Audio Engineering Society

Convention Paper 7473

Presented at the 124th Convention
2008 May 17–20 Amsterdam, The Netherlands

The Design of Ambisonic Decoders for the ITU 5.1 Layout with Even Performance Characteristics

David Moore[1], Jonathan Wakefield[1]

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

ABSTRACT

All previously published Ambisonic decoders for irregular loudspeaker layouts have localisation performance which varies significantly by angle around the listener. This contrasts with decoders designed for evenly spaced arrangements of loudspeakers where performance characteristics are isotropic. Furthermore even localisation performance around the listener is desirable for a number of application areas of 5.1 surround sound. New decoder design criteria are presented which aim to reduce this variation in localisation performance. These criteria are added to a multiobjective fitness function, based on auditory localisation theory, which guides a heuristic search algorithm to derive decoder parameter sets for the ITU 5.1 layout. The derived decoders exhibit a significant improvement in localisation performance variation by angle around the 360º sound stage.

1. INTRODUCTION

The Ambisonic surround sound system has a considerable amount of flexibility. Systems can be designed and optimised for potentially any loudspeaker layout according to models of auditory localisation [1].

One of the positive aspects of Ambisonics is that for regular loudspeaker layouts (i.e. loudspeakers placed at the vertices of regular a polygon), it treats each direction on the 360° sound stage with equal precedence. This results in isotropic performance characteristics that listeners would experience in a real sound field. However, this is not necessarily the case for decoders designed for irregular loudspeaker layouts.

Recent work by the authors [2] shows that the localisation performance of Ambisonic decoders for the irregular ITU 5.1 loudspeaker layout varies by angle around the ideal central listening point. The best performance was found to be in front of the listener, and the worst performance was behind the listener. The difference in performance between these two areas is significant in terms of objective measurements and was also evident in recent listening tests conducted by Lee and Hellar [3].

This paper presents a method for producing Ambisonic decoders for irregular loudspeaker layouts with more even performance by angle around the central listening point. Even localisation performance is important for any application where the decoder designer wishes to give the listener an isotropic listening experience (rather than the frontal-biased experience normally provided for sound to moving picture). Such decoders would have applications in the playback of surround sound mixes of popular music from DVD-A and SACD and reproduction of electroacoustic soundscapes.

In the following section relevant background theory will be reviewed. This will include information on the Ambisonic system and the methodology adopted for designing Ambisonic decoders for irregular loudspeaker layouts. In section 4, a typical first order decoder for the 5.1 layout will be analysed. This analysis highlights the problem areas in terms of sound source localisation and the overall large variation in performance over the 360° sound stage. New decoder design criteria will then be defined which aim to reduce the large performance variation by angle around the listener. Resulting decoders derived using the new criteria are compared alongside the decoder analysed in section 4. The final part of the paper discusses the merits of the new design criteria.

2. BACKGROUND

2.1. Ambisonics

Ambisonics is a system which can capture and reproduce a sound field in two and three dimensions [4]. It is composed of two stages; an encoding stage and a decoding stage. The encoding stage is based upon sampling a sound field using circular (2D) or spherical harmonics (3D).

2.1.1. Encoding

For a basic system which uses up to first order harmonics sound can be encoded using a soundfield microphone. Alternatively, a monophonic sound can be encoded using the following equations:

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

with W, X and Y representing the horizontal encoded Ambisonic signals (known as B-format), S the audio signal and the angle θ denoting the azimuth of the sound source. It should be noted there is an additional component for encoding height (Z) information which has not been included here.

2.1.2. Decoding

Encoded Ambisonic signals are re-combined in weighted amounts to form each loudspeaker's feed:

$$S_i = \alpha_i W + \beta_i X + \gamma_i Y \tag{2}$$

where S_i is the gain of the ith loudspeaker, W, X, and Y are the encoded audio signals and α_i, β_i, and γ_i are the decoder coefficients for the ith loudspeaker.

When the decoder is being designed for a regular loudspeaker layout, the decoder coefficients are straight forward to derive because the decoding equations that need to be solved are linear [1]. However, for irregular loudspeaker arrays the decoding equations become non-linear and complicated to solve mathematically. When this is the case a viable alternative to solving the equations is to use a search algorithm [5]. The search algorithm seeks to find decoder coefficients which best meet the desired performance characteristics. This is the approach adopted in this research.

