$\min E\|\boldsymbol{Ac} - \boldsymbol{b}\|_2^2 \tag{11}$$

Further it can be derived as

$$\begin{aligned} E\|\boldsymbol{Ac} - \boldsymbol{b}\|_2^2 &= E(\bar{\boldsymbol{A}}\boldsymbol{c} - \boldsymbol{b} + \boldsymbol{Uc})^{\mathrm{T}}(\bar{\boldsymbol{A}}\boldsymbol{c} - \boldsymbol{b} + \boldsymbol{Uc}) \\ &= (\bar{\boldsymbol{A}}\boldsymbol{c} - \boldsymbol{b})^{\mathrm{T}}(\bar{\boldsymbol{A}}\boldsymbol{c} - \boldsymbol{b}) + E\{\boldsymbol{c}^{\mathrm{T}}\boldsymbol{U}^{\mathrm{T}}\boldsymbol{Uc}\} \\ &= \|\bar{\boldsymbol{A}}\boldsymbol{c} - \boldsymbol{b}\|_2^2 + \boldsymbol{c}^{\mathrm{T}}\boldsymbol{Pc} \end{aligned} \tag{12}$$

where $\boldsymbol{P} = E\{\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U}\}$ corresponds to mathematical expectation of the autocorrelation matrix of the perturbation matrix $\boldsymbol{U}$. Therefore the statistical robust approximation problem has the form of a regularized least-squares problem[10]

$$\min \|\bar{\boldsymbol{A}}\boldsymbol{c} - \boldsymbol{b}\|_2^2 + \|\boldsymbol{P}^{1/2}\boldsymbol{c}\|_2^2 \tag{13}$$

with analytical solution

$$\boldsymbol{c_{opt}} = (\bar{\boldsymbol{A}}^{\mathrm{T}}\bar{\boldsymbol{A}} + \boldsymbol{P})^{-1}\bar{\boldsymbol{A}}^{\mathrm{T}}\boldsymbol{b} \tag{14}$$

This makes perfect sense: when the matrix $\boldsymbol{A}$ is subject to variation, the vector $\boldsymbol{Ac}$ will have more variation the larger $\boldsymbol{c}$ is, and Jensen's inequality tells us that variation in $\boldsymbol{Ac}$ will increase the average value of $\|\bar{\boldsymbol{A}}\boldsymbol{c} - \boldsymbol{b}\|_2$. Then we need to balance making $\bar{\boldsymbol{A}}\boldsymbol{c} - \boldsymbol{b}$ small with the desire for a small $\boldsymbol{c}$ (to keep the variation in $\boldsymbol{Ac}$ small), which is the essential idea of regularization[13]. The solution of the Tikhonov regularized least-squares problem for crosstalk cancellation in time domain can also be seen a special case in problem minimizing with where $\boldsymbol{U_{ij}}$ are zero mean, uncorrelated random variables.

2. Modeling the random perturbation matrix

In the following, we model the variations of the transfer functions in a way to improve the spatial robustness of crosstalk cancellation from a statistical point of view. Without loss of generality, we model the perturbation $\xi_i^L (i = 1, 2)$ on the transfer function from the loudspeakers to listener's left ear (modeling $\xi_i^R (i = 1, 2)$ similarly) as a statistical variable with zero mean and variance $\sigma_i^L (i = 1, 2)$. The perturbed transfer function is expressed as $u_i^L = \xi_i^L h_i^L (i = 1, 2)$. Further, the perturbation matrix $\boldsymbol{U}$ is denoted as

$$\boldsymbol{U} = \begin{bmatrix} \xi_1^L\bar{\boldsymbol{A}}_{\boldsymbol{1}}^{\boldsymbol{L}} & \xi_2^L\bar{\boldsymbol{A}}_{\boldsymbol{2}}^{\boldsymbol{L}} \\ \xi_1^R\bar{\boldsymbol{A}}_{\boldsymbol{1}}^{\boldsymbol{R}} & \xi_2^R\bar{\boldsymbol{A}}_{\boldsymbol{2}}^{\boldsymbol{R}} \end{bmatrix} \tag{15}$$

