Virtual-loudspeakers-based Multichannel Sound System

P. Y. Wang and S. F. Hsieh

Department of Communication Engineering
National Chiao Tung University
Hsinchu, Taiwan 300
Email: sfhsieh@cc.nctu.edu.tw

ABSTRACT

In this paper, we investigate the 3D virtual-loudspeakers-based multichannel sound system. This system uses the HRTFs (head related transfer functions) as the directional perception cues and makes the transmission paths transparent by using the crosstalk cancellers. We propose both the forward and feedback types of crosstalk cancellation systems and compare their complexities and performance such as equalization and crosstalk suppression factors.

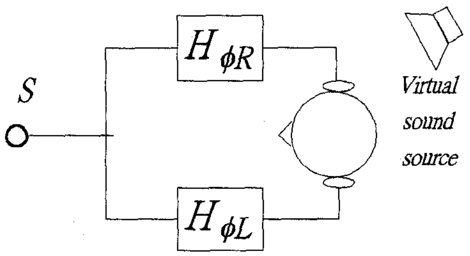

Figure 1: Virtual sound source playback via headphones.

1. INTRODUCTION

In the early stereophonic techniques, virtual sound source could be playbacked on the horizontal plane at a desired azimuth. [1] proposed a stereophonic law of sines with the IID (interaural intensity difference) set to be zero and the position of the virtual sound source can be moved by changing the ITD (interaural time difference). [1] also used the phasor analysis to explain these stereophonic phenomena, but the location of the virtual sound source is blurred because of the crosstalk . To render the 3D sound, the concept of head related transfer function must be included, which can provide sufficient directional perception cues including IID, ITD and spectral cues. Using these transfer functions, we can realize the technology to place the virtual sound source at any 3D position. Figure 1 uses headphones playback to mimic a sound source at location Φ, where $H_{\Phi L}$ and $H_{\Phi R}$ are the left and right HRTFs at location Φ. To get rid of earphones, a pair of loudspeakers can be used for playback as long as a crosstalk canceller is inserted to compensate the loudspeaker/listener paths. The idea can be explained in Figure 2. In this scheme, we can place several virtual loudspeakers at different 3D positions simply using a pair of physical loudspeakers as long as the loudspeaker/listener crosstalk can be cancelled. The problem of crosstalk can be stated as follows. During the transmission through the paths from the two loudspeakers to two ears, the problems rise. The problem can be stated as follows. In a two-loudspeakers system as Figure 2 shows, the signals received at either ear, x_R or x_L, is a combination of the signals played at both loudspeakers. In order to recover the signals, Atal and Schroeder[10] proposed a concept of crosstalk canceller, which can equalize and reduce the crosstalk to suppress the distortion of the received signals. Using matrix notations and referring to Figure 2, we can write:

$$X(z) = G(z) \cdot Y(z)$$

$$\begin{bmatrix} x_R(z) \\ x_L(z) \end{bmatrix} = \begin{bmatrix} g_{11}(z) & g_{12}(z) \\ g_{21}(z) & g_{22}(z) \end{bmatrix} \begin{bmatrix} y_R(z) \\ y_L(z) \end{bmatrix}$$

where $g_{11}(z)$ and $g_{22}(z)$ represent the forward path transfer functions; $g_{12}(z)$ and $g_{21}(z)$ represent the cross path transfer functions. What we should do is finding out the inverse of the transfer function matrix $H(z)$ represented as follows:

$$H(z) = G^{-1}(z)$$

$$= \frac{1}{\Delta(z)} \begin{bmatrix} g_{22}(z) & -g_{12}(z) \\ -g_{21}(z) & g_{11}(z) \end{bmatrix}$$

where $\Delta(z) = g_{11}(z) \cdot g_{22}(z) - g_{12}(z) \cdot g_{21}(z)$, noting that $\Delta(z)$ may be nonminimum-phase so that the elements of the inverse matrix $H(z)$ may be unstable. Besides, there must be proper delays to make elements of the inverse matrix causal. There are several structures for us to implement the inverse filters. In this paper, we investigate two forms of crosstalk cancellers and discuss the complexity of their design procedure and their performances. After finding out the proper structure of crosstalk canceller, we can use it as the main part of our virtual-loudspeakers-based multichannel sound system to make the transmission paths transparent. The basic structure of the virtual-loudspeakers-based multichannel sound system have been shown in Figure 2. In this system, the location of the $j-th$ virtual loudspeaker can be determined by two HRTFs, $H^j_{\phi_R}$ and $H^j_{\phi_L}$. Low-order modeling [12] of the head related transfer functions has been used in order to simulate in real time without losing the directional perception information.

