Audio Engineering Society
Conference Paper
Presented at the 19th International Conference
2001 June 21–24 Schloss Elmau, Germany

A method to convert stereo to multi-channel sound

Roy Irwan and Ronald M. Aarts

Philips Research Laboratories (WY81), Professor Holstlaan 4

Eindhoven, 5656 AA, The Netherlands

ABSTRACT
While stereo music reproduction was a dramatic advance over mono, recently a transition to multi-channel audio has created a more involving experience for listeners. A new method to enhance the conversion from stereo to multi-channel sound reproduction is presented. A better sound distribution to the surround channels is achieved by using cross-correlation technique, and a stable center image is obtained using Principal Component Analysis.

INTRODUCTION

Since the presence of Digital Versatile Disk (DVD) and Super Audio CD (SACD), multi-channel audio has become popular in the sound systems for consumer use today. This paper presents a new method which converts two-channel stereo to multi-channel sound reproduction, using a three dimensional representation (hereafter referred to as *space mapping*).

Although many authors have introduced multi-channel sound systems with a large number of channels, we restrict ourselves to a home cinema setup which has been investigated to be sufficient with five channels. The desired setup is shown in Fig. 1, where L, C, R, S_L, S_R refer to left, right, center, left-surround, and right-surround channel, respectively. This setting is adopted from the ITU multi-channel configuration [1], with three loudspeakers placed in front of the listener, and the other two at the back.

By use of Principal Component Analysis (PCA) [2] we develop an algorithm that produces a vector which indicates the direction of both dominant signal (y) and remaining signal (q) as shown in Fig. 2. Note that these two directions are perpendicular to each other creating a new coordinate system. These two signals are then used as basis signals in the matrix encoding.

PROPOSED METHOD

The algorithm presented in this paper offers two improvements above existing multi-channel techniques. Firstly, a problem associated with channel cross-talk is reduced, and therefore sound localisation is better achieved. The latter gives more space to the listener to enjoy the offered program

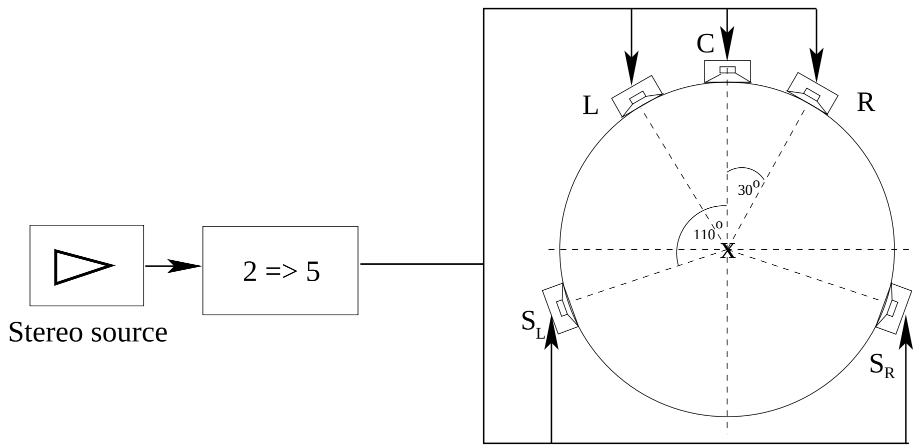

Fig. 1: ITU reference configuration [1]. The reference listening position (sweet spot) is indicated by **x**. Left and right channels are placed at angles $\pm 30°$ from C, and the two surround channels are placed at angles $\pm 110°$ from C.

rather than restricting the listener in the sweet spot[1]

Secondly, a better sound distribution to the surround channels is achieved by using a cross-correlation technique, while maintaining energy preservation. Energy preservation criterion is an important constraint that has been used to design multi-channel matrices [4]. One reason is it is a factor in backward and forward stereo compatibility. Furthermore, the preservation criterion ensures that all signals present in the two-channel transmitted signals are produced at a correct power level, so that the balance between the different signals sounds in the recording is not disturbed.

[1]This algorithm has been a patent pending [3].

The center channel

We consider the three-channel approach first. It is known that the sound quality of stereo sound reproduction can be improved by adding an additional loudspeaker between each adjacent pair of loudspeakers. For example, an additional center loudspeaker C can be fed with the signal $\frac{1}{2}(x_L + x_R)$, where x_L and x_R represent signals from left and right channel, respectively. A major drawback of this approach is that cross talk with left and right channels is inevitable, and therefore narrowing the stereo image.

