DEVELOPMENT OF A SURROUND ENCODING ALGORITHM

David Carugo

Department of Computer Science and Information Systems
University of Limerick, Ireland.
david.carugo@ul.ie

ABSTRACT

The history of surround audio formats is examined, and a low cost solution to multi-channel encoding is explored and developed, taking DSP techniques into consideration . The algorithm developed will process four channels of audio and encode them so that a Dolby Surround decoder (or a Dolby ProLogic decoder) will successfully decode the signals when used as part of a home theatre / home entertainment system. The algorithm is a low cost alternative to the Dolby Encoding System.

1. INTRODUCTION

This project involved an examination of the state-of-the-art in commercial surround sound technology, comparing formats for recording, and delivery. From this it was found that there is a simple way to provide surround encoded material to the listening consumer without obtaining very costly equipment, and without licencing the encoding technology available. A process for encoding surround information (for a basic commercially used surround formats) was explored. Various methods of implementation for the algorithm were also explored, including software implementations and dedicated DSP hardware systems.

1.1. Implementation Issues.

The original idea for this project was to develop a hardware solution. This would involve a DSP integrated circuit to run the surround encoding algorithm. This plan proved to be a little ambitious, as the DSP development systems available in the University were not of a suitable type, having only two channel input capabilities rather than four, and only limited input and output resolution of 12 to 14 bits at best.The option of developing a software solution was also examined, but the lack of availability of a dedicated audio development environment proved to be too large an obstacle. Various audio software code was examined in an attempt to extract the input and output code, and to examine the format in which the data was processed, but this proved unfeasible given the time allocated to the project. A background to surround audio is given below, and the remainder of this text examines the algorithm in more detail..

1.2. The New Direction In Audio Production.

The transition to multichannel surround formats will have an even greater impact on the listening public than the transition from mono to stereo did all those years ago. The introduction of Dolby surround formats in the cinema industry has gotten the public used to hearing sounds from all around them. The development of home theatre systems has seen surround audio technology developed for the home, and people are now using combined audio / video “home theatre” systems for all their home entertainment, both for music playback and for watching films. The recent introduction of digital television will see a huge increase in available bandwidth, and this means that there is room for extra sound delivery channels. Multi channel surround audio formats that can take advantage of the increased bandwidth will become as common as normal two channel stereo is today. The specification for DVD gives multichannel surround audio as one of the selling points of the system.

As more and more audio work for broadcast is being sub-contracted to independent production companies, it follows that there is a need for these companies (many of them small companies with limited budgets) to begin working with surround audio. The standard encoding equipment available is sometimes prohibitively expensive, and therefore not all small studios can afford to begin working in this new format. The purpose of this project has been, therefore, to find a way of encoding just four channels, (the minimum for any surround format), onto a stereo soundtrack, so that consumer decoding systems will decode the four channels properly, thus giving the studio user the ability to work in surround, but without huge expense normally involved.

1.3. A Brief History of Surround Sound.

We have the world of cinema to thank for the continuous development of multichannel audio delivery formats popularly embraced by consumers. Cinema film releases have been using surround sound since the mid 1970’s. The first commonly used format was introduced by Dolby Laboratories. Dolby have been at the forefront of surround audio technology ever since. The few competitors to their almost total market dominance have not made a massive impact, but will be mentioned.

Dolby Stereo.

Dolby Stereo has been used in cinemas since the mid 1970’s. It consists of sound playback from three or four loudspeaker channels in the cinema theatre. These are the Left and Right channels, with an extra ‘effects’ or ‘surround’ channel in the earliest incarnations, and more usually, with a centre or 'dialogue' channel. These four channels are encoded down to two recorded or transmission channels, and decoded in the cinema for playback. The delivery medium is analogue optical audio tracks on 35mm film as used for cinema screenings. Due to the way sound production for films developed, the centre channel is used to this day almost exclusively for dialogue, the left and right normally for directional cues from the on-screen sound effects and music, and the surround channel for ambience effects. This convention on the use of the channels is to give cinema goers the most even sound dispersion experience from most of the seats in an average cinema theatre.