2. CROSSTALK CANCELLERS

2.1. Forward type crosstalk canceller

The structure of the forward type crosstalk canceller is shown in Figure 3. There are four filters to process the signals sent to the loudspeakers. In our approach, we adopt the FIR filter structure and use the least square error method to find out the finite impulse responses of the inverse filters. This can assure the stability of systems. The least square error method

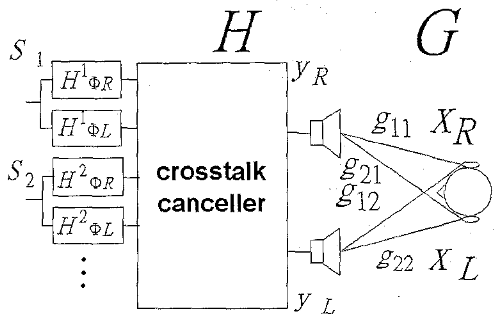

Figure 2: The structure of the virtual-loudspeaker-based multichannel sound system.

can be expressed in the matrix form as follows:

$$\begin{bmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{bmatrix} \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix} = \begin{bmatrix} \underset{\sim}{d} & \underset{\sim}{0} \\ \underset{\sim}{0} & \underset{\sim}{d} \end{bmatrix}$$

where G_{11}, G_{22}, G_{12} and G_{21} represent the convolution matrices of the forward and cross paths' impulse responses, $g_{11}(n)$, $g_{22}(n)$, $g_{12}(n)$ and $g_{21}(n)$, respectively. The desired response $\underset{\sim}{d} = [1\ 0\ 0\ 0...]'$ represents the impulse response in time domain, and $\underset{\sim}{0}$ represents the zero vector. Both lengths of $\underset{\sim}{d}$ and $\underset{\sim}{0}$ are decided according to the order of the FIR inverse filter h_{11}, h_{12}, h_{21} and h_{22}. In practice, the HRTF usually has a transmission delay. To assure the system causality, we would insert a latency into the desired response $\underset{\sim}{d}$. Furthermore, an extra delay would be added to the latency to decrease the minimum least square error in this approach. [2] explains this phenomena in that the head related transfer function is nonminimum phase which would make its inverse filter unstable. Assuring the stability of the inverse filter, nonetheless, would make it noncausal. An extra delay can aid the least square error FIR filter to closely approach the inverse of the HRTF. The choice of the extra delay would depend on the distribution of zeros of the HRTF. When its zeros outside the unit circle are farther away from the unit circle, the smaller number of the extra delay is needed. Usually, the amount of the extra delay is set to be half the order of the FIR inverse filters. Figure 4 shows the signal to crosstalk ratio versus delays. The loudspeakers are located at $\pm 45^o$ azimuth and the signal to crosstalk ratio is defined as $\frac{||x_R||^2}{||x_L||^2}$.

For filters with various orders, the maximum signal to crosstalk ratio occurs at the delay near the sum of the transmission delay(47, in our case) and half of the filter order. Concerning the orders of the FIR filters, the order is affected

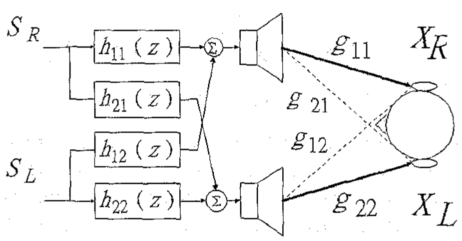

Figure 3: The structure of the forward type crosstalk canceller.

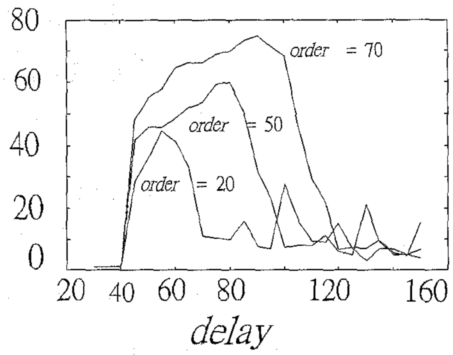

Figure 4: The signal to crosstalk ratio with various order filter versus delay.

