

THE GENERATION OF PANNING LAWS FOR IRREGULAR SPEAKER ARRAYS USING HEURISTIC METHODS.

BRUCE WIGGINS¹

¹ Signal Processing Applications Research Group (SPARG), University of Derby, Derby, UK
b.j.wiggins@derby.ac.uk

Currently, the ITU standard surround sound speaker arrangement is based on an irregular 5 speaker array. However, this may change to an irregular 7 speaker array (as is now the standard on computer hardware) or more in the future. The Ambisonic system, pioneered by Micheal Gerzon, among others, in the late 1960's, is very well suited to situations where the end system speaker configuration is not fixed in terms of number or position while also offering a simple way (via energy and velocity vector analysis) of quantifying the performance of such systems. However, while the derivation of the decoders is well documented for regular speaker arrangements [1], optimising the decoders for irregular layouts is not a simple task, where optimisation requires the solution of a set of non linear simultaneous equations, complicated further by the fact that multiple solutions are possible [2]. Craven [3] extended the system to use higher order circular harmonics and presented a 4th order Ambisonic decoder (9 input channels), although the derivation method used was not presented.

In this paper a semi-automated decoder optimisation system using heuristic methods will be presented that will be shown to be robust enough to generate higher order Ambisonic decoders based on the energy and velocity vector parameters. This method is then analytically compared to Craven's decoder using both energy/velocity vector and head related transfer function based methods.

INTRODUCTION

The standard speaker configuration, as specified by the ITU, is a five-speaker layout, as shown in Figure 1. However, this is likely to be expanded upon in the near future, and other, larger, venues are likely to have more speakers in order to adequately cover a larger listening area.

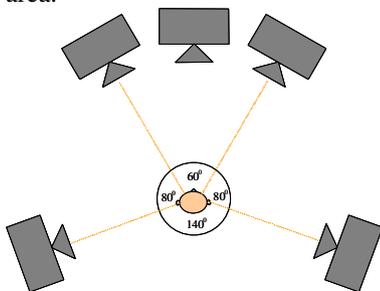


Figure 1 - 5 speaker array as specified by the ITU

This arrangement is, however, likely to change due to the standard in the home computer area being a 7 speaker array. Traditionally, pieces need then be remixed from scratch to transfer the material to the new array, however, a more flexible approach could be used in the creation of multi-channel material, and such a system has been available since the 1960s [4].

1 AMBISONICS

Ambisonic systems are based on a spherical decomposition of the sound field to a set order [5][6]. The main benefit of the Ambisonic system is that it is a hierarchical system, that is, once the sound field is encoded in this way (into four channels for 1st order, and 9 channels for 2nd order) it is the decoder that decides how this sound field is reconstructed using the Ambisonic decoding equations [7]. Essentially, the encoding of the system is carried out by recording, or synthesising microphone polar patterns of each order. In this paper, only the horizontal case is tackled (Ambisonics can also encode and decode height information) which simplifies the number of channels needed to encode sound from all directions to an order, N , to be $2N + 1$. A 0th order microphone is omnidirectional with 1st to 4th order polar patterns shown in Figure 2.

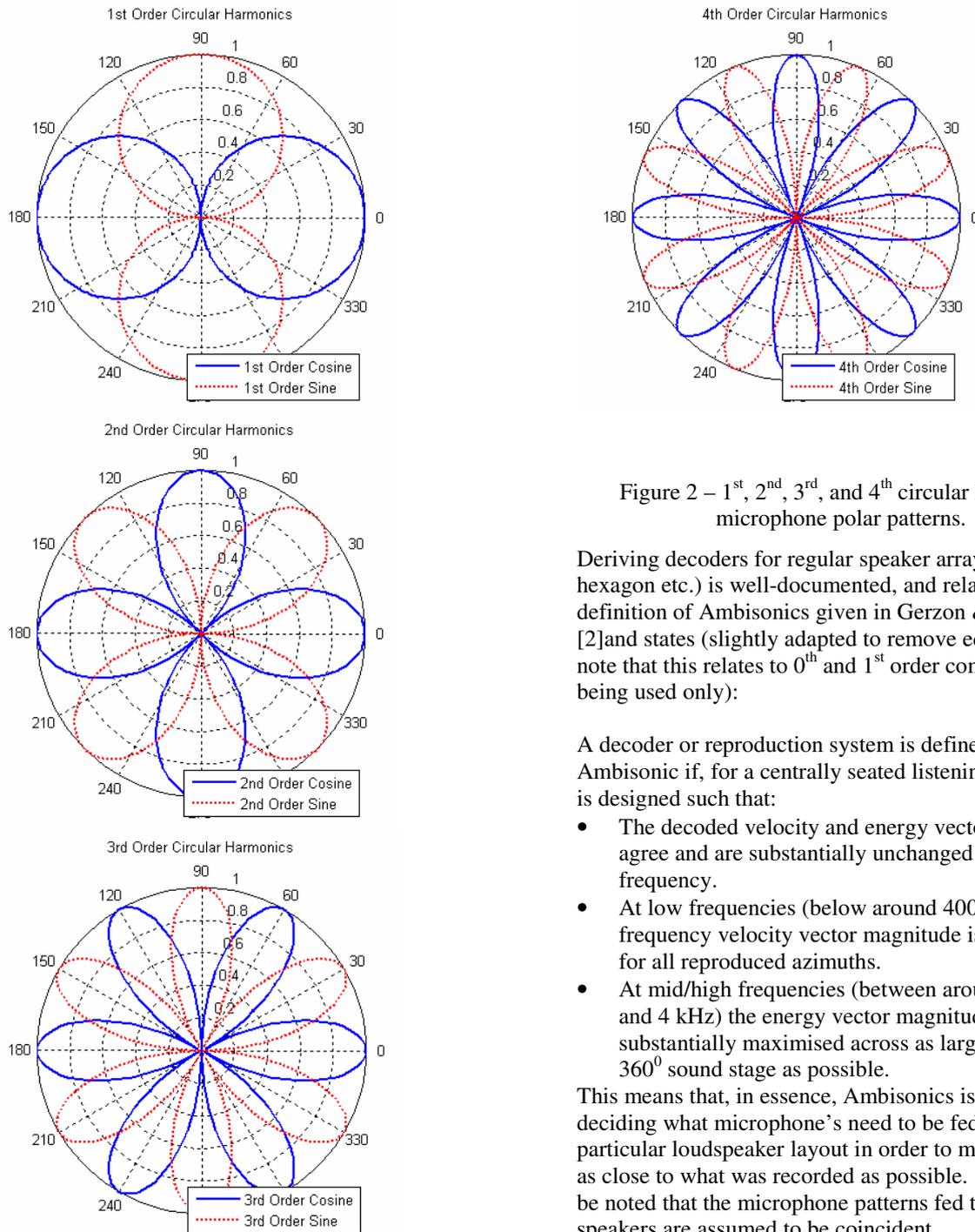


Figure 2 – 1st, 2nd, 3rd, and 4th circular harmonic microphone polar patterns.

Deriving decoders for regular speaker arrays (square, hexagon etc.) is well-documented, and relates to the definition of Ambisonics given in Gerzon & Barton [2] and states (slightly adapted to remove equations, and note that this relates to 0th and 1st order components being used only):

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^o sound stage as possible.

