An Object-Based Audio Rendering System using Spatial Parameters

Kuo-Lun Huang, Tai-Ming Chang, and Pao-Chi Chang
Department of Communication Engineering
National Central University
Jhongli, Taiwan
{klhuang, tmchang, pcchang}@vaplab.ce.ncu.edu.tw

***Abstract*—In this paper, we propose an object-based audio rendering system, in which the audio signal of each object is distributed to multi-channel systems by spatial parameters. The operation of the rendering system is based on the sound localization theories with the Parametric Stereo (PS) coding structure. The performance of the audio surround effect of a sound moving object is evaluated by subjective listening tests. The experimental result clearly shows that the proposed system is capable of delivering sound moving effect to listeners.**

Keywords- Surround; Spatial Audio Object Coding ;Rendering; First-Person Shooter.

I. INTRODUCTION

In the past studies, several 3D audio processing techniques were proposed for loudspeaker systems and headset systems. A well-known spatial coding technique is Parametric Stereo (PS) coding which greatly reduces the bitrate by utilizing the characteristics between two (or more) channels [1]. Head-Related Transfer Function (HRTF) that simulates the response of the sound transmission path to two ears by generating impulse responses with a series of directional coefficients [2].

Moving Picture Experts Group suggested Spatial Audio Object Coding (SAOC) standard in 2010 [3]. SAOC inherits the coding structure of PS and further develops an object-based system. The most important feature of SAOC is that it can re-use the spatial parameters to remix audio by the mixer/renderer. Many applications can profit from the user control of different audio objects at the playback side. Examples are remixing applications, teleconferencing, Karaoke, and on-line gaming [4]. However, problems still exist in above application, such as the difficulty in perfect reconstruction of the up-mix blocks in decoder, and the limitation in the adjustment of mix/renderer. Thus, we propose an Object-based Audio Rendering System (OARS) which uses sub-band analysis filter bank to rebuild the spatial information with a few parameters.

II. THE PROPOSED OBJECT-BASED AUDIO RENDERING SYSTEM (OARS)

Our proposed system mainly consists of four processes: Doppler Effect compensator, analysis/synthesis filter bank, parameter generator, and spatial synthesis. We use re-sampling method to realize the Doppler Effect, and use the Complex-exponential Modulated Quadrature Mirror Filter (QMF) banks as time/frequency analyzer/synthesizer. According to the spectral resolution of Human Auditory System (HAS), the bandwidth of filter banks follows the Equivalent Rectangular Bandwidth scale [5]. The 64-band QMF filter banks result in approximately 344 Hz effective bandwidth per band. The structure of Object-based Audio Rendering System is shown in Fig. 1.

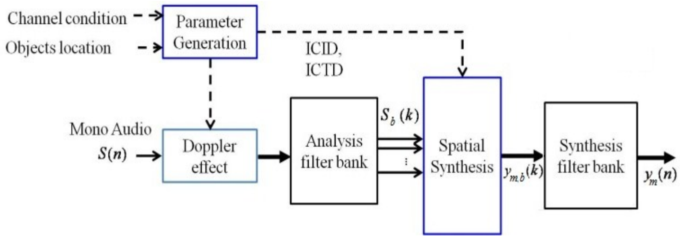

Figure 1. Object-Based Audio Rendering System based on Parametric Coding.

III. RENDERING ALGORITHM

In this section, we discuss the proposed audio rendering system in sub-band domain and use an M-channel playback system to derive the algorithm of spatial synthesis. This playback system has M = 4 as shown in Fig. 2, where θ_b represents the angle of the auditory object in parameter band *b*.

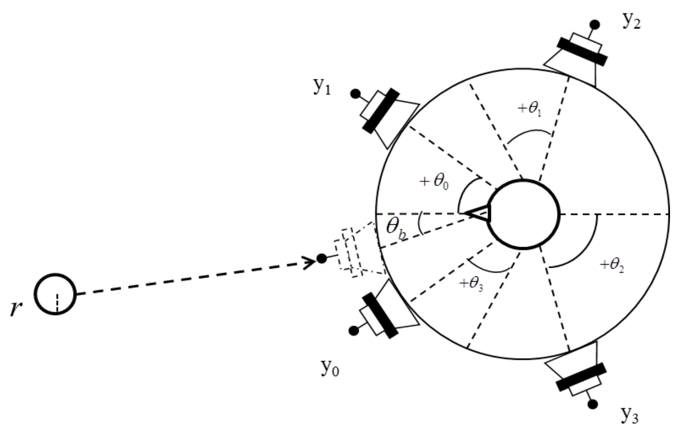

Figure 2. Four-channel playback system.

The Inter-Channel Intensity Difference (ICID) factor A_b is estimated by the “stereophonic law of sines”, as in (1) [6]

$$\frac{a_{m,b}}{a_{m+1,b}} = \frac{\sin\theta_m + \sin\theta_b}{\sin\theta_m - \sin\theta_b} = A_b, \quad (1)$$

where $a_{m,b}$ and $a_{m+1,b}$ denote the power levels of channels y_m and y_{m+1}.