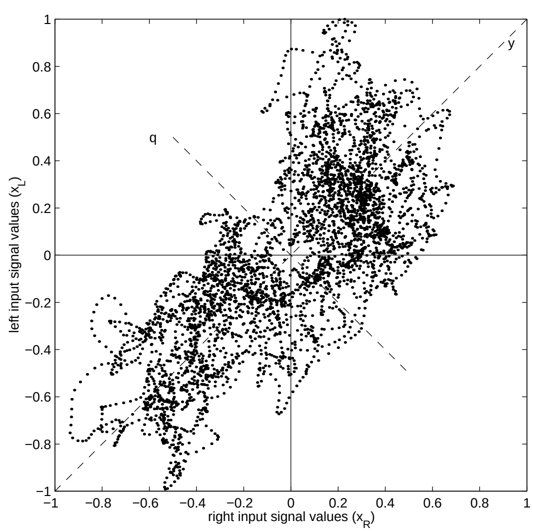

Fig. 2: A Lissajous plot of stereo signal forming. The dashed lines represent a new coordinate system based on both the dominant signal (y) and remaining signal (q).

In the following, we derive a center channel's gain using the direction of a stereo image which is time varying. The dashed line corresponding to the dominant signal (y) shown in Fig. 2 represents the image direction for left and right channel in stereo signals. It can be shown that this direction can be detected by using a simple direction sensing algorithm. At each iteration of the algorithm, the weights at time instant k, i.e. $\mathbf{w}(k) = [w_l(k)\ w_r(k)]^T$, corresponding to the left and right channel respectively, are computed using the following equation:

$$\mathbf{w}(k) = \mathbf{w}(k-1) + \mu(k)\left[\mathbf{x}(k) - \mathbf{w}(k)y(k)\right], \tag{1}$$

where

$$\begin{aligned} y(k) &= \mathbf{w}^T\mathbf{x}(k), \\ \mathbf{x}(k) &= [x_l(k)\ \ x_r(k)]^T, \end{aligned} \tag{2}$$

and the step size is signal dependent as

$$\mu(k) = \frac{C}{y(k)+C}, \tag{3}$$

where C is a constant typically 0.001.

The direction of a stereo image in terms of angle in radians can then be computed using,

$$\alpha(k) = \arctan\left(\frac{w_l(k)}{w_r(k)}\right). \tag{4}$$

Figure 3 shows the fluctuations of the angle α computed from a DVD movie fragment where abrupt changes from one channel to the other are present. Note that the horizontal axis represents the number of blocks (B) where each block contains 512 samples.

Now we can represent a pair of stereo signals using this vector. This is a vector of unit length having the right channel gain in the horizontal axis, and the left channel gain in the vertical axis, as shown in Fig. 4a. To map this stereo vector onto a three-channel vector, doubling the angle α is performed. This transformation can be computed using simple sine and cosine rules:

$$\begin{aligned} c_{lr} &= w_r^2 - w_l^2 \\ c_c &= 2w_l w_r, \end{aligned} \tag{5}$$

which is depicted in Fig. 4b. Note that this transformation works only for nonnegative α. For negative α, a multiplication by a factor two results the vector to be in a lower quadrant, and therefore no gain can be derived for the center channel. To overcome this problem, extra information should be used which is described in the next section.

The surround channels

The surround channels are generally used for creating ambience effects as by music. While for applications in the film industry the surround channels are used for sound effects. A common technique for ambience reconstruction is the use of delayed front channel information for the surround channels. Dolby Pro Logic for instance, has delayed the surround sounds so as to arrive at the listeners' ears at least 10 ms later than the front sounds [4].

Environmental and ambience effects can also be computed by considering left and right channel variation ($x_L - x_R$) in the encoded signals. This variation is usually referred to as the *anti-phase* components, the amount of which can be represented by the remaining signal q (Fig. 2). However, it can be expected that when the amount of the dominant signal equals to that of the anti-phase components, an ambiguity appears since there is no way to determine the direction vector uniquely. This state can be visualised that the distribution in Fig. 2 is no longer an ellipse but a circle ($y = q$), causing α to be arbitrary.

Obviously extra information is necessary when dealing with this sort of ambiguity. In this paper we propose to use a known technique to measure the amount of anti-phase components by use of the correlation coefficient which can be computed recursively by using only a few arithmatic operations as [5].

$$\hat{\rho}(k) = \hat{\rho}(k-1) + \gamma\left[2x_L(k)x_R(k) - \left(x_L(k)^2 + x_R(k)^2\right)\hat{\rho}(k-1)\right], \tag{6}$$

where γ is the step size determining the time constant, and the symbol $\hat{}$ is used to denote that it is an estimation of the true ρ.