This means that, in essence, Ambisonics is a method of deciding what microphone's need to be fed to a particular loudspeaker layout in order to make its output as close to what was recorded as possible. It must also be noted that the microphone patterns fed to the speakers are assumed to be coincident.

2 1ST ORDER REGULAR DECODER DESIGN

Carrying out this optimisation for regular decoders is well documented [1] and relies on the use of the velocity and energetic analysis of the system from the centre point defined by the equations shown in Equ. (1)

$$\begin{aligned}
 P &= \sum_{i=1}^n g_i & E &= \sum_{i=1}^n g_i^2 \\
 V_x &= \sum_{i=0}^n g_i \cos(\theta_i)/P & E_x &= \sum_{i=0}^n g_i^2 \cos(\theta_i)/E \\
 V_y &= \sum_{i=0}^n g_i \sin(\theta_i)/P & E_y &= \sum_{i=0}^n g_i^2 \sin(\theta_i)/E
 \end{aligned}
 \tag{1}$$

Where: g_i represents the gain of a speaker (assumed real for simplicity).

n is the number of speakers.

θ_i is the angular position of the i^{th} speaker.

P is the pressure due to the speakers output

E is the energy due to the speakers output

E_x and E_y are the energy vector

V_x and V_y are the velocity vector

Taking the 0th and 1st order circular harmonic equations (i.e. the recorded/encoded source material) to be as shown in Equ. (2), a 1st order virtual microphone signal can be decoded by using the equation shown in Equ. (3).

$$\begin{aligned}
 W &= 1/\sqrt{2} \\
 X &= \cos(\theta) \\
 Y &= \sin(\theta)
 \end{aligned}
 \tag{2}$$

Where: θ is the source angle

W is the 0th order component (omni)

X is the 1st order cosine component (front/back figure of eight)

Y is the 1st order sine component (left/right figure of eight)

$$\begin{aligned}
 g_w &= \sqrt{2} \\
 g_x &= \cos(\theta) \\
 g_y &= \sin(\theta) \\
 S &= 0.5 \times [(2-d)g_w W + d(g_x X + g_y Y)]
 \end{aligned}
 \tag{3}$$

where: W, X & Y are the signals given in Equ. (2).

S = speaker output

θ = speaker azimuth

d = directivity factor (0 to 2)

The energy and velocity analysis can then be used as a measure to alter the polar pattern of these speaker feeds to optimise the decoder in line with the Ambisonic definitions given in section 1 above. For a regular speaker arrangement, this simply resulted in the design

of a decoder where virtual microphone patterns were derived from a combination 0th and 1st components pointing in the direction of the speakers but with a varying polar pattern between low (<500Hz) and high (>700Hz) frequencies achieved by altering the ‘ d ’ parameter in the decoding equation shown in Equ. (3). The effect of the ‘ d ’ parameter is shown in Figure 3.

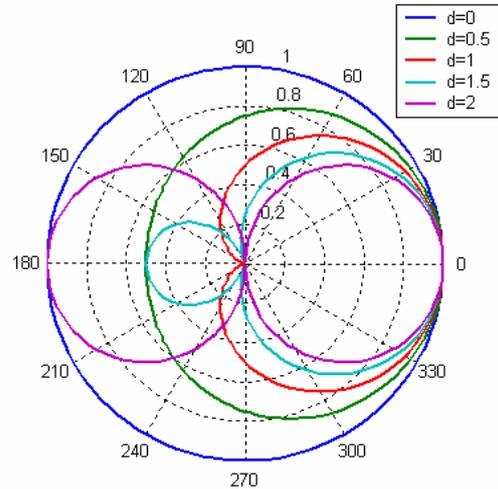


Figure 3 – Ranges of virtual polar pattern obtained to feed a speaker using Equ. (3).

For example, if a cardioid polar pattern is used, then the resulting energy and velocity vector analysis of the decoder is shown in Figure 4. Notice that the encoded and decoded angles match, and that the velocity and energy vector lengths are less than 1 (i.e. sub-optimal).

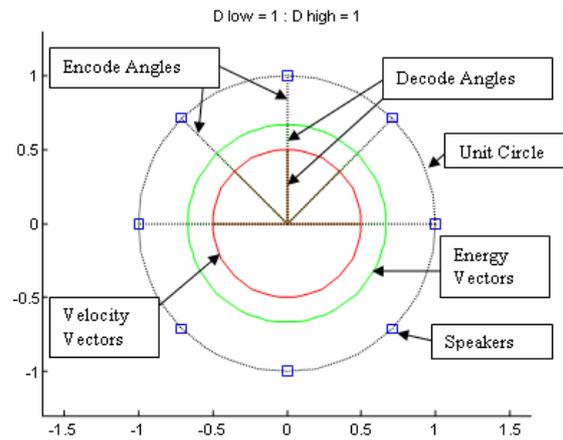


Figure 4 – Energy and Velocity Vector Analysis of a decoder feeding an 8 speaker octagonal array.

However, if the polar patterns of the decoders output are changed to those shown in Figure 5 (the orientation of the polar patterns are only different to increase the plots clarity), then the energy and velocity vector analysis of the decoder is shown in Figure 6.

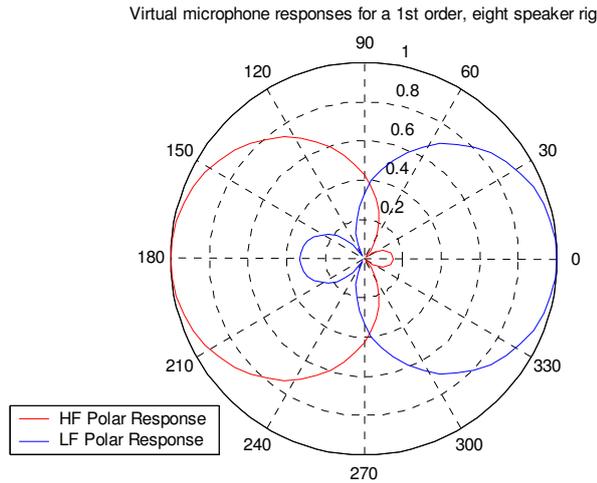


Figure 5 – Virtual microphone patterns used for an ‘optimised’ regular Ambisonic decoder.

Figure 6 shows a velocity vector length of 1 (indication of performance at low frequencies) and an energy vector length that is at its maximum (it can never be 1, unless only one speaker emits the sound source). Increasing the order changes the maximum value of the energy vector length from around 0.7 for a 1st order system to around 0.8 for a 2nd order system, for example.

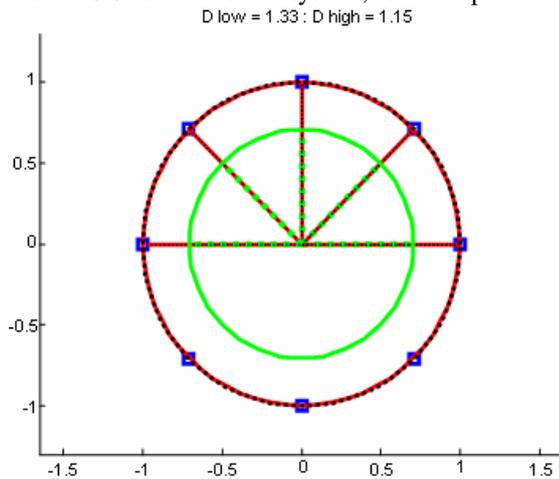


Figure 6 – Energy and velocity vector analysis of an ‘optimised’ regular Ambisonic decoder.