The expectation matrix $\boldsymbol{P}$ of autocorrelation of the perturbation matrix $\boldsymbol{U}$ is expressed as

$$\begin{aligned} \boldsymbol{P} &= E\{\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U}\} \\ &= E\left\{\begin{bmatrix} \xi_1^L\bar{\boldsymbol{A}}_{\boldsymbol{1}}^{\boldsymbol{L}} & \xi_2^L\bar{\boldsymbol{A}}_{\boldsymbol{2}}^{\boldsymbol{L}} \\ \xi_1^R\bar{\boldsymbol{A}}_{\boldsymbol{1}}^{\boldsymbol{R}} & \xi_2^R\bar{\boldsymbol{A}}_{\boldsymbol{2}}^{\boldsymbol{R}} \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} \xi_1^L\bar{\boldsymbol{A}}_{\boldsymbol{1}}^{\boldsymbol{L}} & \xi_2^L\bar{\boldsymbol{A}}_{\boldsymbol{2}}^{\boldsymbol{L}} \\ \xi_1^R\bar{\boldsymbol{A}}_{\boldsymbol{1}}^{\boldsymbol{R}} & \xi_2^R\bar{\boldsymbol{A}}_{\boldsymbol{2}}^{\boldsymbol{R}} \end{bmatrix}\right\} \\ &= \begin{bmatrix} \boldsymbol{P_1^L} & \boldsymbol{P_2^L} \\ \boldsymbol{P_1^R} & \boldsymbol{P_2^R} \end{bmatrix} \end{aligned} \tag{16}$$

where $\boldsymbol{P_1^L} = E\{(\xi_1^L)^2(\bar{\boldsymbol{A}}_{\boldsymbol{1}}^{\boldsymbol{L}})^{\mathrm{T}}(\bar{\boldsymbol{A}}_{\boldsymbol{1}}^{\boldsymbol{L}}) + (\xi_1^R)^2(\bar{\boldsymbol{A}}_{\boldsymbol{1}}^{\boldsymbol{R}})^{\mathrm{T}}(\bar{\boldsymbol{A}}_{\boldsymbol{1}}^{\boldsymbol{R}})\}$, $\boldsymbol{P_2^L}$, $\boldsymbol{P_1^R}$ and $\boldsymbol{P_2^R}$ denoted similarly. Because these perturbations are uncertainty in practice, further assuming all the perturbation random variables Independent and identically distributed (IID) with zero mean and variance σ, and then the anti-diagonal block elements $\boldsymbol{P_2^L}$, $\boldsymbol{P_1^R}$ of the $\boldsymbol{P}$ matrix are reduced to zeros. Finally, the $\boldsymbol{P}$ matrix is expressed as

$$\boldsymbol{P} = \sigma^2\begin{bmatrix} \boldsymbol{P_{11}} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{P_{22}} \end{bmatrix} \tag{17}$$

where $\boldsymbol{P_{11}} = (\bar{\boldsymbol{A}}_{\boldsymbol{1}}^{\boldsymbol{L}})^{\mathrm{T}}\bar{\boldsymbol{A}}_{\boldsymbol{1}}^{\boldsymbol{L}} + (\bar{\boldsymbol{A}}_{\boldsymbol{1}}^{\boldsymbol{R}})^{\mathrm{T}}\bar{\boldsymbol{A}}_{\boldsymbol{1}}^{\boldsymbol{R}}$ and $\boldsymbol{P_{22}} = (\bar{\boldsymbol{A}}_{\boldsymbol{2}}^{\boldsymbol{L}})^{\mathrm{T}}\bar{\boldsymbol{A}}_{\boldsymbol{2}}^{\boldsymbol{L}} + (\bar{\boldsymbol{A}}_{\boldsymbol{2}}^{\boldsymbol{R}})^{\mathrm{T}}\bar{\boldsymbol{A}}_{\boldsymbol{2}}^{\boldsymbol{R}}$.

IV. Experimental Verification and Analysis

In this section, we evaluate the performance of the proposed stochastic robust LMS method in comparison with the traditional LMS method by simulations.

