THE DESIGN OF IMPROVED FIRST ORDER AMBISONIC DECODERS BY THE APPLICATION OF RANGE REMOVAL AND IMPORTANCE IN A HEURISTIC SEARCH ALGORITHM

DAVID MOORE, JONATHAN WAKEFIELD

School of Computing and Engineering, University of Huddersfield, Huddersfield, West Yorkshire, UK
d.j.moore@hud.ac.uk j.p.wakefield@hud.ac.uk

This paper presents improvements to previous work on deriving first order Ambisonic decoders for ITU 5.1. The decoders are derived using a heuristic search method with an objective function based upon Gerzon's metatheory of auditory localisation. An analysis of previously derived decoders shows that they are biased towards particular design objectives due to the nature of the multiobjective function guiding the search. This paper applies a technique called range removal to systematically and logically remove this bias which leads to improved decoder coefficients that better meet all of the objectives. A further technique known as importance is introduced that enables the logical biasing of range-removed objectives. A case study to develop a "max r_E" decoder demonstrates this technique in action.

INTRODUCTION

Reproduced audio quality can be improved by increasing sampling rate and bit resolution. However, increasing the sampling rate and bit resolution are not the only ways of improving the quality of the reproduced audio. The listener's perception of sound location, and hence sound quality, can also be improved by developing better surround sound decoders. This paper presents such a development.

The mathematical design of Ambisonic decoders for irregular loudspeaker arrangements is complicated [1]. A system of non-linear equations needs to be solved in order produce a suitable set of decoder coefficients. A viable alternative to mathematically solving the decoder equations has been introduced by Wiggins et al [2]. It involves using a heuristic search algorithm to search for good values for decoder coefficients.

Heuristic search algorithms are guided to a good set of decoder coefficients by a fitness function that embodies a number of desired performance characteristics i.e. a multi-objective fitness function. In a previous paper, which extended the earlier work of Wiggins et al, a deficiency was identified with the use of multi-objective fitness functions [3]. The crux of the problem lies in each of the fitness function objectives having a different numerical range. When this is the case a search is effectively biased in favour of the objectives with the largest range causing these objectives to dominate the search and become better optimised at the expense of other objectives. This paper applies a technique called range removal to systematically and logically remove this bias, which leads to improved ITU 5.1 decoder coefficients that better meet all of the objectives.

1 BACKGROUND

1.1 Ambisonic encoding

Encoded first-order Ambisonic signals are known collectively as B-Format [1]. 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:

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

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.

1.2 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. 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 [1].

1.3 Velocity and energy vectors

Velocity and energy vectors can be used for designing sound reproduction systems. These vectors are described in detail in a metatheory of auditory localisation by Michael Gerzon [4]. Basically, the velocity vector can be used for predicting the low frequency localisation performance of a sound reproduction system for a centrally seated listener, and the energy vector can be used for predicting the mid to high frequency localisation performance for a centrally seated listener. The vector magnitudes indicate the “quality” of the reproduced sound image and the vector angles indicate the reproduced sound source’s angular position. A magnitude of unity is optimal for both vectors. They are defined 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:

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

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

where P is the pressure, E is the energy, 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 y direction, n is the number of loudspeakers, θ_i is the angular position of the i^{th} loudspeaker and S_i represents the gain of the i^{th} loudspeaker.

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

1.4 Tabu search

It is not feasible to exhaustively search for all possible decoder coefficient values so a heuristic search algorithm must be used [2]. The Tabu search is a form of heuristic search algorithm that explores a search space with the aim of finding the best solution possible. The basic idea of the algorithm is to avoid visiting the same solution more than once during the search by using a Tabu list. In the Tabu list old solutions are stored and used to prevent searches from being guided back towards solutions already found. For a detailed description of the Tabu search algorithm see Glover [5, 6] and Rayward-Smith et al [7].

