HAS AMBISONICS COME OF AGE?

Dr B Wiggins University of Derby, UK

1 INTRODUCTION

Ambisonics was developed in the 1970's as a flexible, psycho-acoustically aware system[1]. Developed at the same time as Quadraphonics[2], Ambisonics is an often mis-understood system that was far ahead of it's time. Due to the ubiquity of surround sound equipment in modern computers and interest in live surround events becoming more widespread, is the time, finally, right for Ambisonics to come into its' own?

In this paper, the definition of what makes a system Ambisonic is clarified with reference made to the traditional energy and velocity vector theory, higher order systems and use in both the live and domestic environment. More recent developments by the author are discussed with respect to irregular Ambisonic decoder design (such as for the ITU 5.1 speaker array) and analysis using Head Related Transfer Function data showing the extra insight this can give into the performance of one, seemingly similar, decoder design over another. The freely available suite of VST plug-ins (comprising of decoders, panners and an Ambisonic reverb) created using this technology are also presented, with case studies of their use in student projects at the University of Derby.

2 AMBISONICS

2.1 Introduction

Ambisonics is often described as a system based on the spherical harmonic decomposition of a soundfield[3]. This description, although true, is not necessarily the most accessible 'way in' to Ambisonics, with a more useful description of Ambisonics found in the 1992 Gerzon & Barton paper[4] and their equivalent U.S. patent regarding Ambisonic decoders for irregular arrays[5], and states (slightly adapted to remove equations):

A decoder or reproduction system is defined to be Ambisonic if, for a centrally seated listening position, it is designed such that:

- The decoded velocity and energy vector angles agree and are substantially unchanged with frequency.
- At low frequencies (below around 400 Hz) the low frequency velocity vector magnitude is equal to 1 for all reproduced azimuths.
- At mid/high frequencies (between around 700 Hz and 4 kHz) the energy vector magnitude is substantially maximised across as large a part of the 360^0 sound stage as possible.

In order to easily meet the demands of the system made above, Ambisonics separate's the encoding (or recording/panning) of the audio material from the decoding (or reproduction) side of the system. This happens in the exactly the same way in which it was first suggested by Blumlein in his famous 1931 patent[6] which documented a system similar in many ways to Ambisonics (a much simplified version of this is the stereo used today).

In the same way that Blumlein used the equivalent of two, coincident, figure of eight microphone patterns which could be mixed together to create a 'virtual' figure of eight response pointing in any direction in 2D space (as shown below in Figure 1), Gerzon's Ambisonics was based on the four, coincident, microphone patterns as shown in Figure 2 (collectively known as B-Format).

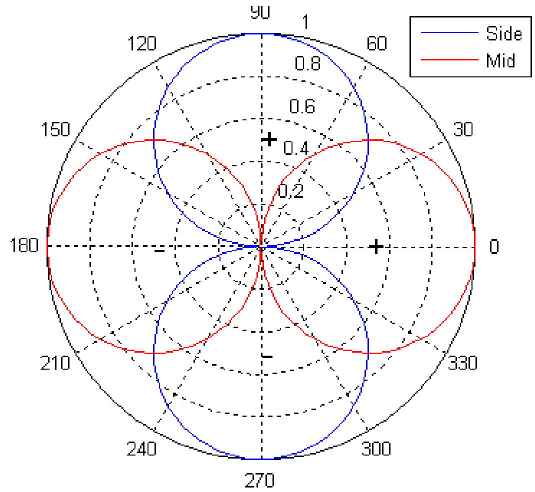

Figure 1 – Blumlein Mid and Side

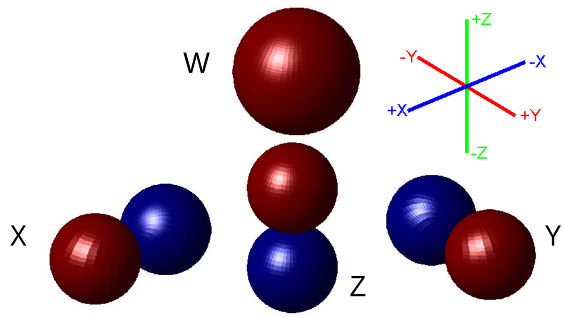

Figure 2 – Ambisonic B-Format Patterns

Using a linear combination of any of the B-format signals, a 'virtual' microphone pattern pointing in any direction in 3D space using any, 1st order, polar pattern could be achieved, and it was the deciding of what signals needed to be fed to any particular speaker array that was at the heart of Ambisonics. A simple 2D example of this virtual pattern extraction is shown in Figure 3 and Equation 1

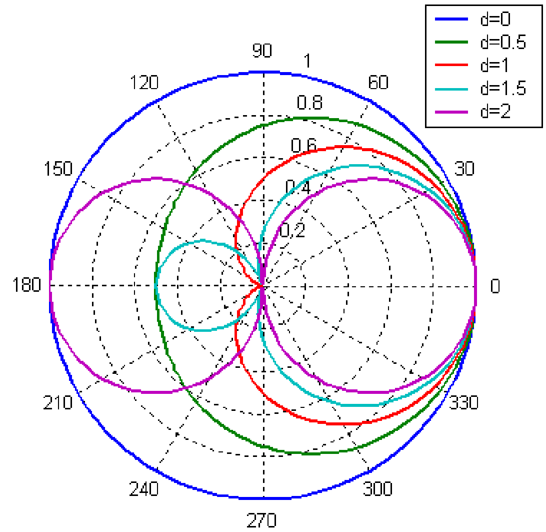

Figure 3 – Summing of Omni and Figure of Eight Coincident microphone signals.

$$g_w = \sqrt{2}$$
$$g_x = \cos(\theta)$$
$$g_y = \sin(\theta)$$
$$S = 0.5 \times \left[(2-d) g_w W + d\left(g_x X + g_y Y\right)\right]$$
$$\text{where}: \mathrm{g_w}, g_y \text{ and } \mathrm{g_z} \text{ are the B-Format mic signals}$$
$$\text{S is the resulting virtual mic or speaker signal.}$$

Equation 1 – Equation used to construct a virtual microphone pattern pointing in direction θ, with a directivity of 'd' from the W, X and Y signals of B-format.