3 IRREGULAR DECODER DESIGN

Optimising a decoder for an irregular speaker array (where the angular spacing is not constant) is a non-trivial task. In order to help visualise the problem, an ITU 5 speaker array is fed with cardioid microphone patterns pointing at 0 degrees, +/-45 degrees and +/-135 degrees (the default settings of a SoundField SP451 decoding unit) with its velocity and energy vector analysis shown in Figure 7.

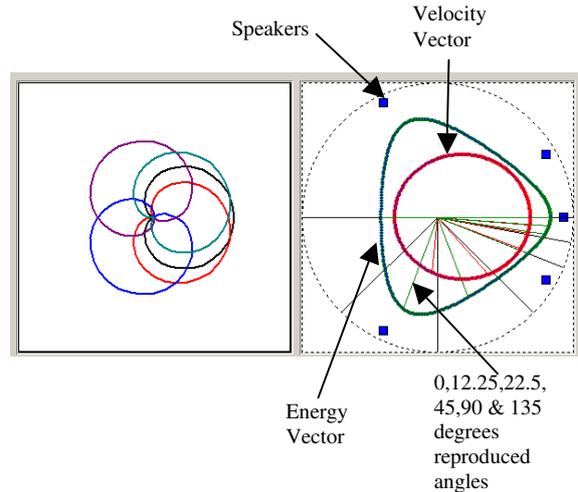


Figure 7 - Energy and velocity vector response of an ITU 5-speaker system, using virtual cardioids.

This figure shows the more complex relationship between the chosen microphone polar patterns used to feed the speaker array and the resulting velocity and energy vector magnitude response. However, in this example, angular distortion is also present, and (although not shown in this figure), it must also be noted that the overall level of the decode is not constant around the unit circle (i.e. sources encoded towards the front are louder due to the greater concentration of speakers).

Up to 1st order, this problem was originally tackled by Gerzon [2] in his 1992 AES Vienna paper (1st order decoders of this type are often referred to as ‘Vienna’ decoders for this reason). Gerzon did not go deeply into his techniques for arriving at the chosen coefficients although he did mention that they were “*very tedious and messy*” [2], and as the ITU standard was not in existence at the time, he tackled five speaker arrangements that he thought could be adopted although none as irregular as the arrangement finally decided upon.

For a 1st order system, the decoding coefficients that need to be provided are shown in Equ. (4).

$$\begin{aligned}
 C_F &= (kW_C \times W) + (kX_C \times X) \\
 L_F &= (kW_F \times W) + (kX_F \times X) + (kY_F \times Y) \\
 R_F &= (kW_F \times W) + (kX_F \times X) - (kY_F \times Y) \\
 L_B &= (kW_B \times W) + (kX_B \times X) + (kY_B \times Y) \\
 R_B &= (kW_B \times W) + (kX_B \times X) - (kY_B \times Y)
 \end{aligned}
 \tag{4}$$

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

- 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

Due to the ITU speaker arrangement being symmetrical about the X axis (taking X to be front/back), then three sets of coefficients need to be obtained, with speaker pairs taken as identical except for a phase inverted Y coefficients (similar to mid/side pairs):

- Centre front speaker W, X and Y coefficients.
- Front left and right W, X and Y coefficients.
- Back left and right W, X and Y coefficients.

From the velocity and energy vectors given in Equation Equ. (1), the angle and magnitude can be found using the equations given in Equ. (5)

$$\begin{aligned}
 M_E &= \sqrt{E_x^2 + E_y^2} & \theta_E &= \tan^{-1}(E_y/E_x) \\
 M_V &= \sqrt{V_x^2 + V_y^2} & \theta_V &= \tan^{-1}(V_y/V_x) \\
 P_V &= \sum_{i=1}^n g_i & P_E &= \sum_{i=1}^n g_i^2
 \end{aligned}
 \tag{5}$$

Where: Magnitude (M) and Reproduced Angle (θ) calculated from the velocity and energy vectors given in Equ. (1).

For an optimal decoder:

- The vector magnitude (M_V and M_E) should be as close to 1 as possible for all values of θ (where θ is the encoded angle).
- $\theta = \theta_V = \theta_E$ for all values of θ .
- $P_V = P_E$ and must be constant for all values of θ .

Gerzon and Barton's original method used a technique where the coefficients were derived, but where the high frequency (energy vector) components were adjusted at the end to bring the P_V and P_E values into alignment. Unfortunately, as reported in Wiggins [9] this actually caused errors to be re-introduced into the decode when comparing θ to θ_V and θ_E . This can be seen in the plot generated from decoder coefficients taken from Gerzon & Barton [2] and shown in Figure 8.

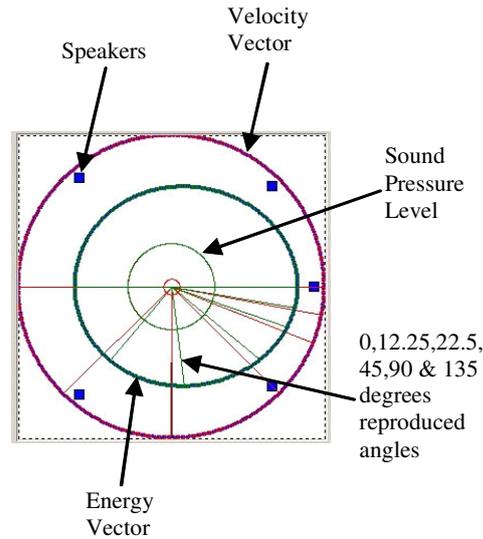


Figure 8 - Energy and velocity vector analysis of an irregular speaker decode optimised by Gerzon & Barton [2].

4 HIGHER ORDER IRREGULAR DECODERS

Craven [3] presented an improved higher order decoder, although did not reveal his derivation technique. This decoder was frequency independent (whereas Gerzon's earlier decoders were designed to have one decoder for low frequencies and another for high), but used higher order circular harmonic components to help improve the system when used as a panning law (no higher order microphones existed at the time). These higher order components are shown in Equ. (6) with each speakers output made up of a linear combination of these virtual polar patterns (multiplied by the source signal, if panning of a mono source is to be used, or multiplied by the outputs of a higher order microphone's signals, if one is available).

$$\begin{aligned}
 W &= 1/\sqrt{2} & C3 &= \cos(3\theta) \\
 X(C1) &= \cos(\theta) & S3 &= \sin(3\theta) \\
 Y(S1) &= \sin(\theta) & C4 &= \cos(4\theta) \\
 C2 &= \cos(2\theta) & S4 &= \sin(4\theta) \\
 S2 &= \sin(2\theta)
 \end{aligned}
 \tag{6}$$

Gerzon's papers have always stated that the highest order used in a decoder should always be less than the number of speakers. More accurately, the number of input channels must be less than the number of speakers (e.g. 4 speakers can only be fed up to 1st order horizontally as this is a three channel system). However, this does not hold true for irregular decoders where the higher order harmonics can essentially be

used to steer the 1st order polar patterns into an asymmetrical shape that will better match the irregular speaker arrangement. A visualisation of Craven's decoder [3] can be seen in Figure 9. Although the velocity and energy analysis shown looks inferior to the one shown in Figure 8 in some respects, it must be noted that the speaker arrangement used in Figure 8 is much more regular than the ITU specification used in Figure 9.