1.5 Fitness function

The Tabu search is guided towards good sets of decoder coefficients by the fitness function. The fitness function in this work evaluates 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 seven objectives based on criteria from Gerzon’s metatheory described earlier. Each of the seven objectives is checked at 180 different angles around the listener. One half of the sound stage is evaluated to reduce the number of calculations as the decoder outputs are symmetrical on either side of the listener. 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 ith and jth degrees respectively.

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

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

where E_{HFVol} is the high frequency volume error, E_i and E_j are the energy at ith and jth degrees respectively.

3) Velocity vector magnitude and an ideal magnitude of unity

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

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

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

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

5) Velocity vector angle and encoded sound source angle

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

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 ith angle.

6) Energy vector angle and encoded sound source angle

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

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 ith angle.

7) Velocity vector angle and energy vector angle

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

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 ith 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.

2 FITNESS FUNCTION OBJECTIVE DOMINANCE

In general when using a multi-objective fitness function the individual objectives will almost always have different effective ranges of possible values [8]. This is true for using a multi-objective fitness function for deriving decoder coefficients. Unless measures are taken to counteract this, particular objectives will dominate the search.

Objective dominance was investigated in the context of the current research. Table 1 displays the mean values of the individual objectives taken from the best 10 sets of decoder coefficients out of 200 decoders derived by the search. These results are for first order Ambisonic decoders that do not use shelf filters and have ITU 5.1 surround speakers angled at 115°.

Objective:	Mean
E_{LFAng}	0.0012
E_{HFAng}	26.1330
$E_{AngMatch}$	26.1330
E_{LFMag}	0.0023
E_{HFMag}	62.5960
E_{LFVol}	1e-005
E_{HFVol}	0.3803
Total fitness	115.2458

Table 1: Mean individual objective values

It can be seen that the individual objective values vary substantially. The best values are achieved for the low frequency objectives (E_{LFAng}, E_{LFMag}, E_{LFVol}). All three have significantly lower values when compared with the other objectives and account for less than 1% of the total fitness error. The good results for the low frequency objectives and the poor values (in comparison) for the others suggest that low frequency objectives are dominating the search for decoder coefficients.

To investigate this further an additional series of searches were undertaken to find the approximate range of the objectives. The ranges were recorded over 200 search runs. Table 2 displays the results with the three largest ranges highlighted.

Objective:	Min	Max	Range
E_{LFAng}	0.0000	557.0200	557.0200
E_{HFAng}	0.0538	529.7900	529.7362
$E_{AngMatch}$	0.0586	522.4900	522.4314
E_{LFMag}	0.4956	5000000.0000	4999999.5044
E_{HFMag}	0.5691	152.7200	152.7200
E_{LFVol}	0.0000	77266.0000	77266.0000
E_{HFVol}	0.0000	326.4800	326.4800

Table 2: Objective ranges

It is clear that the low frequency objectives have the largest range. This suggests that all previous searches were almost certainly biased towards solutions with better low frequency performance.

To resolve the problem of objective dominance previous work by others has tried applying ad hoc weightings to objectives and reported that resulting solutions gave better decoders [9]. However a systematic method of objective range removal would be far superior and lead to improved decoder coefficients that better meet all of the objectives.

3 RANGE REMOVAL AND IMPORTANCE

Objective range removal is not, in itself, a new concept. Bentley and Wakefield have addressed this generic issue in search problems [8]. The range-removal method used in this application domain comes from their work and is known as the sum of global ratios.

Each of the objective values is converted into a ratio by using the globally worst and best objective values encountered in all previous searches. This ensures that no single objective dominates the search because all values are constrained within the range of [0, 1]. Each objective can be formulated thus:

$$FitnessRatio_i = \frac{(objective\_value_i - \min(objective\_value))}{(\max(objective\_value) - \min(objective\_value))} \quad .....(16)$$

where $FitnessRatio_i$ is the ith range-removed objective and objective_value$_i$ is the value of the ith objective.

Once range removal has been implemented, the search can be systematically and logically biased towards specific criteria by placing more or less emphasis on selected objectives. This technique is referred to as importance and simply involves applying weightings to the range-removed objectives:

$$Fitness_i = FitnessRatio_i * w_i \quad (17)$$

where $Fitness_i$ is the ith importance weighted range-removed objective, $FitnessRatio_i$ is the ith range-removed objective, and w_i is the importance weighting for the ith range-removed objective.