2.2 Velocity and Energy Vector Analysis

As mentioned in section 2.1, the 'art of Ambisonics' is actually in the deciding of which microphone patterns pointing in which direction need to be fed to which speakers. In order to simplify this task, Gerzon developed a metatheory which chose two main estimation methods for the performance for a surround sound system[1].

At low frequencies (around less than 400Hz), 1st order Ambisonics has been shown to be, essentially, a 'volume solution' (at least, for living room size listening) and in this case, the velocity vector can be used as a measure of performance[8].

The velocity vector is essentially giving us the signals that need to be fed to the speakers in order to recreate what would have been present in the first place. That is, if you optimise a system using the velocity vector and place your B-format microphone in the middle of the speaker array, you should record exactly the signals that you present to the system.

The velocity vector is derived using the equation shown in Equation 2 and is calculated for single or multiple source positions.

$$P = \sum_{i=1}^{n} g_i$$
$$Vx = \sum_{i=0}^{n} g_i \cos(\theta_i) / P$$
$$Vy = \sum_{i=0}^{n} g_i \sin(\theta_i) / P$$

Where:
g_i represents the gain of the i^{th} speaker (assumed real for simplicity).
n is the number of speakers.
θ_i is the angular position of the i^{th} speaker.

Equation 2 – Velocity Vector

Over around 700Hz, the sweet spot for a 1st order system approaches the size of the human head, and so a better volume solution in this case is to make sure as much energy as possible is coming from the 'correct' direction. In order to do this, the energy vector analysis can be used as shown in Equation 3. A length of one here would indicate that all the sound is coming from one direction, which is only possible if just one speaker is sounding, and so this value is generally (always in Ambisonics!) less than one.

$$E = \sum_{i=1}^{n} g_i^2$$

$$Ex = \sum_{i=0}^{n} g_i^2 \cos(\theta_i)/E$$

$$Ey = \sum_{i=0}^{n} g_i^2 \sin(\theta_i)/E$$

Where:
g_i represents the gain of the i^{th} speaker (assumed real for simplicity).
n is the number of speakers.
θ_i is the angular position of the i^{th} speaker.
Equation 3 – Energy Vector

As an example, if we fed a regular, eight speaker array with signals equivalent to pointing eight coincident cardioid microphones pointing in the direction of those speakers, then the velocity and energy vector analysis shown in Figure 4 would be observed. This cardioid decoder is equivalent to using a directional 'd' parameter of 1 in Equation 1.

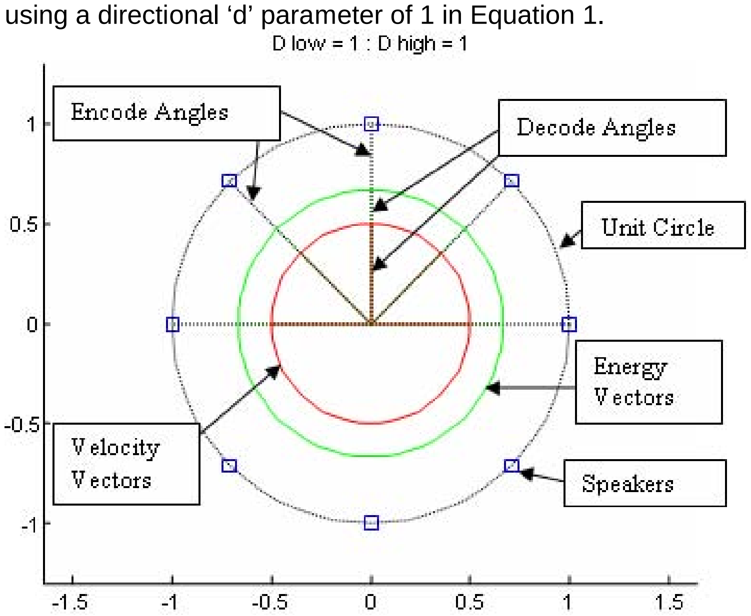

Figure 4 – Velocity and Energy vector analysis of a simple cardioid decoder.

A number of things can be observed from this diagram:

- Neither the velocity vector or energy vector length are 1.
- The decoded angles (vector directions) will always match up with the encoded (or recorded) source directions, as shown by the five 'decode angles' labelled in Figure 4. This will always be true when the speakers are arranged as regular polygons (square, hexagon, octagon etc.).

Using 1st order microphone patterns, the best results, as shown in Figure 5, can be achieved using the microphone patterns shown in Figure 6 (these polar patterns are shown pointing back to back for clarity. The decoder is made up of each of these patterns pointing in the direction of the speakers at low and high frequencies respectively).

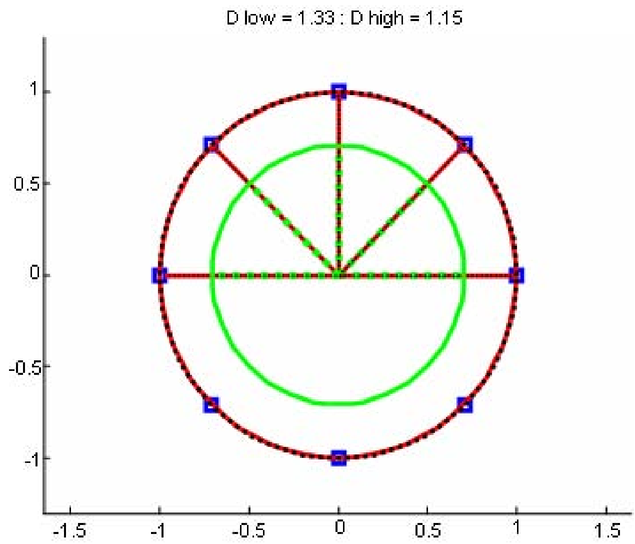

Figure 5 – Analysis of an Ambisonic decoder for an octagon.

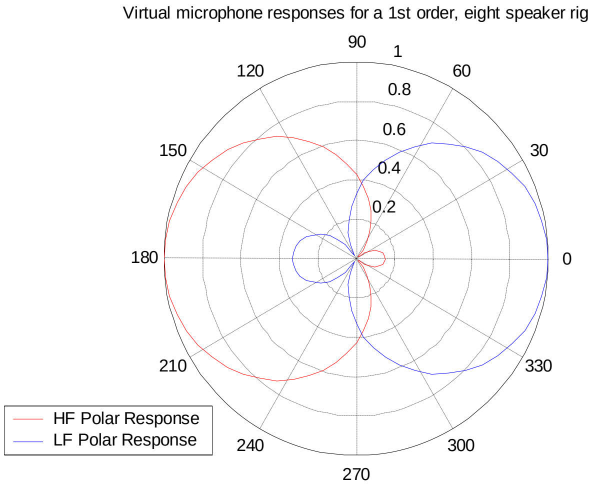

Figure 6 – Low and high frequency virtual microphone polar patterns used to achieve the velocity and energy vector analysis shown in Figure 5.