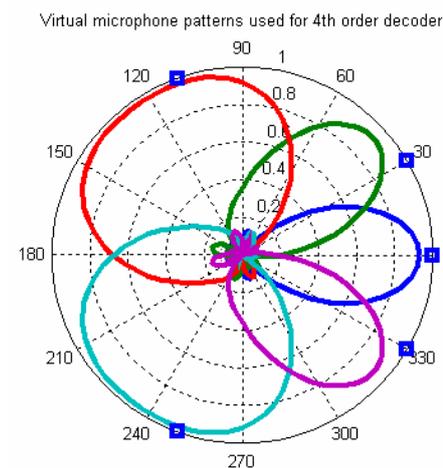
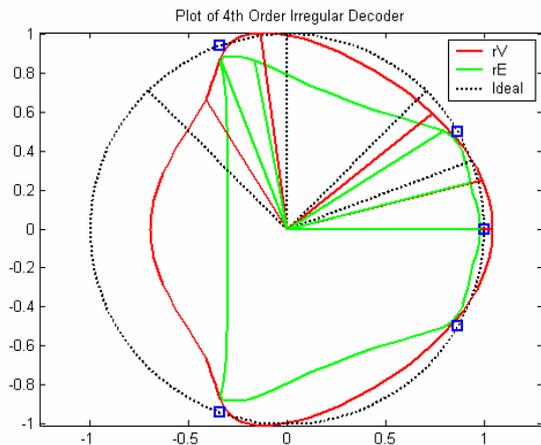


Figure 9 – Velocity and Energy analysis of a 4th order decoder designed for the ITU five speaker arrangement.

One issue with the decoder optimisation problem is that any decoder is a 'best fit' dependant on the criteria already mentioned. This results in the potential for multiple solutions to exist. These then need to be analysed and tested further to find the *actual* 'best fit'. Wiggins [8] showed how a Tabu search could be used to derive 1st order decoders and it's suitability for higher orders will now be shown.

5 HEURISTIC SEARCH METHOD

The decoder derivation problem is one which has a finite search space. That is, the parameters that are to

be derived can be fixed between known bounds. If a 1st order, 5 speaker decoder based upon the ITU standard is considered, then there are 8 coefficients needed to be derived (as shown in Equ. (4)), or eight variables in the search. For 2nd, 3rd and 4th order decoders there are 13, 18 and 23 coefficients needed respectively (each order adds 5 extra variables, 1 for the centre front speaker, 2 for the front speakers and 2 for the rear speakers).

The adapted form of Tabu search used works by having the decoder coefficients initialised at random values (or values of a previous decoder, if these values are to be optimised further). Then the Tabu search program tries changing each of the 'tweakable' coefficients in order, plus or minus the step size (defined by the user). The best result is then kept and the single parameter changed is then restricted to only move in the successful direction for a set number of iterations (again, a value set by the user). This leads us to the most important part of any heuristic search algorithm, and that is, the measure of the decoder's performance, or fitness. Essentially, the fitness of any decoder will be given a single scalar value, which will be used to compare the current decoder with others in the same iteration, and also to compare with the overall best decoder in order to decide which way the current settings are moved in, and whether to store these settings as they are the best so far.

In the case of an irregular Ambisonics decoder, the attributes monitored for fitness will be the same irrespective of the order of spherical harmonics used. The only thing that will change will be the number of parameters that the Tabu search has access to. These six measures of fitness are:

- PFit, the pressure value (low frequency measure of the loudness with respect to encoded source angle)
- MvFit, velocity vector Magnitude, a measure of the localisation quality for low frequencies.
- AvFit, velocity vector angle, a measure of the localisation angle at low frequencies.
- EFit, the energy value (higher frequency measure of the loudness with respect to encoded source angle).
- MeFit, energy vector magnitude, a measure of the localisation quality for higher frequencies.
- AeFit, energy vector angle, a measure of the localisation angle at higher frequencies.

A simple measure of fitness would be to compare the current decoders performance in these six areas with the ideal with respect to the encoded source angle which means that, for example, if a source is taken at every two degrees, 180 results will be found for each of the six measures. The simple fitness equations used in Wiggins [8] are shown in Equ. (7). From these equations it can be seen that the value used for

comparison in each case is 1, i.e., the value is the ideal value on the unit circle and is independent of direction or, in the case of the angle fitness equations, the encoded angle.

$$\begin{aligned}
 PFit &= \sqrt{\frac{\sum_{i=0}^n (1 - P_i)^2}{n}} & EFit &= \sqrt{\frac{\sum_{i=0}^n (1 - E_i)^2}{n}} \\
 MvFit &= \sqrt{\frac{\sum_{i=0}^n (1 - Mv_i)^2}{n}} & MeFit &= \sqrt{\frac{\sum_{i=0}^n (1 - Me_i)^2}{n}} \\
 AvFit &= \sqrt{\frac{\sum_{i=0}^n \left(\theta_i^{Enc} - \theta v_i \right)^2}{n}} & AeFit &= \sqrt{\frac{\sum_{i=0}^n \left(\theta_i^{Enc} - \theta e_i \right)^2}{n}}
 \end{aligned}
 \tag{7}$$

Where: n is the number of encoded angles simulated
 i is the index into the encoded angles

θ^{enc} is the encoded source angle

Once the separate measures of fitness have been calculated, they can then be added together, using a user specified ratio to give an overall value of fitness to be used in the Tabu search. This is necessary as the error values returned aren't directly comparable in terms of the ranges of values that will be typical for a decoder, and the measure with the largest value may well take over in terms of importance if no scaling is used.

So, if the Ambisonic theory is to be followed, then for a low frequency decoder PFit, MvFit, AvFit & AeFit must be optimised and for a high frequency decoder EFit, MeFit, AeFit & AvFit must be optimised for. It should be noted that both velocity and energy vector angles should be in agreement as much as possible if the original theory is to be adhered to (a fact that was not explicitly mentioned in either [8] or [9], and which complicates the optimisation problem for an irregular layout).

A simplified flow chart of the tabu search algorithm used is shown in Figure 10.

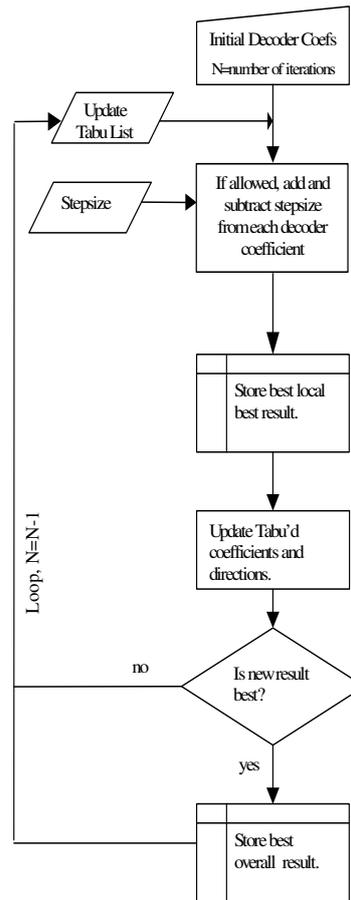


Figure 10 – Simple Tabu Search Flow Chart

A screenshot of the example Tabu search program used to derive 1st to 4th order Ambisonic decoders is shown in Figure 11. In this screenshot a 2nd order optimisation has been carried out (1000 iterations). The weightings have been chosen to concentrate on the energy vector values (En Mag, En Vol and En Ang on the screen shot), but also with the inclusion of the velocity vector magnitude and angle which helps to align the velocity and energy vector angles but also bring up the velocity vector which is necessary for a frequency independent decode. The weightings used can be seen in Table 1.