1. Performance metrics

The Channel separation (CHS) is employed as the evaluation measure for crosstalk cancelation. CHS is defined as the ratio between the desired signal and the crosstalk signal. In our case, because the input right signal is set zero in prior, the signal received by the listener's left ear is the desired signal and the signal received by the listener's right ear is the crosstalk signal. In this case, the CHS is expressed as

$$CHS = 20\lg|\frac{b_L(k)}{b_R(k)}| \tag{18}$$

where k denotes different discrete frequencies. The average CHS is defined as

$$\overline{CHS} = \frac{1}{n_L - n_H + 1}\sum_{k=n_L}^{n_H} CHS(k) \tag{19}$$

where n_L and n_H are the entire frequency ranges of interest.

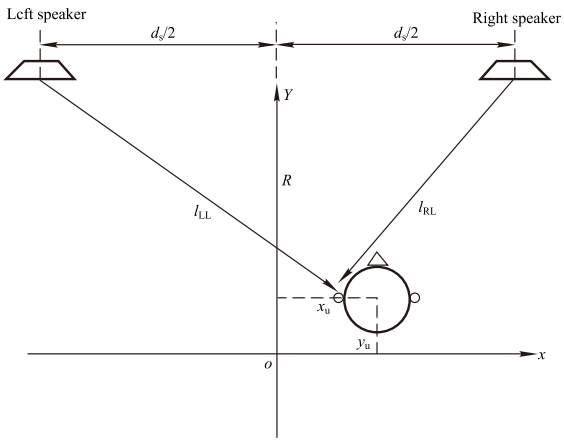


Fig. 2. The schematic of listener's head movement in experiment

2. Experimental setup

In order to verify the robustness of the proposed algorithm, the slight movement of the listener's head is chosen as the disturbance factor among all the disturbance factors. The average CHSs of different listener's head positions are computed and compared with traditional LMS method. The frequency sample rate is 16kHz and the frequency range of interest for computing average channel separation is chosen between 200–5000Hz. The schematic diagram of the experiment is illustrated in Fig.2.

The loudspeakers are separated by a distance of d_s = 0.5m and an $\theta = 5°$ from the default listening position (with the head placed symmetrically between the loudspeakers) corresponding to the "stereo dipole" setup[18]. The vertical distance R between the center of the nominal listener's head position to the two speakers is 0.5m. Because of fundamentally difficult problem to achieve good crosstalk cancellation at low frequencies, the expected $b_L(n)$ was designed to have a high-pass response with a cut-off frequency of 200Hz. The optimum delay for crosstalk cancellation is calculated according to the paper[16]. The region x_u for listener's head slight movement is chosen between (1, 2, 3, 4)cm corresponding to the head movement towards the right (the crosstalk cancellation is more effective as the head moves forwards/backwards than if it move sideways[19] and further, due to the symmetry only consider the right movement and set y_u=0). Under free-field condition, the transfer function (for example, the left speaker to the left ear) in frequency domain is expressed as

$$H_{LL}(\omega) = \frac{1}{4\pi l_{LL}}\mathrm{e}^{-jkl_{LL}} \tag{20}$$

Where l_{LL} is the distance from the left loudspeaker to the listener's left ear, $k = \omega/c_s$ is the wave number and c_s is the sound speed set by 340m/s.

3. The analysis of experimental result

For the proposed stochastic LMS crosstalk cancellation method, the optimal proper σ of the perturbation need to be determined in advance. A series of tests were conducted by changing the variance σ range (0.01–1) with interval 0.005. For $\theta = 5°$, the average CHS designed according to the traditional LMS method and stochastic LMS method with different σ are shown in Fig.3. In Fig.3, the lines from top to bottom describe different head movement positions from x_u=1cm to x_u=4cm with different line styles represent different methods, specifically, thick line for the stochastic LMS crosstalk cancellation method and dot line for the traditional LMS crosstalk cancellation method.

