Moore, John David and Wakefield, Jonathan P.

An enhanced approach to surround sound decoder design

AN ENHANCED APPROACH TO SURROUND SOUND DECODER DESIGN

J.D.Moore, Dr J.P Wakefield
University of Huddersfield, Queensgate, Huddersfield HD1 3DH, UK

ABSTRACT

This paper highlights several key improvements that can be made to current surround sound decoder design techniques. The work focuses on the use of heuristic methods to derive surround sound decoder parameters for irregular loudspeaker layouts. A comparison is made between decoder parameters derived from the new improved algorithm and parameters published by others. The results show that the new parameters derived using this algorithm are superior to other parameters and the new range removal technique introduced in this paper warrants further investigation.

Keywords Ambisonics surround sound heuristic Tabu range-removal 5.1

1 INTRODUCTION

Simple pair-wise panning between speakers, as used in stereophony, does not perform well for surround sound play back (although this method is accepted and widely used) (Pulkki and Karjallainen 2001). Another technique called Ambisonics has numerous advantages over pair-wise panning but despite this it still seems to be only used by audio enthusiasts or in academia. Possibly the main reason this system did not catching on is because of the failure of the first commercial surround sound system 'Quadraphonics' released around the same time (Gerzon 1974).

Ambisonics is built around a mathematical theory of human directional psychoacoustics developed by Michael Gerzon (Gerzon 1992). The system is designed to take into account the fact that human hearing uses different mechanisms for sound localisation in different frequency ranges. This is one of the key advantages that Ambisonics holds over other techniques as it was designed with human perception in mind.

This work will look at using Gerzon's theory with the standard 5.1 surround sound layout specified by the International Telecommunications Union (ITU 1994) which is an irregular speaker layout. The research will look at this standard because it is widely commercially available.

Little work has focused on bringing Ambisonics to irregularly spaced layouts. This is due in part to the complexity of deriving decoder parameters for these layouts. Gerzon admitted himself that the non-linear equations were laborious and difficult to solve mathematically (Gerzon and Barton 1992).

Although similar work has been attempted previously by others (Gerzon and Barton 1992; Wiggins 2004), the initial investigation showed some of their work could be developed further.

Surround sound decoders can be designed by formulating the matter at hand as a search problem. A search algorithm is used to explore a search-space for good combinations of decoder parameters based on Gerzon's theory. This theory is embodied in a fitness function which is used to evaluate the quality of each parameter set. A good parameter set would produce a decoder with good performance for both low and high frequency playback.

Once derived the decoder parameters are used as individual weightings for the each of the Ambisonic encoded audio components known collectively as B-Format. Correctly balancing these components ensures an accurate reconstruction of a recorded or synthesized sound field at the listening point. Each speaker's output can be derived like so:

$$S_i = Wa_i + Xb_i + Yc_i \tag{1}$$

Where i is the speaker number, S is the speaker feed, W,X,Y are the B-Format components and a,b,c are the decoder parameters for each speaker. Thus for 5.1 three parameters are needed for each speaker (excluding the subwoofer). It should be noted that a mono audio signal can easily be converted into B-format for a given angle to the the listener.

Searching exhaustively for the best set of decoder parameters is not feasible. To complete an exhaustive search using a resolution of two decimal places for each parameter would take 2×10^{36} iterations of the search algorithm. This would take 10^{16} years to complete on a 1.8Ghz PC. Therefore

the search method chosen for this research is the Tabu search, a local search algorithm. By using a local search the optimum solution can never have been said to be found. However, the Tabu search has been proven to work well in various well-known combinatorial optimisation problems (Hedar and Fukushima 2006). The basic idea of the Tabu search technique is to avoid search cycling (visiting the same solution) 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.
An adapted form of the Tabu search has been used previously by Wiggins (Wiggins, Paterson-Stephens et al. 2003) for searching for decoder parameters. However, after studying the algorithm it was found that there are several problems inherent in his implementation:

- No Tabu list is kept.
- Fixed number of iterations of the search
- The fitness functions proposed do not fully capture the criteria specified in Gerzon's psychoacoustic theory.

