ROBUST 3-D CROSSTALK CANCELLER DESIGN

*Chun-Rong Huang and Shih-Fu Hsieh**[1]

Department of Communication Engineering
National Chiao Tung University, Hsinchu, Taiwan 300, Republic of China
Tel: 886-3-5731974, E-mail: terry.cm93g@nctu.edu.tw, sfhsieh@mail.nctu.edu.tw

ABSTRACT

It is well known that the effectiveness of loudspeaker-based 3D sound systems is critically dependent on the listener known position. Traditional approaches try to find the optimum loudspeaker location with least sensitivity to the listener movement, or to use multi-loudspeaker to reduce the effect of head movement. In this paper, we propose a region-based crosstalk canceller with only a pair of loudspeakers to provide increased robustness to perturbations such as head movement and head shapes. In the design of the crosstalk canceller, we propose a new common-pole model structure with reduced complexity.

1. INTRODUCTION

As we know, virtual reality technique can be used to render virtual sound sources in three-dimensional (3-D) space around a listener [1]. To achieve good reproduction of 3-D audio, the simplest way is to deliver binaural signals to the listener's ears through headphones. However, headphone reproduction is cumbersome and often suffers in-head localization. To alleviate these drawbacks, a pair of loudspeakers can be used [6,7]. But, the left and right loudspeakers are not coupled directly to the listener's left and right ears. The emitted sound from the right loudspeaker goes to the left ear as well as to the right ear of a listener, and vice versa. This phenomenon is called crosstalk [8]. Therefore, the need for a crosstalk arises. Fig. 1 shows a crosstalk canceller proposed by Atal and Schroeder in 1963 [2]. The aim of the crosstalk canceller is to find the channel matrix inverse. Most design of crosstalk canceller deals with the case of one single listening position which does not change with time. In order to enlarge the listening area, a region-based crosstalk canceller is investigated [3]. To overcome difficulty of the channel-matrix inverse, the stable and causal crosstalk canceller minimizes the least-square error(LSE) within the region. A common pole model is also proposed to reduce complexity.

2. REGION CONCEPT

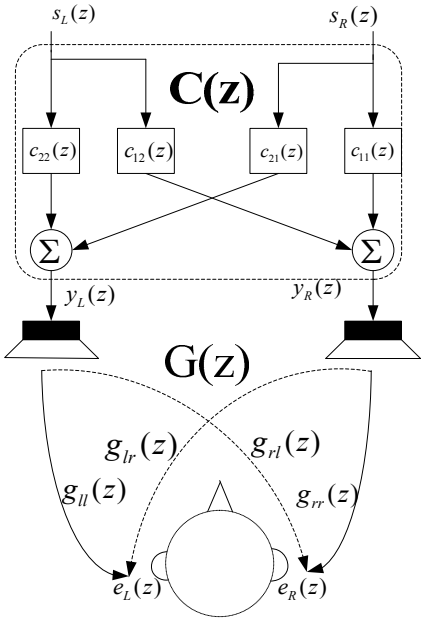

Fig. 1: Atal- Schroeder crosstalk canceller

2.1 Design For One Single Listening Position

First, we start to design the crosstalk canceller with only one point, and then extend to the region concept in the next section. In Fig. 1, $G(z)$ is the channel transfer function matrix comprising two ipsilateral channels $g_{rr}(z)$ and $g_{ll}(z)$, and two contralateral channels $g_{rl}(z)$ and $g_{lr}(z)$. $C(z)$ is the filter matrix of the crosstalk canceller. Our goal is to find *C(z)* so that the received ear signals can be equal to the input signals $s_R(z)$ and $s_L(z)$, i.e.,

$$G(z)\cdot C(z) = z^{-\tau}\cdot I_{2\times 2}(z), \qquad (1)$$

where $I_{2\times 2}(z)$ is a 2 by 2 identity matrix and delay τ is to guarantee the causality. The direct LSE method to design our crosstalk canceller can be written as

$$C(z) = \arg\min\left\{\left\|G(z)\cdot C(z) - z^{-\tau}\cdot I_{2\times 2}(z)\right\|\right\} \qquad (2)$$

[1] This work was supported by the NSC and MediaTek research center at Chiao Tung University.

2.2 Design Over a Region

When the head moves, we get another set of ear signals. Fig. 2 shows the changed transfer function matrix when the head rotates. A region-based concept considers a multiple of listening points. The new crosstalk canceller is designed based on the transfer function matrices in the region.