Dolby Surround.

In the mid 1980's a consumer version of the Dolby Stereo format was developed, called Dolby Surround, and also having Left Right Centre Surround capability. Many television and video film releases were encoded in this format, although it was never really examined for music production. The four channels were again encoded to two for stereo delivery, and could be recovered with a Dolby Surround decoder. In 1987 Dolby introduced the ProLogic decoder, which decoded Dolby Surround encoded material, but with a programme dependent adaptive decoding algorithm which drastically reduced inter-channel crosstalk in adjacent channels.

Dolby Surround-encoded material has the advantage that when replayed on conventional stereo, the sonic elements of the four original channels are all reproduced, but the centre channel signal is now a phantom signal, appearing from between the left and right speakers, while the encoded surround signal has components in both left and right speakers, albeit out of phase. Therefore, when played back on conventional stereo equipment, none of the encoded signals are actually 'lost'.

The above methods and formats have all been designed and implemented on analogue media. There have been a number of developments recently which use a digital transmission system in order to get better channel separation and an increase in the number of delivery channels.

Dolby Digital (SR-D).

Dolby Digital utilises stereo surround channels, i.e. left surround and right surround, and encodes five channels into a digital data stream. This is normally data compressed to get a reasonable data throughput, as the data bandwidth is limited in the transmission channel (cinema film reels). The channel separation is much better than with the older analogue signals, with the advantage of having more directional surround information (due to lack of adjacent channel crosstalk). Another specification of this system is to have a dedicated low frequency 'effects' or 'enhancement' channel (LFE), or a sub-woofer channel. As it is band limited to reproducing frequencies below 120 Hz it is not seen as a full audio channel, and so this playback format is commonly known as 5.1 surround, with the '.1' being the low frequency channel. In the last year or so, the Dolby digital format has been upgraded to Dolby Digital EX , which includes a centre surround channel, known as the EX channel. This is placed high up on the back wall of cinema theatres, or at the back of the ceiling, and is used to reinforce the central and overhead surround image during dynamic audio panning effects in motion pictures. The adoption of this format last year by Lucasfilm for *Star Wars Episode One : The Phantom Menace* led to many cinemas installing the EX system.

DTS.

Of the few competitors to Dolby's supremacy in the field of cinema surround has been Digital Theatre Systems' proprietary surround encoding and transmission system. This DTS system, as it is known, also utilises the 5.1 format, but without printing any digital audio information on the cinema film reel. A timecode is printed on the film reel adjacent to the analogue audio tracks (which are used as backup). This timecode is used to control a CD-ROM player which has the multi-channel data-compressed digital audio on it.

SDDS.

In the last couple of years, the mighty Sony company has been developing the Sony Dynamic Digital Sound (SDDS) system for multi-channel cinema audio delivery. The SDDS format specifies eight channels for playback; left, left-centre, centre, right-centre, right, surround left, surround right, and subwoofer. Thus the system has five loudspeakers (plus a subwoofer) at the front of the soundstage. The SDDS format uses a variant of the Sony ATRAC (Adaptive Transform Acoustic Coding) data compression system, which is a lossy, psychoacoustic masking based data compression system used by Sony for their MiniDisc format[1].

2. THE ENCODING ALGORITHM

2.1. Description of Encoding / Decoding Process.

In order to examine the encoder, we must first see how the decoding is carried out. The four channels, Left (L), Right (R), Centre (C) and Surround (S), are encoded down to two channels , Left Total (Lt) and Right Total (Rt), in an operation known as 'matrixing'. When decoding, the L signal and R signal are just the incoming Lt and Rt signals from the programme material. The Centre channel signal is derived from the Lt and Rt signals, and can be obtained by buffering and summing the siganls. The Surround signal is also derived from the Lt and Rt signals, but in this case is taken as the differential signal between them, that is L–R.

