Audio Engineering Society

Convention Paper 7053

Presented at the 122nd Convention
2007 May 5–8 Vienna, Austria

The Design and Detailed Analysis of First Order Ambisonic Decoders for the ITU Layout

David Moore, Jonathan Wakefield

School of Computing and Engineering, University of Huddersfield, Huddersfield, West Yorkshire, HD1 3DH, UK
d.j.moore@hud.ac.uk

ABSTRACT

Ambisonic decoders for irregular layouts can be designed using heuristic search algorithms. These methods provide an alternative to solving the complex mathematic equations. New fitness function objectives for search algorithms are presented which ensure derived decoders meet the requirements of the Ambisonic system more closely than previous work. The resulting new decoder coefficients are compared to other published coefficients and a detailed performance analysis of first order decoders for the ITU layout is given. This analysis highlights current poor performance characteristics that these decoders hold. Proposed future work will attempt to address these issues by looking at techniques for producing decoders with a more even error distribution around the listener and investigating methods for removing the bias towards meeting certain objectives.

1. INTRODUCTION

Ambisonics is a 3D sound spatialisation technique developed by Michael Gerzon in the early 1970s. It is built on mathematical concepts and allows for the reproduction of a recorded or synthesized soundfield. It is unique because of its ability to recreate the pressure and velocity components of an encoded soundfield at the centre of a loudspeaker array by summation of correctly decoded acoustic signals [1].

A significant amount of the literature on Ambisonic reproduction has been focused on loudspeakers being placed in diametrically opposed pairs [2-4]. However, in the 1990s with the introduction of HDTV, work was developed to bring the advantages of Ambisonics to irregularly spaced loudspeaker layouts [5]. This has been further developed more recently because of the irregularity of the standard surround sound layout specified by the International Telecommunications Union (ITU) [6].

It is well known that the design of Ambisonic decoders for irregular loudspeaker layouts is time consuming and complicated. Methods have been proposed which use the heuristic Tabu search algorithm as a viable alternative to solving the complex mathematical equations [7-8]. While there is work describing the use of heuristic search algorithms for deriving decoders, there is little work that gives a detailed analysis of how these decoders perform. This paper presents new solutions for the ITU layout as well as giving a detailed analysis of these solutions. The heuristic search algorithm known as the Tabu search is used for the derivation of decoder coefficients. The coefficients produced aim to maximise the performance of the decoders for the whole sound stage. Decoders that do and do not use shelf filters are derived and analysed. In addition, decoders with differing speaker configurations (rear speakers at 115° and at 120°) will be studied (see figure 1).

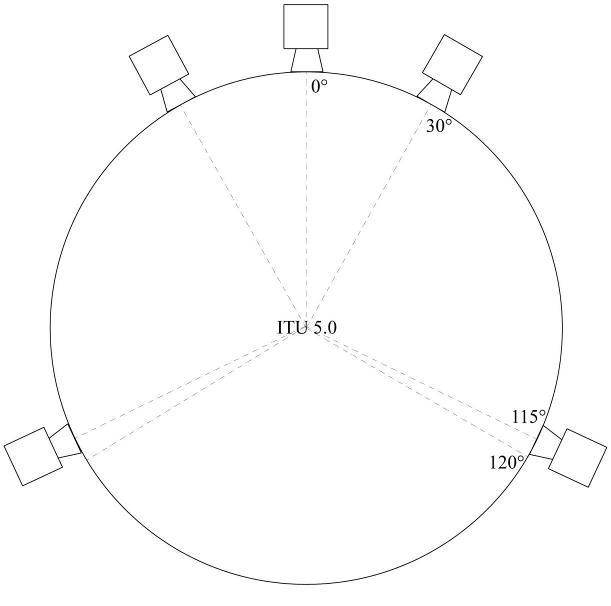

Figure 1: ITU Layout

After an introduction to the theory and techniques used in this work, two analyses are given. The first analysis compares newly derived decoders to previously known decoders. The second analysis looks at the performance of these decoders by reproduced angle. Suggestions for improving the performance of these decoders will be made based on the conclusions drawn from both analyses. It should be noted that although this work focuses on first-order decoders for the ITU system the same principles could be used for decoders of higher-orders or different loudspeaker layouts.

2. BACKGROUND

2.1. Ambisonic encoding

Encoded first-order Ambisonic signals are known collectively as B-Format. In this form they are a representation of a sound field at a single point in space. For first-order horizontal only Ambisonics the equations for encoding a monophonic sound into B-Format are:

$$\begin{aligned} W &= S.\frac{\sqrt{2}}{2} \\ X &= S.\cos\theta \\ Y &= S.\sin\theta \end{aligned} \qquad (1)$$

with W, X and Y representing the horizontal B-format components, S the audio signal and the angle θ denoting the azimuth of the sound source.