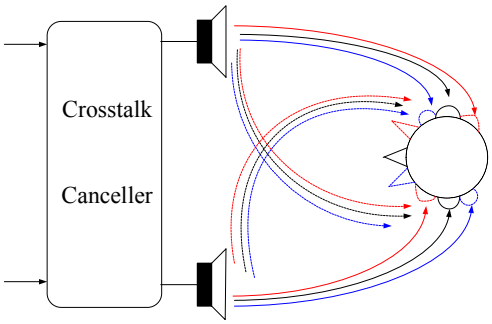

Fig. 2: Head movements

In Fig. 2, $G_o(z)$, $G_+(z)$ and $G_-(z)$ (indicated by black, red and blue lines) represent the transfer function matriices of the fixed, right- and left- rotated human head, respectively. We want that the input signal can be close to, if not equal to, the received ear signals of the head, fixed or rotated. We can express the concept as follows:

$$\left(G(z)+\Delta G(z)\right)\cdot C(z)\approx z^{-\tau}\cdot I_{2\times 2}(z), \tag{3}$$

where $\Delta G(z)$ denotes the change of the head position. Referring to Eq (3), the region-based concept can be formulated as

$$\begin{cases} G_o(z)\cdot C(z)\approx z^{-\tau_o}\cdot I_{2\times 2}(z) \\ G_+(z)\cdot C(z)\approx z^{-\tau_+}\cdot I_{2\times 2}(z), \\ G_-(z)\cdot C(z)\approx z^{-\tau_-}\cdot I_{2\times 2}(z) \end{cases} \tag{4}$$

where τ_o , τ_+ and τ_- are the delays which are dependent on the head positions. Because of different head positions, the distance between sound source and receiving ear will be different. Therefore, the delay of the rotated head should be compensated.

2.3 Delay and Gain Compensations

We know that when the head rotates, the distance between the source and the ears changes. Hence, the time and the gain of the sound source arriving to the ear also change. We will calculate these changes based on a head-centered coordinate model plotted in Fig. 3. We assume the human head is approximate to a sphere of radius r with the ears placed at 90° azimuth. From Fig.3, any source can be represented in three elements, distance R , azimuth angle θ , and elevation angle ϕ . The source and the fixed right ear can be written in vector form as follows:

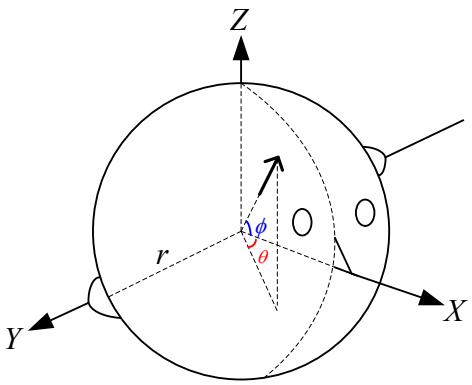

Fig. 3: Head-centered coordinate model

$$\vec{Rs} = R\left(\cos\phi\cos\theta, \cos\phi\sin\theta, \sin\phi\right) \tag{5}$$

$$\vec{re_o} = r(0,1,0) \tag{6}$$

$\vec{Rs}$ and $\vec{re_o}$ denote the source and the fixed right ear, respectively.

If the head rotates $\Delta\theta$ and $\Delta\phi$, the rotated right ear can be written as:

$$\vec{re_\Delta} = r\left(-\cos\Delta\phi\sin\Delta\theta, \cos\Delta\phi\cos\Delta\theta, \sin\Delta\phi\right) \tag{7}$$

If the distance from the source to the head center R is much larger than the radius of the head r . The compensated delay can be calculated by taking the difference of two transmission paths such that:

$$\begin{aligned} delay\_comp(\Delta\theta,\Delta\phi) &= \left(\left|\vec{Rs}-\vec{re_o}\right| - \left|\vec{Rs}-\vec{re_\Delta}\right|\right)\cdot fs/S \\ &\approx r\cdot <\vec{s}, \vec{e_o}-\vec{e_\Delta}>\cdot fs/S, \end{aligned} \tag{8}$$

where fs is the sampling frequency and S is the sound velocity; and $<,>$ denotes vector inner product. Then the $delay\_comp(\Delta\theta,\Delta\phi)$ can be rounded off to an integer.

