Audio Engineering Society

Convention Paper 6683

Presented at the 120th Convention
2006 May 20–23 Paris, France

Virtual Source Location Information based Matrix Decoding System

Han-gil, Moon[1], Manish Arora[1]

[1] Samsung Electronics Co., LTD.
hangil.moon@samsung.com

ABSTRACT

In this paper, a new matrix decoding system using vector based Virtual Source Location Information (VSLI) is proposed as one alternative to the conventional Dolby Pro logic II/IIx system for reconstructing multi-channel output signal from matrix encoded two channel signals, Lt/Rt. This new matrix decoding system is composed of passive decoding part and active part. The passive part makes crude multi-channel signals using linear combination of the two encoded signals(Lt/Rt) and the active part enhances each channel regarding to the virtual source which is emergent in each inter channel. The virtual sources between channels are estimated by inverse constant power panning law

1. INTRODUCTION

Dolby Pro Logic II/IIx is the most popular system to decode the matrix encoded stereo signals into multi-channel signals. Circle Surround and Neo-6 are the alternative systems to Pro Logic II/IIx. All of these matrix decoding systems utilize the active decoding concept in which the gains of the multi-channel output signals are constantly updated during the decoding procedure to improve the inter-channel signal separation of decoded output signals. The decoding algorithms above mentioned update output channel gains using non linear gain control e.g. strong level channels are made stronger and weak channels are weakened further. The drawback of existing matrix decoding systems is that in trying to achieve a high level of inter-channel separation acoustical source perception by the listener is not considered.

In this paper, a new matrix decoding system equipped with vector based Virtual Source Location Information (VSLI) is proposed as an alternative to conventional matrix decoding systems for reconstructing multi-channel output signal from matrix encoded two channel signals Lt/Rt. The proposed matrix decoding system makes use of virtual sources perceived by the listener due to multi-channel playback to enhance the multi-channel output signals. The azimuth information of virtual sources is estimated by means of inverse power panning law and the power of the virtual sources are evaluated by the vector projection method. After estimating the angle information of virtual sources and the power of the virtual sources, each virtual source can be represented in the vector form. Then the global vector which represents the global sound image on whole multi-channel layout can be evaluated by simple vector manipulation of the virtual source vectors. Using the virtual source and global source vectors, logic for channel enhancing system has been devised.

The logic of channel enhancing system consists of two steps. The first step modifies channel powers to improve the degree of inter channel separation in passive decoded signals. The second step enhances the localization of the perceived global sound image by redistributing channel powers on the basis of global vector angle and power. The channels close to global vector position are enhanced by means of a sound image enhancing function and the channels far from global vector position are reduced. Using the two steps the global sound image is localized more clearly and the listener perceives the dynamic effects of moving sources. Also adequate inter channel separation is maintained.

2. VIRTUAL SOURCE LOCATION INFORMATION ESTIMATION

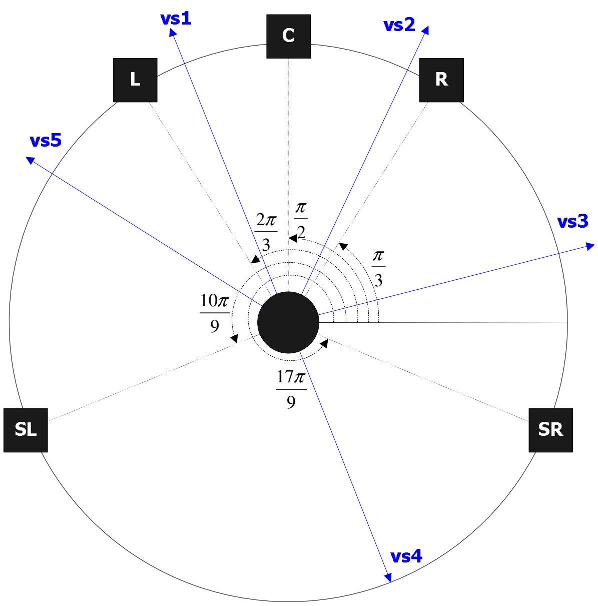

Figure 1 Playback loudspeaker layout

The VSLI is azimuth information which represents the geometric spatial information between power vectors. The cue (VSLI) is extracted under the assumption that the playback layout of multi-channel loudspeakers is fixed as illustrated in Fig. 1. Inter-channel power vectors form a spatial audio image between adjacent speakers. There are five possible number of the existing spatial sound image denoted as vs1,…, vs5 in Fig. 1. One spatial image can be represented by one azimuth information and one power information of corresponding spatial sound image.