2.2. Psychoacoustic 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 [6] and the energy-flow model originates from work by De-Boer [7]. Both of 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 [8]. The models are described in detail in a "metatheory" of localisation proposed by Michael Gerzon [9]. 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 [8]. 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 can be achieved at the expense of localisation in the directions of the loudspeakers.

Optimisation of both vectors can be achieved through the use of shelf filters within the playback system 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 [10]. For more information regarding the design of shelf filters for first order Ambisonic decoders see Lee [11]. The velocity and energy vectors in Cartesian form can be formulated thus:

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

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

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

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

Where:

$$P = \sum_{i=1}^{n} S_i \tag{7}$$

$$E = \sum_{i=1}^{n} S_i^2 \tag{8}$$

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.

Both of these models are used in this research to predict the localisation performance of the decoders derived by the search.

3. DECODER DESIGN METHODOLOGY

The design of an Ambisonic decoder can be formulated as a search problem. The basic principle is to use a search algorithm to find a set of decoder coefficients which best fit the design objectives specified in a fitness function.

3.1. Fitness function

In this work and in all work of a similar nature the following fitness function objectives (or similar) have been employed for measuring a decoder's performance. Each objective is checked at 180 angles around the listener (i.e. half of the left-right symmetrical ITU 5.1 layout). The aim of each objective is to minimise the difference between the following at each angle:

1) Reproduced low frequency volume and reproduced low frequency volume at all other angles i.e.

$$E_{LFVol} = \frac{1}{180^2} \sum_{i=0}^{180} \sum_{j=0}^{180} \left| 1 - P_i / P_j \right| \tag{9}$$

where E_{LFVol} is the low frequency volume error, P_i and P_j are the pressure at *i*th and *j*th degrees respectively.

2) Reproduced high frequency volume and reproduced high frequency volume at all other angles

$$E_{HFVol} = \frac{1}{180^2} \sum_{i=0}^{180} \sum_{j=0}^{180} \left| 1 - E_i / E_j \right| \tag{10}$$

where E_{HFVol} is the high frequency volume error, E_i and E_j are the energy at *i*th and *j*th degrees respectively.

3) Velocity vector magnitude and an ideal magnitude of unity

$$E_{LFMag} = \sum_{i=0}^{180} \left|1 - r_{Vi}\right| \tag{11}$$

where E_{LFMag} is the low frequency magnitude error and r_{Vi} the reproduced velocity vector length at the ith angle.

4) Energy vector magnitude and an ideal magnitude of unity

$$E_{HFMag} = \sum_{i=0}^{180} \left|1 - r_{Ei}\right| \tag{12}$$

where E_{HFMag} is the high frequency magnitude error and r_{Ei} the reproduced energy vector length at the *i*th angle.

5)Velocity vector angle and encoded sound source angle

$$E_{LFAng} = \sum_{i=0}^{180} \left|\theta_i^{Enc} - \theta_i^{V}\right| \tag{13}$$

where E_{LFAng} is the low frequency angle error, θ_i^{Enc} the encoded source angle and θ_i^{V} the velocity vector angle at the *i*th angle.

6) Energy vector angle and encoded sound source angle

$$E_{HFAng} = \sum_{i=0}^{180} \left|\theta_i^{Enc} - \theta_i^{E}\right| \tag{14}$$

E_{HFAng} is the high frequency angle error, θ_i^{Enc} is the encoded source angle and θ_i^{E} the energy vector angle at the *i*th angle.

7) Velocity vector angle and energy vector angle

$$E_{AngMatch} = \sum_{i=0}^{180} \left|\theta_i^{V} - \theta_i^{E}\right| \tag{15}$$

where $E_{AngMatch}$ is the high frequency and low frequency angle match error, θ_i^{V} the velocity vector angle and θ_i^{E} the energy vector angle at the *i*th angle.

The total fitness value is obtained by summing the above objectives and is used by the search as an indicator to the solutions quality. Before summing the objectives to obtain the total fitness, however, it is important that an objective range-removal technique is used to ensure none of the objectives dominate the search. Range-removal and a method of weighting objectives in order of importance is described in detail in [12].

It should be noted that only the volume objectives defined above aim to ensure even performance around the listener. Section 4 will demonstrate that additional criteria is required if decoders with even performance characteristics are to be derived.

3.2. Tabu search

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 [13]). As there are multiple decoder coefficients required for an irregular Ambisonic decoder (8 for the ITU 5.1 layout) searching for the global minimum to any reasonable resolution 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.