It is worth noting that the velocity/energy vector analysis shows all directions to be treated equally, which is one of the redeeming features of an Ambisonic system (the vector lengths are constant). Most other (especially artificially panned) systems will try and get the localisation of a source at a speaker as correct as possible. This is generally done by *only* using that speaker to generate the audio, and results in less error when a source is panned to a speaker location compared to when a source is panned between speakers. This leads to the 'speaker detent' effect, and causes the speakers to be audible sources in the sound field. Ambisonics, on the other hand, makes a source at a speaker position, just as wrong as a source between speaker positions, which generally means that the speakers 'disappear' in the mix, leading to a more natural sounding system.

2.2.1 Speaker Detent Effect

The equal error benefit described above will hold true as long as there are more speakers used in the decode than number of channels. For example, if we are using 1st order Ambisonics the W, X and Y signals can be used to generate feeds for a horizontal only system. Figure 7 shows a velocity and energy vector analysis of both a four speaker and a three speaker decoder. The three speaker decode analysis shows that the energy vector as a source approaches a speaker improves, and then worsens as the source moves to between speakers. The apparent reproduced angles are also affected.

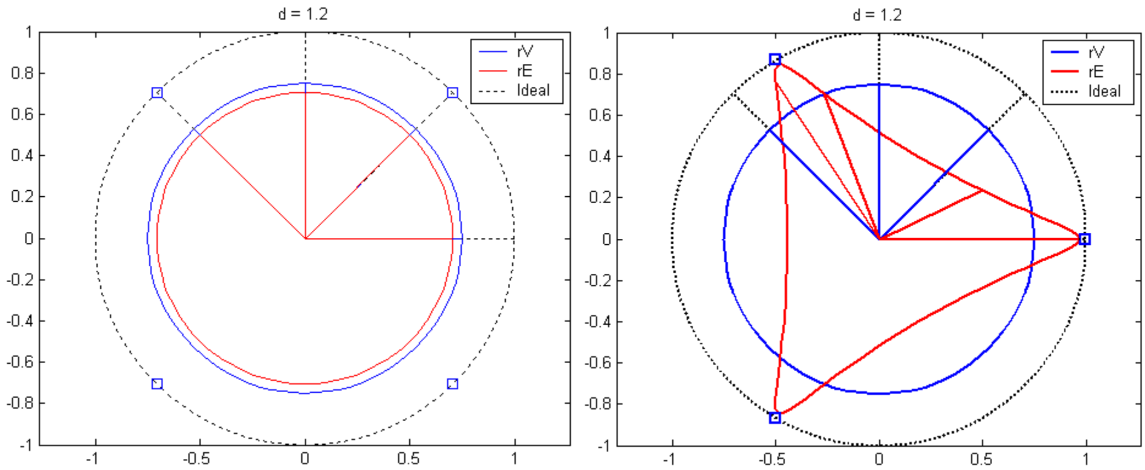

Figure 7 – Velocity and Energy vector analysis of a four and three speaker 1st order decode.

2.3 Irregular Decoder Design

Creating decoders for regular polygons is well documented, although some alternative decoder rationale's have deviated from the strict Ambisonic principles[9] and there are a number of other issues that need to be taken into consideration that won't be described in this paper (notably, near-field compensation, or speaker distance compensation which can have a slight effect at very low frequencies)[12]. However, irregular decoder design (for example, a decoder derived for the ITU 5.1 speaker standard[10] as shown in Figure 8) is far more involved, requiring the solving of a number of non-linear, simultaneous equations[3 & 4].

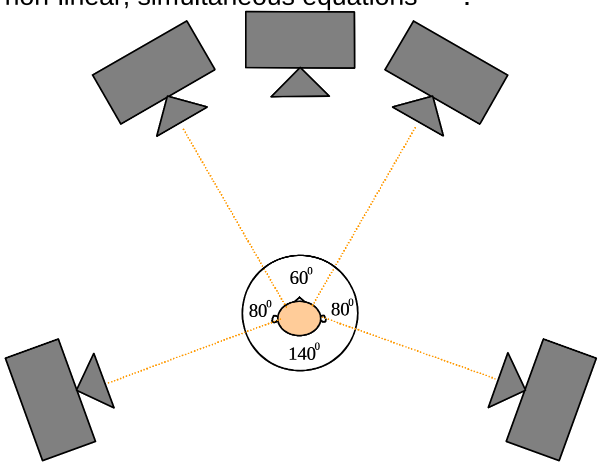

Figure 8 – 5 speaker array as specified by the ITU

The problem can be clarified if we look at a decoder based on pointing cardioid polar patterns in the same directions as the speakers in an ITU array as shown in Figure 9.

The decoder shown in Figure 9 exhibits following problems:

1. The velocity and energy vector lengths are sub-optimal.
2. The reproduced angles do not match up with the encoded source angles, i.e., the sounds will not appear to come from the intended directions.
3. The reproduced angles between the velocity and energy vectors do not agree.

4. Although not shown in this visualisation, the reproduced volume is not constant as a source is panned around the sound stage unless some compensation is applied to take into account the front/back asymmetry of the speaker array.

In regular polygon decoders, only point 1 is an issue, which can be rectified by altering just the polar pattern of the signal fed to each speaker. However, in order to rectify the other points we need to alter the virtual microphone feeds direction and loudness as well.

