3D Audio Processing for Elevated Speakers Using the TI C62 EVM Board

Woon-Seng Gan, See-Ee Tan and Meng-Hwa Er
Digital Signal Processing Laboratory,
School of Electrical & Electronic Engineering
Nanyang Technological University
Singapore 639798
ewsgan@ntu.edu.sg

Abstract

3D audio processing has been used to render virtual surround sound using 2 or more speakers placed at ear level. However, we can extend this 3D audio signal processing to other commonly speakers' setups. Here, we introduce a new technique to bring down sound images produced by elevated loudspeakers. This feature allows a cleaner setup for loudspeakers for home entertainment and virtual reality gaming environments and at the same time preserves the auditory image to match the video display at eye level. In this paper, we address the problem of crosstalk cancellation for multiple speakers' configurations and provide different inverse filter structures and configurations to address this problem.

We have implemented these algorithms using the Texas Instrument TMS320C6201 EVM board that is interfaced to an 8-channel I/O board, built using 4 Crystal CS4218 CODECs. A GUI is also developed on the PC for user to select between different modes and parameters. This allows user to select either a 2, 4 or 6-speakers playback configuration. It also enables the user to control the angle of the elevated speakers. Computational complexity, programming techniques and memory usage for the C62xx DSP are also discussed in this paper.

I. Introduction

In home entertainment and computing system, the proposed number of loudspeakers has been increasing with the demand for better sound effects. There are many coding standards that can support 5.1 channels such as the Dolby Digital, DTS, MPEG. The new THX surround EX can even support up to 7.1 channels. These standard has found its application in Digital Versatile Disk (DVD) and digital television (DTV) system. For computing environment, users are no longer satisfied with the traditional stereo speakers and the four-point surround system is gaining popularity. As the number of loudspeakers increased, the number of cables to these loudspeakers also increased. It is therefore important to conceal these wires to have a tidy appearance in a home environment. However, to conceal these wires might be very costly. Therefore, it is desirable that wires can be concealed in the ceiling or cornes and the loudspeakers placed near the ceiling, which can be more easily achieved without much cost. A problem with this setup is that the speaker is no longer at ear-level and this cause elevated sound image. The problem can result in a mismatch between picture content and sound scene. In order to compensate this distortion, this paper proposes a correction scheme to reproduce the signal as if it is being played back from ear-level loudspeakers.

II. Head Related Transfer Function and Elevation Cues

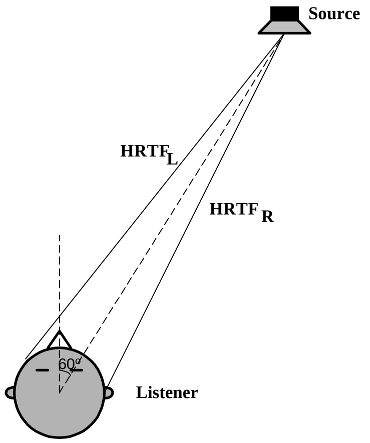

Figure 1. Sound Transmission from Source to Listener

Figure 1 shows a sound source traveling from an azimuth angle of 60^o to the listener. The HRTF [1] is the spatial transformation of a source from a point in free space to the listener's eardrums. Its time domain form is known as the Head Related Impulse

Response (HRIR). In the context of this paper, we shall use HRTF to refer to both time and frequency representations. The three major cues of localization encoded in a HRTF are inter-aural intensity difference (IID), inter-aural time difference (ITD) and modification of spectral profile [2]. These cues are the result of path length difference between the sound source and the two ears, head shadowing, diffraction and reflection of head, outer ear and torso that are linear time-invariant (LTI).

The ITD cue is dominant at low frequency where the head dimension are small compared with the wavelength of the sound. At frequency above 1.5kHz, the IID is the dominant cue [1]. In the median plane and along the cone of confusion, there are no IID and ITD, localization therefore must rely on the spectral profile. Hence, for elevated source the spectral profile is important and cannot be neglected. So, the simple crosstalk cancellation model in [3] implemented for horizontal can no longer be used.