The loudspeakers pair m and m+1 enclosing the original angle θ is first selected, where the left is negative angle and the right is positive angle. Next, we use level decay equations, as in (2)-(4),

$$p(L) = 20\log_{10}\left(\frac{r}{r+L}\right), \quad (2)$$

$$a_{m,b}\cdot 10^{\frac{\Delta p(L)}{20}} + a_{m+1,b}\cdot 10^{\frac{\Delta p(R)}{20}} = a_{s,b} \quad (3)$$

$$L = d_{m,b} \quad R = 2r\sin\theta_{m,b} - L, \quad (4)$$

where r represents the radius the auditory object, $d_{m,b}$ are the distance between the auditory object and loudspeakers y_m, and $a_{s,b}$ is the power level of the auditory object s.

In short, we combine (3) with (1) to obtain the gain of loudspeakers as shown in (5).

$$\begin{cases} a_{m,b} = \dfrac{A_b \cdot a_{s,b}}{C_{m,b}} \\ a_{m+1,b} = \dfrac{a_{s,b}}{C_{m,b}}, \end{cases} \quad (5)$$

where $C_{m,b} = A_b \cdot p(L) + p(R)$.

Refer to (4), the delay time of channels can be calculated as

$$d_{L,b} = \frac{L}{c}, \quad d_{R,b} = \frac{R}{c}, \quad (6)$$

where c is the sound velocity, and $d_{L,b}$ is defined as Inter-Channel Time Difference (ICTD) factor. Finally, each time-frequency tile of the output channels are obtained as in (7).

$$\begin{aligned} y_{m,b}(k) = {} & \delta(l - \mathrm{mod}(l,M))\cdot a_{m,b}\cdot s_b(k - d_{L,b}) + \\ & \delta(l - \mathrm{mod}((l+1),M))\cdot a_{m+1,b}\cdot s_b(k - d_{R,b}), \end{aligned} \quad (7)$$

where l = m+1.

The input information includes the location of objects and setup of the playback system. The information of game objects consists of the initial location, velocity, and the resolution of time spacing.

IV. Subjective Evaluation

We used subjective listening tests to evaluate the proposed OARS with the dynamic audio sources. Eight listeners were invited to carry out these tests. The sampling rate of audio object was 44.1 kHz. Three moving subjects with six different motion directions were tested where FL, FR, SL and SR represents From-Left, Front-Right, Surround-Left and Surround-Right, respectively. The subjective assessment of sound direction, which was modified from the ITU-R general methods [7], were used, as shown in Table I.

TABLE I. The Modified ITU-R 7 Grade Comparison Scale

Bravo (5)	Direction:	Exact
Good (4)	Direction:	Almost no deviation
Satisfactory (3)	Direction:	Slight deviation
Insufficient (2)	Direction:	Confused
Not Executed (1)	I cannot realize the sound!	

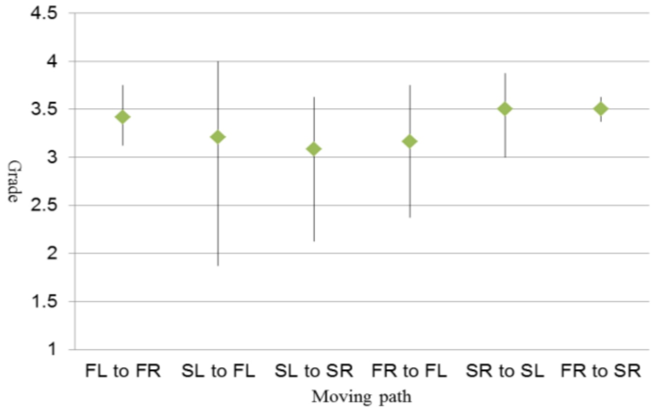

Figure 3. The subjective test results of the moving objects.

Fig. 3 shows the average scores of subjective evaluation of three objects. The average score of six moving direction approximates the scale of "Slight deviation" in a non-soundproof conference room. Nevertheless, listeners can still detect the object moving around easily.

V. Conclusion

This work has proposed an audio rendering system which considers physical characteristics such as distance, direction, and speed of objects to build a lifelike sound field. Experiment results show the proposed OARS system can effectively make a surround sound field with moving objects. Furthermore, OARS applied to online games may greatly reduce the transmission bitrate over networks.

References

[1] ISO/IEC JTC1/SC29/WG11, "Text of ISO/IEC 14496-3:2001/FPDAM2 (parametric coding for high quality audio)," ISO/IEC JTC1/SC29/WG11 N5713, July 2003.

[2] UC Davis the CIPIC interface laboratory - HRTF database: http://interface.cipic.ucdavis.edu/.

[3] ISO/IEC JTC1/SC29/WG11, "Text of ISO/IEC FDIS 23003-2:2010 (Information technology - MPEG audio technologies - Part 2: Spatial Audio Object Coding (SAOC))," Oct. 2010.

[4] J. Engdegard, B. Resch, and etc., "Spatial Audio Coding (SAOC)-The Upcoming MPEG Standard on Parametric Object Based Audio Coding," *124th AES Convention*, Amsterdam, Netherlands, May 2008.

[5] J. Hall and M. Fernandes, "The role of monaural frequency selectivity in binaural analysis," *J. Acoust. Soc. Am.*, vol. 76, pp. 435-439, 1984.

[6] B. B. Bauer, "Phasor analysis of some stereophonic phenomena," *J. Acoust. Soc. Am.*, 33:1536-1539, Nov. 1961.

[7] ITU-R. "General methods for the subjective assessment of sound quality," ITU-R Recommend. BS.1284-1, 2003.