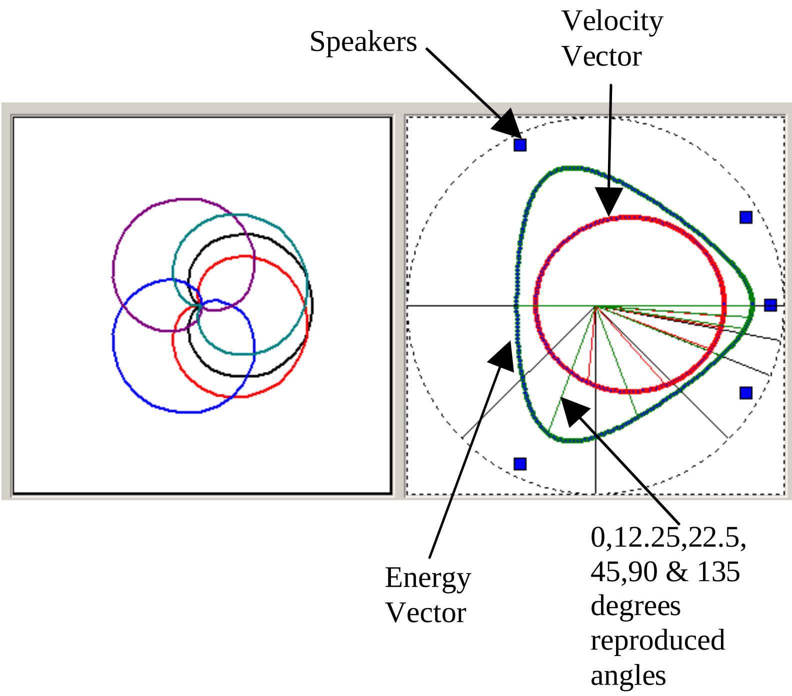

Figure 9 – Velocity and Energy vector of a naive ITU decoder.

For a 1st order system fed with a B-format input, we end up with the unknowns in the system as shown in Equation 4.

$$C_F = (kW_C \times W) + (kX_C \times X)$$
$$L_F = (kW_F \times W) + (kX_F \times X) + (kY_F \times Y)$$
$$R_F = (kW_F \times W) + (kX_F \times X) - (kY_F \times Y)$$
$$L_B = (kW_B \times W) + (kX_B \times X) + (kY_B \times Y)$$
$$R_B = (kW_B \times W) + (kX_B \times X) - (kY_B \times Y)$$

Where: k denotes a decoding coefficient (e.g. kWc represents the weighting given to the W channel for centre front speaker).

- Subscript F, B and C denote front, back and centre speakers respectively.
- W,X and Y represent the incoming 0th and 1st order circular harmonic signals.
- C, L and R denote centre, left and right speakers

Equation 4 – Decoding equations for an irregular 5 speaker array.

Solving this system is laborious at best[4], and up until 2003, no method of deriving decoders for irregular speaker arrays was published[11] until the author presented a heuristic based approach using a modified tabu search algorithm for generating decoders for any arbitrary array automatically.

A decoder optimised for high frequency (energy vector) reproduction using this method is shown in Figure 10 where it can be noticed that some compromises are still present due to the non-optimal nature of the standard ITU 5 speaker array for full 360 degree surround sound audio. Also notice that, as the Ambisonic principles described are predicated on a centrally seated listener the centre front speaker is not necessary, and is actually detrimental to the decoder, as a whole, if used (the optimisation algorithm currently tries to optimise over the full 360 degree soundstage, rather than giving preference to the frontal hemisphere).

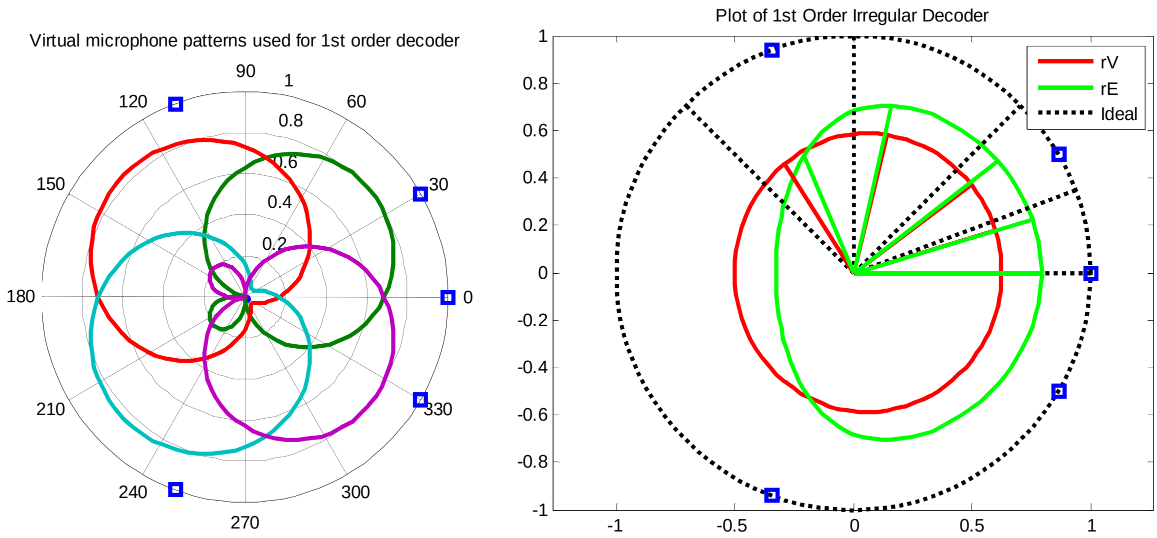

Figure 10 – Optimised 1st Order Ambisonic decoder for the ITU 5 speaker array.

2.4 Higher Order Ambisonics

So far, Ambisonic decoders have been discussed using 1st order input signals (i.e. a mixture of pressure and pressure gradient coincident mics or signals which represent spherical harmonic components of degree 0 and 1). However, the use of higher degree microphone patterns can be used to improve both regular and irregular Ambisonic decoders.

The Furse-Malham set of 2nd and 3rd order equations[7] are often used in sonic applications, although it should be noted that there are a number of amplitude normalisation standards available[12 & 13]. A graphical representation of the 0th, 1st and 2nd degree spherical harmonics normalised using the Furse-Malham coefficients are shown in Figure 11.