To evaluate the angle information of a spatial image, the power panning law [1, 2] is employed. To evaluate the level information, well-known vector manipulation method is employed. For example, the level information L_k^i at time k in channel i and L_k^{i+1} at time k in channel $i+1$ are derived from windowed representation of audio input signal as shown in Fig. 2. To estimate the power vector of a spatial sound image between channels, not only the level information, but also the angle between adjacent channels is indispensable. The angle estimation with these two level information { L_k^i , L_k^{i+1} } of two adjacent channels is performed with the help of power panning law.

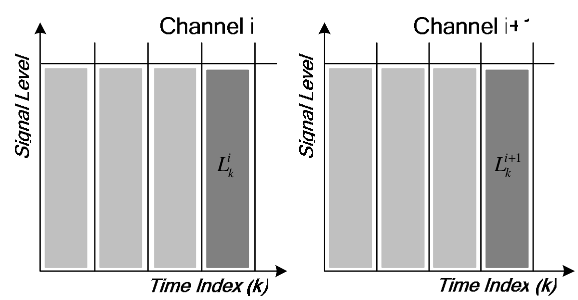

Figure 2 Level information L_k^i and L_k^{i+1} at time index k

To apply the power panning law inversely, the level information of channel i and channel $i+1$ need to be normalized.

$$g_k^i = \frac{L_k^i}{\sqrt{(L_k^i)^2 + (L_k^{i+1})^2}} \tag{1}$$

$$g_k^{i+1} = \frac{L_k^{i+1}}{\sqrt{(L_k^i)^2 + (L_k^{i+1})^2}} \tag{2}$$

Then, the power panning law is applied inversely to the normalized adjacent channel gains at time k { g_k^i , g_k^{i+1} } for azimuth information .

$$\theta_k^{i,i+1} = \cos^{-1}\{g_k^i\} \tag{3}$$

$$\theta_k^{i,i+1} = \sin^{-1}\{g_k^{i+1}\} \tag{4}$$

The evaluated azimuth information (3) or (4) is a relative angle gauged from channel i or channel $i+1$. Therefore, the location information (angle) of virtual source between channel i and channel $i+1$ is evaluated by applying the angle coordinates of channel i and channel $i+1$ to the azimuth information (3) or (4). The proper location information is estimated by the following equation (5).

$$\theta_{vs} = (\theta_k^{i,i+1} \times \frac{2}{\pi}) \times (\theta_{i+1} - \theta_i) + \theta_i \tag{5}$$

The estimated angle value is found to be valid for just one frame k between channel i and channel $i+1$. Therefore, angle values must be updated in a time index with 512 samples.

3. VIRTUAL SOURCE LOCATION INFORMATION BASED MATRIX DECODING SYSTEM

3.1. System overview

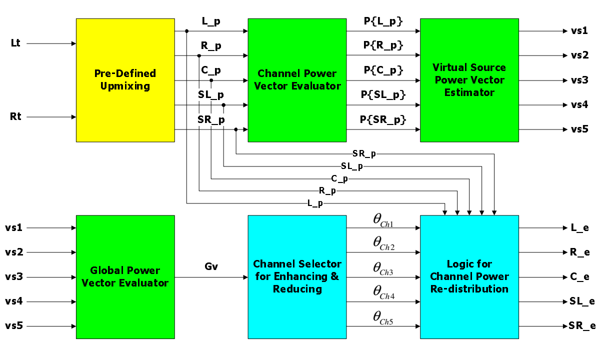

Figure 3 Virtual source vector & Global vector estimation

Figure 3 shows VSLI based matrix decoding system block diagram. The proposed matrix decoding system is composed of 6 discrete blocks.

The first block, "Pre-defined upmixing" block, performs pseudo inverse matrixing to extract crude multi-channel outputs from the matrix encoded stereo inputs. Section 3.2 covers more detail scheme about

The second block, "Channel power vector evaluator", makes multi-channel power vectors corresponding to playback layout using every channel level information at certain time interval and the angular information of the layout. The third block, "Virtual source power vector estimator" estimates inter-channel virtual source power vector. For the estimation channel power vectors are used as the input information and VSLI introduced in section 2 is used as the evaluation tool. The forth block, "Global power vector evaluator" decides the global vector which is related to the perceived main sound source image. The second block, the third block, and the forth block form Virtual source vector & Global vector estimation block set. Section 3.3 covers more detailed description about this block set.

