Proceedings of the Institute of Acoustics

THE PRODUCTION OF STEERABLE BINAURAL INFORMATION FROM TWO-CHANNEL SURROUND SOURCES

D A Keating & M P Griffin

Department of Cybernetics, University of Reading, Whiteknights, PO Box 225, Reading, Berks, RG6 2AY.

1. INTRODUCTION

Conventional stereo recordings when listened to over headphones usually result in poor spacial imaging. 2-channel surround formats can work to a limited extent as the surround information is in anti-phase in the left & right signals which confuses the auditory system of the brain. The brain tends to assume sources are above and behind the head when it is unable to place them due to conflicting auditory clues. The brain is bad at telling front sources from back as the differences tend to be subtle hence these ambiguities are usually resolved by head rotations. In headphone listening head movements do not result in a change in the signals at the ears and thus these ambiguities cannot be resolved. One solution is to use binaural recordings but a problem is that there are a low number of such binaural recordings available compared with stereo. These recordings also give poor imagery when listened to over loudspeakers (the hole in the middle effect).

Virtual reality (VR) and telepresence systems [1] require audio and visual images which remain coherent for large head movements. Such systems are often helmet based and so headphones are used for the audio reproduction. VR head tracking systems are available commercially which give information on head position and orientation. If the head orientation signals could be used to steer surround audio then the imagery obtained from headphones could be good. The effect of the headphones on the conch resonance [2] needs to be compensated for as this is a potential source of lack of realism. What is required therefore is system that produces steerable binaural information from traditional sources. As most domestic audio products are two-channel it is these sources that need to be considered for home use. Nearly all the existing surround formats are horizontal only as full periphonics (with height) really requires more than 2 channels.

2. SURROUND SOUND SOURCES

2.1 Two-Channel Sources

The most common 2-channel source available is conventional stereo. The directional encoding of stereo is not properly defined but could be stated to be that which is perceived when the recording is replayed on two loudspeakers equi-distant from the listener at angles of +/- 30 degrees to the front. Perhaps the second most common source for the home user is the two-channel Dolby surround encoding which is available on a large number of home videos. This is specifically designed as a system to augment a visual image. A system which is designed to be used for audio alone and which is also fully stereo compatible is the 2-channel UHJ ambisonic system. A large number of CDs and vinyl records are available in this format although they are sometimes labelled as such only on the discs themselves and not on the sleeves. Some dummy head binaural recordings are available but these are far from common. There are a number of minor formats such as Q-sound but these will not be considered here as they are often ill- or undefined and available only in very small numbers.

2.2 Multi-Channel Sources

Although in the past there were some multi-channel quadraphonic formats these are now obsolete and for home use there is now little in the way of multi-channel formats available. One pseudo multi-channel source is that derived from the Dolby pro-logic system. The pro-logic decoder produces four output signals which differ from those of a conventional Dolby surround decoder when there is a detectable dominant direction. Steering logic is used to enhance the directional information present in the Dolby surround signal. Although Dolby surround sources can be used without a Dolby decoder better results are obtained if a decoder is used. This is true even if straight surround signals are required as the surround information is subject to a modified B-type noise reduction process before encoding. To recover the correct surround information therefore the surround signal should be 'decoded' by a suitable modified B-type processor. This is of course carried out by the commercially available decoders which are licensed by Dolby Laboratories. A disadvantage of some of these decoders is that the surround signal is subject to a time delay which is unwanted in our application and although this is adjustable it may not be able to be reduced to zero. This delay is added to compensate for the close proximity of the surround speakers to the audience in most practicable domestic speaker setups.

Other multi-channel film formats such as the discrete four channel Dolby system or five or six channel digital Dolby SR-D or the THX system etc. are at present unavailable to the home user. The obvious choice of audio multi-channel format is that of the ambisonic B-format. This uses three channels for horizontal surround and an additional fourth channel for with height periphonics [3]. Unfortunately although B-format master tapes exist for some of the 2-channel ambisonic UHJ recordings these are not available to the home user.