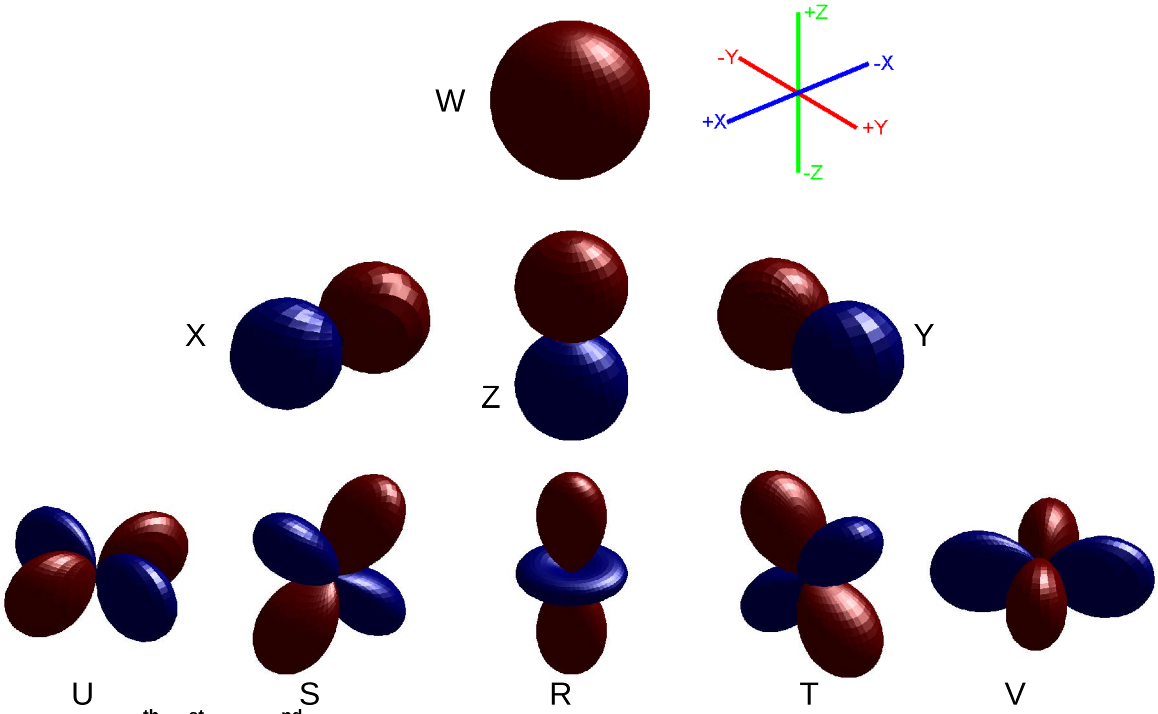

Figure 11 – 0th, 1st and 2nd degree spherical harmonics

The use of higher degree spherical harmonics allows for tighter virtual mic polar patterns to be constructed (examples shown in Figure 12). These can be used to improve the performance of the Ambisonically decoded material in a number ways:

- The tighter mic patterns allow for the useful frequency range of the velocity decode to increase. i.e. the decoder works as a volume solution for more of the frequency range[8&12]
- An energy vector optimised decode can achieve a value closer to 1.
- However, more speakers must be used in order to avoid the 'speaker detent' effect as described in section 2.2.1.

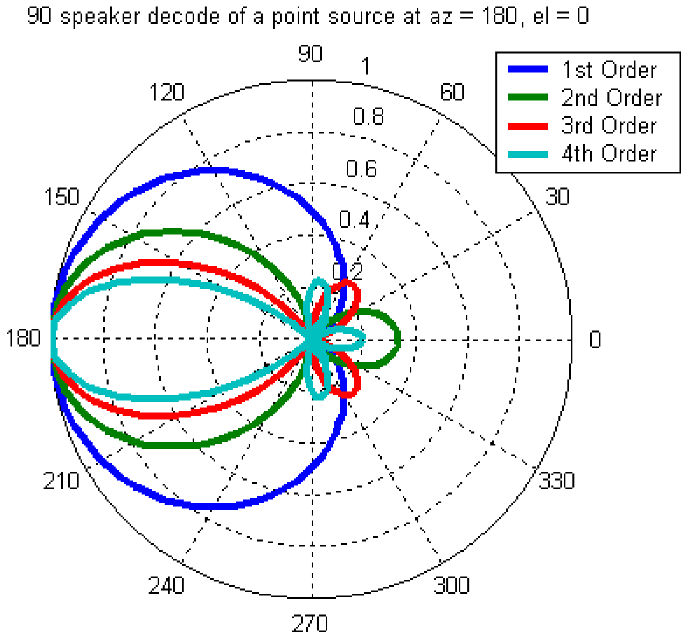

Figure 12 – Example 1st, 2nd, 3rd and 4th order Virtual Microphone patterns.

One, less obvious, advantage of including higher degree spherical harmonic components is that by using spherical harmonics greater than degree 1, irregular (asymmetrical) virtual microphone patterns can be derived. This greatly aids in the creation of decoders or panners for use with irregular speaker arrays such as the ITU standard 5 speaker layout[14&3]. For example, decoders designed using 2nd and 4th order Ambisonic principles are shown below in Figure 13.

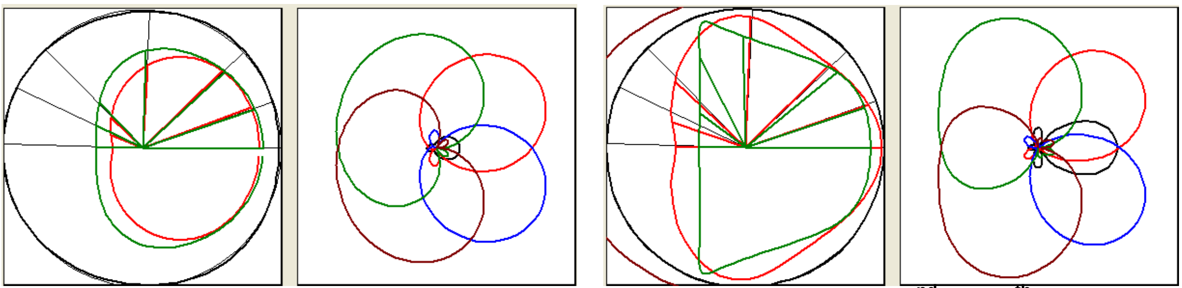

Figure 13 – Energy (green)), Velocity (red) and virtual mic patterns of a 2nd and 4th order decoder for the ITU 5 speaker arrangement shown in Figure 8.

2.5 HRTF Analysis

Velocity and energy vector analysis have been used as the basis for Ambisonics since its inception, but sound fields can be analysed using HRTF data in order to try and gain some extra insight into the performance of, seemingly, similar decoders. This is particularly of use when deriving decoders for irregular speaker arrays as there are multiple solutions and compromises that can be made in order to generate decoding coefficients[3].