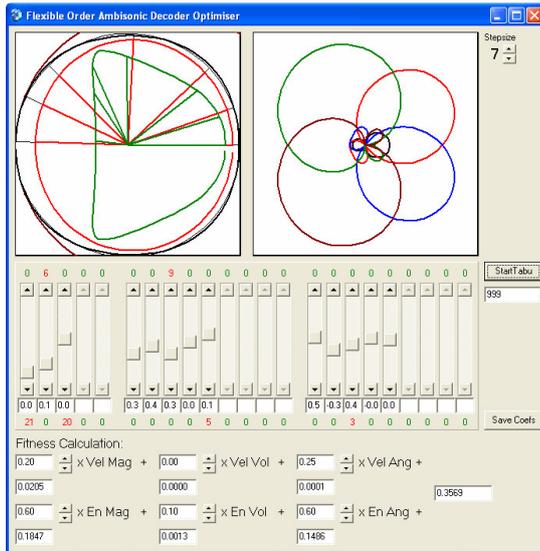


Figure 11 – Screenshot of the Tabu search application. Example shown is a running a 2nd order optimisation with weightings adjusted for a frequency independent decode.

From developing this application the following observations have been made:

- As would be expected, optimising the energy parameters is more difficult than the velocity parameters (as the energy measures depend on squared terms making them non-linear).
- 1st order decoders are difficult to produce for the standard ITU arrangement. Extending the order to 2nd order and above makes it much easier for the tabu search algorithm to find ‘good’ results.
- Although some assumptions can be made for the 1st order coefficients (the rear speakers will only have polar patterns pointing behind, meaning that the X coefficient is always negative, for example), this is not the case for the higher order components. The higher order components are basically used to contaminate the 1st order mic patterns to make them asymmetrical, and the direction in which they need to be contaminated by any particular order cannot be assumed.
- Having an editable step-size and fitness equation weighting is absolutely necessary for this program. As the orders increase, the step-size needs to increase in order to stop the tabu search algorithm getting ‘stuck’ and oscillating between a few values. Also, the fitness weightings can be adjusted depending on how you want the final decoder to perform (i.e. very accurate angular matching, or best localisation quality etc.). This is especially useful as the tabu search optimisation is very fast.
- Having real-time graphical feedback showing the current state of the optimisation process helps in the tweaking of the fitness weightings as the parameter

that the heuristic search algorithm is having trouble optimising can be easily observed, and the weightings adjusted accordingly.

Figure 12 shows an example low and high frequency optimisation that could be used to implement a frequency dependant 2nd order Ambisonic decoder. The weightings used are shown in Table 1. Notice the constantly high weighting of the energy vector angle needed as this parameter is difficult for the search algorithm to optimise.

Decoder type	PFit (Vel Vol)	MvFit (Vel Mag)	AvFit (Vel Ang)	EFit (En Vol)	MeFit (En Mag)	AeFit (En Ang)
Freq. independent	0.00	0.20	0.25	0.10	0.60	0.60
Low frequency	0.25	0.50	0.50	0.00	0.00	0.60
High Frequency	0.00	0.00	0.15	0.75	0.15	0.60

Table 1 – Fitness weightings used for the 2nd order decoder optimisations shown in Figure 12.

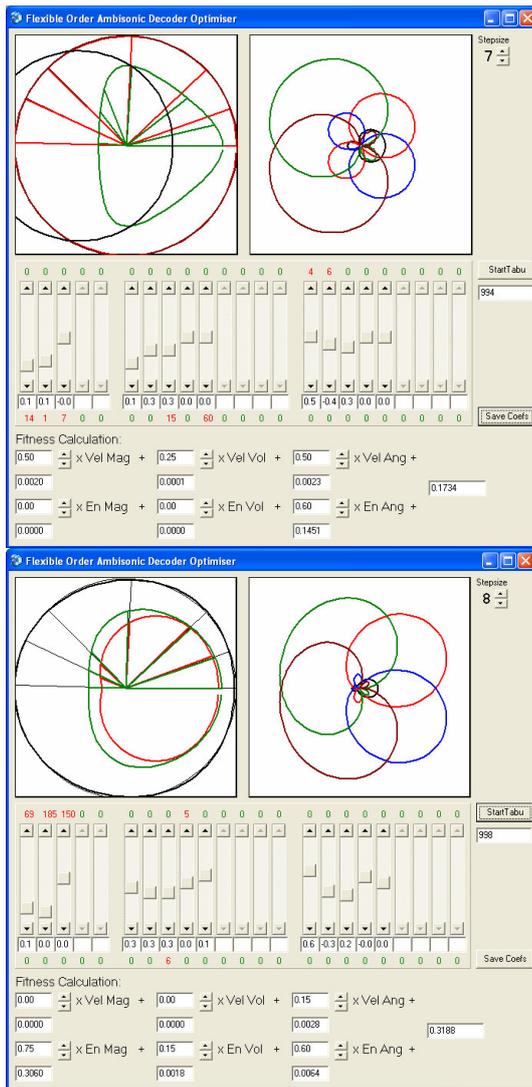


Figure 12 – Low and high frequency optimisations of a 2nd order decoder for the ITU 5 speaker arrangement.

Optimising a 4th order decoder for an irregular 5 speaker array presents 23 tweakable parameters to the tabu search program. However, optimisation in the same way as the 2nd order example above can quickly be achieved. The starting point of the parameters is shown below, and is essentially Craven’s decoder coefficients [3], but with the W gain out by a factor of 1.414 (Craven used a W signal gain of 1 whereas the standard for Ambisonics is 0.707, which is used in this program). The starting point is shown in Figure 13. Using starting values that are further from a good decoder just takes more iteration for an optimal result to be found.

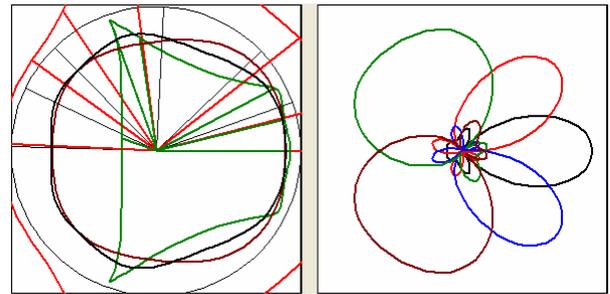


Figure 13 – Starting point for the Tabu search program.

The actual response of the Craven (2003) decoder is shown in Figure 14 and, as a frequency independent decode the following can be observed:

- Both the energy and velocity vector length have been maximised, particularly for the front
- The angles match well near the centre front, but are mismatched for an encoded angle of 45 degrees and do not match the encoded source angle.
- The overall energy is reasonably constant, but the level does drop off around the frontal hemisphere, with the energy level being greater towards the rear.

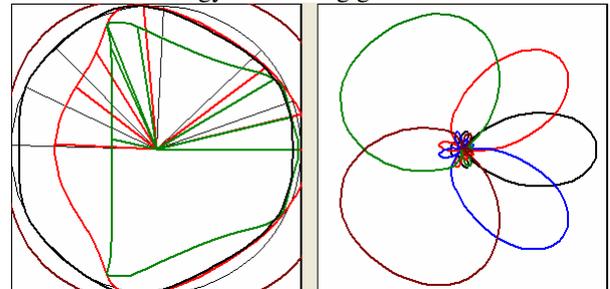


Figure 14 – Actual response of the Craven (2003) decoder as visualised by the Tabu search program.