It should be noted that the scaling of the omnidirectional W component is based upon a design implementation in the equivalent acoustic encoder known as the Soundfield microphone [9].

2.2. Ambisonic decoding

To playback the encoded audio a re-composition is made that takes into account the location of each loudspeaker. For first-order horizontal Ambisonics there are three constant gain coefficients needed for deriving each loudspeaker feed. This increases to six for decoders that use shelf filters (i.e. three coefficients for low frequencies and three for high frequencies). The output of each loudspeaker is a weighted sum of the encoded B-Format audio:

$$S_i = \alpha_i W + \beta_i X + \gamma_i Y \qquad (2)$$

where S_i is the gain of the ith loudspeaker, W, X, and Y are the encoded B-Format audio signals and α_i, β_i, and γ_i are the constant gain coefficients for the ith loudspeaker. When loudspeaker arrays are regular the gain coefficients can be derived analytically. However, for irregular arrays it becomes more time consuming due to the designer having to generate coefficients numerically [5]. This work seeks to find good values for the constant gain coefficients for the ITU layout using a search algorithm.

2.3. Velocity and Energy Models

The velocity and energy models can be used to predict the localisation performance of an Ambisonic system. They describe the acoustic particle velocity and energy-flow of a soundfield respectively. The velocity model originates from work by Makita [10] and the energy-flow model from work by De-Boer [11]. Both of the these models can be used as an indicator of the Interaural Time Difference (ITD) and Interaural Level Differences (ILD) respectively and broadly represent low and mid/high frequency localisation [12]. The models are described in detail in a "metatheory" of localisation proposed by Michael Gerzon [13]. In his metatheory Gerzon derived a "localisation" vector for both of these models with the angle of each vector being used to show the direction of a reproduced sound source and the magnitude an indicator to the quality of the reproduced sound image. A nominal value of one for the magnitude of both vectors is equivalent to a real single point sound source, less or more than this can be interpreted as a lack of precision in sound localisation. If both vectors are the same for a reproduced sound source as they are for the real sound source then the reproduced sound source should be perceived to be the same as a real sound source. Daniel has shown that it is possible to recreate an ideal velocity vector using an array of loudspeakers [12]. However, this is not possible with the energy vector unless the sound is coming from a single point sound source (i.e. one loudspeaker). Therefore, one of the aims of the Ambisonic system is to ensure that the magnitude of the energy vector is maximised for the entire sound stage and this is at the expense of localisation in the directions of the loudspeakers.

Optimisation of both vectors can be achieved through the use of shelf filters with the crossover frequency usually being between 300Hz and 700Hz. It has been shown that shelf filters should be used for first order systems in order to maximise the performance of the system [2]. For more information regarding the design of shelf filters for first order Ambisonic decoders see Lee [14]. The velocity and energy vectors in Cartesian form can be formulated thus:

$$r_V^x = \sum_{i=1}^{n} S_i \cos(\theta_i) / P \qquad (3)$$

$$r_V^y = \sum_{i=1}^{n} S_i \sin(\theta_i) / P \qquad (4)$$

$$r_E^x = \sum_{i=1}^{n} S_i^2 \cos(\theta_i) / E \qquad (5)$$

$$r_E^y = \sum_{i=1}^{n} S_i^2 \sin(\theta_i) / E \qquad (6)$$

Where:

$$P = \sum_{i=1}^{n} S_i \qquad (7)$$

$$E = \sum_{i=1}^{n} S_i^2 \qquad (8)$$

r_V^x is the velocity vector in the x direction, r_V^y is the velocity vector in the y direction, r_E^x is the energy vector in the x direction, r_E^y is the energy vector in the x direction, n is the number of loudspeakers, θ_i is the angular position of the ith loudspeaker and S_i represents the gain of the ith loudspeaker. These models are used in this research to predict the localisation performance of the decoders derived by the search.

2.4. Tabu search

The Tabu search is a form of local search that explores a search space with the aim of finding the best solution possible. The algorithm is 'intelligent' in that it enhances its performance by using memory structures. One of these memory structures is known as the 'Tabu' list which is a list of previous moves which are designated out-of-bounds, or 'Tabu' (hence the name).

Starting from a random point in the search space the search is allowed to step in positive and negative directions along each coordinate axis (the coordinate axes correspond to each decoder parameter in this work). Each of the neighbouring moves is evaluated by a fitness function with the algorithm selecting the move with the best fitness score. This process is repeated starting from the newly selected current best point until a fixed number of bad moves has been reached. The search will have then reached a local minimum.

The Tabu list is used to guide the search away from previously visited areas in the search domain preventing search cycling. For a detailed description of the Tabu search algorithm see Glover [15] and Rayward-Smith et al [16].