The Tabu search is an 'intelligent' algorithm 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). The Tabu list prevents the search from visiting previously visited areas in the search domain avoiding the problems of search cycling. For a detailed description of the Tabu search algorithm see Glover [14] and Rayward-Smith et al [15].

Starting from a random point in the search space the search moves locally along each coordinate axis (the coordinate axes correspond to each decoder parameter in this work). Each of the neighbouring moves is evaluated by the fitness function with the algorithm selecting the move with the best fitness score. This process is repeated starting from the newly selected best point until a fixed number of bad moves have been reached.

4. PERFORMANCE ANALYSIS

Figure 1 plots the combined performance error for all objectives by angle for a typical first order Ambisonic decoder derived for the ITU 5.1 layout. This particular decoder is known as a “max r_E” decoder which requires that the energy vector magnitude is maximised for the whole 360° sound stage. Max r_E decoders are often implemented when shelf filters are not used as optimisation of the energy vector can result in better performance in off-centre positions [8]. This decoder was derived using the fitness function defined in section 3 with range-removal and importance weightings applied to the objectives in the search. The mean performance error is included for reference.

Figure 1 is divided into five segments, each representing the spacing between a pair of loudspeakers in the 5.1 layout. As might be expected, each segment has a different performance which is directly related to the angular spacing between the loudspeakers (i.e. the best performance is in the frontal segments where the loudspeaker angular spacing is narrowest, and the worst performance in the rear segment where the loudspeaker angular spacing is widest).

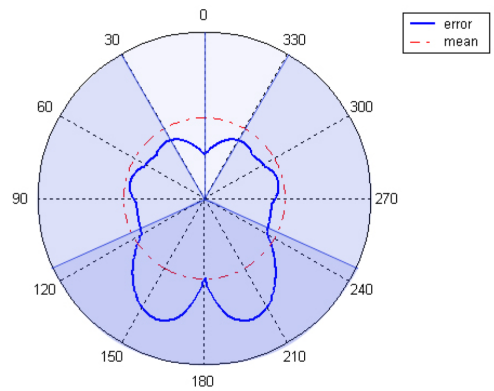

Figure 1: Performance error by angle

Figure 2 plots each fitness function objective across the 360° sound stage. The volume objectives have been omitted from this figure as their error was negligible. The contribution of each objective to the total performance error is also given in table 1 as a percentage. Please note that these percentages will largely reflect the importance weightings given to each of the objectives in the search.

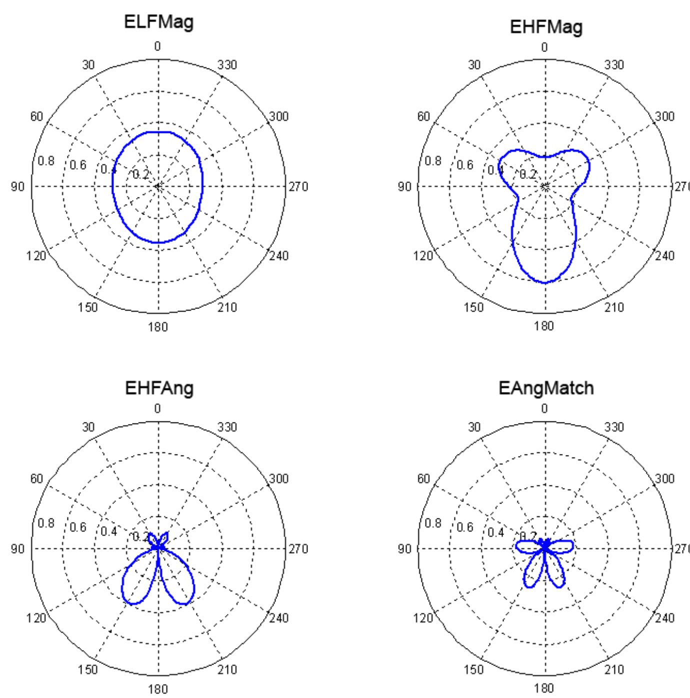

Figure 2: Individual objective errors by angle

	E_{LFVol}	E_{HFVol}	E_{LFMag}	E_{HFMag}	E_{LFAng}	E_{HFAng}	$E_{AngMatch}$
% of total error	0.0767	0.0001	32.5846	30.7850	10.8681	14.1505	11.5344

Table 1: Objective errors as a percentage of the total error

For this particular decoder the magnitude objectives (E_{LFMag} and E_{HFMag}) have the highest percentage of the total fitness so have the largest bearing on the overall trend seen in figure 1.