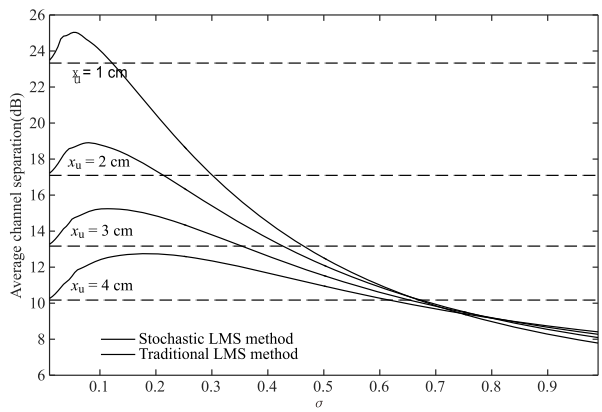


Fig. 3. Comparison of the average CHSs at different head positions: from top to bottom x_u=1cm, 2cm, 3cm, 4cm for $\theta = 5°$ with different σ (thick line for the stochastic LMS method and dot line for the traditional LMS method)

Strictly speaking, for the traditional LMS crosstalk cancellation method, the average CHS of different head position is a specific value and does not vary as the σ varies. For comparison, it is described as a straight line. It's clear from the Fig.3 that in the vicinity of the $\sigma = 0.1$, the average CHS of the proposed stochastic LMS crosstalk

cancellation method is higher than the corresponding traditional LMS method, demonstrating that the proposed method is robust against the listener's slight movement. Further, when the variance σ=0.1, the average CHSs of the two methods are illustrated in Table 1 along with the improvement.

Table 1. The comparison of CHS with σ=0.10 for $\theta = 5°$

Head position (cm)	Traditional (dB)	Proposed (dB)	% improvement
1	23.33	24.05	3.07
2	17.09	18.80	9.94
3	13.17	15.23	15.64
4	10.17	12.41	21.99

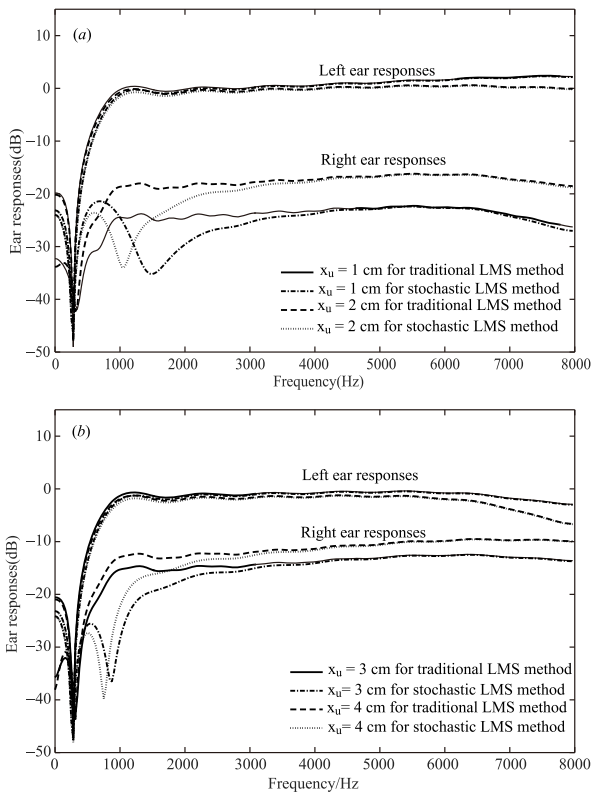


Fig. 4. Ear responses at head positions with different line styles representing the stochastic LMS method and traditional LMS methods for $\theta = 5°$: (*a*) x_u=1cm, 2cm; (*b*) x_u=3cm, 4cm

For showing the improvement clearly, the listener's ear responses are drawn in Fig.4 with different head position. Ideally, the left ear response should be unity (above 200Hz) and the right ear response should be zero. As can be seen from the Fig.3 and Fig.4, compared with the traditional LMS method, introducing the perturbation brings improved CHS in the vicinity of 1000Hz. As can be seen from Eq.(6) and Eq.(13), its analytical solution involves matrix inversion. The condition number of the inverse matrix implies the sensitiveness to perturbations. For the traditional LMS method, the condition number is expressed as $\kappa_1 = cond(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}) = \frac{\sigma_{\max}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})}{\sigma_{\min}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})}$, and its value is 2002. For the proposed stochastic robust algorithm, the condition number is expressed as $\kappa_2 = cond(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} + \boldsymbol{P}) = \frac{\sigma_{\max}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}+\boldsymbol{P})}{\sigma_{\min}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}+\boldsymbol{P})}$, and its value is 182 and less than κ_1, which further confirms the enhanced robustness of proposed method against perturbations.