When there are multiple variables involved in the search problem, using the brute force method of searching exhaustively for the best solution is just not feasible (this point was highlighted by Wiggins in [8]). Finding the global minimum would take too long using present day computational processing power. Therefore, using a local search algorithm like the Tabu search is appropriate as it intelligently navigates the search space with the aim of finding a local minimum. If the algorithm is run repeatedly numerous local minima can be found and the best selected. However, there is no guarantee that this minimum is the global minimum.

2.5. Fitness function

The fitness function embodies the key elements of an optimisation problem. It is the value of this function which the search seeks to minimise or maximise by systematically choosing the values of real or integer variables from within an allowed set. The function is always problem dependent. For example, in this work the aim is to minimise the total error of the decoder's localisation performance for a centrally seated listener. The measure of the decoder's performance is encapsulated in a set of objectives based on criteria from Gerzon's metatheory discussed earlier. The representation of the solution in this context is a set of decoder coefficients, namely α_i ,β_i, γ_i for values of i from 1 to n where n is the number of loudspeakers (see equation 2). Speed of execution is of the upmost importance because the function will be called many times during the search.

3. DEVELOPED FITNESS FUNCTION

The fitness function designed for the present work is a multi-objective function that encapsulates criteria from the velocity and energy models. The function algorithm is based on the algorithm proposed by Wiggins in [7], however, some of the objectives have been improved to match Gerzon's specification for the Ambisonic system more closely.

Coefficients generated by the search are passed to the function for evaluation. Each of the coefficients are used as constant gain weightings for the encoded audio (B-Format). The performance of a set of coefficients is assessed at multiple angles around the reproduction array by a number of different function objectives. In the original work by Michael Gerzon on decoders for irregular layouts he checked mathematically seven angles equally spaced around the reproduction array [5]. More recently though due to increases in computational processing power it is now possible to check multiple objectives at multiple angles. In this work seven objectives are checked at one hundred and eighty angles (i.e. half of the sound stage).

When deriving decoders that use shelf filters, it is the view of the authors that low and high frequency coefficients should be calculated simultaneously. Although both sets of coefficients are different they should not be considered independent from one another. It is stated by Gerzon that it is important to make sure both the velocity and energy vectors match up for reproduced source angle up to around 4 KHz [5]. The velocity vector is determined by the low frequency coefficients and the energy vector is determined by the high frequency coefficients. In order to compare them in the fitness function we need to derive them simultaneously.

The function objectives were made computationally efficient because the function will be called many times in the search. For example, taking the absolute value of the objective error was preferred to the root mean square method previously suggested by Wiggins [7] to reduce computational complexity.

3.1. Vector angle objectives

For diametric decoders Gerzon states that the velocity and energy vector localisation will coincide if:

- All speakers are the same distance from the centre of the layout
- Speakers are placed in diametrically opposed pairs
- The sum of the two signals fed to each diametric pair is the same for all diametric pairs [3]

As only the first of these conditions will be met with an ITU decoder it can be taken that the localisation vectors will not coincide. The following objectives are proposed to ensure this performance error is minimised for the whole 360° sound stage. They can be formulated thus:

$$E_{LFAng} = \sum_{i=0}^{n} \left| \theta_i^{Enc} - \theta_i^{V} \right| \qquad (9)$$

$$E_{HFAng} = \sum_{i=0}^{n} \left| \theta_i^{Enc} - \theta_i^{E} \right| \qquad (10)$$

where E_{LFAng} is the error between the encoded source angle and the velocity vector angle, E_{HFAng} is the error between the encoded source angle and the energy vector angle, n is the number of angles to check around the unit circle, θ_i^{Enc} is the encoded source direction at i degrees, and θ_i^{V} and θ_i^{E} are velocity and energy vector angles at i degrees respectively.

3.2. Vector angle match objective

When applying these models to decoder design Gerzon, states that it is important for the velocity and energy vector angles to match up to around 4KHz [5]. In this work a further objective has been added to ensure this which is given by:

$$E_{AngMatch} = \sum_{i=0}^{n} \left| \theta_i^{V} - \theta_i^{E} \right| \qquad (11)$$

where $E_{AngMatch}$ is the error between the velocity and energy vector angles, θ_i^{V} and θ_i^{E} are velocity and energy vector angles at i degrees respectively.

3.3. Vector magnitude objectives

As previously highlighted a localisation vector length of one is optimum. Therefore, the aim of both objectives is to minimise the error at each angle between the ideal length and the reproduced length. The magnitude objectives can be formulated thus:

$$E_{LFMag} = \sum_{i=0}^{n} \left| 1 - R_V \right| \qquad (12)$$

$$E_{HFMag} = \sum_{i=0}^{n} \left| 1 - R_E \right| \qquad (13)$$

where E_{LFMag} is the error between and ideal velocity vector length and the reproduced velocity vector length, E_{HFMag} is the error between the ideal energy vector length and the reproduced energy vector length, n is the number of angles to check around the unit circle, R_V and R_E are the magnitudes of the velocity and energy vector at i degrees respectively.