3. STEERING SURROUND SIGNALS

3.1 Steering Two-channel Sources

If we are considering a generalised surround signal steering block then the major problem is the large number of different formats which must be catered for. In addition to this each of the formats has its own set of problems. We cannot, for example, rotate a stereo signal and produce a stereo format output as stereo is not a surround format. In Dolby surround attempts to mix between speaker feeds (L, R, C & S) to achieve rotation yields strange results as the front and surround signals are simple combinations of left and right. Rotating the sound image by 180 degrees would in any case be impossible as there is no side information at the rear due to the speaker layout used.

Steering UHJ in 2-channel format is difficult as the direction information is encoded using amplitude and phase difference, thus a rotation is not a simple function of the two signals. The same is true to an even greater extent with binaural sources. Not only must the amplitudes and phases be manipulated in a complex fashion but the generalised head related transfer function (HRTF) must be known. This makes steering these formats directly very difficult.

3.2 Steering Multi-channel Sources

Multi-channel formats fair little better as they are defined in terms of speaker feeds. This is the classic quadraphonic mistake being repeated twenty years later. The one format which is a notable exception to this is B-format. This consists of three signals (for horizontal only), one of which 'W' is a reference signal of gain 0.7071 regardless of direction, and two are vectors. 'Y' is a left-right vector defined as Signal multiplied by sin(A) and 'X' which is a front-back vector defined as Signal multiplied by cos(A). Where A is the angle of the source measured counter clockwise from the front.

Steering these signals is simplicity itself [4]. To rotate a signal 'B' degrees counter clockwise the following equations are used:

$$newW := W \quad ...(1)$$
$$newX := X.\cos(B) - Y.\sin(B) \quad ...(2)$$
$$newY := Y.\cos(B) + X.\sin(B) \quad ...(3)$$

In other words W (the reference) is unchanged and the new X and Y terms are combinations of the old terms modified by gains in the range -1 to +1. Tilt may also be added if required producing a further output 'Z' (up-down) and resulting in a full periphonic output. As it is easy to manipulate the position of sources using B-format it makes sense to use this method of steering our surround signals. This, however, creates two new problems. The first is that of converting the many surround formats into B-format and the second is converting B-format into binaural.

4. CONVERTING SURROUND SOURCES TO B-FORMAT

Fortunately there already exist methods of producing pseudo B-format signals from stereo and UHJ sources. These are not 'pure' B-format as the directional information is contained in both amplitude and phase. The steering algorithms quoted earlier, however, are unaffected by this and give the same results as they would with 'pure' B-format.

4.1 Converting Dolby Surround to B-fomat

It remains necessary to develop methods of converting Dolby surround, binaural and the multi-channel formats into B-format. Although it would be possible to decode the 2-channel Dolby surround signals directly the resultant surround channel (L-R) requires decoding by a modified B-type noise reduction decoder. It is therefore better to take the centre and surround channels from an existing Dolby-surround decoder with zero delay on the surround channel. To test the compatibility of an ambisonically processed Dolby-surround signal with that of the original a model set-up consisting of six loudspeakers was tested. The speakers were arranged in a regular hexagonal form and polar diagrams of the speaker amplitude vs sound source angle were produced. In the case of Dolby-surround the rear three speakers (LB, CB and RB) were all fed with the surround signal whilst the front three were fed with Left, Centre and Right.

The ambisonic decoder was a standard six speaker decoder with modified decoder equation to produce the B-format signals. The decoder equations used were as follows :

STEERABLE BINAURAL FROM TWO-CHANNEL SURROUND

$$W = 0.5\,(C - jS) \quad \text{...(4)}$$
$$X = 0.7\,(C + jS) \quad \text{...(5)}$$
$$Y = S \quad \text{...(6)}$$

These are similar to the 'super stereo' or 'enhanced' equations usually used but with increased difference signal. It is not suggested that these are optimal but they give reasonable results all the same.

The resulting speaker feeds for the Dolby decoder are shown in figure 1, whilst those for the ambisonically processed surround signal are shown in figure 2.

As can be seen there is a great deal of similarity between the two. The ambisonically processed signals are slightly more diffuse than the Dolby ones but there is some side information to the rear. Perhaps most importantly in both cases the rear speaker gives no output for frontal signals. It is not our purpose to suggest an alternative way of decoding speaker feeds for surround listening merely that we can go via a B-format type set of signals and end up with a very similar result. The B-format signals are of course present within an ambisonic decoder and steering of the signals may take place at this point.