Using HRTF data, a speaker array and Ambisonic decode can be simulated, and the response of the system to Ambisonically panned sound sources analysed.

One simple form of analysis is to use the group delay and amplitude differences between the 'ears' of the HRTF data, and compare this to the inter-aural delay and amplitude differences for a real source (the HRTF data for that position). The HRTF analysis for the 1st order Ambisonic decoder introduced in Figure 10 is shown in Figure 14. In this figure, 'G Format' is the Ambisonically panned and decoded source, and 'Real Source' is taken from a single pair of HRTFs. It can be seen that the amplitude differences available at the ears of a centrally seated listener are a very good match when compared to a real source, but the low frequency time (measured using group delay) plot doesn't match as well. This decoder was optimised for its high frequency (energy vector) performance.

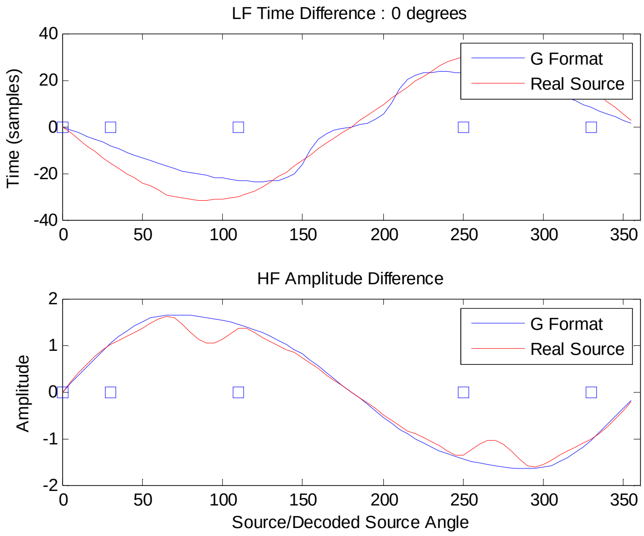

Figure 14 – Low frequency time and mid frequency amplitude differences from HRTF data due to a real and Ambisonically panned and decoded sound source.

In order to try and gain information on the robustness of the decoder under test, the head turning of a listener can be simulated. Figure 15 shows such an analysis on a 1st order and a 4th order ITU decoder. If the system is perfect, then the blue line should still match the red line as the head is rotated. If the source/decoded angle of 0 degrees is compared, marked differences between the two decoders are apparent. As the head has rotated, the amplitude and time differences at the ears of a listener for a sound source at 0 degrees should show the values that were previously observed for a source or -45 degrees (315 degrees in Figure 14). This isn't the case and, interestingly, the matching generally gets worse as the listener turns, but some decoders perform better than others in this respect. It can be seen in the right hand graph of Figure 15 that a sound source at 0 degree is still exhibiting a time and amplitude difference of around 0. This suggests that as the head is turned, a sound source at 0 degrees will, in fact, track with that listener, giving unstable imagery. This problem is not exhibited by the 1st order decoder shown on the left of Figure 15.

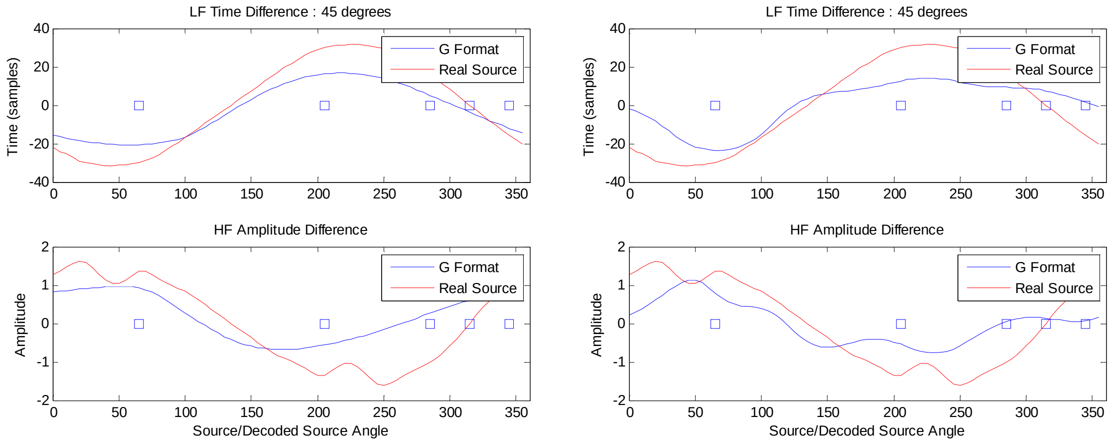

Figure 15 – HRTF analysis of a 1st order ITU decoder (left) and a 3rd order ITU decoder (right) when the listener is rotated 45 degrees from facing front.

3 WIGWARE AMBISONIC TOOLS

Plug-ins and guides describing how to mix and decode Ambisonics have been around for many years (for example see the article by Richard Elen[15] and the VST Plug-ins by Dave Malham[16]). However, working in Ambisonics has always been problematic, mainly due to the various host software packages available being too prescriptive in terms of audio channels and there being a lack of Ambisonic aware effects.

I wanted my students to be able to mix, present and distribute their work using Ambisonic principles and so started developing VST plug-ins to fill in, what I saw, as the major gaps in the Ambisonic toolset. Missing at this point was the ability to decode, Ambisonically, to the ITU 5 speaker array and there were no Ambisonic reverb effects available. I had already introduced a method that could be used to derive Ambisonic ITU decoders, and so a VST plug-in that publishes 7 such decoders for the user to test and use was produced. A 1st order Ambisonic reverb based on the excellent, open source, freeverb[17] has also been created and made available. A list of the freely available tools are given below (which can be downloaded from my University website[18]), with sample screenshots shown in Figure 16. Currently, the available plug-ins are:

- 1st, 2nd and 3rd order panners/encoders
    - Using polar co-ordinates
    - Using Cartesian co-ordinates
- 1st and 2nd order decoders.
    - For regular speaker arrays (cube, square, octagon etc.)
    - For the ITU irregular speaker array
- Ambisonic Reverb
    - 1st Order (four channel) reverb, which can also be used with higher order decoders.

