THE DESIGN AND OPTIMISATION OF SURROUND SOUND DECODERS USING HEURISTIC METHODS.

BRUCE WIGGINS, I. PATERSON-STEPHENS, VAL LOWNDES & S. BERRY

University of Derby
Derby, United Kingdom
B.J.Wiggins@derby.ac.uk

Abstract: Surround sound has, for a number of years, had the standard of an irregular five-speakers layout (as defined by the ITU), but this is most likely set to expand to 7,9 or more, speaker configurations. The Ambisonic system, pioneered by Micheal Gerzon in the late 1960's, is very well suited to situations where the end system speaker configuration is not fixed in terms of number or position. However, while designing Ambisonic decoders for a regular (e.g. hexagonal) layout is well documented, optimising the decoders for irregular layouts is not a simple task, when optimisation requires the solution of a set of non linear simultaneous equations [1 – Gerzon & Barton]. This paper describes an alternative approach to the determination of these "optimised coefficients". This approach, based on a Tabu Search methodology [2 – Berry & Lowndes], efficiently determined sets of alternative optimal settings which were better (in terms of the reviewed parameters) than the results obtained using the standard analytical methods.

INTRODUCTION

Since the introduction of the DVD (both video and audio) surround sound has become an affordable luxury, and surround sound mixing and reproduction equipment is also in widespread use. The standard speaker configuration, as specified by the ITU, is a five-speaker layout, as shown in figure 1. However, this is likely to be expanded upon in the near future, and other, larger, venues are likely to have more speakers in order to adequately cover a larger listening area.

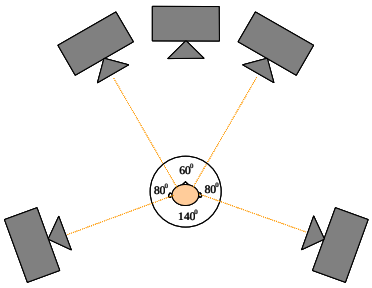

Figure 1 - Recommended loudspeaker layout, as specified by the ITU.

Due to the likelihood of ever changing reproduction layouts a more portable approach should be used in the creation of multi-channel material, and such a system has been available since the 1960s [3 - Borwick].

Ambisonic systems are based on a spherical decomposition of the sound field to a set order (typically 1st or 2nd order [4 – Malham, 5 – Leese). The main benefit of the Ambisonic system is that it is a hierarchical system, i.e., once the sound field is encoded in this way (into four channels for 1st order, and 9 channels for 2nd order) it is the decoder that decides how the sound field is reconstructed using the Ambisonic decoding equations [6 - Gerzon]. The Ambisonic system was largely researched and developed by Gerzon, and in 1992 papers were published proposing a method for the optimisation of Ambisonic decoders for irregular speaker arrays [1]. This was necessary because the original decoding equations were difficult to solve for irregular speaker arrays in the conventional way (i.e. using shelving filters).

IRREGULAR AMBISONIC DECODING

In order to quantify decoder designs Gerzon chose two main criteria for designing and evaluating multi-speaker surround sound systems in terms of their localisation performance. The two criteria represent the energy and velocity vector components of the sound field [7 - Gerzon]. The vector lengths represent a measure of the 'quality' of localisation, with the vector angle representing the direction that the sound is perceived to originate from. A vector length of one indicates a good localisation effect. These are evaluated using the equations shown in equation 1.

For regular speaker arrays, designing an optimised Ambisonics decoder is simply a case of using one virtual microphone response for low frequencies and a slightly different virtual microphone response for

the mid and high frequencies by the use of shelving filters [8 – Farina & Ugolotti] as shown in figures two and three.