3.4. Volume objectives

For irregular arrays of loudspeakers one needs to make sure the perceived volume is equal all the way around the listener. The volume objectives proposed by Wiggins compare the volume at every angle against the volume at zero degrees [7-8]. However, this does not necessarily find solutions where the difference between the volumes at each angle is similar. In this work the volume at every angle is compared to the volume at all other angles to ensure reproduced volume error is reduced. The low and mid/high frequency volume objectives are given by:

$$E_{LFVol} = \frac{1}{n^2} \sum_{i=0}^{n} \sum_{j=0}^{n} \left| 1 - P_i / P_j \right| \qquad (14)$$

$$E_{HFVol} = \frac{1}{n^2} \sum_{i=0}^{n} \sum_{j=0}^{n} \left| 1 - E_i / E_j \right| \qquad (15)$$

where E_{LFVol} is the absolute error difference of the Pressure, E_{HFVol} is the absolute error difference of the Energy, n is the number of angles to check around the unit circle, P_i and P_j are the pressure at i and j degrees respectively, and E_i and E_j are the energy at i and j degrees respectively.

3.5. Overall fitness

The overall fitness value is obtained by summing the above objectives. This is used by the search as an indicator to the solutions quality.

4. IMPLEMENTATION

Three different types of decoder will be derived and analysed. For the benefit of the presentation of results each of these decoders will be referred to as type A, B and C decoders. They are defined:

Type A
No shelf filters with rear speakers angled at ± 120°.

Type B
No shelf filters with rear speakers angled at ± 115°.

Type C
Shelf filters with rear speakers at angle ± 115°.

It should be noted that type A and B decoders are a compromise for localisation because they do not use shelf filters so higher error values are expected for these decoders.

The Tabu search was run a total of two hundred times for the three different types of decoder. The coefficients were written to a comma separated value (CSV) file. The objective errors by angle for each set of coefficients were also recorded for the analysis given in section 5.2.

5. RESULTS

5.1. Comparison of decoder performance

Tables 1 to 3 present three of the best solutions derived by the search for each type of decoder. The tables show the individual objective values as well as the minimum, maximum and mean values for each of the objectives. The minimum, maximum and mean were taken from the top ten solutions of each type of decoder generated by the search. Additional solutions derived by others are provided for comparison in Tables 1 and 2.

5.1.1. Type A decoders

All three new type A decoders have total fitness scores that are significantly better than the previously published Furse/Malham decoder (see Table 1). When comparing the individual objectives the new decoders have better fitness scores for 6 out of 7 individual fitness scores. The only objective that was not improved upon was the high frequency magnitude objective (E_{HFMag}). However, there is only 5% difference for this objective.

5.1.2. Type B decoders

The three type B decoders presented all give a better total fitness than the solution provided by Wiggins with 3 out of the 7 individual objectives giving better performance (see table 2). However, it should be noted that Wiggins' solution aims to maximise the energy vector and is known as a 'max Re' decode. The good performance shown for the high frequency objectives are at the cost of poorer performance for others. It is likely the search was weighted in favour of these objectives.

5.1.3. Type C Decoders

As might be expected, Type C decoders (i.e. decoders that use shelf filters) gave much lower error values overall (see Table 3). This is due to the separate optimisation of both low and high frequency objectives. Previous work has also confirmed these decoders perform better subjectively [2]. However, implementation is more complicated than for a non-shelf filter decoder due to the need for linear phase shelf filters.

Error for:	FM	MW 1	MW 2	MW 3	Min	Max	Mean
E_{LFAng}	0.0068	0.0003	0.0010	0.0015	1.3e-005	0.0036	0.0011
E_{HFAng}	35.2040	20.5070	20.4950	20.4910	20.4310	20.5830	20.4860
$E_{AngMatch}$	35.2040	20.5070	20.4950	20.4910	20.4310	20.5830	20.4860
E_{LFMag}	0.0198	0.0009	0.0005	0.0015	0.0004	0.0068	0.0020
E_{HFMag}	59.1480	62.3130	62.3350	62.3430	62.1710	62.4700	62.3580
E_{LFVol}	8.2e-005	3.5e-006	4.4e-006	4.7e-006	9.4e-007	4.7e-005	1e-005
E_{HFVol}	0.3899	0.3101	0.3117	0.3109	0.3071	0.3170	0.3124
Total Fitness	129.9700	103.6400	103.6400	103.6400			

Table 1 Individual objective values for type A decoders. FM (Furse/Malham), MW (Moore/Wakefield).