The standard deviation of all the fitness function objectives is given in table 2. This reveals how the objectives vary across the 360° sound stage about their mean values. All the objectives for this particular decoder (apart from the volume objectives which were originally designed to ensure even error) have a certain amount of variability. The objectives with the greatest overall variation are the energy vector magnitude and energy vector angle objectives (E_{HFMag} and E_{HFAng}). This large fluctuation in energy vector performance is likely to have a significant impact on the even listening experience for this type of decoder (i.e. max r_E).

	E_{LFVol}	E_{HFVol}	E_{LFMag}	E_{HFMag}	E_{LFAng}	E_{HFAng}	$E_{AngMatch}$
Standard deviation	0.0001	0.0001	0.0241	0.1173	0.0615	0.1262	0.0765

Table 2: Standard deviation of the objective errors

Further examination of each objective in each segment confirmed that the lowest deviation from the mean is found in the frontal segment (±30°), and the highest deviation from the mean is found in the rear segment (115° – 180°).

In summary, relying on the summed error by angle for producing a good decoder is not always appropriate. It is clear that some of the fitness function objectives are easier to meet in areas of the system where loudspeakers are closer together. This results in a search producing decoders which are strongly biased towards better frontal playback. Although good frontal sound stage performance is desirable in some circumstances (e.g. audio for moving picture), if a more natural sound field is to be perceived, then equal importance should be given to every direction in the sound stage (a trait which occurs naturally for Ambisonic decoders designed for regular loudspeaker arrangements). The following section introduces new criteria to the fitness function to reduce the overall performance variation of each objective across the 360° sound stage.

5. EVEN PERFORMANCE CRITERIA

In the fitness function only the volume objectives are already designed to ensure even performance around the 360° sound stage. In order to produce more even performance for the velocity and energy vectors additional objectives need to be added to the fitness function. The following four objectives have been designed to address this problem:

$$E_{LFAngEven} = \sqrt{\frac{1}{180}\sum_{i=0}^{180}\left(E_{LFAng\,i} - \overline{E_{LFAng}}\right)^2} \quad (9)$$

$$E_{HFAngEven} = \sqrt{\frac{1}{180}\sum_{i=0}^{180}\left(E_{HFAng\,i} - \overline{E_{HFAng}}\right)^2} \quad (10)$$

$$E_{LFMagEven} = \sqrt{\frac{1}{180}\sum_{i=0}^{180}\left(E_{LFMag\,i} - \overline{E_{LFMag}}\right)^2} \quad (11)$$

$$E_{HFMagEven} = \sqrt{\frac{1}{180}\sum_{i=0}^{180}\left(E_{HFMag_i} - \overline{E_{HFMag}}\right)^2} \quad (12)$$

where $E_{LFAngEven}$, $E_{HFAngEven}$, $E_{LFMagEven}$, $E_{HFMagEven}$ are the standard deviation of the corresponding objectives defined in section 3.1.

Each of the new objectives uses the standard deviation to measure the variation of the performance around the loudspeaker layout for the corresponding objective. If the optimum value is met for each of these objectives there will be no deviation from the mean and hence no variation for the corresponding objective.

6. EVEN ERROR DECODERS

6.1. Implementation

The new even error objectives were incorporated into the fitness function and a number of searches were undertaken. The Tabu search was started from a random position in the search space each time and a range-removal technique used to ensure none of the objectives dominated the search.

In order to find suitable importance weightings for each objective a number of pilot searches were undertaken. This was necessary because of the difficultly in determining relationships between objectives. For example, it was often the case that adjusting the importance of one objective had a direct effect on another. This was particularly apparent when selecting importance weightings for even error vector magnitude objectives ($E_{LFMagEven}$, $E_{HFMagEven}$). Applying a high weighting to either of these objectives led to a decrease in performance for the vector magnitude objectives (E_{LFMag}, E_{HFMag}) and vice versa.

Once suitable importance weightings were found, two decoders with different performance characteristics were derived. In the following analysis they are compared to the max r_E decoder analysed in section 3. The importance weightings for all three decoders are provided in table 3.

