A Real-Time Audio Upmixing Method from Stereo to 7.1-Channel Audio

Chan Jun Chun[1], Young Han Lee[1], Yong Guk Kim[1],
Hong Kook Kim[1], and Choong Sang Cho[2]

[1] School of Information and Communications
Gwangju Institute of Science and Technology (GIST), Gwangju 500-712, Korea
{cjchun,cpumaker,bestkyg,hongkook}@gist.ac.kr
[2] Multimedia-IP Research Center
Korea Electronics Technology Institute, Seongnam-si, Gyeonggi-do 463-816, Korea
ideafisher@keti.re.kr

Abstract. In this paper, we propose a new method of upmixing stereo signals into 7.1-channel signals in order to provide more auditory realism. The proposed upmixing method employs an adaptive panning and a decorrelation technique for making more channels and reproducing natural reverberant surround sounds, respectively. The performance of the proposed upmixing method is evaluated using a MUSHRA test and compared with those of conventional upmixing methods. It is shown from the tests that 7.1-channel audio signals upmixed by the proposed method are preferred, compared to not only their original stereo audio signals but also 7.1-channel audio signals upmixed by conventional methods.

Keywords: Audio upmixing, multi-channel audio, conversion of stereo to 7.1-channel audio, adaptive panning, decorrelator.

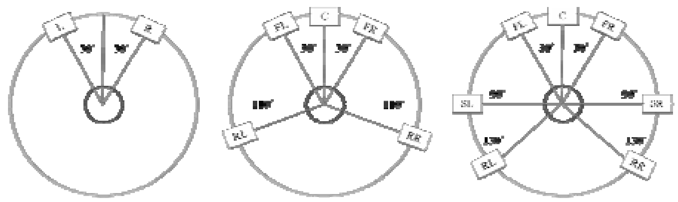
1 Introduction

Due to the rapidly increasing demand for audio applications, researchers have been investigating many audio fields. Among such fields, multi-channel, rather than stereo, audio systems utilize additional speakers to present a more realistic sound. Specifically, such audio systems not only improve ambient effects but also widen the sound.

In multi-channel audio systems, the number of channels for playing audio signals should be identical to that for recording in order to take full advantage of the system. If audio signals with smaller number of channels as in a playing-out speaker configuration are available, then the auditory realism cannot be expected. However, by using audio upmixing, i.e., conversion of stereo signals into multi-channel audio signals, this drawback can be mitigated. Thus, we can utilize mono or stereo audio content for multi-channel audio systems, providing more realistic sound.

There exist numerous multi-channel audio systems; typically, stereo, 5.1-channel, and 7.1-channel speaker configurations are shown in Figs. 1(a), 1(b), and 1(c), respectively, which are defined by ITU-R Recommendation BS.775-1 [1]. Although stereo

T.-h. Kim et al. (Eds.): FGCN 2010, Part II, CCIS 120, pp. 162–171, 2010.



(a) (b) (c)

Fig. 1. Speaker configurations defined by ITU-R Recommendation BS.775-1: (a) stereo, (b) 5.1-channel, and (c) 7.1-channel

or 5.1-channel audio content has been popularly available, 7.1-channel audio content is still relatively rare. Therefore, it is necessary to convert audio content with stereo or 5.1-channel into that suitable for a 7.1-channel speaker system.

In this paper, we propose an audio upmixing method that converts stereo audio signals to 7.1-channel audio signals. To this end, we first review several upmixing methods that have been applied for converting stereo signals into 5.1-channel signals. The methods to be reviewed here include a passive surround decoding method [2], an least mean square (LMS)-based method [3], a principal component analysis (PCA)-based method [4], and panning methods [5]-[7]. After comparing quality of such methods, we adopt adaptive panning to derive the center channel signal for 7.1-channel upmixing. Furthermore, a decorrelator is employed in order to reproduce reverberant effects in surround channel signals.

This paper is organized as follows. Following the introduction, we shortly review conventional upmixing methods for multi-channel speaker systems in Section 2. Next, we propose an upmixing method from stereo to 7.1-channel audio signals using adaptive panning and a decorrelator in Section 3. In order to compare the performance of the proposed method with those of the conventional methods, we conduct subjective tests based on multiple stimuli with hidden reference and anchor (MUSHRA) tests [8] in Section 4. Finally, we conclude this paper in Section 5.