Not including the Tabu list is a substantial fault. The Tabu list allows the search to escape from a 'local' minimum and is a fundamental feature of the Tabu search, the absence of this decreases the chances of converging on a 'better' solution. This is a minimum requirement for a Tabu search and is how the search got its name (Glover 1989). Without this, solutions already found can be returned to, resulting in cycling. The above three points are flaws which have been addressed in this research.

2 WORK COMPLETED

An MFC Windows application was developed to implement correctly the Tabu search algorithm. Key papers from the inventor of the search were referred to ensuring that the algorithm was correct (Glover 1990). The search is encouraged to escape from local minima by allowing moves to solutions that are worse than the current solution. With the addition of a Tabu list the search is guided away from previously visited areas in the search space. This is a step-by-step guide of the Tabu search algorithm used in this research:

1. Start at a random point in the search space.
2. Evaluate the start point and store as the current solution
3. If the current solution is better than the best solution so far, store it as the new best solution.
4. Add the current solution to the Tabu list; remove the oldest item on the Tabu list if it contains more than N items.
5. Move locally around the current solution to generate temporary solutions.
6. Rank the temporary solutions by fitness.
7. If the highest ranked temporary solution is the better than the current solution then jump to step 9.
8. Eliminate the temporary solutions which are 'Tabu'.
9. Select the highest ranked temporary solution that was not eliminated as the new current solution, unless all temporary solutions where eliminated.
10. If the termination criterion is not satisfied, repeat from step 3.

An example from a recent search run can be seen in Figure 1. A worse move was taken at iteration 15 which led to 'better' solutions being found. The fitness function developed in this research is a multi-objective function based on three important design aspects from Gerzon's theory. The first of these ensures the reproduced sound source angle matches the intended angle of a sound source. Secondly, the magnitude of the vector calculated from summing the speaker gain vectors must be as close to unity as possible (a real sound source gives a value of one). Finally, the perceived volume must be equal all the way around to give a natural reproduction. It was found in Wiggins' work that these requirements were not closely matched so the following improvements were made:

2.1 Simultaneous calculation of low and high frequency parameters

Ambisonic decoders have separate sets of parameters for low and high frequencies to take into account the different mechanisms used by humans for perceiving sound at low and high frequencies. Although the sets are different they should not be considered independent from one another. It is important to make sure both sets match up for reproduced source angle and volume. Wiggins evaluated the low and high frequency parameters separately and did not compare them which is insufficient to accurately encapsulate Gerzon's theory. In this work parameters for both frequency ranges are calculated at the same time to allow for this.

2.2 Volume

Wiggins compares the volume at every angle against the value at zero degrees. However, this does not necessarily find solutions where the difference between the volume at each angle is similar. In this work the volume at every angle is compared to the volume at all angles to ensure reproduced volume is equal. The new volume objective is given by

$$Volume\_Fitness = \sum_{i=0}^{n} \sum_{j=0}^{i-1} \left| Vol_i - Vol_j \right| \qquad (2)$$

Where n is the number of angles and Vol is the volume at the angle i or j.

2.3 Low and high frequency angle match

Wiggins compared the high frequency reproduced angle with the actual sound source angle and the low frequency reproduced angle with the actual sound source angle. However it is important that high and low frequency reproduced sound angles match when replayed over an Ambisonics system. Wiggins' fitness function does not try to minimise this. A further objective was added to ensure that high and low frequency angles match ensuring that the reproduced sound image is well defined and not smeared. The additional objective can be written as

$$Overall\_Angle\_Fitness = \sum_{i=0}^{n} \left| Ang_i^{LF} - Ang_i^{HF} \right| \qquad (3)$$

Where n is the number of angles, Ang^{LF} and Ang^{HF} are respective reproduced angles for the actual angle i