Error for:	WIG	MW 1	MW 2	MW 3	Min	Max	Mean
E_{LFAng}	9.1222	0.0005	0.0011	0.0002	6.6e-005	0.0039	0.0012
E_{HFAng}	24.7760	26.1320	26.1190	26.1130	26.0340	26.2160	26.1330
$E_{AngMatch}$	21.4170	26.1320	26.1190	26.1130	26.0340	26.2160	26.1330
E_{LFMag}	40.0770	0.0033	0.0021	0.0009	8.5e-005	0.0070	0.0023
E_{HFMag}	57.9260	62.5910	62.6160	62.6340	62.4370	62.7960	62.5960
E_{LFVol}	0.2281	7.4e-006	4.5e006	4.5e-006	3.8e-007	3.9e-005	1e-005
E_{HFVol}	0.0026	0.3793	0.3808	0.3787	0.3753	0.3855	0.3803
Total Fitness	153.5500	115.2400	115.2400	115.2400			

Table 2 Individual objective values for type B decoders. WIG (Wiggins)

Error for:	MW 1	MW 2	MW 3	MW 4	Min	Max	Mean
E_{LFAng}	0.0002	0.0009	0.0028	0.0038	0.0003	0.0277	0.0047
E_{HFAng}	0.0044	0.0047	0.0167	0.0270	0.0013	0.0683	0.0148
$E_{AngMatch}$	0.0047	0.0051	0.0162	0.0270	0.0018	0.0666	0.0147
E_{LFMag}	0.0080	0.0033	0.0055	0.0091	0.0011	0.0390	0.0069
E_{HFMag}	66.6640	66.6180	66.7210	66.6640	66.6180	67.2820	66.9710
E_{LFVol}	1.2e-006	2.4e-005	1.5e-006	7.7e-005	3.7e-007	0.0002	4e-005
E_{HFVol}	1.0234	1.0773	0.9624	1.0000	0.8732	1.1144	0.9889
Total Fitness	67.7050	67.7090	67.7250	67.7290			

Table 3 Individual objective values for type C decoders

5.1.4. Discussion

Type A decoders gave a lower total fitness value than type B. It is likely that this lower value is due to the loudspeakers being slightly more evenly spaced. This suggests the obvious in that a pentagon arrangement of loudspeakers is the optimum arrangement for five loudspeakers when each direction on the sound stage is treated equally. Although this does not represent a significant finding it may warrant further investigation. For example, movement of the speaker angles could be introduced as a variable in the search.

The objectives were ranked in order of difficulty to obtain a good fitness by calculating the average error value for each objective. The average was taken from thirty solutions (the best ten solutions derived for each type of decoder defined in section 4). The order in which they are ranked is not unexpected (easiest first).

1) Low frequency volume (E_{LFVol})
2) Low frequency angle (E_{LFAng})
3) Low frequency magnitude (E_{LFMag})
4) High frequency volume (E_{HFVol})
5) Angle match ($E_{AngMatch}$)
6) High frequency angle (E_{HFAng})
7) High frequency magnitude (E_{HFMag})

The low frequency objectives are easier to meet because of the nature of the velocity vector equations. As highlighted earlier work has shown it is possible to recreate an ideal velocity vector. For all types of decoder the objective that gave the highest error value was the energy vector magnitude (E_{HFMag}). This result was expected because of the nature of the equations used (i.e. the difficulty in producing an energy vector with unit magnitude). The next objective which was the most difficult to minimise was the energy vector angle (E_{HFAng}). The high error value obtained for this objective also greatly influenced the error for the angle match objective ($E_{AngMatch}$). To ameliorate this, new searches were undertaken with an ad hoc weighting applied in favour of the energy vector angle objective and also the angle match objective. As might be expected, this improved the performance for both of these objectives at the cost of the meeting of others. This highlights a problem with using multi-objective functions for this application; namely, some of the objectives are dominating the search at the expense of meeting other objectives. This is caused by each objective having a different effective range. This means that a bad error value for one objective could be a good error value for another and leads to a search that is dominated by the objective with the largest range. A logical method of removing this domination from these objectives is needed so that all the objectives have an equal impact in the fitness function. This would then lead to coefficients being generated that better meet all objectives or can be logically biased towards meeting certain criteria.

5.2. Performance of decoders by reproduced angle

A more detailed analysis of a number of decoders by angle around the listening point is given. The aim of this analysis is to find common error characteristics in decoders for the ITU system so future decoders derived using a search can be improved. It should be noted that although the objectives are examined individually there is a strong interdependency between. Determining the dependency would be complicated and has not been tackled in this paper.

A total of thirty first order decoders will be examined (the best ten solutions derived for each type of decoder defined in section 4). The data for this analysis was taken from the errors given by each objective at each angle around the listening point. During this analysis the solutions will be divided into classes. The class the solutions belong to is dependant on their peak error angles. The decision to do this was based on the initial observation that many of the decoders gave comparable performance at similar angles. All of the plots show the objective error values by angle and for clarity the error was only shown over the range of 0° to 180° (each layout is left/right symmetrical). The lower the error values for the objectives, the better the performance for that aspect of the decoder. 0° is the front of the ITU layout.