For encoding , therefore, the L and R channels pass on to the Lt and Rt unchanged. The Centre signal is reduced by 3dB and fed to both Lt and Rt equally. The Suround channel signal must be sent to the Lt and Rt channels in a differential form. This means that the two components of the surround signal in the Lt and Rt outputs will be 180 degrees out of phase. This can be done by applying a +90 degree phase shift to the surround signal as it goes to the Lt output, and a -90 degree phase shift as it goes to the Rt output. Thus the two components of the surround signal are recoverable by the decoder, using a differential amplifier stage. As the decoded surround signal is derived from the difference between Rt and Lt, there will be no centre channel crosstalk in it, as the centre channel appears equally in Lt and Rt. There will be no surround crosstalk in the centre channel either , as the Centre channel consists of Rt + Lt, and the two out of phase surround components will cancel.

There will be some L and R channel components in the reproduced Centre channel , however, but this will only serve to narrow the stereo spread, and can be compensated for while mixing. Any sound panned fully left will be reproduced in the

Left channel, and also 3dB down in the Centre channel. The same occurs for sounds panned Right. There will also be the two components of the surround signal present in the L and R reproduced channels, but as these are out of phase, there will be some cancellation of the surround image from the front. The surround signal will be louder in the surround channel, and this crosstalk is generally not much of an issue.

Sounds panned right and left will have components of their signals recovered by the differential stage. However, the decoder generally delays the reproduced Surround channel by approximately 20 milliseconds to take advantage of the Precedence Effect (also known as the Haas Effect), by which the brain normally only takes directional cues for a sound in the first 15 to 20 ms after receiving the first components of the sound[2]. The brain is fooled into disregarding the slightly delayed components of the L and R signals in the Surround channel, and so the sound appears to come from just the Left or Right speaker.

2.2. Deviations and Omissions from the Dolby System

The Dolby Surround encoding process is patented and also makes use of a modified Dolby B-type noise reduction. Thus when manufacturing an encoding unit for Dolby Surround (and marketing the unit as such), manufacturers have to arrange licencing of the exact process from Dolby Laboratories, and this obviously accounts for some of the high prices for the encoder units. However, in order to have a Dolby Surround *compatible* encoder, the exact Dolby process does not have to be used and the unit does not have to be approved by them. By using simple time delays, phase shifts and signal panning (which are common features of design available to any engineer) , the Dolby encoding process can be quite effectively approximated. Some deviations from the Dolby encoding process are outlined below.

The Dolby Surround encoding specification specifies that the surround channel is also encoded with a modified Dolby B-Type noise reduction. This is a variable frequency, programme dependent compansion process. The encoded surround channel should also be band limited from 100 Hz to 7kHz. The Dolby decoder will also apply the band limiting to the surround channel on reproduction, so it is not entirely necessary to band limit during the encoding process. It is generally not possible to disable the B-Type noise reduction that is applied by the decoder, and this noise reduction has the subjective effect of slightly 'dulling' the signal, i.e. lowering the level of high frequency content. Thus by applying a high frequency boost of a few dB during encoding, the decoder's noise reduction is compensated for. Obviously the effect is not as good as with the proper modified Type B noise reduction, but the results are acceptable.

2.3. The Phase Shift Problem in Detail

The surround channel should ideally be processed so that the components of the signal in the two encoded Left Total (Lt) and Right Total (Rt) channels are at +90 degrees and -90 degrees phase shifts from the original surround signal. This is because of the need for dynamic panning effects, for example causing a sound to sweep over the heads of the audience, front to rear. If no plus and minus 90 degree phase shift is applied, then the Surround channel would have to be panned fully left, and then phase inverted and panned fully right. Then the following would occur during a front to rear sweep:

When the signal is in the Centre channel, it consists of L + R. When the signal is in the Surround (i.e. rear) channel, it consists of L - R. For a moving sweep from front to rear, the level must be reduced from full to zero level over time in the Centre channel, and increased from zero level to full level over the same amount of time in the Surround channel. However , the problem occurs when the level of the Centre channel is the same as the level of the Surround channel. The '+R' component in the Centre channel and the '-R' component in the Surround channel will cancel each other out completely, leaving the signal only in the Left reproduction channel. Thus the sound will appear to sweep from front to the left and then to the rear, which is not the desired effect.