III. Elevated Image Correction for Two Loudspeakers

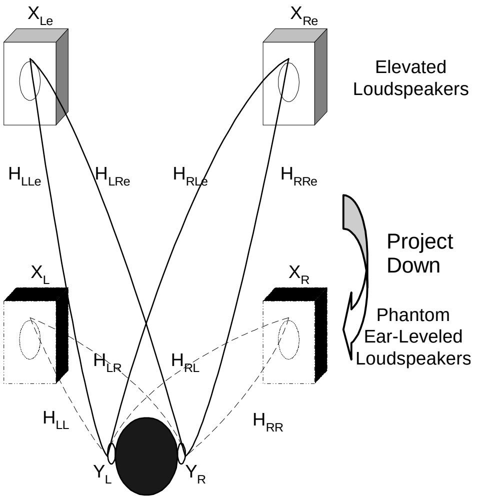

Figure 2. Elevated Loudspeaker Setup and the Desired Phantom Loudspeakers

Figure 2 shows the correction of the elevated loudspeakers image of a two-channel system by creating two phantom loudspeakers at ear-leveled via a psycho-acoustics signal processing explained as follows. The elevated loudspeaker system can be described by the following equations:

$$\mathbf{y_e} = \mathbf{H_e}\mathbf{x_e} \qquad (1)$$

$$\mathbf{y_e} = \begin{bmatrix} Y_{Le} \\ Y_{Re} \end{bmatrix}$$

$$\mathbf{x_e} = \begin{bmatrix} X_{Le} \\ X_{Re} \end{bmatrix}$$

$$\mathbf{H_e} = \begin{bmatrix} H_{LLe} & H_{RLe} \\ H_{LRe} & H_{RRe} \end{bmatrix}$$

where, $\mathbf{y_e}$ is a column vector representing the signal at the eardrums of the listener contributed by the elevated loudspeakers, $\mathbf{x_e}$ is a column vector representing the signal output to the elevated loudspeakers and $\mathbf{H_e}$ is a matrix representing the transmission path from the elevated sound source to the listener's eardrums, which are modeled by the respective HRTF (H_{LLe}, H_{RLe}, H_{LRe} and H_{RRe}) at 60^o elevation. In this project, all speakers are placed at an elevation of 60^o with reference to the listener.

Similarly, the ear-leveled system can be described by the following equations:

$$\mathbf{y} = \mathbf{H}\ \mathbf{x} \qquad (2)$$

$$\mathbf{y} = \begin{bmatrix} Y_L \\ Y_R \end{bmatrix}$$

$$\mathbf{x} = \begin{bmatrix} X_L \\ X_R \end{bmatrix}$$

$$\mathbf{H} = \begin{bmatrix} H_{LL} & H_{RL} \\ H_{LR} & H_{RR} \end{bmatrix}$$

where, $\mathbf{y}$ is a column vector representing the signal at the eardrums of the listener contributed by the ear-leveled loudspeakers, $\mathbf{x}$ is a column vector representing the signal output to the ear-leveled loudspeakers and $\mathbf{H}$ is a matrix representing the transmission path from the ear-leveled sound source to the listener's eardrums, which are modeled by the respective HRTF (H_{LL}, H_{RL}, H_{LR} and H_{RR}) at 0^o elevation (i.e. listener's eye level).

In order to create the phantom ear-leveled loudspeakers, the signal at the listener's eardrum produced using the physical elevated loudspeakers should be identical to that when the loudspeakers are placed at ear-leveled (i.e. equating Egn (1) to (2)). Hence,

$$\mathbf{y_e} = \mathbf{y} \qquad (3)$$

$$\mathbf{H_e}\mathbf{x_e} = \mathbf{H}\ \mathbf{x}$$

$$\begin{aligned}\mathbf{x_e} &= \mathbf{H_e^{-1}H}\ \mathbf{x} \\ &= \mathbf{H_k x}\end{aligned} \qquad (4)$$

$$\text{where } \mathbf{H_k} = \begin{bmatrix} H_{11k} & H_{12k} \\ H_{21k} & H_{22k} \end{bmatrix} \qquad (5)$$

Therefore, in order to produce the desired signal at ear-level using the phantom ear-leveled loudspeakers with input signal **x**, the signal $\mathbf{x_e}$ applied to the elevated loudspeakers should be preprocessed according to Eqn (4).