4.2 Converting Dolby Pro-logic to B-format

Using a Dolby Pro-Logic decoder presents some problems as the outputs may or may not be 'steered' electronically by the decoder depending on whether there is any dominant direction present. A converter must therefore give acceptable results when presented with simple surround sound signals or (in the extreme case) only one active output. If for example there is only dialogue present the pro-logic decoder will reduce the left, right and surround output signal amplitudes considerably, leaving only the centre channel active.
A set of decoder equations which give the same results as those given earlier for surround signals but also copes with Pro-Logic signals is as follows :

$$W := 0.5\cdot\left[\left(\frac{C}{2}+\frac{L+R}{2\cdot\sqrt{2}}\right)-(j\cdot(S))\right]$$

$$X := 0.7\cdot\left(\frac{2\cdot C}{3}+\frac{L+R}{3\cdot\sqrt{2}}+0.7\cdot j\cdot S\right) \qquad \text{...(7, 8 \& 9)}$$

$$Y := 0.7\cdot(L-R)$$

The result of feeding this decoder with synthesised pro-logic signals is shown in figure 3. Despite the unusual shape of the polars these are roughly compatible with those produced by a pro-logic decoder direct. In the absense of any dominant detectable direction the polars will revert to those shown in figure 2.

4.3 Converting UHJ to B-format

The results of decoding UHJ signals are shown in figure 4. As can be seen the direction is encoded in a regular manner and apart from a slight front bias the decoding is very good.

The set of polars given in figure 5 show the effect of steering the source by 60 degrees. This was achieved by modifying the X and Y signals in the manner given in equations 2 and 3. It can be seen that the speaker feeds have just moved round one place and thus the rotation is perfect. Other rotations yield similar results including +/ - 180 degrees.

4.4 Converting Binaural to B-format

The problem format is binaural. In this, unlike UHJ, there is no definition of the Left and Right signals. These will in general be similar but will depend on the exact recording technique employed. Although side information is at high frequencies carried in the relative amplitudes of the two signals this reverts to time difference (or phase) at frequencies below about 700 Hz. In addition front-back information is carried in subtle clues caused by the head and ear shapes. It is thus very difficult to decode or translate the direction information present in binaural recordings and for this reason has not yet been considered by the authors.

5. CONVERTING B-FORMAT SIGNALS TO BINAURAL

The translation of B-format signals to binaural could potentially be as difficult as binaural to B-format but fortunately we can cheat. It would be very difficult to derive a generalised transfer function which translated B-format signals directly to left and right ear signals. The complete head related transfer function (HRTF) [5] would have to be known and filters derived from this to modify the left and right ear signal amplitudes and phases. Determining the HRTF over a full sphere experimentally would require a great deal of time in an anechoic chamber; whilst implementing the translation functions thus derived would require a powerful DSP system or massive amounts of analog electronics.

The HRTF may of course be computed [6] rather than measured but once again the implementation of such a complex transfer function is difficult. The solution is to only measure or compute the function at a few fixed angles. An ambisonic decoder driving, say, six speakers placed around the listener is capable of giving a very good illusion of direction. It can be argued therefore that if the HRFT is known for these six positions then that is all the information that is required. (In fact symmetry reduces this to four). Calculating the cumulative effect of these speakers at the ears results in right and left ear signals in terms of the six speaker feeds. Each speaker feed is defined in terms of the B-format signals W, X and Y (for horizontal only) and so an overall definition of the right and left ear signals in terms of the B-format signals may be derived very simply. This then gives us the translation of B-format to binaural. The number of speakers is of course arbitrary and a greater number may give better results. This method would of course also work for periphonics thus allowing a source to be tilted as well as rotated.

6. EXPERIMENTAL RESULTS

Initial experiments by the authors used an ambisonic decoder driving four speakers placed around a dummy head, which in turn was driving a pair of headphones. In this simple experiment the B-format signals were not steered. Switching the headphone signal from direct feed from the source to the decoded feed gave a considerable improvement, however, in the sence of spaciousness even from stereo sources. The decoder is now being implemented using DSP to replace the 'acoustic' decoder used in the initial experiment. This will eliminate any unwanted effects due to room resonances or imperfections in the loudspeakers or dummy head.