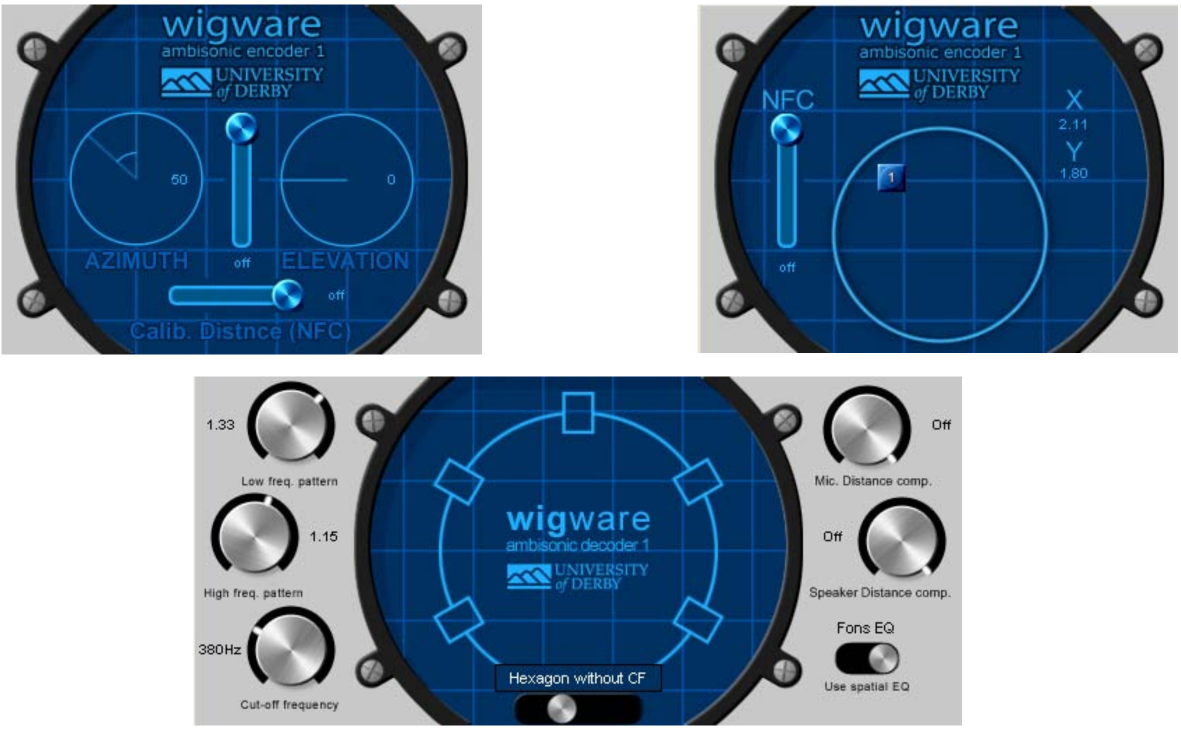

Figure 16 – Screenshots of the Wigware Ambisonic VST Plug-ins

On its own, this development, although useful for modular, flexible audio programs such as Audiomulch and Plogue Bidule, did not fulfil the goal of allowing others to mix, present and distribute their work using Ambisonics. However, at the same time, three other major developments have now helped to bring about this possibility:

- The development of the Ambisonia website[19], which adopted the standard proposed by Richard Dobson for an Ambisonic wave file format[20].
- The creation of DirectShow filters (plug-ins) by myself to allow people to listen to Ambisonic Wave Files using Windows Media Player and other directx aware media players[18].
- The Reaper audio sequencing and creation software[21].

Ambisonia and the .AMB wave file format, combined with the free DirectShow decoders has provided an avenue for creators of Ambisonic content to distribute their material and a way for people to actually listen to it. Everything on the website is currently free to download (via bit-torrent), with contributors using it as a place to show-case their work. As far as I am aware, this is the only Ambisonic library currently available.

“REAPER is a fully featured multitrack audio and MIDI recording, editing, processing, mixing, and mastering environment.”[21]. Using Reaper’s inbuilt functionality, it is stereo only, but its routing options are so powerful and flexible that, finally, a host for Ambisonic plug-ins has arrived! Ironically, many of the currently available commercial DAWs (such as Cubase, Logic and Pro-tools, for example) although capable of surround sound mixing and playback, are often too prescriptive in terms of the number of channels per audio track, and what plug-ins can be inserted onto surround tracks/busses. For example, Logic 8 has a limit of 8 channels per track, which means that only 1st order ‘with-height’ Ambisonics can be realised. Reapers current channels per track limit is 64.

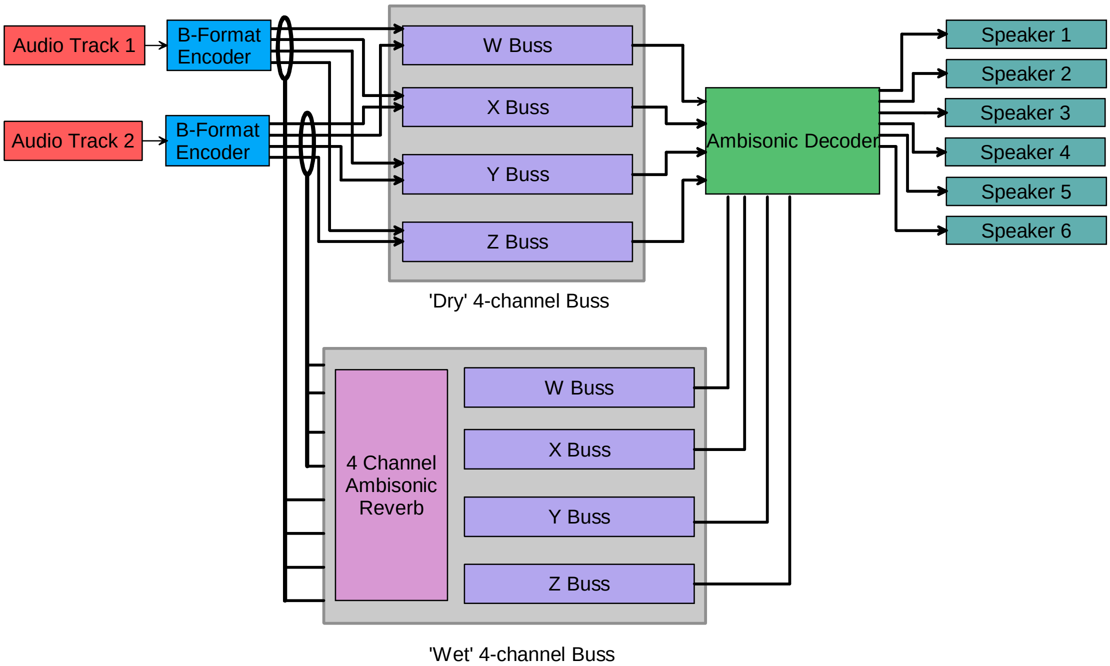

Figure 17 – Example channel diagram for a 1st order Ambisonic mix.