In 3D audio rendering and crosstalk cancellation, certain form of equalization [3] is required to obtain the desired directional information without the distortion caused by a non-flat microphone and loudspeaker response. For headphone, usually a diffused-field equalization is performed on the HRTF used. In loudspeaker reproduction, a free-field equalization is preferred. It can be shown that for loudspeakers pair with non-identical response, the loudspeakers' response does not cancel out each other through based on the relation $\breve{\mathbf{H}}_{\mathbf{e}}^{-1}\breve{\mathbf{H}}$. However, if the differences in the responses are small, then $\breve{\mathbf{H}}_{\mathbf{e}}^{-1}\breve{\mathbf{H}} \approx \mathbf{H}_{\mathbf{e}}^{-1}\mathbf{H}$ can be easily deduced from (8). For microphone, even non-identical response can be cancelled off in the filter design process.

$$\breve{\mathbf{H}} = \mathbf{M\,H\,S} \qquad (6)$$

$$\breve{\mathbf{H}}_e = \mathbf{M\,H}_e\,\mathbf{S} \qquad (7)$$

where $\mathbf{S} = \begin{bmatrix} S_L & 0 \\ 0 & S_R \end{bmatrix}$ and $\mathbf{M} = \begin{bmatrix} M_L & 0 \\ 0 & M_R \end{bmatrix}$ is the response of the loudspeaker and microphone respectively.

$$\breve{\mathbf{H}}_e^{-1}\breve{\mathbf{H}} = \begin{bmatrix} H_{11k} & H_{12k}\dfrac{S_R}{S_L} \\ H_{21k}\dfrac{S_L}{S_R} & H_{22k} \end{bmatrix} \qquad (8)$$

Further, we can show that the HRTF does not require any further equalization if we assume the response of the left and right loudspeakers and also microphones are identical and the measuring setup for all the HRTF remains the same, which is generally the case. The HRTF measured with combined distortion caused by loudspeaker and microphone response, t can be denoted as:

$$\tilde{\mathbf{H}} = t\,\mathbf{H} \qquad (9)$$

$$\tilde{\mathbf{H}}_e = t\,\mathbf{H_e} \qquad (10)$$

By inverting Eqn (10) and post-multiplying it with Eqn (9), the distortion, t is cancelled out.

$$\tilde{\mathbf{H}}_e^{-1}\tilde{\mathbf{H}} = \mathbf{H}_{\mathbf{e}}^{-1}\mathbf{H} \qquad (11)$$

IV. Extension to Multiple Loudspeakers

The discussion above has been based on a two-channel system that can be modelled by a 2x2 matrix. If the number loudspeakers are increased to n, then the system can be described by a 2xn matrix. In such a system where the number of loudspeakers is greater than the number of ears, many solutions for $\mathbf{H_e}$ exist [4]. For such cases, by computing the pseudoinverse $\mathbf{H_e^+}$, a minimum power solution can be obtained.

$$\mathbf{H_e^+(z) = H_e^T(z^{-1})\left[H_e(z)H_e^T(z^{-1})\right]^{-1}} \qquad (12)$$

However, in this paper another approach is adopted by exploiting the linear time invariant property of the system. By partitioning the loudspeakers into pairs, the signal at the eardrums can be described as:

$$\begin{aligned}\mathbf{y_1} &= \mathbf{H_{k1}x_1} \\ \mathbf{y_2} &= \mathbf{H_{k2}x_2} \\ &\mathrm{M} \\ \mathbf{y_m} &= \mathbf{H_{km}x_m}\end{aligned} \qquad (13)$$

$$\begin{aligned}\mathbf{y_T} &= \mathbf{y_k} + \mathbf{y_2} + \Lambda + \mathbf{y_m} \\ \mathbf{y_T} &= \sum_{i=1}^{m}\mathbf{H_{ki}x_i}\end{aligned} \qquad (14)$$

where, $\mathbf{y_i}$ is the signal at eardrums contributed by the ith loudspeaker pair, $\mathbf{x_i}$ is the input signal to the ith loudspeaker pair, $\mathbf{H_{ki}}$ is the 2x2 matrix that defined the transfer function of the transmission path from the ith pair loudspeaker to the eardrums of the listener and m is the total number of loudspeaker pairs. The total signal at the eardrums, $\mathbf{y_T}$ is the sum of contribution from all the m loudspeaker pairs.