2 Conventional Upmixing Methods

The following subsections describe several conventional upmixing methods including passive surround decoding, LMS-based, PCA-based, adaptive panning, constant power panning and speaker-placement correction amplitude panning (SPCAP) methods.

2.1 Passive Surround Decoding Method

The passive surround decoding (PSD) method is an early passive version of the Dolby Surround Decoder [2]. In this method, a center channel is obtained by adding the original left and right channels. On the other hand, a surround channel can be derived by subtracting the right channel from the left channel. That is, the center and the surround channels are obtained as

$$Center(n) = \left(x_L(n) + x_R(n)\right)/\sqrt{2}, \tag{1}$$

$$Rear(n) = \left(x_L(n) - x_R(n)\right)/\sqrt{2} \tag{2}$$

where $x_L(n)$ and $x_R(n)$ denote the left and the right audio sample at the time index n, respectively. Note that in order to maintain a constant acoustic energy, the center and the surround channel are lowered by 3 dB, which is implemented by multiplying 1/2 to the center and the surround channel signals.

However, there are two surround channels in the 5.1-channel configuration as shown in Fig. 1(b). In this paper, a discrete Hilbert transform is used to generate such channels [9]. By using a finite-duration impulse response (FIR) approximation having a constant group delay, we can implement the discrete Hilbert transform. In particular, the approximation is done using a Kaiser window which is defined as

$$h(n) = \begin{cases} \dfrac{I_0\left(\beta\left(1-\left[\dfrac{n-n_d}{n_d}\right]^2\right)^{1/2}\right)}{I_0(\beta)} \cdot \dfrac{\sin\left(\pi\dfrac{n-n_d}{2}\right)}{\pi\dfrac{n-n_d}{2}}, & 0 \le n \le M \\ 0, & otherwise \end{cases} \tag{3}$$

where M is the order of the FIR discrete Hilbert transform, and n_d is equal to $M/2$. In this case, M and β are set to 31 and 2.629, respectively.

2.2 LMS-Based Upmixing Method

The LMS-based upmixing method creates the center and surround channels using the LMS algorithm [3]. In this method, one of the original stereo channels is taken as a desired signal, $d(n)$, of the adaptive filter, and the other is considered as an input, $x(n)$. Then, the error signal, $e(n)$ is the difference between the output, $y(n)$, of the filter and the desired signal, $d(n)$. Finally, $y(n)$ is defined as a linear combination of the input signals such as

$$y(n) = \mathbf{w}^T(n)\mathbf{x}(n) = \mathbf{w}(n)\mathbf{x}^T(n) \tag{4}$$

where $\mathbf{x}(n) = \left[x(n)\ x(n-1)\ \Lambda\ x(n-N+1)\right]^T$ and $\mathbf{w}(n) = \left[x_0\ w_1\ \Lambda\ w_{N-1}\right]^T$. In Eq. (3), $\mathbf{w}(n)$ is a coefficient vector of the N-tapped adaptive filter that is obtained based on the LMS algorithm as

$$\mathbf{w}(n+1) = \mathbf{w}(n) + 2\mu e(n)\mathbf{x}(n) \tag{5}$$

where μ is a constant step size, which is set to 10^{-4} in this paper. As a result, $y(n)$ and $e(n)$ are the signals for the center and the surround channel, respectively. Similarly, a discrete Hilbert transform using a Kaiser window is utilized to determine surround channels as shown in Eq. (3).