A typical track/channel/buss layout that is easily realised in Reaper is shown in Figure 17. Two tracks of audio are shown, along with 4-channel dry and reverb busses, which are then routed to a 4-in 8-out Ambisonic decoder. Reaper allows any track to host any plug-in regardless of the number of channels available in the track with the channels being able to be assigned in any order. A simple 4-in, 8-out decoder plug-in with Reaper channel routing being shown in Figure 18.

Figure 18 – Example VST Plug-in Routing in Reaper

Using this configuration, mixing in Ambisonics is now relatively straight forward, and we have already integrated the possibility of creating Ambisonic mixes into one of our 2nd year undergraduate modules, Computer Music Production, giving them an introduction and experience of some of the more advanced topics they will be studying in their final year. A number of screencasts were made documenting how to go about using the Wigware plug-ins in Reaper, and these can be found at my Wiki site constructed to help document our Multi-channel Research Lab at the University of Derby[22]. These screencasts have also allowed staff and students from our Popular Music & Music Technology and Multi-Media Technology & Music Production degrees to create and showcase their work at the recent Space-net funded Spatial Sound and Music International

Workshop[23] where a larger scale 3D 15 speaker array was driven with 2^{nd} order Ambisonic compositions and material.

4 CONCLUSIONS

It has taken many years, but technology (software as much as hardware) has finally caught up with the functionality and flexibility that Michael Gerzon envisaged in his Ambisonic system. True, future-proof surround sound audio can be created, distributed and auditioned easily using a combination of Reaper, Wigware Ambsionic plug-ins, the Ambisonia website and Richard Dobson's .AMB file format and associated command line tools with tutorial screencasts documenting the process on-line.

Ambisonics has *finally* come of age!

5 REFERENCES

1. Gerzon, M. A. (1974a) Sound Reproduction Systems. Patent No. 1494751.
2. Gerzon, M. A. (1974b) What's wrong with Quadraphonics. Retrieved: September, 2008 from http://www.audiosignal.co.uk/What's%20wrong%20with%20quadraphonics.html
3. Wiggins, B., (2007) The Generation of Panning Laws for Irregular Speaker Arrays Using Heuristic Methods. *Proceedings of the 31st International AES conference,* London, UK
4. Gerzon, M. A. & Barton, G. J. (1992) Ambisonic Decoders for HDTV. Proceedings of the 92nd International AES Convention, Vienna. 24 – 27 March. Preprint 3345.
5. Gerzon, M.A, Barton, G.J. (1998) Surround Sound Apparatus. U.S. Patent No. 5,757,927
6. A D Blumlein. UK Patent 394325 (December 14, 1931).
7. Malham, D. (2002) Second and Third Order Ambisonics. Retrieved: September, 2008, from http://www.york.ac.uk/inst/mustech/3d_audio/secondor.html.
8. Bamford, J.S. (1995) An Analysis of Ambisonic Sound Systems of First and Second Order, Master of Science thesis, University of Waterloo, Ontario, Canada.
9. Furse, R. (N.D.) First and Second Order Ambisonic Decoding Equations. Retrieved: September, 2008 from http://www.muse.demon.co.uk/ref/speakers.html
10. ITU-R BS. 775 (1994) Multichannel Stereophonic Sound System With and Without Accompanying Picture
11. Wiggins, B., Paterson-Stephens, I., Lowndes, V., Berry, S. (2003) The Design and Optimisation of Surround Sound Decoders Using Heuristic Methods. Proceedings of UKSim 2003, Conference of the UK Simulation Society p.106-114
12. Daniel, Jérôme, (2001) Représentation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimedia. Thèse de doctorat de l'Université Paris 6, 31 juillet 2001. Retrieved September, 2008 from http://gyronymo.free.fr/audio3D/downloads/These.pdf
13. Green, Robin. 2003, *Spherical Harmonic Lighting: The Gritty Details*. Retrieved: September, 2008 from http://www.research.scea.com/gdc2003/spherical-harmonic-lighting.pdf
14. Craven, P. (2003), Continuous Surround Panning for 5-speaker Reproduction, AES 24th International Conference, Banff, Canada.
15. Elen, R. (1983), *Ambisonic Mixing – An Introduction*. Retrieved: October, 2008 from http://www.ambisonic.net/ambimix.html.
16. Malham, D. (2002), *Swiss Center for Computer Music VST Ambisonic Tools*. Retrieved: October, 2008 from http://www.dmalham.freeserve.co.uk/vst_ambisonics.html.
17. Freeverb. Retrieved: October, 2008 from http://www.dsprelated.com/dspbooks/pasp/Freeverb.html
18. University of Derby (2008) *Bruce Wiggins Staff Page*,. Retrieved: October, 2008 from http://www.derby.ac.uk/staff-search/dr-bruce-wiggins
19. Deleflie, E. (N.D.) *Ambisonic Surround Sound. Ambisonics, 5.1, audio recordings*. Retrieved: October, 2008 from http://www.ambisonia.com/.
20. Dobson, R. (1999) *WAVE-EX and B FORMAT*. Retrieved: October, 2008 from http://people.bath.ac.uk/masrwd/bformat.html.

21. Cockos Inc. (2005) *Reaper* **|** *About*. Retrieved: October, 2008 from http://reaper.fm/.
22. Wiggins, B. (2008) *Using Ambisonics in Reaper*. Retrieved: October, 2008 from http://uod-true-multi-channel-mixing.wikispaces.com/Reaper+Tutorial+Videos.
23. *Spatial Audio Creative Engineering Network (Space-net)*. Retrieved: October, 2008 from http://space-net.org.uk/.