	E_{LFAng}	E_{HFAng}	$E_{AngMatch}$	E_{LFMag}	E_{HFMag}	E_{LFVol}	E_{HFVol}	$E_{LFAngEven}$	$E_{HFAngEven}$	$E_{LFMagEven}$	$E_{HFMagEven}$
Max r_E	1.0	2.2	1.4	1.0	2.6	1.0	1.6	-	-	-	-
Decoder A	0.5	1.0	1.0	0.75	1.5	0.0	1.0	1.7	1..9	0.7	0.8
Decoder B	0.5	0.9	0.9	0.5	0.9	0.0	0.9	1.0	1.0	1.0	1.0

Table 3: Importance weightings applied to the objectives

6.2. Analysis of decoders

Figure 3 plots the performance error by angle for all three decoders. The mean and the standard deviation of the total error are also provided for reference. In terms of overall localisation performance Decoder A is similar to the max r_E decoder. However, the localisation performance of Decoder A is more even at the front and the sides of the system (0° - 120°). Decoder B has the most even error distribution out of all three decoders. However, it should be noted that the increase in even performance has been at the cost of a reduction in overall performance

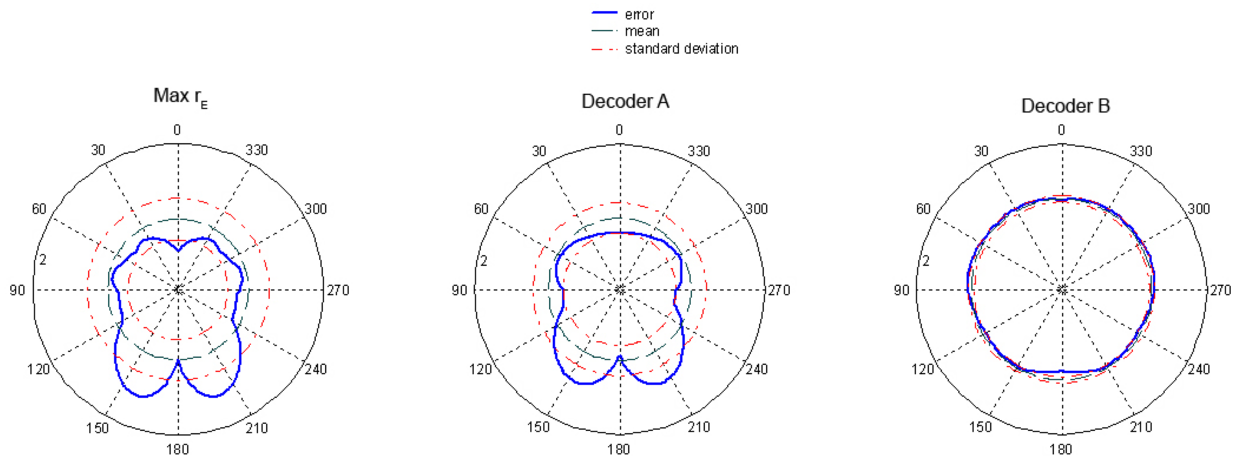

Figure 3: Performance error by angle for all three decoders

Figure 4 plots the individual objectives values by angle for each decoder. It is clear that both Decoder A and Decoder B have more even performance for all objectives when compared to the max r_E decoder. This is highlighted in table 4 which gives the standard deviation for all objectives for each of the decoders. Among the most interesting points from these plots is the significant increase in even performance for the energy vector magnitude for Decoder B. The energy vector magnitude performance is now much smoother around the 360º sound stage and no longer peaks in the direction of the loudspeakers.