4 METHODOLOGY

The range removal technique was incorporated into the multi-objective function. In order to derive the fitness ratios the minimum and maximum values of each objective were dynamically updated and saved to files during each search. Two-hundred solutions were derived with all fitness function objectives given equal importance in the search. The solutions are for decoders that do not use shelf filters and have ITU 5.1 surround speakers at 115°.

5 RESULTS

Table 3 presents the individual objective values for the four best solutions produced by the range removal search. For comparison, Table 4 gives the individual objective values for the four best solutions derived in previous work (i.e. without range removal and therefore producing solutions with low frequency objectives dominating the search). In both sets of results the minimum, maximum and mean values were taken from the top ten solutions generated by the search.

Objective:	MW1	MW2	MW3	MW4	Min	Max	Mean
E_{LFAng}	0.0415	0.0403	0.0462	0.0484	0.0393	0.0565	0.0488
E_{HFAng}	0.0153	0.0202	0.0216	0.0217	0.0153	0.0291	0.0250
$E_{AngMatch}$	0.0683	0.0755	0.0743	0.0735	0.0683	0.0807	0.0758
E_{LFMag}	3.0e-006	2.7e-006	2.8e-006	2.8e-006	2.3e-006	3.1e-006	2.7e-006
E_{HFMag}	0.4695	0.4604	0.4536	0.4522	0.4419	0.4695	0.4486
E_{LFVol}	2.5e-006	2.7e-006	2.4e-006	2.3e-006	2.0e-006	2.7e-006	2.3e-006
E_{HFVol}	0.0020	0.0020	0.0015	0.0016	8.5e-005	0.0036	0.0016
Total Fitness	0.5967	0.5967	0.5972	0.5975			

Table 3: Individual objective values from range-removed search (MW = Moore/Wakefield)

Objective:	MW 1	MW 2	MW 3	MW 4	Min	Max	Mean
E_{LFAng}	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004
E_{HFAng}	0.0324	0.0324	0.0324	0.0323	0.0321	0.0327	0.0324
$E_{AngMatch}$	0.1061	0.1060	0.1060	0.1060	0.1057	0.1064	0.1061
E_{LFMag}	1.6e-010	1.0e-010	4.7e-011	2.8e-011	4.2e-012	3.5e-010	1.1e-010
E_{HFMag}	0.5070	0.5072	0.5074	0.5075	0.5058	0.5087	0.5071
E_{LFVol}	1.1e-010	1.1e-010	6.7e-011	1.4e-010	1.3e-011	5.9e-010	1.5e-010
E_{HFVol}	0.1613	0.1619	0.1611	0.1622	0.1596	0.1640	0.1617
Total Fitness	0.8065	0.8072	0.8064	0.8076			

Table 4: Individual objective values from a previous non range-removed function

The new solutions are better than previously published solutions because they better meet all of the objectives simultaneously. This is confirmed by the total fitness values given by the new range-removed objective function for the new solutions and previous solutions. However, it should be noted that in order to produce better decoder coefficients across all objectives the search has compromised the objectives it was previously biased towards, namely the low frequency objectives (E_{LFAng}, E_{LFMag}, E_{LFVol}).

6 CASE STUDY

6.1 Introduction

This case study investigates the use of importance and aims to derive a decoder that maximises the mid to high frequency performance. The decoder is known as a "max r_E" decoder and requires that the energy vector magnitude is maximised for the whole 360° sound stage [10].

6.2 Methodology

Six groups of 200 searches were undertaken (i.e. 6 x 200). In the first group the importance of the energy vector magnitude (E_{HFMag}) was increased. For each following group of searches the importance weightings were adjusted according to the results achieved in the previous group of searches.

6.3 Results

Table 5 gives the importance weightings used for the six groups of searches. Table 6 presents the best solution derived for each of the six groups of searches.