Under the free-field condition, the analysis of the transfer function matrix[18] concluded that, for a given loudspeaker angle, its robust frequency range is determined by the so-called "Ring frequency" (RF), which is inversely proportional to the angle. It indicates that the crosstalk cancellation is inherent non-robust in the frequency range above the RF. The RF of the "stereo dipole" setup is about 11kHz, which is beyond the scope of the frequency (8kHz) considered in our experiment. For further comparison, the loudspeaker angle is increased to 10° where the RF is about 5.6kHz. Repeat the experiment with other parameters keeping the same.

Similar to Fig.3, for $\theta = 10°$, the CHS designed according to the traditional LMS method and stochastic LMS method with different variance σ and head positions are shown in Fig.5.

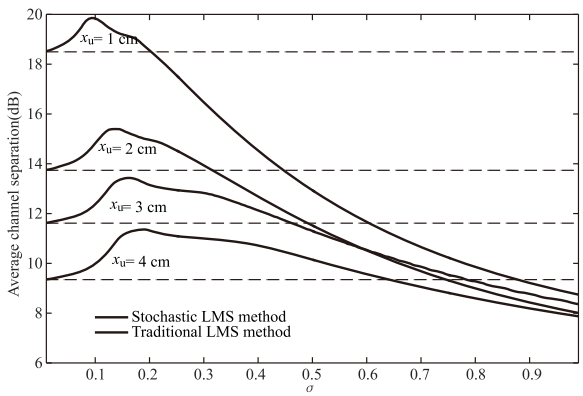


Fig. 5. Comparison of the average CHSs at different head positions: from top to bottom x_u=1cm, 2cm, 3cm, 4cm for $\theta = 10°$ with different σ (thick line for the stochastic LMS method and dot line for the traditional LMS method)

Table 2. The comparison of CHS with σ=0.15 for $\theta = 10°$

Head position (cm)	Traditional (dB)	Proposed (dB)	% improvement
1	18.49	19.20	3.78
2	13.73	15.35	11.72
3	11.61	13.39	15.27
4	9.35	11.13	19.02

As can be seen from Fig.5, in the vicinity of the σ = 0.15, the average CHS of the proposed stochastic LMS crosstalk cancellation method is still higher than the traditional LMS method, which shows good agreement with the first experiment. When the variance $\sigma = 0.15$, the average CHS of the two methods is illustrated in Table 2

along with the improvement. Also, the listener's ear responses are drawn in Fig.6 with different head movement.

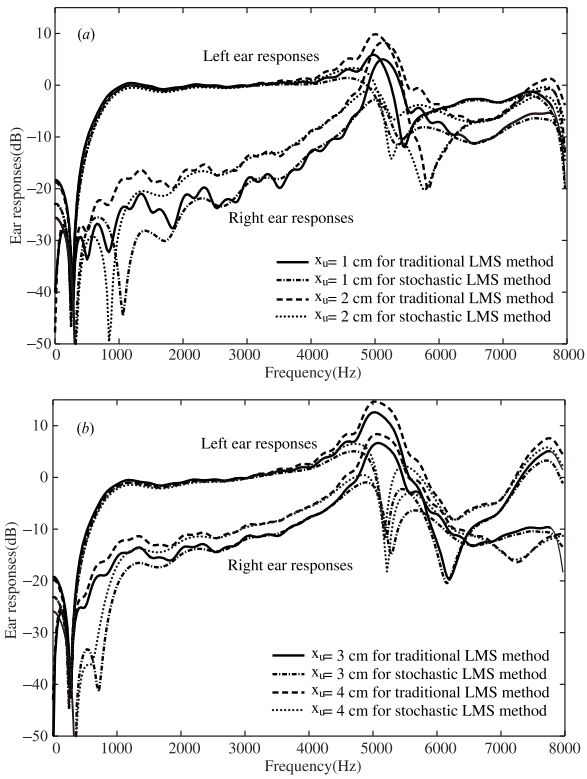


Fig. 6. Ear responses at head positions with different line styles representing the stochastic LMS method and traditional LMS methods for $\theta = 10°$: (*a*) x_u=1cm, 2cm; (*b*) x_u=3cm, 4cm