5.2.1. Velocity vector angle

The solutions were grouped into two classes depending on their error characteristics. The first class held 24 out of the 30 decoders and was characterised by having two peak error points P1 and P2 and one minimum M1 (see figure 2).

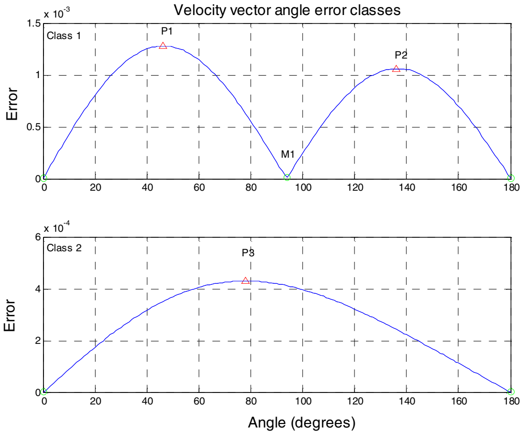

Figure 2: Velocity vector angle classes

40% of type A and C decoders belonged to class one. 60% of the type B decoders belonged to class two. The position variation of all significant peak and minimum error points from each class are presented in table 4.

	Low angle	High angle	Min error	Max error
P1	8	57	0.00001	0.00596
P2	121	157	0.00010	0.00843
P3	60	122	0.00015	0.03510
M1	13	139	0	0.00001

Table 4 Range of peak and minimum error positions for velocity vector angle

In all decoders the error for this objective was lower around the front and back of the system. This is due to the deformation of the velocity vector angle at the sides introduced by the irregularity of the ITU loudspeaker arrangement. However, this deformation is only very slight as highlighted by the low error values given for this objective.

It can be seen that all decoders belonging to class one exhibit better velocity vector angle performance at the sides at the cost of poorer performance elsewhere (see point M1). In all of the decoders the error varied by angle and was never constant.

5.2.2. Energy vector angle

All of the thirty decoders gave a very similar error trend for the energy vector angle objective (see figure 3).

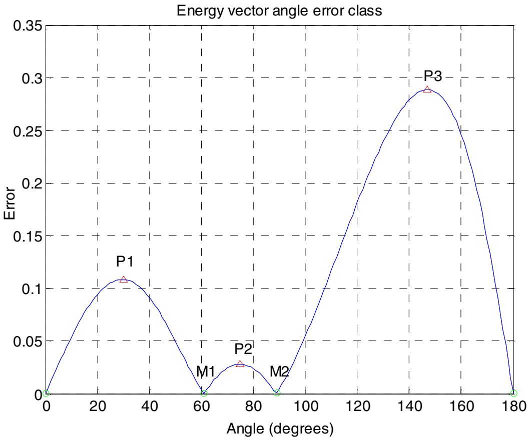

Figure 3: Energy vector angle classes

As with the velocity vector angle, the best performance is at the front and the back of the system. The area that gave the worst performance was towards the rear, as marked by peak error position three (P3) in table 5. The angular ranges of the two larger error peaks (P1 and P3) broadly correlate with the peaks from the class one of the velocity vector angle solutions suggesting common problems for performance in this area.

	Low angle	High angle	Min error	Max error
P1	20	35	0.0020	0.1631
P2	69	103	0.0003	0.0676
P3	139	157	0.0020	0.3851
M1	39	85	0.0001	0.0023
M2	88	138	0.0001	0.0022

Table 5 Variation of peak and minimum error positions for the energy vector angle

5.2.3. Angle match

From the analysis given in section 5.1 it is clear that the large error value from this objective is directly influenced by the large error given by the energy vector objective. This interrelation leads to an analysis which is very similar to that of the energy vector.

5.2.4. Velocity vector magnitude

The decoders where divided into three classes (see figure 4).

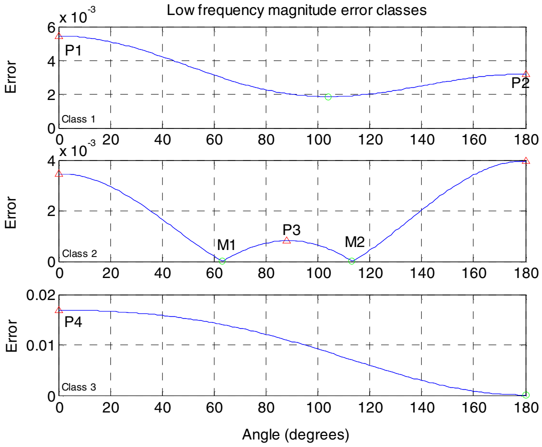

Figure 4: Velocity vector magnitude classes

Class one holds 14 of the 30 solutions and is characterised by having peak error points. Class two holds 13 of the solutions which have three peak error points. In both class one and two the performance for this objective improved around the sides of the system. Class three holds the remaining solutions and despite its poor performance at the front it gradually improves when moving around the angles to the rear of the system. Table 6 displays all significant peak and minimum error positions.

