논문 2010-47SP-5-3

가상 음원 위치 정보를 이용한 능동 메트릭스 디코더

(A Perception Based Active Matrix Decoder with Virtual Source Location Information)

문 한 길*

(Han-gil Moon)

요 약

본 논문에서는 돌비 프로로직 II/IIx를 대체하기 위한 가상 음원 위치 정보 기반의 새로운 메트릭스 디코더 시스템을 제안하고자 한다. 제안하는 신규 메트릭스 디코더는 역행렬 계산을 통해 얻어지는 수동 메트릭스 디코딩부와 수동 메트릭스 디코딩을 통해서 얻은 신호들을 멀티채널 신호의 채널간 이미지 특성에 따라서 적응적으로 가변시키는 능동 메트릭스 디코딩부로 구성된다. 멀티채널 환경에서 채널 간에 형성되는 다수의 이미지는 실제 청각 시스템에 의해서 인지되어 만들어지는 가상의 사운드 이벤트와 연결이 되어 있다. 따라서 이 이미지의 위치와 크기에 기반하여 멀티채널 신호를 적응적으로 가변시키면, 인지적인 관점에서 우수한 성능의 메트릭스 디코더를 설계할 수 있다. 더불어 채널간 분리도를 향상시키기 위해서 비선형 삼각함수의 조합을 사용하였다.

Abstract

In this paper, a new matrix decoding system using vector based Virtual Source Location Information (VSLI) is proposed as an alternative to the conventional Dolby Pro logic II/IIx system for reconstructing multi-channel output signals 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. Since the virtual sources are related to the perceptual sound images in virtual sound field, the reconstructed multi-channel sound results in good dynamic perception and stable image localization. Moreover, the good channel separation is maintained with nonlinear trigonometric enhancing function.

Keywords : Multichannel audio, Matrix decoder, Active, VSLI

I. Introduction

Dolby Pro Logic II/IIx[1~2] is the most popular system to decode the matrix encoded stereo signals into multi-channel signals. Circle Surround[3] and Neo-6[4] 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 seamlessly updated during the decoding procedure to improve the inter-channel signal separation of decoded output signals.

In this paper, a new matrix decoding system equipped with vector based Virtual Source Location Information (VSLI) is proposed as an alternative to the conventional matrix decoding systems for reconstructing multi-channel output signals from matrix encoded two channel signals. The proposed matrix decoding system makes use of virtual sources perceived by the listener to enhance the

* 정회원, 삼성전자 DMC 연구소
(DMC R&D center, Samsung Electronics CO., LTD.)
접수일자: 2010년7월13일, 수정완료일: 2010년8월8일

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. Estimating the angle information of the 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 the multi-channel layout can be evaluated by simple vector manipulation of the virtual source vectors. Using the virtual source vectors and the global source vector, 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. 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. Therefore, the listener can get a good perception of the dynamic effects of moving source. Also adequate inter channel separation is maintained with the proposed method.

II. 본 론

1. Virtual Source Location Information Estimation

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

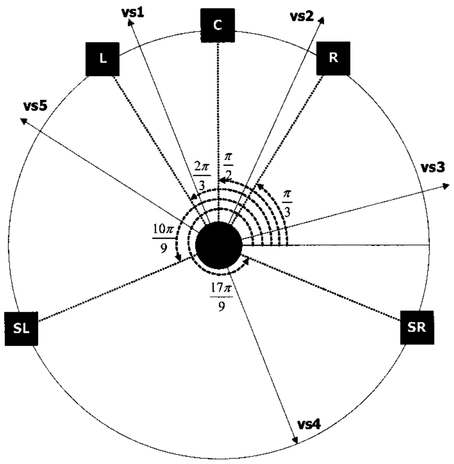


그림 1. 재생시 스피커 배치도
Fig. 1. Playback loudspeaker layout.

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[5~6] 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.

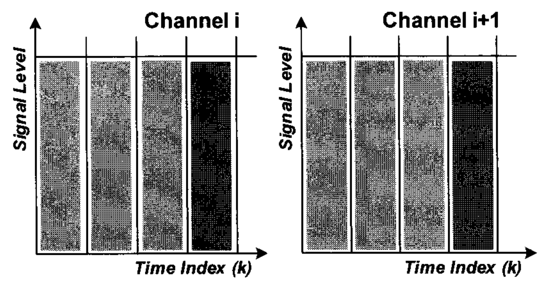


그림 2. 시간 k에서 L_k^i과 L_k^{i+1}의 레벨 정보
Fig. 2. Level information L_k^i and L_k^{i+1} at time index k.

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

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

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

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

$$\theta_k^{i,i+1} = \arccos(g_k^i) \quad (3)$$

$$\theta_k^{i,i+1} = \arcsin(g_k^{i+1}) \quad (4)$$

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). θ_i means the angle coordinate of channel i and θ_{i+1} means that of channel $i+1$.