Since

$$-1 \le \rho \le 1, \tag{7}$$

it is possible that the anti-phase components exceed the dominant signal ($|y| < |q|$). In this case we treat the input signals as uncorrelated, and it is therefore

$$\rho_0 = \begin{cases} \rho & 0 \le \rho \le 1 \\ 0 & \text{otherwise.} \end{cases} \tag{8}$$

Space mapping

To avoid ambiguity discussed in the previous section, use of both direction of the stereo image and the correlation coefficient is necessary. The latter is included in the mapping (Fig. 4) by, for example, putting the surround channels in the vertical plane, as shown in Fig. 5.

The angle β can be defined to represent the actual surround information by means of the adaptive correlation coefficient, for example, by using the following expression

$$\beta(k) = \arcsin(1 - \rho_0(k)), \tag{9}$$

and hence

$$0 \le \beta(k) \le \frac{\pi}{2}. \tag{10}$$

Thus as the amount of anti-phase components increases (input signals are weakly correlated), the angle β also increases, which reduces the total distribution to the front channels. On the other hand, when the input signals are strongly correlated (mono signals), β approaches zero, producing a larger distribution to the front channels. This principle makes the energy preservation criterion work since the matrix (Section **Matrixing**) can be proved to be orthogonal. In other words, the total output energy is preserved regardless of the encoding process.

Since the direction vector on the horizontal plane is now tilted with an angle β, recalculation for the projections is necessary. Using straightforward trigonometry we obtain,

$$\begin{aligned} c'_{lr} &= \cos\beta \times c_{lr} \\ c'_c &= \cos\beta \times c_c \\ c_s &= \sin\beta. \end{aligned} \tag{11}$$

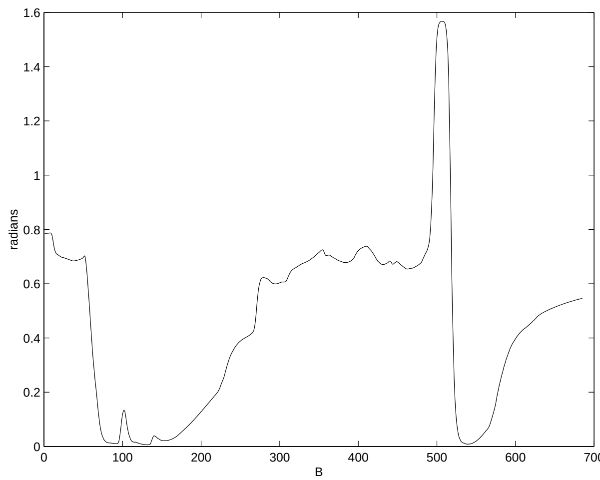

Fig. 3: Fluctuation of the direction, α, computed from a DVD fragment where a car is passing by with a high speed from one channel to the other. The horizontal axis represents the number of blocks B, where B = 512. Note that the sampling frequency used in this experiment is equal to 44.1 kHz, and therefore the above fragment has taken 8 seconds approximately.

Matrixing

The mapping process, also known as matrixing, is a key issue when dealing with multi-channel sound systems. The system as described so far decodes four channel signals as L, C, R and S. The stereo surround in the rear channels can be accomplished by using a decorrelator.

Our adaptive matrix to produce four channel signals is described as:

$$\begin{bmatrix} u_l(k) \\ u_r(k) \\ u_c(k) \\ u_s(k) \end{bmatrix} = \begin{bmatrix} c_l(k) & w_r(k) \\ c_r(k) & -w_l(k) \\ c_c(k)' & 0 \\ 0 & c_s(k) \end{bmatrix} \begin{bmatrix} y(k) \\ q(k) \end{bmatrix} \tag{12}$$

where the base signals are computed using a rotation of the input signals:

$$\begin{aligned} y(k) &= w_l(k)x_l(k) + w_r(k)x_r(k) \\ q(k) &= w_r(k)x_l(k) - w_l(k)x_r(k), \end{aligned} \tag{13}$$