	Low angle	High angle	Min error	Max error
P1	0	74	0.000130	0.01920
P2	93	180	0.000052	0.01140
P3	69	114	0.000006	0.00670
P4	0	97	0.000001	0.00001
M1	11	110	0	0.00009
M2	105	166	0	0.00006

Table 6 Variation of peak and minimum error positions for velocity vector magnitude

5.2.5. Energy vector magnitude

The error for the energy vector magnitude was divided into two types of class (see figure 5).

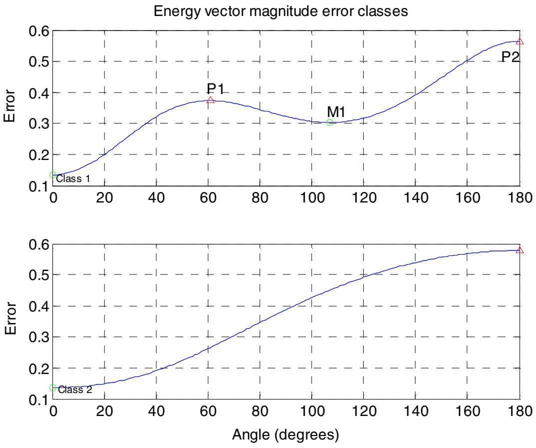

Figure 5: Energy vector magnitude classes

Class one holds 20 of the 30 solutions. All of the type A and B decoders belong to this class. Both classes of decoder have similar performance with their best positions being at the front of the system. Performance got progressively worse from front to rear. Again, this can be explained by the larger angular spacing between the loudspeakers at the rear of the ITU system. Table 7 shows the variation of peak and minimum error positions for the class 1 decoders.

	Low angle	High angle	Min error	Max error
P1	55	61	0.3405	0.3878
P2	180	180	0.5614	0.6649
M1	98	108	0.2337	0.3139

Table 7 Variation of peak and minimum error positions for the energy vector magnitude

5.2.6. Low frequency volume

This objective was the easiest to obtain a good fitness for. All of the decoders showed almost identical error values with the worse performance around the front and rear.

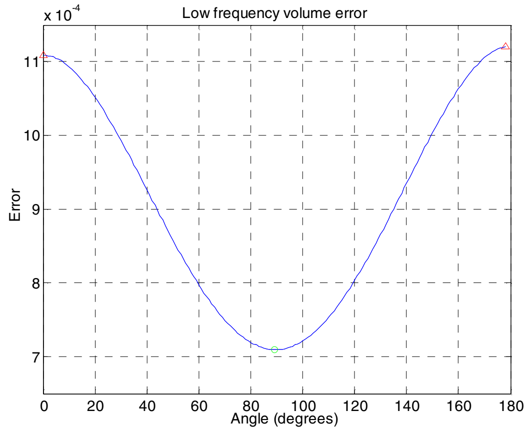

Figure 6: Low frequency volume class

	Low angle	High angle	Min error	Max error
P1	0	0	8.2e-009	8.1e-007
P2	180	180	8.2e-009	8.1e-007
M1	91	91	5.2e-009	2.8e-007

Table 8 Variation of peak and minimum error positions for the low frequency volume

5.2.7. High frequency volume

On the whole this objective gave comparable performance characteristics to the low frequency volume objective (see figure 7).

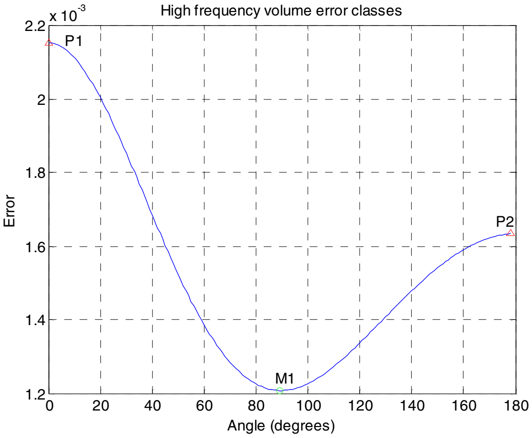

Figure 7: High frequency volume classes

However, there was more of a variation between the error at the front and rear of the system. The range of these values is given in table 9.

	Low angle	High angle	Min error	Max error
P1	0	0	0.0019	0.0032
P2	180	180	0.0022	0.0120
M1	98	108	0.2337	0.3139

Table 9 Variation of peak and minimum error positions for high frequency volume

5.2.8. Overall performance

All type A and B decoders belong to class one with class two holding type C decoders. Overall all, the decoders performed at their best for frontal playback (see figure 8). This trait was expected because of the larger number of speakers in this area. It is known that a smaller angular spacing between speakers allows for the reproduction of better velocity and energy vectors.