by the characteristics of transfer functions and the performance requirement, we use a simple example to illustrate how the HRTFs affect the choice of the filter order. Assume $g_{11}(z) = g_{22}(z) = 1$ and $g_{12}(z) = g_{21}(z) = \alpha \cdot Z^{-D}$,where D is the ITD of the two transfer functions and α is the reciprocal of the IID of the two transfer functions. We have

$$h_{11} = h_{22} = \frac{1}{1 - \alpha^2 \cdot Z^{-2D}}$$

$$h_{12} = h_{21} = \frac{-\alpha \cdot Z^{-D}}{1 - \alpha^2 \cdot Z^{-2D}}$$

If we rewrite the transfer functions in Taylor series form as follows:

$$h_{11} = 1 + \alpha^2 \cdot Z^{-2D} + \alpha^4 \cdot Z^{-4D} + \alpha^6 \cdot Z^{-6D} + ...$$

$$h_{12} = -\alpha \cdot Z^{-D}(1 + \alpha^2 \cdot Z^{-2D} + \alpha^4 \cdot Z^{-4D} + \alpha^6 \cdot Z^{-6D} + ...)$$

From the Taylor series expansion expression, we find that when the IID increases and the ITD remains the same, the decaying rate of the series also increases. It means that when the IID increases and ITD remains the same, the required order of the FIR inverse filters can be reduced. To the contrary, if the ITD increases and the IID remains the same, the necessary order of the FIR inverse filter increases. The orders increases, because the distance between each impulse of the inverse filers increase with the ITD and in order to approximate the IIR responses exactly, the order of inverse filters must be increased. It is should be mentioned that when the ITD decreases, the IID also decreases since the difference of distances from source to two ears decreases. If we want to find the proper positions to locate the loudspeakers, there will be a tradeoff between IID and ITD to make the filter order small.

2.2. Feedback type crosstalk canceller

From section 1, we know the signals sent to the loudspeaker can be represented as follows:

$$Y = H \cdot S \quad (*)$$

$$\begin{bmatrix} y_R \\ y_L \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix} \begin{bmatrix} s_R \\ s_L \end{bmatrix}$$

To derive the filters of feedback type crosstalk canceller, we can rewrite (*) as follows:

$$Y = F \cdot Y + S$$

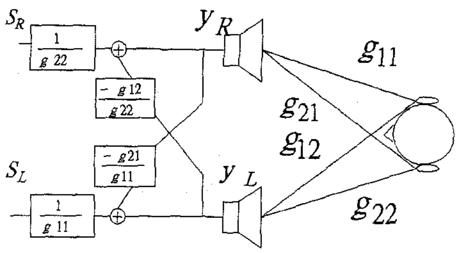

Figure 5: The basic structure of the feedback type crosstalk canceller.

$$Y = H \cdot S$$
$$H^{-1} \cdot Y = S \quad (**)$$
$$(I - F) \cdot Y = S$$
$$Y = F \cdot Y + S$$
$$F = I - H^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} g_{11}(z) & g_{12}(z) \\ g_{21}(z) & g_{22}(z) \end{bmatrix}$$

To simplify the structure of the feedback type crosstalk canceller, we can multiply both side of (**) a diagonal matrix with two diagonal elements $\frac{1}{g_{11}(z)}$ and $\frac{1}{g_{22}(z)}$ to zero-out the diagonal element of F, namely,

$$F = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} 1 & \frac{g_{12}(z)}{g_{22}(z)} \\ \frac{g_{21}(z)}{g_{11}(z)} & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 0 & \frac{-g_{12}(z)}{g_{22}(z)} \\ \frac{-g_{21}(z)}{g_{11}(z)} & 0 \end{bmatrix} = \begin{bmatrix} 0 & f_1(z) \\ f_2(z) & 0 \end{bmatrix}$$

The scheme of the simplified feedback type crosstalk canceller is shown in Figure 5. To design $f_1(z)$ and $f_2(z)$, we would use the zero-pole signal modeling approaches to reduce the filter orders. In the design procedure, the HRTFs are first modeled as their minimum phase parts with transmission delays. This step enables $f_1(z)$ and $f_2(z)$ to generate stable impulse responses for signal modeling but without losing directional perception cues. When we model the responses, the stabilities of the models can be achieved by adding zeros to the response end. Achieving the stabilities of the signal model does not guarantee the stability of the feedback type crosstalk canceller. To see the relationship between the signal models and the stability of the feedback type crosstalk canceller, we attempt to express the transfer functions of the crosstalk canceller in $f_1(z)$ and $f_2(z)$ as follows:

$$\frac{y_R(z)}{x_R(z)} = \frac{y_L(z)}{x_L(z)} = \frac{1}{1 - f_1(z) \cdot f_2(z)}$$
$$\frac{y_L(z)}{x_R(z)} = \frac{f_2(z)}{1 - f_1(z) \cdot f_2(z)}$$
$$\frac{y_R(z)}{x_L(z)} = \frac{f_1(z)}{1 - f_1(z) \cdot f_2(z)}$$