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

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.

2. Virtual Source Location Information Estimation based Matrix Decoding System

A. System overview

Fig. 3 shows VSLI based matrix decoding system block diagram. The proposed matrix decoding system is composed of 4 discrete blocks. The first block, "Passive Matrix Decoder" block, performs pseudo inverse matrixing to extract crude multi-channel outputs from the matrix encoded stereo inputs. Section 2.B covers more detail scheme about.

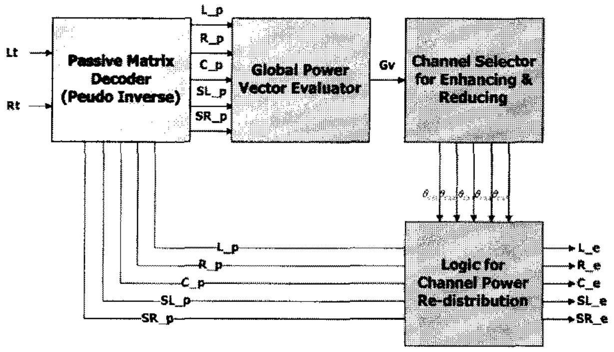


그림 3. VSLI 기반의 메트릭스 디코딩 시스템 블럭도
Fig. 3. VSLI based matrix decoding system block diagram.

The second block, "Global power vector evaluator" decides the global vector which is related to the perceived main sound source image. This block is composed of "Channel power vector evaluator" and "Virtual source power vector estimator".

The first sub-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 and the second sub-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 1 is used as the evaluation tool. Section 2.C covers more detailed description about this block set.

The third 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 forth 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. This block also increases the channel separation ratio by filtering each channel output with nonlinear function.

B. Passive Matrix Decoding

The encoding matrix, M, to down-mix 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). Since the encoding matrix, M, is under determined system, the inverse of the encoding matrix, M, is usually not specified by normal matrix inversion process. Therefore, the inverse of encoding matrix should be evaluated by Moore-Penrose generalized inverse[7~8]. 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.

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

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

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

C. 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 the virtual source vectors as illustrated in Fig. 4. As explained in section 2, the virtual source vectors can be determined by inverse power panning law. The equation for global vector estimation is as follows.

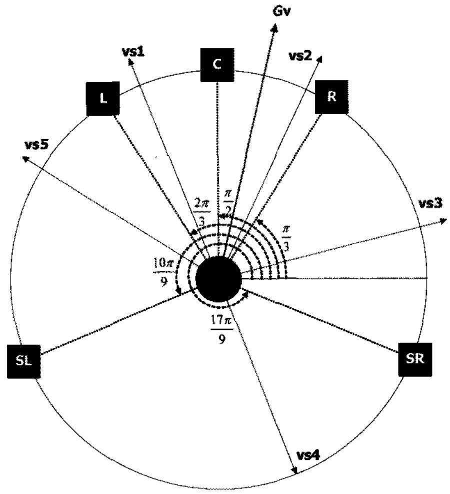


그림 4. 가상 음원 벡터 & 전역 벡터 추출

Fig. 4. Virtual source vector & Global vector estimation.

$$Gv = \sum_{i=1}^{N} vs_i \tag{8}$$

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).

As explained above, the global vector can be assumed to be an approximation of the perceived sound source vector[5~6, 9]. 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.

$$|Gv| = \left| \sum_{i=1}^{N} vs_i \right| \tag{9}$$

$$\angle Gv = \angle \sum_{i=1}^{N} vs_i \tag{10}$$

D. 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).

$$f(\theta_{ch}) = a \cdot \cos^n\theta_{ch} + b \cdot \sin^m\theta_{ch} \quad (11)$$

$$\theta_{ch} = |\angle Gv - \angle Ch_i| \quad (12)$$

In equation (11), θ is defined by the angular difference between a specific channel (Ch_i) and global vector (Gv) as shown in equation (12). Ch_i means position vector of each channel such as $R(i=1)$, $L(i=2)$, $C(i=3)$, $SL(i=4)$, $SR(i=5)$ on the standard ITU-R recommendation BS.775-1[10]. Therefore, $\angle Ch_1$ returns $\frac{\pi}{3}$ as shown in fig. 4. The constants, a, b, n, and m, are integer values which are chosen to stabilize the perceived sound images. To guarantee a good channel separation, each channel power is stretched by nonlinear equation (13). In equation (13), the constant k is a positive integer value. Both c and d are values between 0 and 1. P is the sum of the whole channel power.

$$g(x) = \begin{cases} x \cdot k & (c \cdot P < x^2) \\ x \cdot k^{10 \cdot (\frac{x^2}{P} - d)} & (d \cdot P < x^2 \le c \cdot P) \\ x & (x^2 \le d \cdot P) \end{cases} \quad (13)$$

By combining equation (11) and equation (13), channel power enhancing logic is determined as shown in equation (14).

$$Enhanced\_Ch_i = f(\theta_{Ch_i}) \cdot g(abs(Ch_i)) \quad (14)$$

With this signal adaptive method, 5 channel signals are updated in every 512 sample.