The non-overlapping grouping scheme for loudspeaker pairs is shown in Figure 3 for the case of 6 speakers, arranged at an angle of 60^o apart. In this scheme, the filter design can be made very efficient if we assume a symmetrical listening condition by implementing the "shuffler" structure proposed in [5] for group 1 and 2.

For group 3, the symmetry results in a simple SISO system,

$$\mathbf{y}_{\mathbf{FC}} = \frac{H_{FCL}}{H_{FCLe}}\mathbf{x}_{\mathbf{FC}}\,,\quad \mathbf{y}_{\mathbf{SC}} = \frac{H_{SCL}}{H_{SCLe}}\mathbf{x}_{\mathbf{SC}} \qquad \text{(15a),(15b)}$$

since,

$$H_{FCL} = H_{FCR}\,, H_{SCL} = H_{SCR} \qquad \text{(16a),(16b)}$$

$$H_{FCLe} = H_{FCRe}\,, H_{SCLe} = H_{SCRe} \quad \text{(17a),(17b)}$$

Therefore,

$$\mathbf{x}_{\mathbf{e}} = \begin{bmatrix} \frac{H_{FCL}}{H_{FCLe}} & 0 \\ 0 & \frac{H_{FCL}}{H_{FCLe}} \end{bmatrix}\mathbf{x} \qquad (18)$$

$$= \mathbf{H}_{\mathbf{k3}}\mathbf{x}$$

where, the subscript $_{FCL}$ and $_{FCR}$ denotes front-centre for the left and right ear respectively, and subscript $_{SCL}$ and $_{SCR}$ denotes surround-centre for the left and right ear respectively. Eqn (15a)-(17b) are used base on the symmetry condition since the transmission paths from centre loudspeaker to both ears are identical.

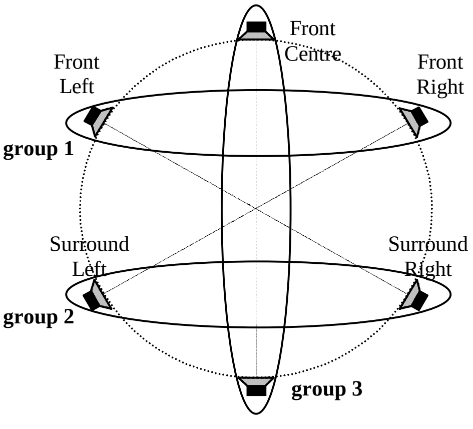

Figure 3. Pairwise Scheme for 6 Loudspeakers

The pairwise loudspeaker method does not guarantee a minimum power system but has the advantage of resolving front-back reversal problem provided that the loudspeaker pairs are partitioned properly.

V. Implementation

In order to support multichannel I/O, an extension board has to be built. The board is built using four Crystal CS4218 codecs[6]. This allows up to 8 inputs as well as 8 outputs at a sampling rate 44.1kHz for each channel if a 11.2896 MHz clock is used. However, if only six more channels are required, the TI's TLV320AIC27 [7] can be used. The part is AC97 Rev. 2.1 compliant, it has analog 3D stereo enhancement and have attractive resolution of 18 bits. In both cases, the Codec is interfaced to the McBSP of the EVM. Figure 4 shows the connection between the EVM and the CS4218 codecs.

The algorithm is designed to be compact and can be easily integrated into post-processing phase of a decoder such as the open DSP software architecture for consumer audio proposed by Chen et al [8]. For 2-speaker case, the algorithm consists of four filters as stated in Eqn (5). Each of the left and right signals is passed through two of the filters. The results are then summed together to form the final output.