We note that the gain of sound is proportional to the inverse square of distance such that:

$$gain\_comp(\Delta\theta,\Delta\phi) = \left(\frac{\left|\vec{Rs}-\vec{re_o}\right|}{\left|\vec{Rs}-\vec{re_\Delta}\right|}\right)^2 \tag{9}$$

The gain compensation will be omitted if $R \gg r$.

According to Eq (8), the compensated delay can be found. For example, the desired signal to the fixed head is τ_o and then the head rotates right side about $\Delta\theta_+$ and $\Delta\phi_+$. The delay to the rotated-right head becomes

$$\tau_+ = \tau_o + delay\_comp(\Delta\theta_+, \Delta\phi_+) \tag{10}$$

In the same way, if the head turns to the left side about $\Delta\theta_-$ and $\Delta\phi_-$, the desired signal to the rotated-left head becomes

$$\tau_- = \tau_o + delay\_comp(\Delta\theta_-, \Delta\phi_-) \tag{11}$$

Therefore, the compensated delays in Eq (4) can be found.

3. ROBUST FILTER DESIGN

3.1 Design in FIR Form

When the compensated delay has been found, it can be incorporated to implement the region-equalized concept. The LSE criterion is given by:

$$C(z)=\arg\min\left\{\left\|\begin{bmatrix}G_o(z)\\G_+(z)\\G_-(z)\end{bmatrix}\cdot C(z)-\begin{bmatrix}z^{-\tau_o}\cdot I_{2\times2}(z)\\z^{-\tau_+}\cdot I_{2\times2}(z)\\z^{-\tau_-}\cdot I_{2\times2}(z)\end{bmatrix}\right\|^2\right\} \quad (12)$$

We can find the above equation consists of three elements as follows:

$$E_x(z)=G_x(z)\cdot C(z)-z^{-\tau_x}\cdot I_{2\times2}(z), \quad (13)$$

where $x\in\{o,+,-\}$. We can handle the first column of $C(z)$ such that:

$$\mathbf{e}_x(z)=\begin{bmatrix}g_{rrx}(z) & g_{lrx}(z)\\ g_{rlx(z)} & g_{llx}(z)\end{bmatrix}\begin{bmatrix}c_{11}(z)\\c_{21}(z)\end{bmatrix}-\begin{bmatrix}z^{-\tau_x}\\0\end{bmatrix} \quad (14)$$

If $c_{11}(z)$ and $c_{21}(z)$ are designed in FIR forms, the above equation can be written in convolution matrices and vectors [4],

$$\mathbf{e}_x=\begin{bmatrix}\mathbf{G}_{rrx} & \mathbf{G}_{lrx}\\ \mathbf{G}_{rlx} & \mathbf{G}_{llx}\end{bmatrix}\begin{bmatrix}\mathbf{c}_{11}\\\mathbf{c}_{21}\end{bmatrix}-\begin{bmatrix}\mathbf{d}_x\\\mathbf{0}\end{bmatrix}, \quad (15)$$

where $\mathbf{G}_{tyx}$ is the convolution matrix of channel $g_{tyx}(z)$ and $\mathbf{d}$ is the vector of the delay impulse. Plug Eq (15) into Eq (12) to get

$$\begin{bmatrix}\mathbf{c}_{11}\\\mathbf{c}_{21}\end{bmatrix}=\arg\min\left\{\left\|\begin{bmatrix}\mathbf{G}_o\\\mathbf{G}_+\\\mathbf{G}_-\end{bmatrix}\cdot\begin{bmatrix}\mathbf{c}_{11}\\\mathbf{c}_{21}\end{bmatrix}-\begin{bmatrix}\mathbf{q}_o\\\mathbf{q}_+\\\mathbf{q}_-\end{bmatrix}\right\|^2\right\}, \quad (16)$$

where $\mathbf{G}_x=\begin{bmatrix}\mathbf{G}_{rrx} & \mathbf{G}_{lrx}\\ \mathbf{G}_{rlx} & \mathbf{G}_{llx}\end{bmatrix}$ and $\mathbf{q}_x=\begin{bmatrix}\mathbf{d}_x & \mathbf{0}\end{bmatrix}^T$. The filter coefficient can be readily solved by the normal equation,

$$\begin{bmatrix}\mathbf{c}_{11} & \mathbf{c}_{21}\end{bmatrix}^T=\left(\mathbf{W}^T\mathbf{W}\right)^{-1}\mathbf{W}^T\mathbf{v} \quad (17)$$

where $\mathbf{W}=\begin{bmatrix}\mathbf{G}_o & \mathbf{G}_+ & \mathbf{G}_-\end{bmatrix}^T$ and $\mathbf{v}=\begin{bmatrix}\mathbf{q}_o & \mathbf{q}_+ & \mathbf{q}_-\end{bmatrix}^T$.