As a comparison, three optimised decoders will be presented here:

- Two decoders optimised for maximum energy and velocity vector length while also trying to maintain energy and velocity angle matching with a constant energy around the full 360^o sound stage.
- One decoder that tries to make the reproduced angles as accurate as possible.

The fitness equation weightings used are shown in Table 2.

Decoder type	PFit (Vel Vol)	MvFit (Vel Mag)	AvFit (Vel Ang)	EFit (En Vol)	MeFit (En Mag)	AeFit (En Ang)
Max Me Mv 1	0.00	0.25	0.15	0.10	0.50	0.60
Max Me Mv 2	0.00	0.15	0.15	0.10	0.90	0.60
Max Ae Av	0.00	0.15	0.25	0.10	0.50	0.70

Table 2 – Fitness weightings used for the three 4th order decoder optimisations

Using these settings for 2000 iterations produced the decoders shown in Figure 15, Figure 16 and Figure 17. These decoders have the following properties when compared to the decoder from Craven [3] shown in Figure 14:

- The energy level is more accurate over the full 360° sound stage.
- The maximum values of Mv and Me are slightly lower, but more consistent across the sound stage.
- The decoded angles are better matched and closer to the encoded source angle, especially across the frontal hemi-sphere.
- The decoders Max Me Mv 1 & 2 are very similar, although Max Me Mv 1 has higher vector magnitudes for both the energy and velocity vectors in the rear hemi-sphere.

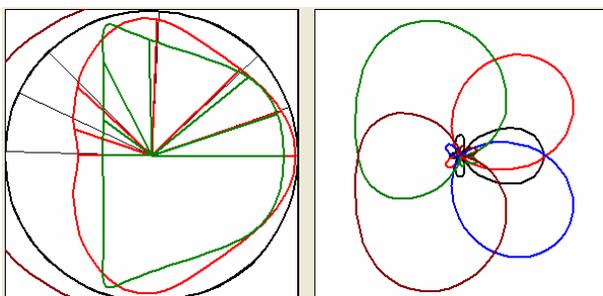


Figure 15 – Max Me Mv 1 4th Order Decoder

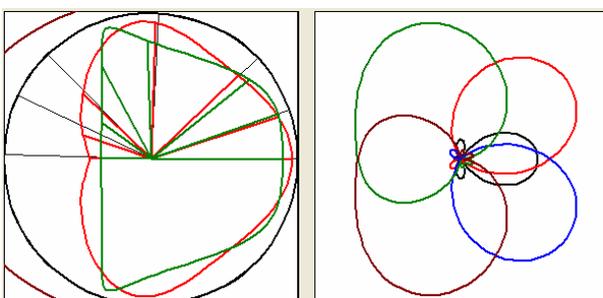


Figure 16 – Max Me Mv 2 4th Order Decoder

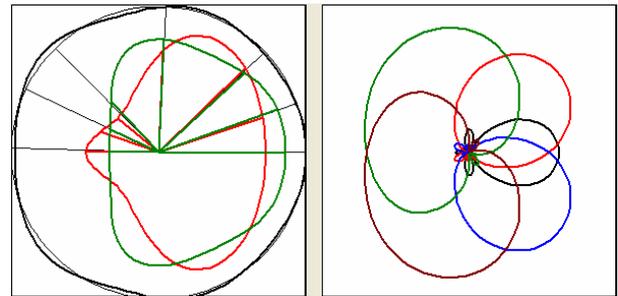


Figure 17 – Max Av Ae 4th Order Decoder

6 HRTF ANALYSIS

As shown in Wiggins [9], HRTF analysis can be used to compare these decoders in more detail. Wiggins [9] showed that there was good agreement with the velocity and energy vector analysis when compared with the two simple lateralisation parameters of time and amplitude differences between the ears of a listener. However, the inclusion of head turning was also introduced in order to quantify the stability of a decoders performance still further, as multiple solutions for each optimised decoder exist. The 4th order decoders already presented will also be compared with an optimised 1st order decoder with respect to the lateralisation parameters of a centrally seated, forward facing listener, and a listener facing 45° to the left. The HRTF data used are those measured by Gardner and Martin [10].

The 1st order optimised Ambisonic decoder can be seen in Figure 18 and has been designed to be frequency independent with matching Av and Ae where possible, although it should be noticed that the angular reproduction of this decoder is skewed slightly towards the front (as the Craven decoder is).

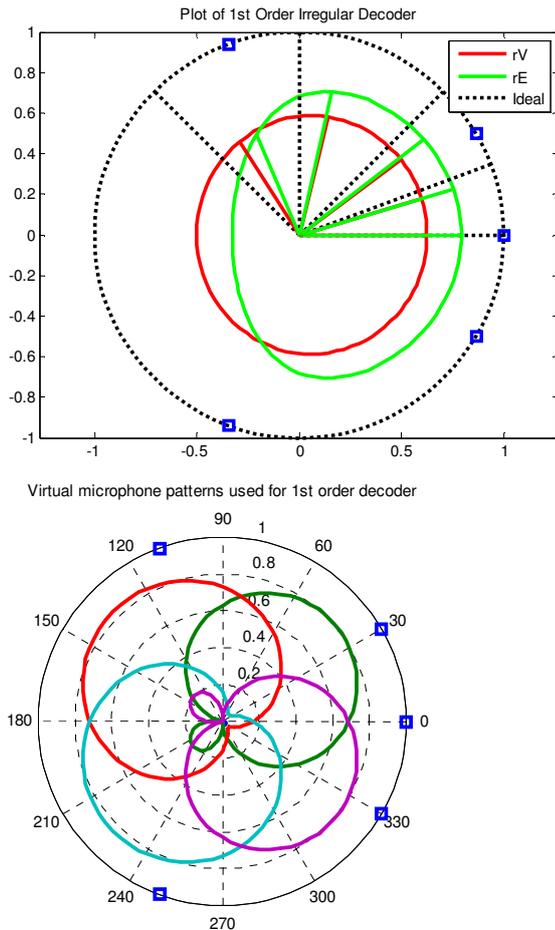


Figure 18 – 1st Order Optimised Decoder.