Assume the models of $f_1(z)$ and $f_2(z)$ are $\frac{Z_1(z)}{P_1(z)}$ and $\frac{Z_2(z)}{P_2(z)}$ respectively, the stability of the feedback type crosstalk canceller can be guaranteed as long as all zeros of $P_1(z)P_2(z) - Z_1(z)Z_2(z)$ locate in the unit circle. In this paper, we only discuss the symmetry case, which assumes g_{11} is identical to g_{22} and g_{12} is identical to g_{21}. Under this assumption, $f_1(z)$ and $f_2(z)$ are the same and can be expressed as the model $\frac{Z(z)}{P(z)}$. We have assumed anechoic condition, and use the HRTFs from the MIT media laboratory as the transfer function of the room. Then the stability of the feedback crosstalk canceller can be guaranteed when $\frac{1}{(P(z)+Z(z))\cdot(P(z)-Z(z))}$ is stable. Hence, a lemma follows to guarantee stability of systems.

LEMMA

For an asymmetric structure of feedback crosstalk canceller, the stability can be guaranteed as long as all zeros of $P_1(z)P_2(z) - Z_1(z)Z_2(z)$ locate in the unit circle; as to the symmetric structure, the stability can be guaranteed as long as both zeros of $P(z) + Z(z)$ and $P(z) - Z(z)$ locate in the unit circle.

In order to efficiently reduce the orders of the model, we use pole-zero models to approximate the impulse response of $f_1(z)$ and $f_2(z)$ and it should be mentioned that when we model the response of $f(z)$, we extract the ITD of g_{11} and g_{12} out. What we model is the remainder, and the ITD will be restored after modeling.

3. COMPUTER SIMULATIONS

In our simulation, we have the assumptions as follows:

- $g_{11}(z)$ is the same as $g_{22}(z)$ and $g_{12}(z)$ is the same as $g_{21}(z)$, so the structure of the crosstalk canceller is symmetric.
- We use the HRTFs from MIT Media Lab as our database of the simulation.
- To compare the crosstalk-suppression performances of these crosstalk canceller, we assume the transfer function $g_{11}(z) = 1$ for simplicity and the other HRTFs should be normalized by $g_{11}(z)$.

To justify the performance of crosstalk cancellers, we propose the crosstalk suppression factor (CSF) defined as follows:

$$CSF = 10\log\left|\left|\frac{x_R}{x_L}\right|\right|^2\Big|_{withH's} - 10\log\left|\left|\frac{x_R}{x_L}\right|\right|^2\Big|_{withoutH's}$$

To justify the performance of the crosstalk canceller, we not only consider the crosstalk suppression, but also the equalization result. Here, we come up with the equalization factor (EQ) to see the equalization performance in frequency domain. At first, we define the average variance δ of x_R.

$$\delta = \left[\frac{1}{M}\sum_{m=0}^{M-1}(20\log|x_R(m)| - AV)^2\right]^{\frac{1}{2}}$$

where $AV = \frac{1}{M}\sum_{m=0}^{M-1} 20\log|x_R(m)|$ and $x_R(m)$ is the M-point DFT of x_R, then the equalization factor can be defined as the average variance with crosstalk cancellation subtracted by that without crosstalk cancellation. In the assumption, we have assumed the HRTFS are normalized by g_{11}, so the smaller the CF be the better performance crosstalk canceller do.