3.2 Design in IIR Form

Now we replace FIR by IIR in Eq (14), then the direct IIR error can be rewritten as

$$\mathbf{e}_x(z)=\begin{bmatrix}g_{rrx}(z) & g_{lrx}(z)\\ g_{rlx(z)} & g_{llx}(z)\end{bmatrix}\begin{bmatrix}b_{11}(z)/a_{11}(z)\\b_{21}(z)/a_{21}(z)\end{bmatrix}-\begin{bmatrix}z^{-\tau_x}\\0\end{bmatrix} \quad (18)$$

However, minimization of the error will lead to a set of nonlinear equations. To get around this difficulty, our approach is to impose a common-pole constraint:

$$a_{11}(z)=a_{21}(z)=a_1(z) \quad (19)$$

The block diagram is shown in Fig. 4.

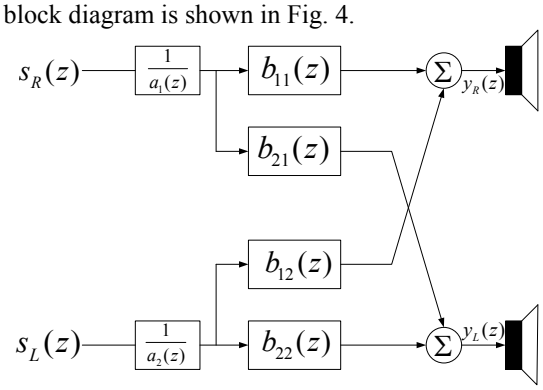

Fig. 4: Common-pole Structure

The direct IIR error can be rewritten as follows:

$$\mathbf{e}_x(z)=\begin{bmatrix}g_{rrx}(z) & g_{lrx}(z)\\ g_{rlx(z)} & g_{llx}(z)\end{bmatrix}\cdot\begin{bmatrix}b_{11}(z)\\b_{21}(z)\end{bmatrix}-a_1(z)\cdot\begin{bmatrix}z^{-\tau_x}\\0\end{bmatrix} \quad (20)$$

The above equation can be rewritten by convolution matrices and vectors.

$$\mathbf{e}_x=\begin{bmatrix}\mathbf{D}_x & \mathbf{G}_{rrx} & \mathbf{G}_{lrx}\\ \mathbf{0} & \mathbf{G}_{rlx} & \mathbf{G}_{llx}\end{bmatrix}\begin{bmatrix}-\mathbf{a}\\\mathbf{b}_{11}\\\mathbf{b}_{21}\end{bmatrix}-\begin{bmatrix}\mathbf{d}_x\\\mathbf{0}\end{bmatrix}, \quad (21)$$

where $\mathbf{D}_x$ is the convolution matrix of the delay signal $z^{-(\tau_x+1)}$ and $\mathbf{a}_1=[1\ \mathbf{a}^T]^T$.

Referring to Eq (12), we formulate the robust common pole as follows:

$$\begin{bmatrix}-\mathbf{a}\\\mathbf{b}_{11}\\\mathbf{b}_{21}\end{bmatrix}=\arg\min\left\{\left\|\begin{bmatrix}\mathbf{Q}_o\\\mathbf{Q}_+\\\mathbf{Q}_-\end{bmatrix}\begin{bmatrix}-\mathbf{a}\\\mathbf{b}_{11}\\\mathbf{b}_{21}\end{bmatrix}-\begin{bmatrix}\mathbf{q}_o\\\mathbf{q}_+\\\mathbf{q}_-\end{bmatrix}\right\|^2\right\}, \quad (22)$$

where $\mathbf{Q}_x=\begin{bmatrix}\mathbf{D}_x & \mathbf{G}_{rrx} & \mathbf{G}_{lrx}\\ \mathbf{0} & \mathbf{G}_{rlx} & \mathbf{G}_{llx}\end{bmatrix}$. In the same way, we can use the LSE criterion to find out the filter coefficients.