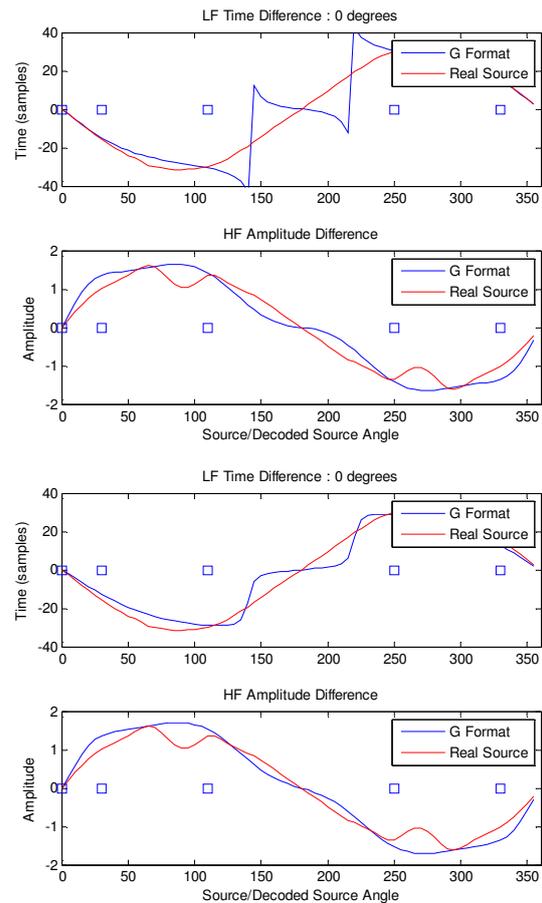


Figure 19 – HRTF analysis of the Max Mv Ma 1 & 2 decoders for a front facing listener

6.1 Forward Facing HRTF Analysis

The time and amplitude differences for the five decoders under test can be seen in Figure 19, Figure 20 and Figure 21. Simple observation suggests that the Craven decoder is the best performing, closely followed by the 2nd Max Mv Ma decoder according to the HRTF analysis (in that the graphs more closely match those of a real source). However, it is interesting to note that it is, in fact, the 1st order decoder that possesses the best fit for the amplitude differences between the ears of a centrally seated, forward facing listener. Another easily observed artefact of these decoders is the sub-optimal panning in the rear section of the sound stage. This is most easily observed in the low frequency time difference plots to various degrees and is perceived as the sound jumping across the rear of the sound field. This is an expected result of having such a large gap between the two rear speakers and been shown to be the case even when a ‘correct’ velocity vector response has been designed [9] showing a weakness in the velocity vector analysis.

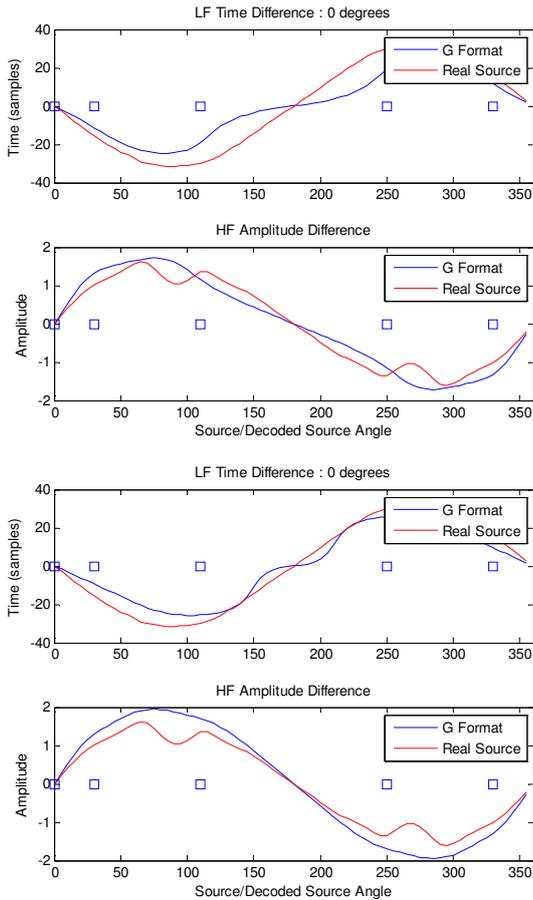


Figure 20 – HRTF analysis of the Max Av Ae and the Craven decoder for a front facing listener

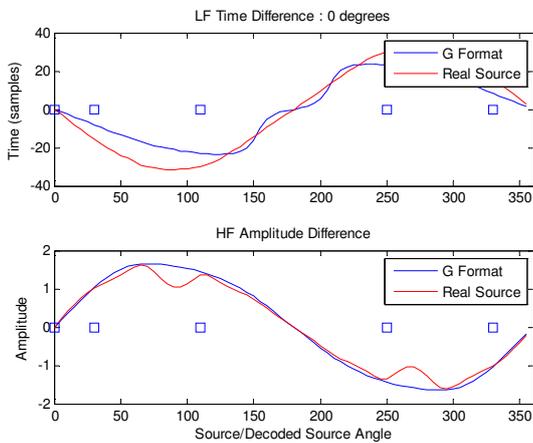


Figure 21 – HRTF analysis of the optimised 1st order decoder for a front facing listener

6.2 Rotated Listener HRTF Analysis

However, when looking at the analysis of a listener whose orientation is rotated (although it should be noted that in the simulation it is, in fact, the speakers that are

rotated around the listener which in an anechoic situation is the same thing) some interesting results are discovered. In this case it is the Craven and the Max Av Ae decoders that perform worse and, surprisingly, also worse than the 1st order optimised decoder. When looking at the Max Av Ae decoder, it can be seen that a source encoded at 0⁰ with respect to the speakers still has a time and amplitude difference of 0. This means that the source is tracking with the listener as they rotate their head. However, in the Craven decode it can be seen that this effect is actually exaggerated in that the source overshoots the rotation by the listener almost resulting in a complete reversal (especially when observing the amplitude difference graph). The two decoders shown in Figure 22 perform better, however, although this isn't apparent when comparing these two decoders to the Craven decoder using purely velocity and energy vector analysis.

It must be noted that this test is purely with regard to the stability of the decoder in the sweet spot and no measure, using HRTF data, of any other attributes have yet been suggested, such as source focus, for example, which would be expected to be better on the higher order decoders when compared to the 1st order variety.

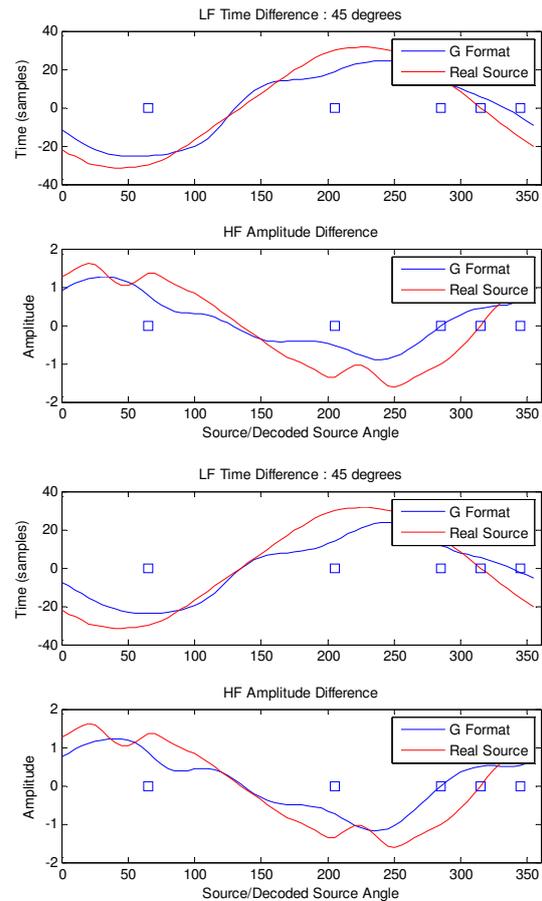


Figure 22 – HRTF analysis of the Max Mv Ma 1 & 2

decoders for a listener 45^0 to the left.