7. CONCLUSIONS

In this paper we have considered the problems of providing headphone listeners with a satisfactory image of spaciousness when listening to stereo and surround recordings. We have suggested that important auditory direction clues are lost if the natural effects of head rotation are not reproduced. The problems associated with providing this information have been outlined and a possible solution suggested. A system has been described which can take surround sound sources and provide binaural output. The system is configured such that a head tracking system may be added in order to modify the outputs and thus take account of the listeners head movements.

8. REFERENCES

[1] Griffin, M and Keating, D "Factors Affecting Sound Synthesis and Presentation Within Virtual Reality Systems", Proc. I.O.A., **14**, Part 5 (1992).

[2] Moore, B.C.J, "An Introduction to the Psychology of Hearing", 3rd Edition, Academic Press, London, 1989.

[3] Gerzon, M.A., "Practical Periphopny: The Reproduction of Full-Sphere Sound", Proc. 65th AES Convention, London, Feb 1980.

[4] Gerzon, M.A., "Panpot and Soundfield Controls", NRDC Ambisonic Technology Report No 3, Aug 1975.

[5] Wenzel, E.M., "Localisation in Virtual Acoustic Displays", Presence, **1**, No 1, 1992.

[6] Rasmussen, K.B. and Juhl, P.M., "The Effect of Head Shape on Spectral Stereo Theory", J. A.E.S., **41**, No 3, March 1993.

STEERABLE BINAURAL FROM TWO-CHANNEL SURROUND

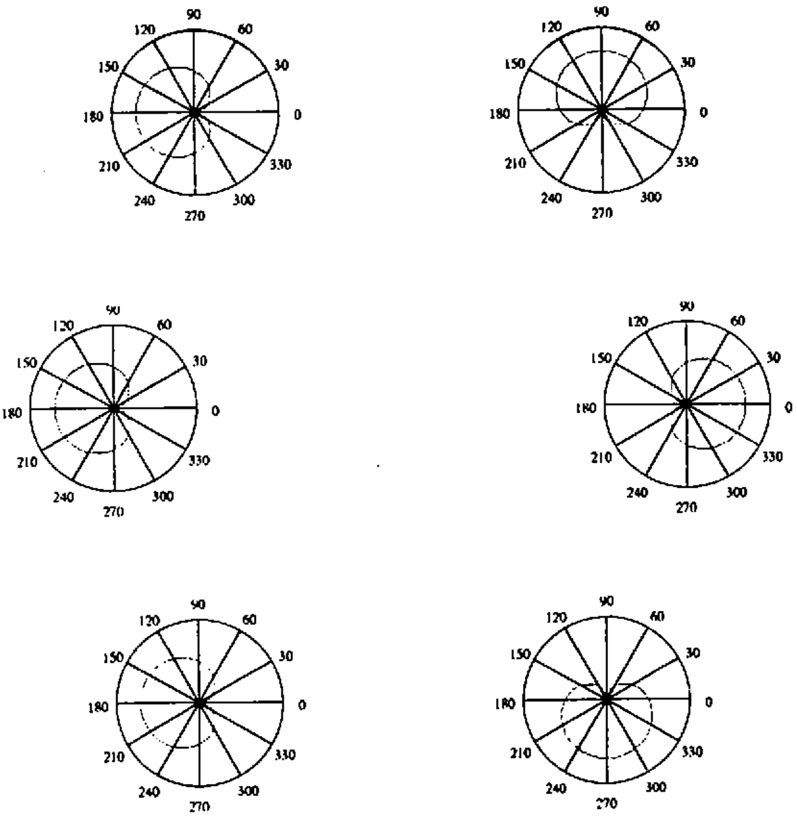

Figure 1. Speaker feed ploars for Dolby surround source, Dolby surround decode.