4. RESULTS

We first use the robust FIR model to simulate the performance of the robust crosstalk canceller. Here, our channels used in the simulations are the HRTFs from the database of MIT Media Lab [5]. Our simulation is that an impulse signal $\delta[n]$ is inputted at s_R and a zero signal is inputted at s_L in Fig. 1. Therefore, the received desired signal of right ear is a delayed impulse and the left ear is

zeros. We will calculate the error energy(EN) between the received ear signals and desired signals,

$$EN = \sum_{n=0}^{\infty} | \delta(n - delay) - e_R(n) |^2 + | e_L(n) |^2$$

We first use the 200-tap FIR and the loudspeaker is placed at $\pm 5^o$. The head is rotated to right and left by 5^o and the delay compensations can be computed to be ± 1. *EN* of different head position is found out and plotted in Fig. 5.

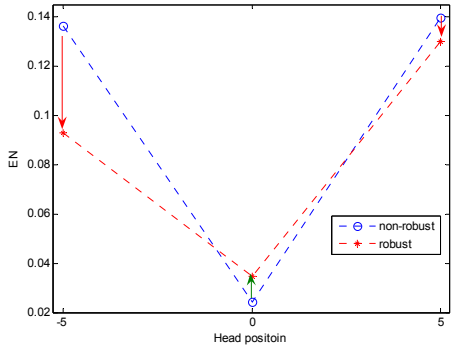

Fig. 5: EN between robust and non-robust designs

In Fig. 5, the horizontal axle is the head position (0: head fixed; 5: head rotated right 5^o; -5: head rotated left 5^o). The non-robust design is expressed in blue line and the robust design is red line. Simulation results agree well with our expectation. In non-robust design, we will use the channels of the fixed head to design the crosstalk canceller. In blue line, we can find that if the head moves, the error energy *EN* will increase. In red line, we can find that EN of the fixed head becomes bad, but those of the rotated heads become better. The total *EN* (sum of *ENs* at three positions) of robust design is smaller than the non-robust design. Because our method is to strike a balance in a region, we do not handle a specific place. Our goal is to minimize the total error in the region. Fig. 6 shows that the total *EN* of robust design gets smaller than non-robust design with different taps.

Next, we test the total *EN* by robust FIR and robust common pole IIR with same system taps. The results are listed in Table 1. We can see that the total EN of robust common-pole is smaller than the robust FIR, in other words, the performance of robust common pole model is better.

Table 1: Total EN between robust LSE FIR and robust common pole IIR

System Taps	100	200	400
LSE FIR	0.3908	0.2959	0.2579
Common-pole	0.383	0.2879	0.2574

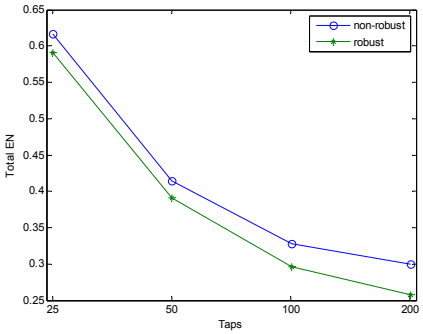

Fig. 6: Total *EN* between robust and non-robust designs

5. CONCLUSIONS

The robust crosstalk canceller design over a region can be eliminate the effect of the head movements. The performance of the common pole model is better than the LSE FIR model.

6. REFERENCES

[1] D.R. Begault, *3-D Sound for Virtual Reality and Multimedia,* AP Professional, 1994.

[2] M.R. Schroeder and B.S. Atal, "Computer simulation of sound transmission in rooms," *IEEE Int. Conv. Record, 7,* 1963.

[3] S.F. Hsieh and Y.F. Chow, "Multichannel sound field equalization," *International Symposium on Communications,* pp. 1086-1093, Taipei, Dec 27-29, 1995.

[4] P.Y. Wang and S.F. Hsieh, "Virtual-loudspeakers-based multichannel sound system," *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, Mohonk Mountain House, Oct 19-22, 1997.

[5] B. Gardner and K. Martin, "HRTF measurements of a KEMAR dummy-head microphone," MIT Media Lab Perceptual Computing Technical Report #280, May 1994.

[6] C.P. Brown and R.O. Duda, "A structural model for binaural sound synthesis", *IEEE Transactions on Speech and Audio Processing*, Vol.6, No. 5, pp. 476-488, Sep 1998.

[7] 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, pp. 387–395, May 1998.

[8] T. Takeuchi and P. A. Nelson, "Robustness to head misalignment of virtual acoustic imaging systems," *J. Acoust. Soc. Am.* 109(3), pp. 958-971, Mar 2001.