2.3 PCA-Based Upmixing Method

The PCA-based upmixing method decomposes the original stereo signals into two different signals, where one is highly correlated but the other is somewhat uncorrelated [4]. In other words, to derive the center and the surround channels, a 2 x 2 covariance matrix, $\mathbf{A}$, is obtained as

$$\mathbf{A} = \begin{bmatrix} \mathrm{cov}(x_L, x_L) & \mathrm{cov}(x_L, x_R) \\ \mathrm{cov}(x_R, x_L) & \mathrm{cov}(x_R, x_R) \end{bmatrix} \tag{6}$$

where $\mathrm{cov}(x_p, x_q)$ is the covariance of x_p and x_q, and p and q could be the left channel, L, or the right channel, R. The covariance matrix in Eq. (5) has two eigenvectors, which become the basis vectors for a new coordinate system. These eigenvectors are then used as weight vectors corresponding to the left and right channels to generate the center and surround channels, such as

$$Center(n) = c_L x_L(n) + c_R x_R(n), \tag{7}$$

$$Rear(n) = s_L x_L(n) + s_R x_R(n) \tag{8}$$

where $[c_L \;\; c_R]$ is the eigenvector corresponding to the greatest eigenvalue and $[s_L \;\; s_R]$ is the other eigenvector.

2.4 Adaptive Panning Method

The adaptive panning (ADP) method generates the center and surround channels by panning the original stereo signals [5]. A weight vector for ADP is recursively estimated using the LMS algorithm. Let us define $y(n)$ to be a linear combination of the original stereo signals, $\mathbf{x}(n) = [x_L(n) \; x_R(n)]^T$, with a weight vector, $\mathbf{w}(n) = [w_L(n) \;\; w_R(n)]^T$. Then, $y(n)$ is represented as

$$y(n) = \mathbf{w}^T(n)\mathbf{x}(n) = \mathbf{w}(n)\mathbf{x}^T(n) \tag{9}$$

where $w_L(n)$ and $w_R(n)$, the elements of the weight vector corresponding to the left and the right channels, respectively, are then estimated using the LMS algorithm as

$$w_L(n+1) = w_L(n) - \mu\, y(n)[x_L(n) - w_L(n)y(n)], \tag{10}$$

$$w_R(n+1) = w_R(n) - \mu\, y(n)[x_R(n) - w_R(n)y(n)] \tag{11}$$

where μ is a constant step size and set to 10^{-10} in this paper. Finally, the center and surround channels can be determined as

$$Center(n) = w_L(n)x_L(n) + w_R(n)x_R(n), \tag{12}$$

$$Rear(n) = w_R(n)x_L(n) - w_L(n)x_R(n). \tag{13}$$

Finally, in order to derive surround channels, we also use a discrete Hilbert transform as described in Section 2.1.

2.5 Constant Power Panning Method

In this method, audio signals for additional channels are determined by panning original signals. If each additional channels is mixed with stereo audio signals, denoted as L and R shown in Fig. 2, we have

$$y(n) = g_L x_L(n) + g_R x_R(n) \tag{14}$$

where $y(n)$ is the audio signal for an additional channel. The g_L and g_R are panning gains for the left and right channels, $x_L(n)$ and $x_R(n)$, respectively. In order to estimate panning gains, θ_m, is calculated as

$$\theta_m = \begin{cases} \dfrac{\theta_i - \theta_1}{\theta_4^{`} - \theta_1} \cdot 90, \text{ if } \theta_1 \geq \theta_i \\ \dfrac{\theta_i - \theta_1}{\theta_4 - \theta_1} \cdot 90, \text{ if } \theta_4 \leq \theta_i < \theta_1 \\ \dfrac{\theta_i - \theta_1^{`}}{\theta_4^{`} - \theta_1} \cdot 90, \text{ if } \theta_i \geq \theta_4 \end{cases} \tag{15}$$

where θ_i as shown in Fig. 2 is the placement angle for the i-th additional channel and it is mapped to θ_m. Then, the panning gains, g_L and g_R, are determined as

$$g_L = \cos\theta_m \textit{ and } g_R = \sin\theta_m. \tag{16}$$

When stereo audio signals are converted into 5.1-channel audio signals, Table 1 shows panning gains for each additional channel. Finally, we determine the center and surround channels by using the equation of

$$Center(n) = 0.7071 \cdot x_L(n) + 0.7071 \cdot x_R(n), \tag{17}$$

$$RL(n) = 0.9135 \cdot x_L(n) + 0.4067 \cdot x_R(n), \tag{18}$$

$$RR(n) = 0.4067 \cdot x_L(n) + 0.9135 \cdot x_R(n). \tag{19}$$

In order to create LFE channel, we employ an FIR low-pass filter having a cut-off frequency of 200 Hz. By filtering audio signals for the center channel, audio signals for the LFE channel are derived.