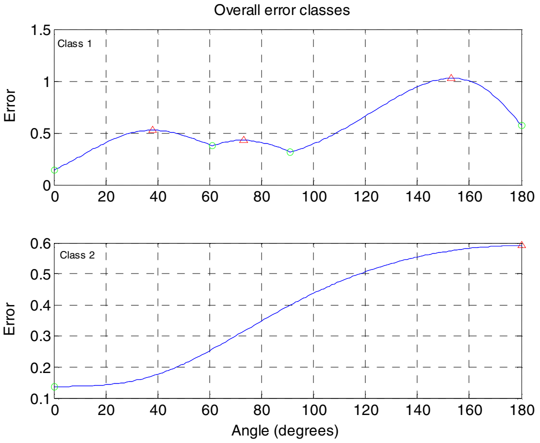

Figure 8: Overall error

This analysis highlights that performance for all ITU decoders varies significantly by angle.

6. CONCLUSIONS

A new multiobjective fitness function for guiding a heuristic search has been presented. The individual objectives that make up this function match the requirements of the Ambisonic system more closely. New solutions derived using this function give better overall performance than those presented in previous work.

Two different analyses were given in which the performance of the decoders were examined. The first showed that some objectives are easier to minimise than others. An additional search was undertaken with an ad hoc weighting applied to poor performing objectives. As expected this improved their overall values at the cost of others. This additional search highlights a common problem inherent when using multiobjective functions in that each of the objectives are likely to have a different range of values potentially leading to certain objectives dominating the search. Part two of the analysis showed that all individual objective errors and the overall error vary by angle. This type of analysis has not previously been presented and highlights a deficiency in relying upon the summed angle errors in guiding the search to a good decoder solution. A decoder with a more even error distribution may have a worse overall fitness using the current fitness function but would perform better in practice having more homogeneous performance characteristics similar to the diametric Ambisonic systems [17].

7. FUTURE WORK

It has been shown that certain objectives are easier to minimise and suggested that certain objectives are dominating the search due to their different range of values. A method for removing this bias will be investigated in a future paper.

Techniques for producing decoders with a more even error distribution around the listener will be investigated.

8. REFERENCES

[1] Gerzon, M. A. Surround-sound psychoacoustics. Wireless World, Vol. 80, pp. 483-486, (1974 December)

[2] Benjamin, E., R. Lee, and A. Heller, Localization in Horizontal-Only Ambisonic Systems, presented at the 121st Audio Engineering Society Convention, San Francisco, CA, USA, 2006 October 5-8.

[3] Gerzon, M., Practical Periphony: The Reproduction of Full-Sphere Sound, presented at the 65th Audio Engineering Society Convention, London, England, 1980 February.

[4] Gerzon, M., Multi-System Ambisonic Decoder. Wireless World, Vol. 83(1499) pp. 43-47, (1977)

[5] Gerzon, M.A. and G.J. Barton, Ambisonic Decoders for HDTV, presented at the 92nd Audio Engineering Society Convention, Vienna, Austria, 1992 February.

[6] ITU, Multichannel stereophonic sound system with and without accompanying picture, 1994.

[7] Wiggins, B., et al. The Design and Optimisation of Surround Sound Decoders Using Heuristic Methods, presented at the Conference of the UK Simulation Society, Emmanuel College, Cambridge, 2003.

[8] Wiggins, B., An Investigation into the real-time Manipulation and Control of Three-Dimensional Sound Fields, in SPARG. 2004, University of Derby: Derby.

[9] Farrer, K., Sound field Microphone. Wireless World, Vol. 85, pp. 48-50, (1979 October)

[10] Makita, Y., On the Directional Localisation of Sound in the Stereophonic Sound Field, E.B.U Review, Vol. 73(Part A Technical), pp. 1536 – 1539, (1962)

[11] K. de Boer. Stereophonic Sound Production. Phillips Technical Review, Vol. 5, pp. 107-144, (1940)

[12] Daniel, J., Représentation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia, in Department 6. 2001, University of Paris.

[13] Gerzon, M.A., General Metatheory of Auditory Localisation, presented at the 92nd Audio Engineering Society Convention, Vienna, Austria, 1992 February.

[14] Lee, R., Shelf Filters for Ambisonic Decoders. http://www.ambisonicbootlegs.net/members/ricardo/SHELFs.doc/view

[15] Glover, F., Tabu Search - Part 1. ORSA Journal of Computing, Vol. 1(3), pp. 190-206, (1989)

[16] Rayward-Smith, V.J., I.H. Osman, et al. Modern Heuristic Search Methods, John Wiley & Sons, 1996.

[17] Malham, D.G., Homogeneous and Nonhomogeneous Surround Sound Systems, in Second Century of Audio, London, England, 1999 June 7-8.