STEERABLE BINAURAL FROM TWO-CHANNEL SURROUND

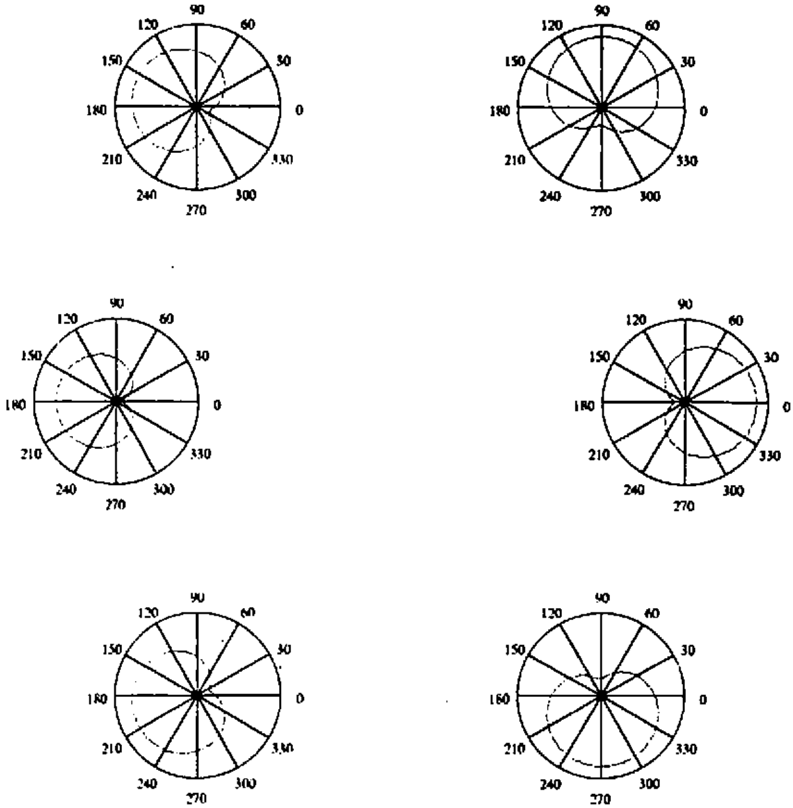

Figure 2: Speaker feed polars for Dolby surround source, ambisonic decode.

STEERABLE BINAURAL FROM TWO-CHANNEL SURROUND

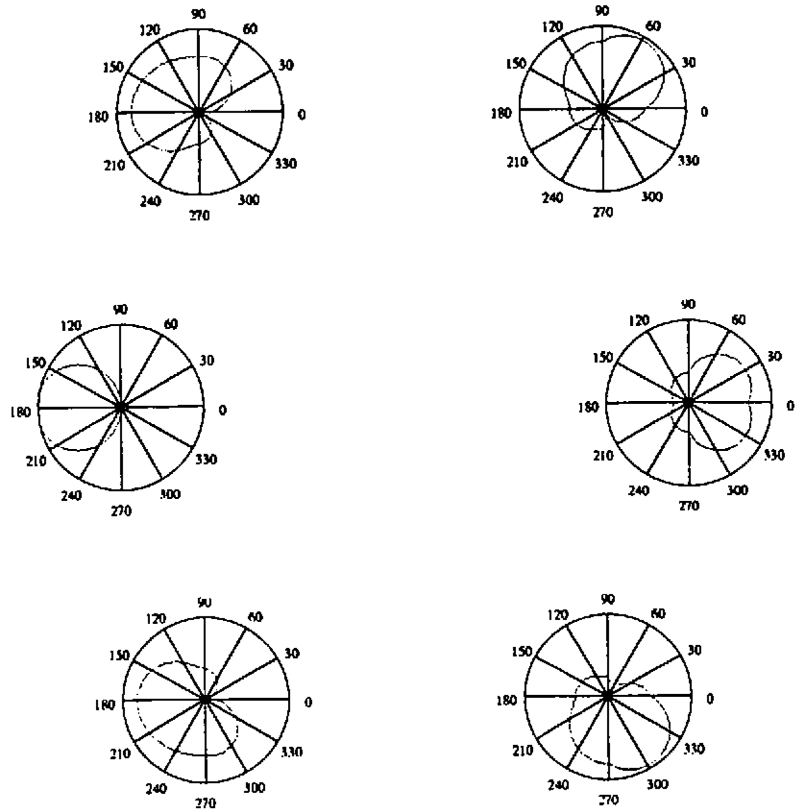

Figure 3. Speaker feed polars for Dolby pro-logic source, ambisonic decode.