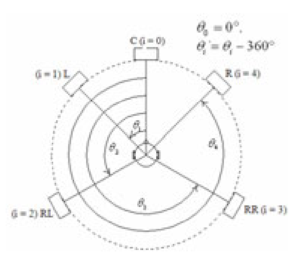


Fig. 2. Angles for computing panning gains used in Eq. (14) for the constant power panning method

Table 1. Panning gains for constant power panning from stereo to 5.1-channel audio

Channel	g_L	g_R
C (i=0)	0.7071	0.7071
L (i=1)	1	0
RL (i=2)	0.9135	0.4067
RR (i=3)	0.4067	0.9135
R (i=4)	0	1

2.6 Speaker-Placement Correction Amplitude Panning (SPCAP) Method

Similarly to the constant power panning method, the SPCAP method derives additional channels by panning original signals [7]. In SPCAP, however, a cosine-weighted panning method is used for calculating a panning value. If stereo audio signals are upmixed from stereo to 5.1-channel audio signals using SPCAP, two panning values are estimated as

$$p_L = \frac{1}{2}\left[1 + \cos(\theta_i - \theta_L)\right] \text{ and } p_R = \frac{1}{2}\left[1 + \cos(\theta_i - \theta_R)\right] \tag{20}$$

where L and R are the left and right channel, respectively, and θ_i is the placement angle for the additional channel as shown in Fig. 2. In order to conserve power, the panning values are normalized to obtain two panning gains, g_L and g_R, as

$$g_L = \frac{p_L}{\beta} \text{ and } g_R = \frac{p_R}{\beta} \tag{21}$$

where $\beta = p_L + p_R$. By using Eq. (21), audio signals for the additional channels are derived as

$$Center(n) = 0.5 \cdot x_L(n) + 0.5 \cdot x_R(n), \tag{22}$$

$$RL(n) = 0.8338 \cdot x_L(n) + 0.1662 \cdot x_R(n), \tag{23}$$

$$RR(n) = 0.1662 \cdot x_L(n) + 0.8338 \cdot x_R(n). \tag{24}$$

Similarly, we use an FIR low-pass filter having a cut-off frequency of 200 Hz to derive the LFE channel, considering the input of the low-pass filter as the center channel.

3 Upmixing from Stereo to 7.1-Channel Audio

In this section, we propose a new upmixing method from stereo to 7.1-channel audio signals. Fig. 3 shows an overall structure of the proposed method. As shown in the figure, the proposed method basically combines two upmixing methods; upmixing from stereo to 5.1-channel signals, and upmixing from 5.1-channel to 7.1-channel signals. The stereo-to-5.1 upmixing block is adopted from one of the upmixing methods described in Section 2. On the other hand, the 5.1-to-7.1 upmixing block employs a decorrelator to generate the surround channels for the 7.1-channel configuration.

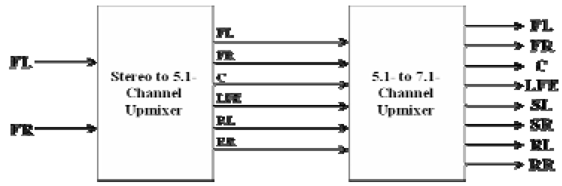
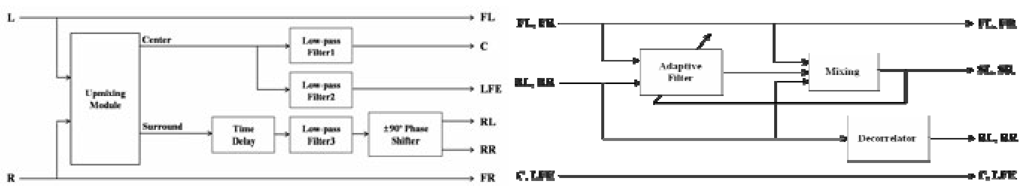


Fig. 3. Overall structure of upmixing stereo to 7.1-channel audio signals

3.1 Stereo to 5.1-Channel Upmixing

Fig. 4(a) shows a detailed block diagram for upmixing stereo to 5.1-channel audio signals based on the adaptive panning method described in Section 2.4. Note here that the adaptive panning method is selected from the exhaustive subjective tests. In the figure, each channel is labeled as FL (front left), FR (front right), C (center), LFE (low frequency enhancement), RL (rear left), or RR (rear right).