III. Experiment

To evaluate objective performance of the proposed method, the channel separation rate was measured by matrix encoded stream which is composed of sine singal in left channel and zeros in other channels with following coefficients. a=1, b=0, c=0.3, d=0.2, k=2, n=3, m=0. As shown in Table 1, minimum channel separation rate is 36dB which is better than that of prologic II/IIx. Most sound images evaluated by (8) are concentrated in frontal sound stage. This means that the enhancing function (14) enhances front channels better than rear channels. This is the reason why the rear channel separation rate is not improved even though the center channel separation rate is improved.

Subjective assessment was performed to evaluate the preference of the listeners. For the subjective assessment, down-mixed multi-channel sound track contents from 5 DVD titles in Table 2 were used as the test materials.

The blind triple-stimulus test[11] was adopted as a perceptual preference test to grade the perceived quality difference.

The test materials decoded by proposed method and those decoded by prologic II/IIx were labeled "T1" and "T2" respectively. The listening panels

표 1. DPLII와 채널 분리도 비교표
Table 1. Channel separation rate comparison with DPLII.

	Proposed	Prologic II/IIx
R	∞	∞
C	36dB	32dB
Ls	44dB	46dB
Rs	∞	∞

표 2. 테스트 음원
Table 2. Test materials.

A	Gladiator
B	Matrix
C	Pearl Harbor
D	Star Wars
E	Saving Private Ryan

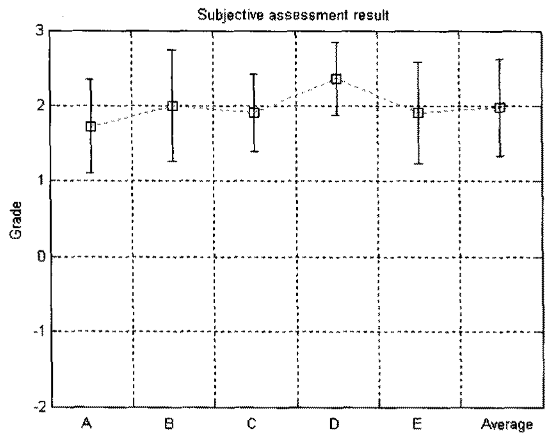


그림 5. 주관평가 결과
Fig. 5. Subjective assessment result.

were requested to choose one of the seven grades (+3 to −3) from the point of the perceived multi-channel moving sound effect.

As shown in Fig. 5, the proposed method induced a better moving sound effect for the listeners than prologic II/IIx. The average mean of the five test sequences was 1.98 and the standard deviation of all test sequences was 0.64.

IV. Conclusion

The proposed matrix decoder in this paper 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 enhance channel separation rate, nonlinear filter with trigonometric function is used for the all channels. By combining these nonlinear filter and channel re-distribution function with pre-defined up-mixer, both channel separation rate and moving sound perception are improved.

References

[1] Dolby labs, "Method for apparatus for audio matrix decoding," US Patent no. 7280664, Oct., 2007.

[2] Dolby labs, "Matrix-encoded surround-sound channels in a discrete digital sound format," US Patent no. 6624873, Sep., 2003.

[3] Lexicon Inc., "Sound reproduction," US Patent no. 5136650, Aug., 1992.

[4] DTS Inc., "Method of decoding two-channel matrix encoded audio to reconstruct multi-channel audio," WO Patent Pub. No. WO02/032186, Aug., 2003.

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

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

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

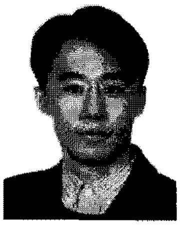
[8] J. A. Fill and D. E. Fishkind, "The Moore-Penrose generalized inverse for sums of matrices", *SIMAX*, vol. 21, no. 2, pp 629-635, 1999.

[9] 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 Consumer Electronics*, vol. 51, no. 4, pp. 1253-1259, Nov. 2005.

[10] ITU-R Recommendation, "Multi-channel stereophonic sound system with and without accompanying picture", *International Telecommunication Union*, BS. 775-1 1998.

[11] ITU-R Recommendation, "Subjective Assessment of Sound Quality", *International Telecommunication Union*, BS. 562-3, Geneva, 1990.

저 자 소 개



문 한 길(정회원)
2000년 서울대학교 전기공학부 학사 졸업.
2002년 서울대학교 전기·컴퓨터 공학부 석사 졸업.
2005년 서울대학교 전기·컴퓨터 공학부 박사 졸업.
2005년~현재 삼성전자 DMC 연구소 책임연구원
<주관심분야 : A/V 신호처리, 멀티채널 오디오, 오디오 코딩, A/V 전송 및 미디어 처리>