The fifth block, "Channel selector for enhancing & reducing", decides channels to be enhanced or to be reduced by the location information (angle) of global power vector. The sixth block, "Logic for channel power re-distribution", adjusts the whole channel powers based on the angle difference between global vector and a specific channel power vector. The sixth block also increases the channel separation ratio by filtering each channel output with nonlinear function.

3.2. Pre-defined upmixing

The encoding matrix, M , to downmix multi-channel audio signals to stereo (Lt/Rt) signals can be expressed by equation (6). Theoretically, the decoding matrix, N , to upmix the matrix-encoded stereo signals to multi-channel audio signals is the inverse of equation (6). The inverse of the encoding matrix, M , is usually not specified by normal matrix inversion process because the encoding matrix, M , is underdetermined system. Therefore, the inverse of encoding matrix should be evaluated Moore-Penrose generalized inverse [3, 4]. It can be expressed by equation (7). However, the inversed matrix is not an optimal version of decoding solution. This matrix should be refined by input stream dependent channel power enhancing algorithm to enhance the output channel separation and to give listeners better multi-channel sound images.

$$\begin{bmatrix} L_t \\ R_t \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & a_3 & a_4 & a_5 \\ b_1 & b_2 & b_3 & b_4 & b_5 \end{bmatrix} \times \begin{bmatrix} L \\ R \\ C \\ SL \\ SR \end{bmatrix} \quad (6)$$

$$M = \begin{bmatrix} a_1 & a_2 & a_3 & a_4 & a_5 \\ b_1 & b_2 & b_3 & b_4 & b_5 \end{bmatrix}$$

$$N = M^T \times (M \times M^T)^{-1} = M_R^+ \quad (7)$$

3.3. Virtual source vector & Global vector estimation

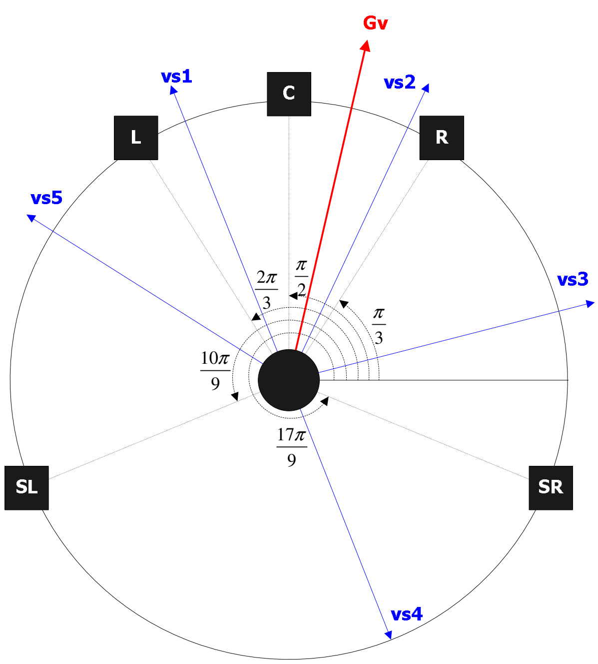

Figure 4 Virtual source vector & Global vector estimation

In proposed matrix decoding system, channel power enhancing logic and channel power re-distribution logic are devised not only for the better channel separation but also for guaranteeing good moving source effect. To utilize these two logics, global vector which can be approximated to the main perceived sound source vector should be evaluated. The global vector is evaluated by vector manipulation of virtual source vectors as illustrated in Fig. 4. As explained in section 2, the virtual source vectors can be determined by inverse constant power panning law. The equation of global vector estimation is as follows.

$$Gv = vs1 + vs2 + vs3 + vs4 + vs5 \quad (8)$$

Therefore, the power information of perceived main sound source can be represented by equation (9) and the location information (angle) of perceived main sound source can be represented by equation (10).

$$|Gv| = |vs1 + vs2 + vs3 + vs4 + vs5| \quad (9)$$

$$\angle Gv = \angle(vs1 + vs2 + vs3 + vs4 + vs5) \quad (10)$$

As explained above, the global vector can be assumed to be an approximation of the perceived sound source vector [1, 2, 5]. The power of global vector can be thought as the power (loudness) of perceived sound image and the angle of global vector can be considered as the absolute location of perceived sound image on the playback layout system.