Weightings:	E_{LFAng}	E_{HFAng}	$E_{AngMatch}$	E_{LFMag}	E_{HFMag}	E_{LFVol}	E_{HFVol}
1	1.0000	1.0000	1.0000	1.0000	5. 0000	1.0000	1.0000
2	1.0000	1.0000	1.0000	1.0000	3. 0000	1.0000	1.0000
3	1.0000	1.1000	1.3000	1.0000	2.5000	1.0000	1.1000
4	1.0000	1.2000	1.4000	1.0000	2.6000	1.0000	1.2000
5	1.0000	1.4000	1.4000	1.0000	2.6000	1.0000	1.2000
6	1.0000	2.2000	1.4000	1.0000	2.6000	1.0000	1.6000

Table 5: Importance weightings for 6 searches

Search:	E_{LFAng}	E_{HFAng}	$E_{AngMatch}$	E_{LFMag}	E_{HFMag}	E_{LFVol}	E_{HFVol}
Best non range-removed	0.0004	0.0324	0.1061	1.6e-010	0.5070	1.1e-010	0.1613
Best range-removed	0.0415	0.0153	0.0683	3.0e-006	0.4695	2.5e-006	0.0020
1	0.5697	0.7502	0.1757	9.5e-007	0.0446	1.5e-005	0.0225
2	0.0134	0.0860	0.1637	2.7e-006	0.4009	9.1e-007	0.0002
3	0.0607	0.0313	0.0781	2.7e-006	0.4309	1.9e-006	0.0009
4	0.0612	0.0311	0.0762	2.7e-006	0.4283	1.8e-006	0.0050
5	0.0604	0.0299	0.0756	2.7e-006	0.4293	1.8e-006	0.0052
6	0.0574	0.0267	0.0751	9.3e-007	0.4354	5.9e-007	0.0004

Table 6: Individual objective values for the best "max r_E" decoders from each search group (the best non range-removed and range-removed solutions included for comparison)

Initially, it was found that the importance weighting for the energy vector magnitude objective (E_{HFMag}) was too heavy in comparison with the other objectives. This caused the search to produce odd decoders with very good vector magnitudes (E_{LFMag}, E_{HFMag}) but very poor performance for velocity vector angle and energy vector angle (E_{LFAng}, E_{HFAng}) (see search 1 in Table 5 and 6). This is likely to have occurred due to the interdependency between the objectives. Examination of these decoders in use revealed that sound was limited to coming from a single loudspeaker. Carefully adjusting the importance of objectives in each following group of searches resulted in a decoder with favourable performance characteristics being derived (see search 6 in Table 5 and 6).

When compared to a "max r_E" decoder provided by Bruce Wiggins of the University of Derby (personal email communication), the "max r_E" decoder derived performs better overall and for four out of the seven objectives (i.e. $E_{AngMatch}$, E_{HFMag}, E_{LFVol}, E_{HFVol}) (see table 7). The energy vector performance (E_{HFMag}) has been increased without a significant loss to the other three objectives. The new decoder gives an average energy vector magnitude (r_E) of 0.7021 for the whole sound stage compared to 0.6800 for Wiggins' decoder.

Search results by:	E_{LFAng}	E_{HFAng}	$E_{AngMatch}$	E_{LFMag}	E_{HFMag}	E_{LFVol}	E_{HFVol}	Total fitness
Moore/Wakefield	0.0574	0.0587	0.1051	9.3e-007	1.1320	5.9e-007	0.0007	1.3539
Wiggins	0.0273	0.0587	0.1114	6.5e-007	1.2161	9.3e-006	0.0014	1.4148

Table 7: Comparison of "max r_E" solutions

The gain in performance for the energy vector magnitude (r_E) has been increased at the front and the rear of the system (see Fig.1) with a slight loss in performance between 45° and 90°.