In order to alleviate this effect of leftwards sweeping, the plus and minus phase shifts are added in order to de-correlate the -R component of the panned sound in the Surround channel from the +R component of the panned sound in the Centre channel. Therefore the component of the sound in the +R part of the Centre channel is not the exact inverse of the sound in the -R part of the Surround channel, and thus the sound is not cancelled in the Right reproduction channel when the levels of the Centre and Surround channels are the same during the front to rear panning. The actual way to implement these phase shifts was examined in some detail, however, the solution used was to delay the surround signal before sending it and its inverse to the Left Total and Right Total channels respectively.

If the Surround signal is delayed by a time interval equal to 90 degrees of the wavelength of a frequency which is in the centre of the bandwidth occupied by the surround signal, then the phase shift is 90 degrees at that frequency, and deviates from 90 degrees as the frequency gets higher and lower than the centre frequency. At this point the signal can be approximated as having been phase shifted by 90 degrees. The delayed signal can be lowered by 3 dB, and sent to the Left Total channel, then the signal can be inverted and sent to the Right Total channel. In this way, there is an approximate phase shift of plus and minus 90 degrees, and the signal is encoded differentially onto the two transmission channels Left Total and Right Total. The time delay causes the surround channel's -R component to be de-correlated from the centre channel's +R component, in a similar fashion to the way the pure plus and minus 90 degree phase shifts de-correlate these signal components. 800Hz seems to be a good mid-frequency point on the 100Hz to 7kHz bandwidth (taking into account the logarithmic frequency law). At this frequency, a full wavelength phase shift is achieved by a time delay of 1/800 seconds. This is equal to 1.25 milliseconds. Thus a 1/4 wavelength phase shift is equal to 0.3125ms. The lowest frequency in the reproduced surround channel is 100Hz. This has a period of 10 ms, and thus a 1/4 wavelength period of 2.5 ms. The time delay value chosen should then be a multiple of the value for 1/4 wavelength of 800Hz , that is it should be a multiple of 0.3125ms. It should also be greater than 2.5ms , in order to effectively de-correlate the surround signal effectively at the lower frequencies. A time delay of 7.8 ms was found to be most effective during algorithm testing. This is the equivalent to a delay of 6.25 wavelengths at 800Hz.

2.4. Implementation Issues

The system was originally to have been a stand-alone hardware unit, with onboard power supply, four balanced analogue inputs, and two balanced analog outputs. There was to be a DSP integrated circuit in the unit run the surround encoding algorithm.

The hardware system would be divided into three sections; input, DSP, and output. The input and output sections could be combined onto one circuit board, especially if the connectors chosen were PCB mounted, as well as attached to the rear panel of the unit. The DSP section includes the A/D converters and D/A converters, and these would have to have 16 bit resoluotion at least. The DSP would be programmable in C/C++ and would have some sort of on-board emulation for the development of the system to take place on a host computer.

Each input to the unit would have to be unity gain buffered, possibly with a differential op-amp as the inputs would be balanced signals. From each input, the signal would be sent to the A/D converter. The data stream from the converter would be passed to the DSP integated circuit, processed (i.e. encoded), and sent to the D/A converters. The signal after D/A conversion would be sent to the outputs of the unit via differential line driver op-amps, to keep the analogue signal balanced when it exits the unit. The unit would have a dual rail power supply for the op-amps and the DSP unit, of a normal commercially available type.