2.4 Optimised code

Wiggins' implementation of his fitness functions is quite inefficient. Given that the fitness function is called with every iteration of the search algorithm, this adds a significant amount of time to the search process. Several changes were made to the fitness function to make it more efficient. The fitness function was then redesigned to avoid computationally expensive operations.
A test of the speed of both algorithms was made. Both functions were coded in c++ and were run using the same Tabu search application. The results produced show the new fitness function runs much faster. Wiggins' fitness function takes approximately 14 seconds (just low frequency), whereas the new function takes approximately 9 seconds for 5,000 search iterations on a 1.8GHz PC. This is a significant increase in speed taking into account that the new fitness function calculates high and low frequency simultaneously and also has more objectives. This allows more potential solutions to be evaluated and could therefore lead to finding a better solution.

2.5 Removed objective range-dependence

Different objectives that make up the fitness function may have different ranges of values. This can potentially lead to certain objectives dominating the search. For example, a bad fitness value for one objective could be a good fitness value for another making the objective with the largest range dominate the search. It is therefore desirable to make each objective range independent. Wiggins had tried some ad hoc weighting of the search objectives to cope with certain objectives biasing the search. However, this is not ideal and a method is needed to address this.
In this work a technique has been implemented in the fitness function that was proposed by Bentley et al (Bentley and Wakefield 1997) to make each objective range independent. Removing the range gives equal importance to all objectives thus allowing them to be weighted in order of importance. For the first time this gives the user of the application accurate control over the relevant importance of each of Gerzon's criteria. In the search application the weightings for each of the search objectives can be adjusted by the user (see figure 2).

3 EXPERIMENTAL RESULTS

3.1 Results without range independence of objectives

Initial results show that there are a very large number of 'good' solutions. Already through using the new fitness function better results have been found than in previous work (see figure 3). Two of the solutions given in this figure are provided by Wiggins in his PhD thesis (Wiggins 2004). The other solution is given by the company Sound Field Ltd. Sound Field have developed many products based on Ambisonics technology. As can be seen a significant improvement has been made using the Moore/Wakefield parameters (it should be noted Michael Gerzon doesn't give figures for the ITU layout as his work looks at more regular layouts).

3.2 Results with range independence of objectives

Figure 4 shows fitness values for the magnitude objective with and without weighting. It can clearly be seen that when weighting the search in favour of this objective the solutions produced surpass many of the results given from an un-weighted search. The values also tend not to vary as much as the un-weighted objective suggesting that much lower values for magnitude are more difficult to obtain. However, weighting in favour of single objectives causes other objectives to give poorer results. Future research will look at the relative importance of each objective in producing a good decoder.

4 CONCLUSIONS

The decoder parameters derived using the modified algorithm show good potential for improving playback. Simultaneous calculation of low and high frequency parameters allows for the introduction of a new fitness objective for low and high frequency reproduced angle match. This new objective will ensure decoder solutions will match Gerzon's criteria for the Ambisonic system more closely.
The range-removal technique introduced in this paper warrants further investigation as it shows good potential for widening decoder design options.

5 FUTHER WORK

If a set of decoder parameters closely meet the requirements of the theory (as embodied in the fitness function) then it should be the case that the decoder will perform well in reality. However some listening tests will be performed to confirm the quality of the decoder solutions. A decoder audition application has been developed for this purpose. The software allows easy comparisons to be made between parameter sets (see figure 5).

REFERENCES

Bentley, P. J. and J. P. Wakefield (1997). Finding Acceptable Solutions in the Pareto-Optimal Range using Multiobjective Genetic Algorithms, in Chawdhy, P.K., Roy, R., Pant, R.K. (Eds),*Soft Computing in Engineering Design and Manufacturing*, Springer Verlag, London, 2nd Online World Conference on Soft Computing in Engineering Design and Manufacturing, pp.231-40.

Gerzon, M. (1974). "What's wrong with quadrophonics." Studio Sound **16**(5): 51-56.

Gerzon, M. A. (1992). General Metatheory of Auditory Localisation. 92nd Audio Engineering Society Convention. Vienna, AES.