(a) (b)

Fig. 4. Block diagram for upmixing: (a) stereo to 5.1-channel audio signals, (b) 5.1-channel to 7.1-channel audio signals

3.2 5.1 to 7.1-Channel Upmixing

As shown in Figs. 1(b) and 1(c), the channel configuration for 7.1-channel is different from that of 5.1-channel. In other words, the surround channels in 5.1-channel look like being split into two pairs of stereo channels such as one pair of side channels, SL (side left), SR (side right), and the other pair of real channels, RL (rear left) and RR (rear right). The side channels go frontier than the surround channels in 5.1-channel, but the rear channels go back.

Fig. 4(b) shows a block diagram for performing the 5.1-to-7.1 channel upmixing. Similarly to the block diagram shown in Fig. 4(b), the adaptive panning method is also applied to create SL and SR for 7.1-channel. Here, SL and SR are determined by panning the front and rear channels as

$$SL(n) = w_{FL}(n)FL(n) + w_{RL}(n)RL(n), \tag{25}$$

$$SR(n) = w_{FR}(n)FR(n) + w_{RR}(n)RR(n). \tag{26}$$

The weight vectors are recursively estimated using the LMS algorithm as

$$w_{FL}(n+1) = w_{FL}(n) + \mu SL(n)[FL(n) - w_{FL}(n)SL(n)] \tag{27}$$

$$w_{RL}(n+1) = w_{RL}(n) + \mu SL(n)[RL(n) - w_{RL}(n)SL(n)] \tag{28}$$

$$w_{FR}(n+1) = w_{FR}(n) + \mu SR(n)[FR(n) - w_{FR}(n)SR(n)] \quad (29)$$

$$w_{RR}(n+1) = w_{RR}(n) + \mu SR(n)[RR(n) - w_{RR}(n)SR(n)] \quad (30)$$

where μ is a constant step size, set at 10^{-10}.

In order to add reverberation effects to the rear channels, we employ a decorrelator that is designed by randomizing the phase response in the frequency domains. The following subsections further describe the decorrelator design and a mixing method using the decorrelator in detail.

3.2.1 Decorrelator Design

One approach of designing a decorrelator is to employ the magnitude and phase randomization. Initially, the time-domain original audio signals are transformed into the frequency-domain ones using a Fourier transform. Then, the magnitude and phase responses of the transformed audio signals are obtained. Subsequently, we randomize the magnitude and phase responses, but unwanted discontinuity in the response boundaries could be occurred. Therefore, we employ a cosine interpolation to eliminate this discontinuity with the weight value shown in Table 2. Finally, we determine decorrelated audio signals using an inverse Fourier transform.

Table 2. Weights in the phase response

(kHz)	< 2	2 – 4	4 – 8	8 – 16	> 16
Weight	1.0	0.5	0.25	0.125	0.625

3.2.2 Mixing Method

After the decorrelation process, original and decorrelated audio signals are mixed to generate the rear left and right channel signals, such as

$$RL(n) = (0.7071 \cdot SL(n) + 0.7071 \cdot DL(n))/2, \quad (31)$$

$$RR(n) = (0.7071 \cdot SR(n) + 0.7071 \cdot DR(n))/2 \quad (32)$$

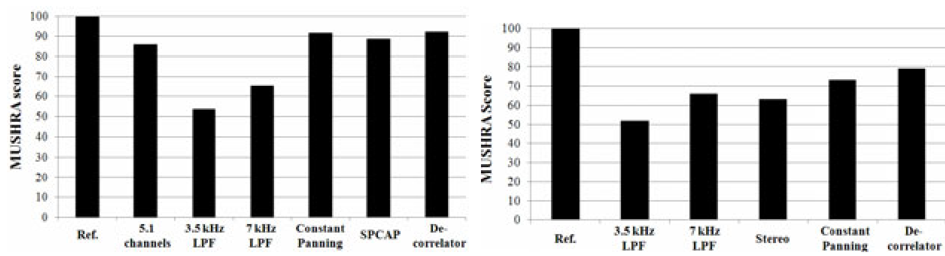
where $DL(n)$ and $DR(n)$ are decorrelated audio signals from $SL(n)$ and $SR(n)$, respectively. Note that in order to match the energy of the original audio signal and that of the upmixed audio signal, the rear channel signals are lowered by 6 dB, which is implemented by multiplying $1/2$.