From discussions described above, there exists non-robust frequency point corresponding the RF in the vicinity of 5000Hz. The perturbation introduced by the listener's slight head movement results in the rapid decrease of its performance and the spectral distortion. The proposed stochastic robust LMS method not only improves channel separation in the vicinity of 1000Hz, but also greatly reduces the spectral distortion around the RF. This confirms the expectation that proposed stochastic LMS crosstalk cancellation method provides enhanced robustness. Further, using the condition number as a measure index, the values of κ_1 and κ_2 are 1145 and 252, respectively. It also demonstrates the enhanced robustness of the stochastic LMS method.

V. Conclusions

A novel stochastic robust LMS crosstalk cancellation method based statistical modeling was proposed for the designing of crosstalk cancellation filters. A random perturbation matrix modeling the variation due to perturbations is introduced and lied in parallel to the actual nominal transfer matrix during the design process. Simulation results demonstrated that the proposed method is robust against listener's slight head movement.

References

[1] Y. Huang, J. Chen and J. Benesty, "Immersive audio schemes" *IEEE Signal Processing Magazine*, Vol.28, No.1, pp.20–32, 2011.

[2] B.S. Xie, *Head Related Transfer Function and Virtual Auditory*, National Defense Industry Press, Beijing, China, pp.330–324, 2008. (in Chinese)

[3] Y.C. Zhang, J. Chen and Y.T. Wang, "City-scale location services based on mobile augmented reality", *Chinese Journal of Electronics*, Vol.42, No.8, pp.1503–1508, 2013. (in Chinese)

[4] W.G. Gardner, "3-D audio using loudspeakers", *Ph.D.Thesis*, Massachusetts Institute of Technology, USA, 1997.

[5] R.S. Xia, J.F. Li, C.D. Xu, *et al.*, "A sound image externalization approach for headphone reproduction by simulating binaural room impulse responses", *Chinese Journal of Electronics*, Vol.23, No.3, pp.527–532, 2014.

[6] E. Choueiri, "Optimal crosstalk cancellation for binaural audio with two loudspeakers", available at *https://www.princeton.edu/3D3A/Publications/BACCHPaperV4d.pdf*, 2008/2015-10-10.

[7] B.B. Bauer, "Stereophonic earphones and binaural loudspeakers", *Journal of the Audio Engineering Society*, Vol.9, No.2, pp.148–151, 1961.

[8] B.S. Atal and M.R. Schroeder, "Apparent sound source translator", *Patent*, 3236949, USA, 1966-2-22.

[9] P. Nelson, H. Hamada and S.J. Elliott, "Adaptive inverse filters for stereophonic sound reproduction", *IEEE Transactions on Signal Processing*, Vol.40, No.7, pp.1621–1632, 1992.

[10] O. Kirkeby, P. Nelson, H. Hamada, *et al.*, "Fast deconvolution of multichannel systems using regularization", *IEEE Transactions on Speech and Audio Processing*, Vol.6, No.2, pp.189–194, 1998.

[11] T. Takeuchi and P.A. Nelson, "Optimal source distribution for binaural synthesis over loudspeakers", *The Journal of the Acoustical Society of America*, Vol.112, No.6, pp.2786–2797, 2002.

[12] J. Zheng, J. Lu and X. Qiu, "Linear optimal source distribution mapping for binaural sound reproduction", *Proc. of INTER-NOISE and NOISE-CON Congress and Conference, Institute of Noise Control Engineering*, Vol.249, No.7, pp.939–946, 2014.

[13] S. Boyd and L. Vandenberghe, "Convex optimization", *Cambridge University Press*, Cambridge, UK, pp.332–333, 2004.

[14] H.I. Rao, V.J. Mathews and Y.C. Park, "A minimax approach for the joint design of acoustic crosstalk cancellation filters", *IEEE Transactions on Audio, Speech, and Language Processing*, Vol.15, No.8, pp.2287–2298, 2005.