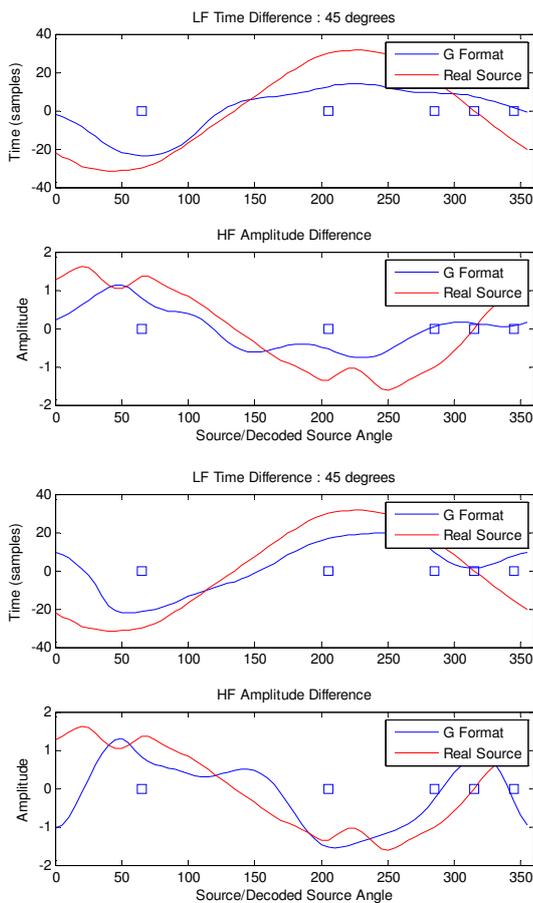


Figure 23 – HRTF analysis of the Max Av Ae and the Craven decoder for a listener facing 45^0 to the left.

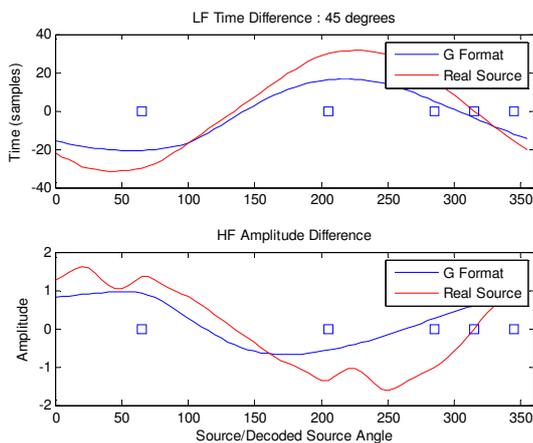


Figure 24 – HRTF analysis of the optimised 1st order decoder for a listener facing 45^0 to the left.

7 CONCLUSIONS AND FURTHER WORK

This paper has demonstrated a simple and robust technique for the derivation of higher order Ambisonic

decoders for the ITU 5 speaker configuration. A HRTF analysis has also been used in order to gain further insight into the performance of the optimised decoders including the use of a simple head turning model to further differentiate the decoders' performance as first demonstrated in Wiggins [9].

This work is, clearly, leading up to listening tests to validate, or otherwise, the hypothesis formed due to this extra insight when compared to using just the velocity and energy vector model. As head movement is used in normal spatial hearing by the ear/brain system, this cue should be an important one.

REFERENCES

- [1] Gerzon, M.A. (1985) Ambisonics in Multichannel Broadcasting and Video. *Journal of the Audio Engineering Society*, Vol. 33, No. 11, p. 851-871.
- [2] Gerzon, M. A. & Barton, G. J. (1992) Ambisonic Decoders for HDTV. *Proceedings of the 92nd International AES Convention*, Vienna. 24 – 27 March. Preprint 3345.
- [3] Craven, P. (2003), Continuous Surround Panning for 5-speaker Reproduction, *AES 24th International Conference*, Banff, Canada.
- [4] Borwick, J. (1981) Could 'Surround Sound' Bounce Back. *The Gramophone*, February, p 1125-1126.
- [5] Malham, D. (2002) *Second and Third Order Ambisonics*. Retrieved: August, 2003, from http://www.york.ac.uk/inst/mustech/3d_audio/secondor.html.
- [6] Leese, M. (n.d.) *Ambisonic Surround Sound*. Retrieved: August, 2003, from http://members.tripod.com/martin_leese/Ambisonic/
- [7] Gerzon, M. A. (1977b) Multi-system Ambisonic Decoder, parts 1 & 2. *Wireless World*, July & August. p. 43 – 47 & p. 63 – 73.
- [8] Wiggins, B. *et al.* (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*.
- [9] Wiggins, B. (2004), An Investigation into the Real-time Manipulation and Control of Three-dimensional Sound Fields, PhD thesis,

University of Derby, Derby, UK

- [10] Gardner B., Martin K. (1994) *HRTF Measurements of a KEMAR Dummy-Head Microphone*, Retrieved: April, 2007, from <http://sound.media.mit.edu/KEMAR.html>.

8 APPENDIX

Coefficients for the 4th order decoders generated for this paper are given below.

8.1 4th Order Max Me Mv 1

Settings used:

VMag	VVol	VAng
0.25	0.00	0.15
EMag	EVol	EAng
0.50	0.10	0.60

	C	FL	BL	BR	FR
W	0.0000	0.2750	0.5350	0.5350	0.2750
C1	0.2250	0.4050	-0.3700	-0.3700	0.4050
S1	0.0000	0.3100	0.4050	-0.4050	-0.3100
C2	0.1200	0.0250	-0.0550	-0.0550	0.0250
S2	0.0000	0.1750	0.0450	-0.0450	-0.1750
C3	0.0550	-0.0100	-0.0000	-0.0000	-0.0100
S3	0.0000	0.0450	-0.0500	0.0500	-0.0450
C4	-0.0050	0.0300	-0.0150	-0.0150	0.0300
S4	0.0000	0.0050	0.0300	-0.0300	-0.0050

8.2 4th Order Max Me Mv 2

Settings used:

VMag	VVol	VAng
0.15	0.00	0.15
EMag	EVol	EAng
0.90	0.10	0.60

	C	FL	BL	BR	FR
W	0.0950	0.3300	0.5650	0.5650	0.3300
C1	0.2200	0.3300	-0.3500	-0.3500	0.3300
S1	0.0000	0.2800	0.4050	-0.4050	-0.2800
C2	0.1850	0.0300	-0.0400	-0.0400	0.0300
S2	0.0000	0.1950	-0.0100	0.0100	-0.1950
C3	0.0600	-0.0200	-0.0200	-0.0200	-0.0200
S3	0.0000	0.0600	-0.0600	0.0600	-0.0600
C4	-0.0150	0.0300	-0.0250	-0.0250	0.0300
S4	0.0000	-0.0050	0.0300	-0.0300	0.0050

8.3 4th Order Max Ae Av

Settings used:

VMag	VVol	VAng
0.15	0.00	0.25
EMag	EVol	EAng
0.50	0.10	0.70

	C	FL	BL	BR	FR
W	0.1450	0.3500	0.6350	0.6350	0.3500
C1	0.2500	0.2950	-0.3150	-0.3150	0.2950
S1	0.0000	0.2750	0.3300	-0.3300	-0.2750
C2	0.2100	-0.0100	-0.0300	-0.0300	-0.0100
S2	0.0000	0.1400	0.0800	-0.0800	-0.1400
C3	0.1000	-0.0100	0.0400	0.0400	-0.0100
S3	0.0000	0.0550	-0.0300	0.0300	-0.0550
C4	-0.0500	0.0200	-0.0000	-0.0000	0.0200
S4	0.0000	-0.0150	-0.0100	0.0100	0.0150