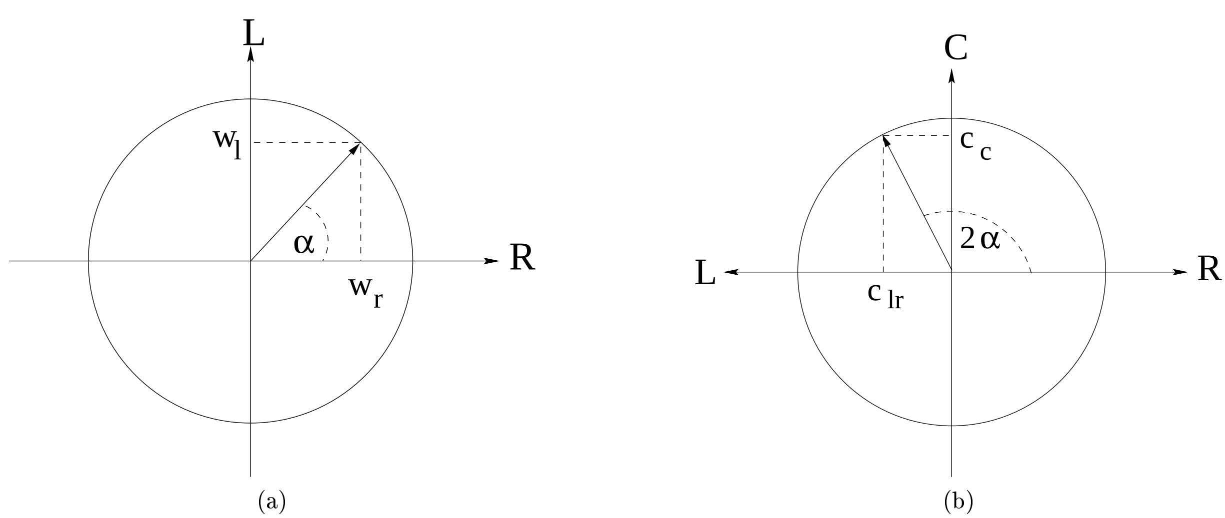

Fig. 4: (a) Direction vector plots of stereo signals, (b) The corresponding 3-channel representation by doubling the angle α.

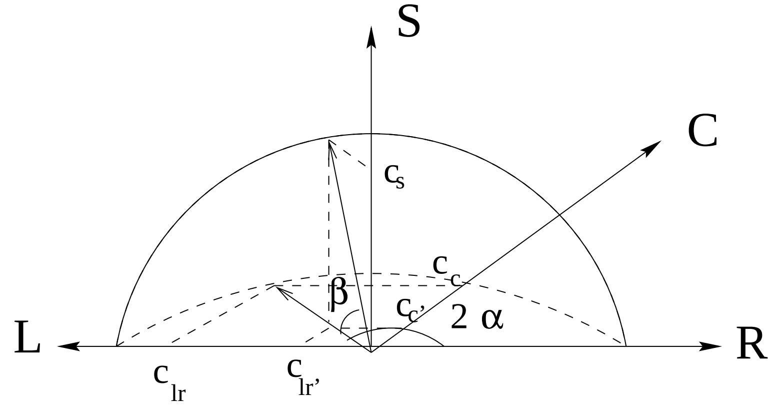

Fig. 5: Space mapping showing front channels (horizontal plane) and surround (vertical plane). A parameter β determines the level of surround information with respect to the front channel sounds.

and

$$c_l = \begin{cases} -c'_{lr} & c'_{lr} < 0 \\ 0 & \text{otherwise} \end{cases} \tag{14}$$

$$c_r = \begin{cases} c'_{lr} & c'_{lr} \geq 0 \\ 0 & \text{otherwise.} \end{cases}$$

In this paper, the Lauridsen [6] decorrelator is used to obtain stereo surround because of its simplicity. This decorrelator can be viewed as FIR comb filters with two taps each for surround left and surround right. A time delay of 10 ms is used between the taps, which is determined experimentally.

CONCLUSION

A new decoder for multi-channel reproduction of stereo sound based on a novel space mapping is presented. The three dimensional representation can be used not only as a visual analysis tool for matrix surround sound systems, but also to manipulate such systems to offer user interactivity possibilities.

A new parameter computed using correlation coefficient has been included in the decoder to determine the rear channels distribution. This parameter also solves the ambiguity problem when the direction of a stereo image is indeterminate.

ACKNOWLEDGEMENT

The authors thank David Roovers for initiating the project.

REFERENCES

[1] ITU-R Recommendation BS.1116, "Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems", International Telecommunication Union, Geneva, Switzerland, (1994).

[2] Te-Won Lee, "Independent Component Analysis: Theory and Applications", Kluwer Academic publisher, (1998).

[3] R. Irwan, Ronald M. Aarts, "Multi-channel stereo converter for deriving a stereo surround and/or audio centre signal", PHNL000401 EP.P, Filing Date: 19-7-2000.

[4] R. Dressler, "Pro logic surround decoder principles of operation", http:www.dolby.com, Dolby Laboratories, (1997).

[5] Ronald M. Aarts, R. Irwan, A. J. E. M. Janssen, "Efficient tracking of the cross-correlation coefficient", *submitted to IEEE Trans. Speech and Audio Processing*, (1999).

[6] H. Lauridsen, "Experiments concerning different kinds of room-acoustics recording", Ingenioren (*in Danish*), **47**, (1954).