3.4. Channel power enhancing

Estimating the power and the angle of global vector, channels for enhancing or reducing are determined to adjust the whole channel power according to the global vector in a each time window. For example, if the estimated global vector lies in the region between center channel and left channel as depicted in Fig. 4, the channel power re-distribution function, $f(\theta)$, re-distribute overall gains of the whole channels with respect to the angle difference between a specific channel and global vector to enhance the level of center channel and left channel and to reduce the levels of surround channels. The channel power re-distribution function is defined by nonlinear combination of trigonometrical functions (11).

$$f(\theta) = a \cdot \cos^n \theta + b \cdot \sin^m \theta \quad (11)$$

In equation (11), θ is defined by the difference between a specific channel and global vector. a, b, n, and m are integer values which is chosen to stabilize the perceived sound images. To guarantee a good channel separation, each channel power is stretched by nonlinear equation (12). In equation (12), the constant k is a positive integer value. Both l and m are values between 0 and 1. P is the sum of the whole channel power.

$$g(x)=\begin{cases} k\cdot x & (m\cdot P\le x^2) \\ x\cdot k^{10\cdot(\frac{x^2}{P}-l)} & (l\cdot P\le x^2\le m\cdot P) \\ x & (x^2\le l\cdot P) \end{cases} \quad (12)$$

By combining equation (11) and equation (12), channel power enhancing logic is determined as shown in Fig. 5.

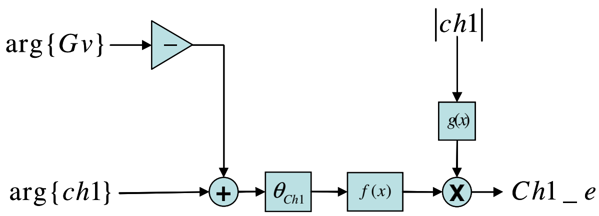

Figure 5 Logic for channel power enhancing

4. PERFORMANCE

The channel separation is measured by matrix encoded stream which is composed of sine signal in only left channel and zero in other channels. In the matrix decoded output signals, other channel outputs are less than decoded left channel by over 36dB. In other cases (e.g. sine signal in only left surround channel and zero in other channels), the similar channel separation ratio can be resulted. For the subjective assessment, down-mixed multi-channel sound track contents from Star Wars, Matrix, Saving Private Ryan, Gladiator, and Pearl Harbor were used as the test materials. The portions of the sound track selected have strong moving sound images. Seven of ten listeners preferred proposed matrix decoded stream to the conventional DPLIIx decoded stream.

5. CONCLUSION

The proposed matrix decoder is devised based on sound image perception in multi-channel layout. The dominant sound image perceived by listeners in multi-channel layout is estimated by global vector which is formed by virtual source vectors. With the power information and the angle information of this dominant global vector, every channel power is re-distributed to enhance the dominant sound images. The enhanced dominant sound source gives listeners better sound images specially in case of moving source. To improve channel separation, nonlinear filter with logarithmic function is used for the all channels. By combining these nonlinear filter and channel re-distribution function with pre-defined upmixer, virtual source location information based matrix decoder is structured.

6. REFERENCES

[1] V. Pulkki, "Localization of Amplitude-Panned Virtual Sources II: three-dimensional panning," *J. Audio Eng. Soc.*, vol. 49, no. 9, pp. 753-767, September 2001.

[2] West J. R, "Five-channel panning laws: an analytic and experimental comparison", Master's Thesis, Music Engineering, University of Miami, 1998.

[3] G. Strang, *Linear algebra and its application*, Harcourt, 1988.

[4] J. A. Fill and D. E. Fishkind, "The Moore-Penrose generalized inverse for sums of matrices", SIMAX, vol. 21, no. 2, pp 629-635, 1999.

[5] H. Moon, J. Seo, S. Beack, K. Sung "A multi-channel audio compression method with virtual source location information for MPEG-4 SAC," IEEE trans on CE, vol. 51, no. 4, pp. 1253-1259, November 2005.