To design the filter f in the feedback type canceller, we use the Prony method to model the response. Since the response of $\frac{-g_{12}}{g_{11}}$ decays quickly, we can use small order to model the response but the lemma of stability should be satisfied. Figure 6 shows the results of a feedback type canceller with denominator and numerator orders being 5 and 9, respectively, and the loudspeakers are at $\pm 60^o$ azimuth. The CSF is 5.3613dB and EQ is -0.0947dB in this example. Figure 7 shows both CSF and EQ of the forward type crosstalk cancellers with different filter orders and the loudspeakers are also located at $\pm 60^o$ azimuth. Both CSF and EQ are worst before filter order is 50 due to the large ITD (21 samples), which means that we should use larger order to perform better. The reasons have been discussed in section 2.1. Compare Figure 6 and Figure 7, we find that the feedback type crosstalk can achieve the largest EQ of forward type with a much smaller order. To accomplish better CSF, we can enlarge the order of feedback type crosstalk canceller but the lemma of stability would be guaranteed with more difficulty. In the introduction, we have mentioned that the crosstalk canceller play a role equivalent to the inverse matrix of the transfer functions matrix. [3] use the condition number to identify the frequency hard to be inverted. The condition numbers in [3] would varies with frequency. The larger the condition number be, the more difficult the transfer function matrix be inversed. In order to perform better, we would find out the proper transfer function which has smaller condition numbers.

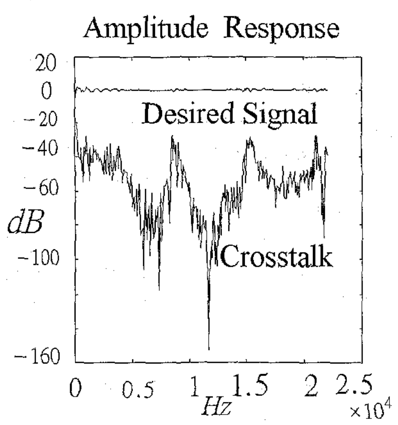

Figure 6: The equalized and crosstalk-cancelled signal of the feedback type canceller.

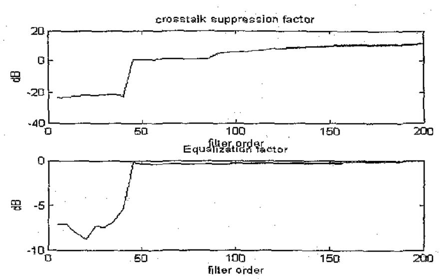

Figure 7: The CSF and EQ of forward type crosstalk canceller with different filter order.

4. CONCLUSION

In HDTV multichannel sound system, our poposed virtual-loudspeaker-based system with crosstalk cancellation can greatly reduce the cost of required physical loudspeakers. By appropriate modeling of HRTFs and taking into account of filter structures, complexity and stability issues, it appears to be very promising in the application of multichannel sound system.

References

1. Benja B Bauer, "Phasor Analysis of Some Stereophonic Phenomena." J. Acoust. Soc. Am., vol. 33, no. 11, pp .1536-1539 (1961)
2. B. Widrow and S.D Stearns, *Adaptive Signal Processing*, Engle wood Cliffs, NJ: Prentice-Hall, 1985, CH. 10.
3. Bruno Korst-Fagundes, Jun Xie Martin Snelgrove, "Multipoint Equalization with the Condition Number," 1996
4. Naraji Sakamoto, Toshiyuki Gotoh, Takuyo Kogure, Masatoshi Shimbo, "Controlling Sound-Image Localization in Stereophonic Reproduction I," JAES, vol. 29, no.11, pp. 794-798, 1981 November.
5. Naraji Sakamoto, Toshiyuki Gotoh, Takuyo Kogure, Masatoshi Shimbo, "Controlling Sound-Image Localization in Stereophonic Reproduction II," JAES, vol. 30, no. 10, pp. 719-721, 1982 October.
6. Philip A. Nelson, Hareo Hamada, and Stephen J. Elliott, "Adaptive Inverse Filter for Stereophonic Sound Reproduction," IEEE Transaction on signal processing, vol. 40, no. 7, July 1992.
7. F. Asano, Y. Suzuki. T. Sone, "Sound Equalization using Derivative Constraints," ACUSTICA acta acustica, vol. 82, 1996.
8. Chou Yun Fu, "Acoustical Spatial Equalization," NCTU Master These 1994.
9. Durand R. Begault, *3D Sound*, AP Professional, 1994
10. Manfred R. Schroeder, "Models of Hearing," Proceedings of the IEEE, vol. 63, no. 9, 1975 September.
11. Kistler, D. J, "A model of head-related transfer functions based on principal component analysis and minimum phase reconstruction," J. Acoustic. Soc. Am., vol. 91, pp. 1637-1647
12. Kulkarni, A.; Colburn, H. S., "Efficient Finite Impulse Response Models of the Head Related Transfer Functions," J. Acoustic. Soc. Am.,vol. 97 , pp. 3278