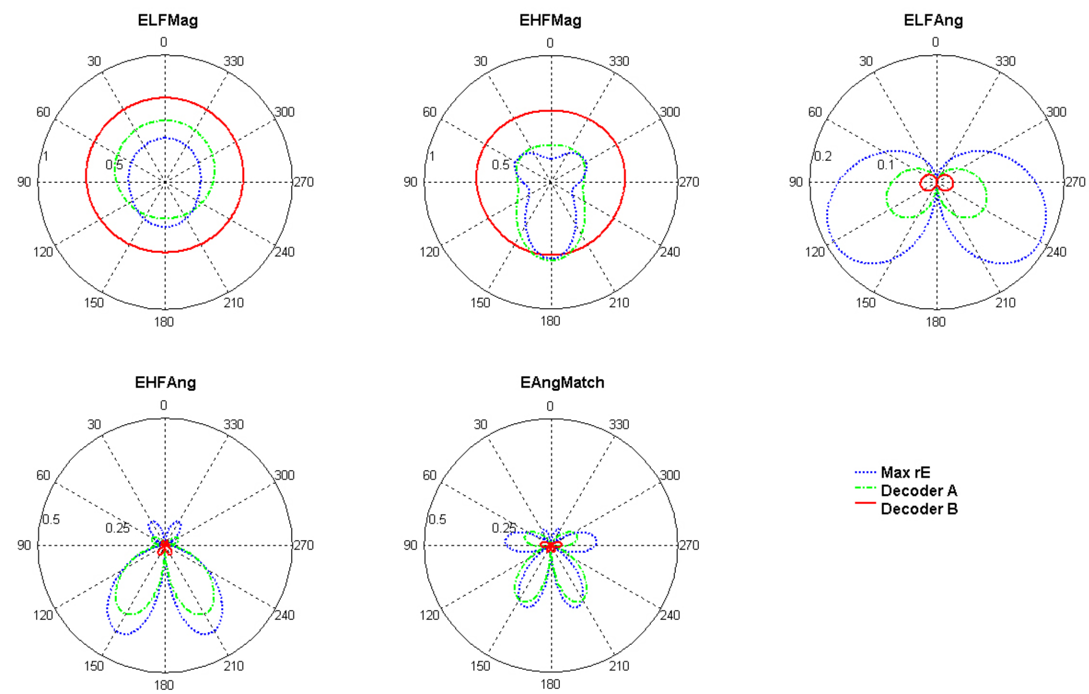

Figure 4: Objective error by angle for all three decoders (please note the change in scale for the different plots)

	E_{LFAng}	E_{HFAng}	$E_{AngMatch}$	E_{LFMag}	E_{HFMag}	E_{LFVol}	E_{HFVol}
Max r_E	0.0615	0.1262	0.0765	0.0241	0.1173	0.0000	0.0000
Decoder A	0.0567	0.1104	0.0693	0.0077	0.1040	0.0001	0.0000
Decoder B	0.0080	0.0122	0.0126	0.0382	0.0102	0.0001	0.0000

Table 4: Standard deviation of objective error for all three decoders

An interesting objective inter-relationship became apparent when conducting these searches. When a low error value was obtained for the vector angle objectives, a high error value was obtained for vector magnitudes and vice versa (this is particularly apparent for Decoder B). Future searches which incorporate even error design criteria should take this into consideration during the weighting of the objectives.

7. CONCLUSIONS

The localisation performance of Ambisonic decoders for irregular loudspeaker layouts varies by angle around the listening point. In an attempt to reduce this variation new decoder design objectives have been incorporated into a search. The new objectives are based on the standard deviation and are specifically designed to reduce the performance variation of the velocity and energy vector magnitudes and angles around the 360° sound stage.

The analysis shows that the new objectives have been successful in this respect, however, consideration should be made when determining their importance weightings in the search. It was found that there is a direct trade-off between choosing good overall performance and good even performance by angle for each of the objectives. However by adjusting the importance weighting between the original objectives and the newly added objectives a decoder designer can achieve the required balance between good overall decoder performance and even performance for all angles.

8. REFERENCES

[1] Gerzon, M.A. and G.J. Barton, "Ambisonic Decoders for HDTV", presented at the AES 92nd Convention, 1992 March 24-27.

[2] Moore, J.D. and J.P. Wakefield, "The Design and Detailed Analysis of First Order Ambisonic Decoders for the ITU Layout", presented at the AES 122nd Convention, Vienna, Austria, 2007 May 5-8.

[3] Lee, R. and A.J. Hellar, "Ambisonic Localisation – Part 2", presented at the 14th International Congress on Sound and Vibration, Cairns, Australia, 2007 July 9-12.

[4] Gerzon, M.A. "Periphony (with-height sound reproduction)", presented at the AES 2nd Central European Convention, Munich, Germany, 1972 March 14-16.

[5] 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, UK, 2003.

[6] 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)

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

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

[9] Gerzon, M.A., General Metatheory of Auditory Localisation, presented at the AES 92nd, Vienna, Austria, 1992 February.

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

[11] Lee, R., Shelf Filters for Ambisonic Decoders. http://www.ambisonia.com/Members/ricardo

[12] Moore, J.D. and J.P. Wakefield, "The Design and Improved First Order Ambisonic Decoders by the Application of Range-Removal and Importance in a Heuristic Search Algorithm", presented at the AES 31st International Conference, London, UK, 2007 June 25-27.

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

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

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