Gerzon, M. A. and G. J. Barton (1992). Ambisonic Decoders for HDTV. 92nd Audio Engineering Society Convention, AES.

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

Glover, F. (1990). Tabu Search: A Tutorial. The Institute of Mangement Sciences, Interfaces, **20**(4):74-94

Hedar, A.-R. and M. Fukushima (2006). "Tabu Search directed by direct search methods for nonlinear global optimization." European Journal of Operational Research **170**(2): 329.

ITU (1994). Multichannel stereophonic sound system with and without accompanying picture, International Telecommunications Union, Geneva, Switzerland

Pulkki, V. and M. Karjallainen (2001). "Localisation of Amplitude Panned Virtual Sources 1: Stereophonic Panning." Journal of the Audio Engineering Society **49**(9): 739-752.

Strybel, T. Z., C. L. Manglias, et al. (1992). "Minimum Audible Movement Angle as a Function of the Azimuth and Elevation of the Source." Human Factors **34**: 267-275.

Wiggins, B. (2004). An Investigation into the real-time Manipulation and Control of Three-Dimentional Sound Fields. SPARG. Derby, University of Derby. **Doctor of Philosophy:** 370.

Wiggins, B., I. Paterson-Stephens, et al. (2003). The Design and Optimisation of Surround Sound Decoders Using Heuristic Methods. Conference of the UK Simulation Society 2003 Emmanuel College, Cambridge.