2.5. Algorithm Testing

It was decided to test the algorithm using existing hardware, in the form of a general purpose digital mixer: a Yahama model 01/V. The testing of the algorithm involved routing the signals under test internally in the mixer, processing them to be in the same format that a dedicated encoder would. In normal use, four of the bus outputs from the mixer would be used to send the four signals (L,C,R,S) into a hardware encoder. For test purposes, a mono signal was used, and the test involved panning the signal between the four reproduced channels. The signals were monitored through a Shure model HTS 5000 Dolby Surround decoder (kindly supplied by GFD Communications Ltd, Dublin) , thus ensuring that any ‘encoding’ processes could be confirmed to be working correctly.

The routings and connections used for the test were as follows : The Left and Right outputs of the mixer (acting as Left Total and Right Total) were connected to the Lt and Rt inputs of the Surround Decoder. The Left and Right outputs of the decoder were amplified and connected to the left and right monitor loudspeakers, placed in the same position that normal stereo monitor loudspeakers would occupy. The Centre output from the decoder was amplified and sent to the center speaker, and the Surround output from the decoder was amplified and sent to the two surround loudspeakers (connected in parallel). The centre speaker was placed directly between the left and right speakers, and at the same radial distance from the listening position. The L, C, and R speakers used were all JBL Control 1 models. The surround speakers used were Sony SS X2A models, and these were placed 2m behind the listening position, at a height of 1.5m, and at a seperation of 1.5m. These loudspeakers were placed facing the listening position (from the rear of the listener).

The test was performed in the live room in the University Recording Studio. The room is quite large, but is partially acoustically treated, and placing the monitoring system off-centre gave a reasonable simulation of a typical room used for home theatre systems.

The signal under test was input into channel 1 of the mixer, and routed to main busses 1,2,3, and 4 only, and not to the stereo bus. The bus outputs were fed to channels 2, 3 and 4. Channel 2 was used for the front image, as it was panned centrally, and channels 3 and 4 supplied the surround encoding as follows: channel 3 was delayed by 7.8ms (to de-correlate the signal from that in channel 1) and panned fully left. Channel 4 was also delayed by 7.8ms but was panned fully right and phase inverted. Thus channels three and four were 180 degrees out of phase, and appeared at the left and right output channels. When the test signal was fed to those channels only, it appeared in the decoded surround channel only. When the test was in channel 2 only, it appeared only in the centre decoded channel. Panning this channel to the left and right made the decoded signal pan left and right also. There noticable crosstalk in the decoded surround channel. The front to rear panning effect was then tested. The signal was panned centrally in channel 2 and the channel fader brought up in level. The sound then appeared in the centre decoded channel. The dynamic pan was performed by raising the channel 3 and 4 faders (increasing the surround level) while simultaneously bringing down the channel 2 fader (decreasing the centre level). The timing of the two fades must be carefully controlled to achieve the effect. The front to rear panning was successful in that the sound does not appear to travel from the front to the left and then to the rear, but rather seems to travel around the listening position on both sides. The surround speakers were not very high behind the listening position. The purpose was not to create the ‘flying overhead’ effect, but to alleviate the ‘leftward sweep’ that arises when the surround channel encoding does not de-correlate the left surround component and the centre channel.

3. CONCLUSIONS

The original proposal for this project outlined a finished unit that could be used in the Recording Studio in the University. This was unfeasible given the time frame for the project, and the unit is still in the development stages. Block diagrams and basic circuit ideas / diagrams for the unit have been developed, and the algorithm itself has been tested, and shown to be perfectly adequate for the purposes of basic surround mixing. I am still of the opinion that such a unit could be constructed for sale for much less than the asking price for a very basic Dolby Surround Encoder (i.e. an encoder manufactured and sold by Dolby Laboratories). The previously detailed dynamic panning effect, where the sound is panned from front to rear in a continuous motion, also works with the basic encoding method.

4. REFERENCES

[1] Yamamuchi, H., Saito, E., and M. Kohut., “The SDDS system for digitising film sound.” In: Vijay K. Madisetti., and Douglas B Williams, eds. *The digital signal processing handbook.* CRC Press / IEEE Press, Florida, 1998 .

[2] Benade, Arthur H., *Fundamentals of musical acoustics.* 2nd ed. New York: Dover Publications. 204-207, 1990.