4 Performance Evaluation

We compared the quality of the proposed upmixing method in two aspects such as upmixing from 5.1-channel to 7.1-channel and upmixing from stereo to 7.1-channel audio signals. Thus, we conducted MUSHRA tests in compliance with the ITU multi-channel configuration standard defined by ITU-R Recommendation BS.775-1 [1]. The audio contents, sampled at 44.1 kHz, to be compared in the test were as follows:

- Hidden reference
- 3.5 kHz low-pass filtered anchor
- 7 kHz low-pass filtered anchor
- 5.1-channel audio signals (in case of upmixing from stereo to 7.1-channel, stereo audio signals)
- Upmixed audio signals obtained by the conventional upmixing methods, and
- Upmixed audio signals obtained by the proposed method.

The MUSHRA test results for upmixing from 5.1-channel to 7.1-channel and from stereo to 7.1-channl are shown in Figs. 5(a) and 5(b), respectively. It was shown from the figures that the upmixed 7.1-channel audio signals were preferred, compared to the original audio signals. Moreover, the proposed upmixing method outperformed the conventional methods.



(a) (b)

Fig. 5. Comparison of MUSHRA test scores for the audio signals upmixed by different methods: (a) upmixing from 5.1-channel to 7.1-channel audio signals, (b) upmixing from stereo to 7.1-channel audio signals

5 Conclusion

In this paper, we proposed an upmixing method based on adaptive panning and decorrelation. The proposed upmixing method could convert stereo to 7.1-channel signals. Moreover, comparing the performance of the proposed method with those of conventional methods in terms of MUSHRA test scores, it was shown that 7.1-channel audio signals generated by the proposed upmixing method were preferred rather than those by the conventional methods.

Acknowledgement

This work was supported in part by "Fusion-Tech. Developments for THz Info. & Comm." Program of GIST in 2010, and in part by the Ministry of Knowledge Economy (MKE), Korea, under the Information Technology Research Center (ITRC) support program supervised by the National IT Industry Promotion Agency (NIPA) (NIPA-2010-C1090-1021-0007).

References

1. ITU-R BS.775-1: Multi-Channel Stereophonic Sound System with or without Accompanying Picture (1994)
2. Dolby Laboratory, `http://www.dolby.com/professional/getting-dolby-technologies/index.html`
3. Bai, M.R., Shih, G.-Y., Hong, J.-R.: Upmixing and downmixing two-channel stereo audio for consumer electronics. IEEE Trans. on Consumer Electronics 53, 1011–1019 (2007)
4. Chun, C.J., Kim, Y.G., Yang, J.Y., Kim, H.K.: Real-time conversion of stereo audio to 5.1 channel audio for providing realistic sounds. International Journal of Signal processing, Image processing and Pattern Recognition 2(4), 85–94 (2009)
5. Irwan, R., Aarts, R.M.: Two-to-five channel sound processing. J. Audio Eng. Soc. 50, 914–926 (2002)
6. West, J.R.: Five-channel Panning Laws: an Analytical and Experimental Comparison. M.S. Thesis, Department of Music Engineering, University of Miami, Coral Gables, Florida (1998)
7. Sadek, R., Kyriakakis, C.: A novel multichannel panning method for standard and arbitrary loudspeaker configurations. In: Proc. of 117th AES Convention, Preprint 6263, San Francisco, CA (2004)
8. ITU-R BS. 1534-1: Method for the Subjective Assessment of Intermediate Quality Levels of Coding System (2003)
9. Bosi, M., Goldberg, R.E.: Introduction to Digital Audio Coding and Standards. Kluwer Academic Publishers, Massachusetts (2002)