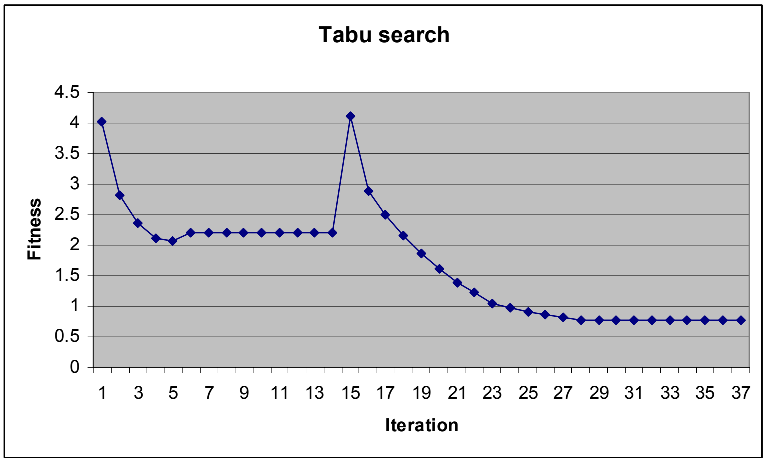

Figure 1: Tabu search moves

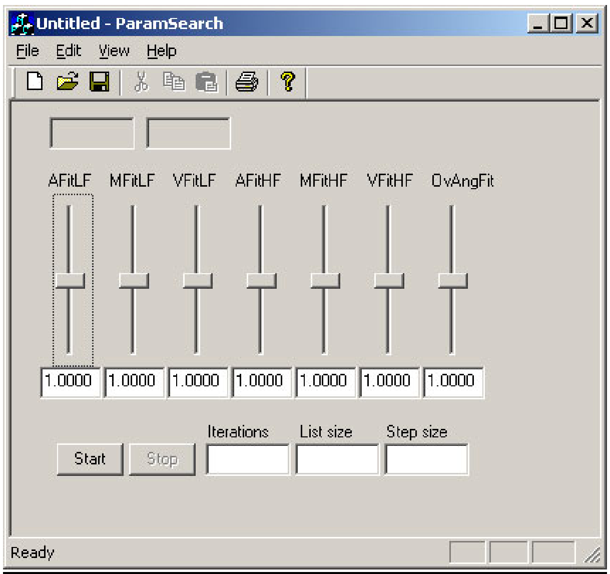

Figure 2: Tabu search application

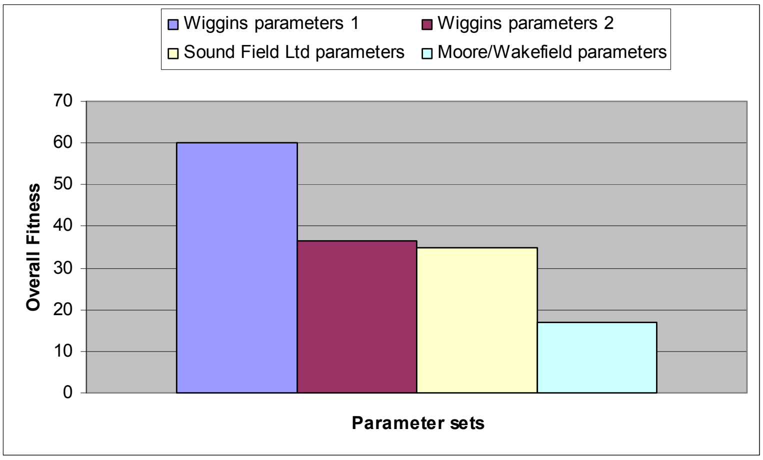

Figure 3: Overall fitness of decoder parameters (the lower the fitness value the better the result)

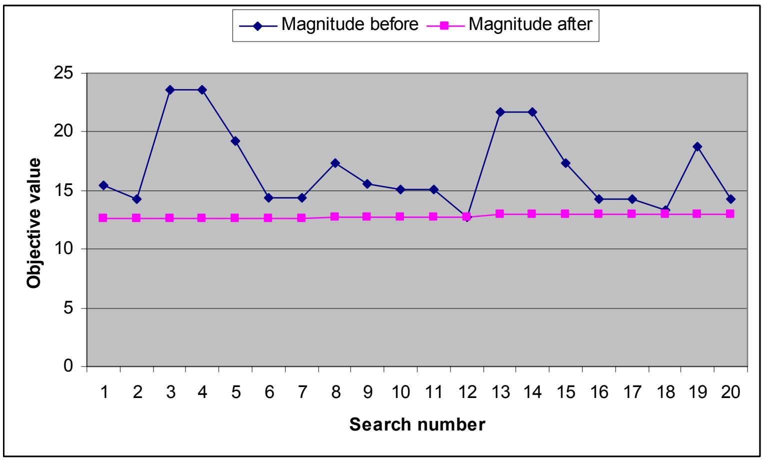

Figure 4: Tabu fitness for the magnitude objective

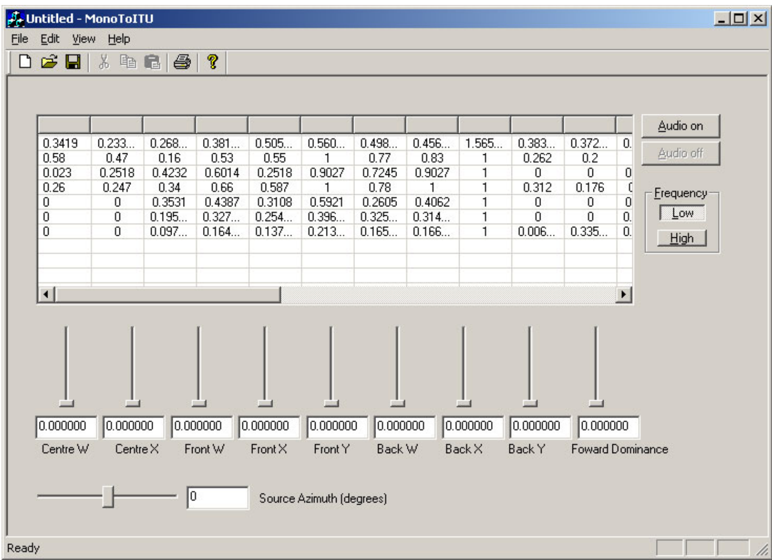

Figure 5: Decoder audition tool