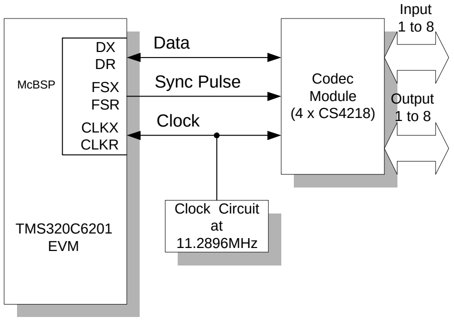

Figure 4.Connection between EVM and Codec Board

In order to optimize for speed, the filter routine employs loop unrolling up to a factor of 8. The data is processed in block of 32 samples using circular buffering. This gives a delay of less than 1ms that will not pose any problem to the synchronization with picture content. Using two LDW instructions [9] in parallel to load four short data or coefficients into the registers. This reduces the memory accesses by half.

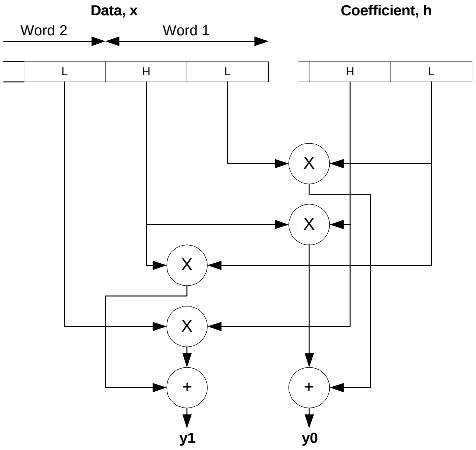

Figure 3 Using Word access for Short Data

Memory bank conflict resulting in memory hits in FIR can also be easily resolved by careful partitioning and allocation of data and coefficients. By further making the outer loop parallel with the inner loop epilog and prolog, better performance is obtained. The outer loop is also made conditionally executed with inner loop.

For the summation, the ADD2 instruction [9] can be used to reduce the number of additions by half. The ADD2 instruction adds the upper half and lower half of the first operand to the upper half and lower half of the second operand respectively.

Based on a filter length of 128, the number of cycles required for the algorithm per sampling interval is show in Table 1. The memory usage requirement is shown in Table 2. However, by further reducing the filter length a better performance can be obtained.

Function	No. of Cycles
2 Speakers	257
4 Speakers	514
6 Speakers	771

Table 1. Functions and their required Cycles

Function	Type	Memory (Bytes)
2 Speakers	Buffer	1536
	Coefficients	1024
4 Speakers	Buffer	2560
	Coefficients	2048
6 Speakers	Buffer	3584
	Coefficients	3072

Table 2. Functions and memory requirements

To enable change of different setup and aid testing, a GUI is built using MS Visual Basic. The simple GUI allows users to select a 2, 4 or 6-speakers playback configuration. A dynamic link library (DLL) is also compiled using stdcall so that the EVM board access functions can be used in Visual Basic.

VI. Conclusion

We have proposed an algorithm for correction of elevated speaker image using 3D audio processing. Although the algorithm can also be implemented in other family such as the TMS320C54xx, this paper focuses on the implementation using TMS320C62xx. We discuss the hardware implementation as well as the algorithm aspect. The algorithm has been implemented efficiently on C62 by applying various programming techniques described in this paper.

VII. References

1. J. Blauert, "Spatial Hearing. The psychophysics of human sound localization", revised edition, MIT Press, Cambridge MA, 1997
2. S. Carlile, "The Physical and Psychophysical Basis of Sound Localization", in Virtual Auditory Space: Generation and Applications, S. Carlile, Ed. Austin, TX: R. G. Landes, 1996, pp 27-78
3. W. G. Gardner, "Transaural 3-D Audio", MIT Media Lab Technical Report No.342, 1995
4. J. L. Bauck and D. H. Cooper, "Generalized Transaural Stereo and Applications", J. Audio Eng. Soc., 1996, vol. 44, pp 683-705
5. D. H. Cooper and J. L. Bauck, " Prospects for Transaural Recording", J. Audio Eng. Soc., 1989, vol. 37, pp 3-19
6. "Crystal Semiconductor Audio Data Book", 1994
7. "TLV320AIC27 Stereo Audio Codec", 2000, Texas Instruments, SLAS253
8. T. Chen, J Datta and B Karley, "An Open DSP Software Architecture For Consumer Audio", IEEE Transaction on Consumer Electronics, 1999, 45 No. 4, pp 1253-1258
9. "TMS320C62x/67x CPU and Instruction Set", 1998, Texas Instruments, SPRU198C