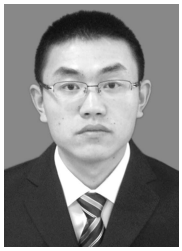
[15] L.E. Ghaoui and H. Lebret, "Robust solutions to least-squares problems with uncertain data", *SIAM Journal on Matrix Analysis and Applications*, Vol.18, No.4, pp.1035–1064, 1997.

[16] D.B. Ward, "Joint least squares optimization for robust acoustic crosstalk cancellation", *IEEE Transactions on Speech and Audio Processing*, Vol.8, No.2, pp.211–215, 2000.

[17] J. Wang, Q.H. Ye, C.S. Zhen, *et al.*, "A robust algorithm for binaural audio reproduction using loudspeakers", *Proc. of International Conference on Measuring Technology and Mechatronics Automation, ICMTMA*, Vol.1, pp.318–321, 2010.

[18] O. Kirkeby, P.A. Nelson and H. Hamada, "The "stereo dipole: a virtual source imaging system using two closely spaced loudspeakers", *Journal of the Audio Engineering Society*, Vol.46, No.5, pp.387–395, 1998.

[19] D.B. Ward and G.W. Elk, "Optimum loudspeaker spacing for robust crosstalk cancellation", *Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing*, Vol.6, pp.3541–3544, 1998.



XU Huaxing was born in 1988. He received the B.E. degree in electronic engineering degree in Communication Engineering from JiLin University in 2012. He is now a Ph.D. candidate of the Key Lab of Speech Acoustics and Content Understanding, Institute of Acoustics, Chinese Academy of Sciences. His research interests include 3D audio processing. (Email: xuhx1314@sina.com)

WANG Qia She received the B.E. degree from Beijing Forestry University in Applied Mathematics in 2008 and the Ph.D. degree from Academy of Mathematics and Systems Science, Chinese Academy of Sciences in Applied Mathematics in 2013. Her research interests include mathematical optimization, operational research and control.

XIA Risheng was born in 1978. He received the Ph.D. degree from Institute of Acoustics, Chinese Academy of Sciences in 2013. He has worked as a embedded system designer for several years, mainly about speech communication products. His research interests include 3D audio processing, speech coding, and embedded system design.

LI Junfeng (corresponding author) was born in 1979. He received the B.E. degree from Zhengzhou University and the M.S. degree from Xidian University both in Computer Sciences in 2000 and 2003. He received the Ph.D. degree in Information Science from Japan Advanced Institute of Science and Technology (JAIST) in March 2006. From April 2006, he was a post-doctoral research fellow at Research Institute of Electrical Communication (RIEC), Tohoku University. From April 2007 to July 2010, he was an Assistant Professor in School of Information Science, JAIST. Since August 2010, he has been a Professor in Institute of Acoustics, Chinese Academy of Sciences. His research interests include psychoacoustics, acoustic signal processing and 3D audio technology. Dr. Li received the Best Student Award in Engineering Acoustics First Prize from the Acoustical Society of America in 2006, and the Best Paper Award from JCA2007 in 2007, and the Itakura Award from the Acoustical Society of Japan in 2012. Dr. Li is now serving as the Subject Editor for Speech Communication and the Editor for IEICE Trans. on Fundamentals of Electronics, Communication and Computer Sciences. (Email: lijunfeng@hccl.ioa.ac.cn)

YAN Yonghong was born in 1967. He received the B.E.degree from the Electronic Engineering Department of Tsinghua University in 1990, and Ph.D. degree in Computer Science and Engineering from Oregon Graduate Institute of Science and Engineering in 1995. From 1995 to 1998, he worked in OGI as an Assistant Professor, Associate Director and Associate Professor of the Center for Spoken Language Understanding. From 1998 to 2001 he worked as the Principal Engineer of Intel Microprocessors Research Lab, Director and Chief Scientist of Intel China Research Center. In 2002 he returned to China to work for Chinese Academy of Sciences. He is a professor and director of Key Laboratory of Speech Acoustics and Content Understanding, Institute of Acoustics, Chinese Academy of Sciences. His research interests include large vocabulary speech recognition, speaker/language recognition and audio signal processing. He has published more than 100 papers and holds 40 patents.