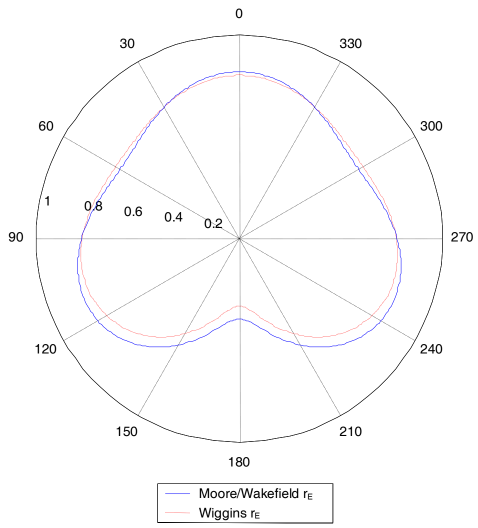

Figure 1: Energy vector magnitude (r_E) comparison by angle

Figure 2 displays the total error by angle. It can be seen that the new decoder performs better overall at the rear of the system.

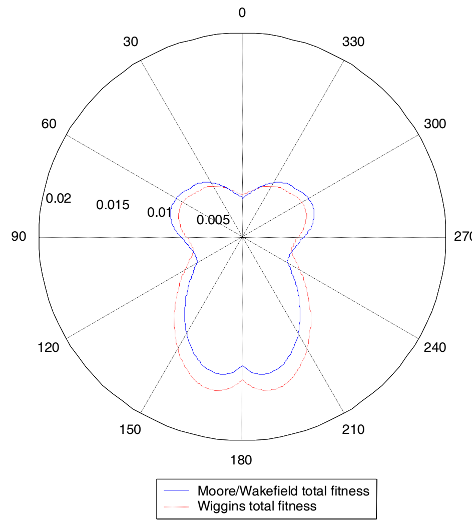

Figure 2: Total fitness error comparison by angle

The new methods have shown good potential for designing Ambisonic decoders with specific performance characteristics. A "max r_E" decoder has been derived for the ITU 5.1 layout with good energy vector magnitude performance after fine tuning the importance weightings.

7 CONCLUSIONS

New first-order Ambisonic decoder coefficients for ITU 5.1 have been derived that better meet all of the design objectives for a decoder by using the technique known as range removal.

Decoders biased towards specific design objectives can be derived by making use of the concept known as importance.

REFERENCES

[1] Gerzon, M.A. and G.J. Barton, *Ambisonic Decoders for HDTV*, in *92nd Audio Engineering Society Convention*. 1992, AES: Vienna, Austria.

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

[3] Moore, J.D. and J.P. Wakefield, *The Design and Detailed Analysis of First Order Ambisonic Decoders for the ITU layout*, in *The 122nd Audio Engineering Society Convention*. 2007, AES: Vienna, Austria.

[4] Gerzon, M.A., *General Metatheory of Auditory Localisation*, in *92nd Audio Engineering Society Convention*. 1992, AES: Vienna, Austria.

[5] Glover, F., *Tabu Search - Part 1*. ORSA Journal of Computing, 1989. **1**(3): p. 190-206.

[6] Glover, F., *Tabu Search- Part 2*. ORSA Journal of Computing, 1990. **2**(1): p. 4-32.

[7] Rayward-Smith, V.J., et al., *Modern Heuristic Search Methods*. 1996: John Wiley & Sons

[8] Bentley, P.J. and J.P. Wakefield, Finding Acceptable Solutions in the Pareto-Optimal Range using Multiobjective Genetic Algorithms, in Soft Computing in Engineering Design and Manufacturing, P.K. Chawdhry, R. Roy, and R.K. Pant, Editors. 1998, Springer Verlag London Limited. p. 231-240.

[9] Wiggins, B., *An Investigation into the real-time Manipulation and Control of Three-Dimensional Sound Fields*, in *Signal Processing Applications Research Group*. Ph.D. Thesis, 2004, University of Derby: Derby.

[10] Daniel, J., J.-B. Rault, and J.-D. Polack, *Ambisonics Encoding of Other Audio Formats for Multiple Listening Conditions*, in *105th Audio Engineering Society Convention*. 1998, AES: San Francisco, CA.