STEERABLE BINAURAL FROM TWO-CHANNEL SURROUND

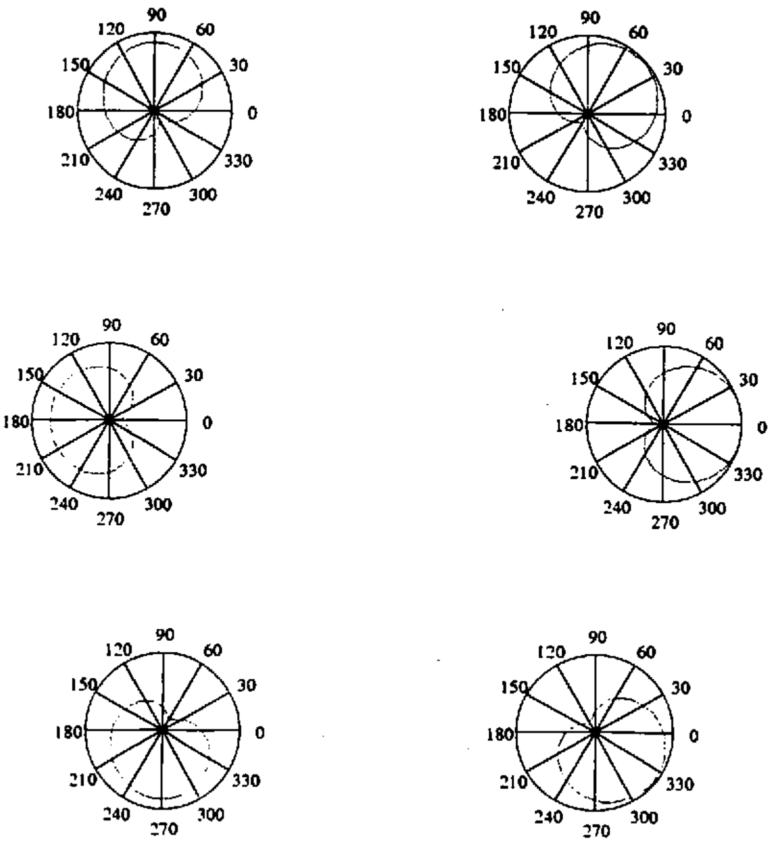

Figure 4. Speaker feed polars for UHJ source, UHJ decode.

STEERABLE BINAURAL FROM TWO-CHANNEL SURROUND

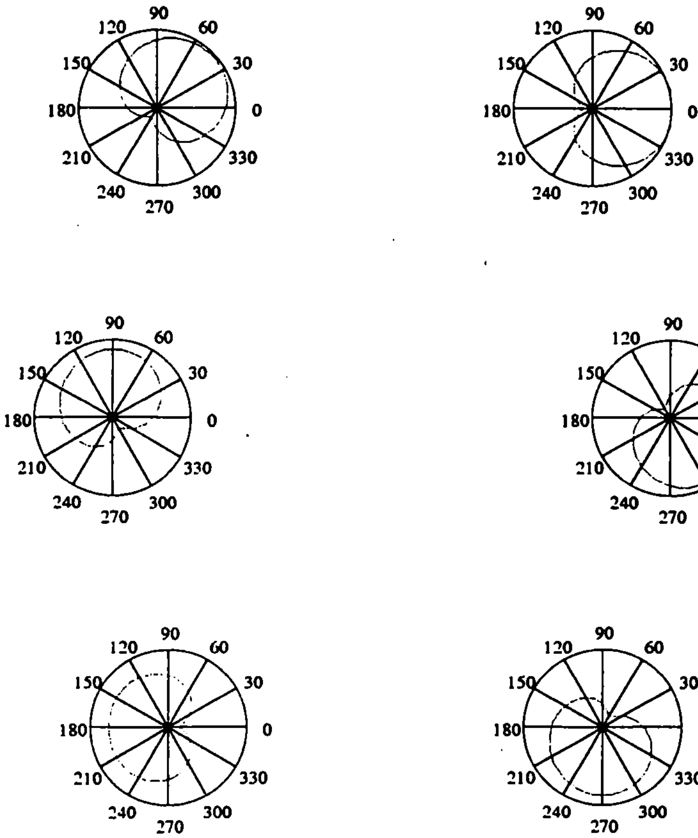

Figure 5. Speaker feed polars for UHJ source, UHJ decode rotated by 60 degrees.