As long as the virtual microphone patterns are the same for each speaker, the estimated localisation angle is always the same as the encoded source angle, with only the localisation quality (length of the vector) affected by changing the polar patterns.

$$P = \sum_{i=1}^{n} g_i \qquad E = \sum_{i=1}^{n} g_i^2$$

$$Vx = \sum_{i=0}^{n} g_i \cos(\theta_i)/P \qquad Ex = \sum_{i=0}^{n} g_i^2 \cos(\theta_i)/E$$

$$Vy = \sum_{i=0}^{n} g_i \sin(\theta_i)/P \qquad Ey = \sum_{i=0}^{n} g_i^2 \sin(\theta_i)/E$$

Where: g_i represents the gain of a speaker (assumed real for simplicity).
n is the number of speakers.
θ_i is the angular position of the i^{th} speaker.

Equ 1 – Velocity and Energy vector equations

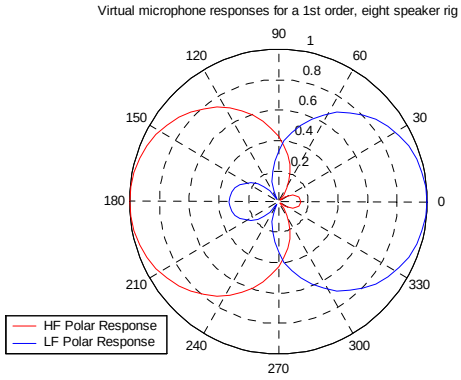

Figure 2 – Virtual micropone polar diagrams that satisfy Equ 1 for a 1st order, eight speaker rig.

When irregular speaker arrays are used, not only does the vector magnitudes need compensation, but the replay angle and overall volume of the decoded sound need to be taken into account, otherwise excessive decoding artefacts will be observed.

For example, consider the non-uniform speaker configuration of the ITU five speaker layout. If all speakers have the same polar pattern then a sound encoded to the front of a listener will be louder than a sound emanating from the rear. Also, the perceived direction of the reproduced sound will be distorted, as shown in figure 4.

These decoding artefacts are not a problem when the audio is produced for a fixed setup (for example, amplitude panned 5.1) since the material is mixed to sound correct on the chosen speaker layout. This is in contrast to a truly hierarchical system in which, ideally, it would be possible to reproduce the audio material accurately regardless of the configuration of the output system. Such a hierarchical system requires corrections to be made at the decoding stage where the speaker layout is known.

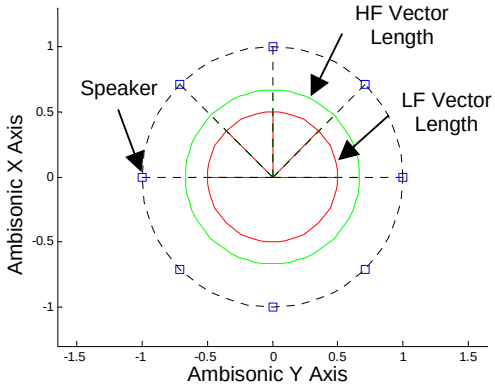

Figure 3a – Velocity and energy localisation vectors. Magnitude plotted over 360^0 and angle plotted at five discrete values. Inner circle represents energy vector, outer circle represents velocity vector. Using virtual cardioids.

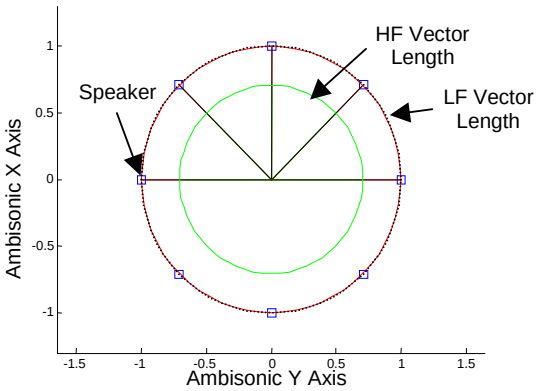

Figure 3b – Velocity and energy localisation vectors. Magnitude plotted over 360^0 and angle plotted at five discrete values. Inner circle represents energy vector, outer circle represents velocity vector. Using virtual patterns from figure 2.

Due to the added complexity of the speaker arrays response to an Ambisonic-type decode (see the

reproduction angle discrepancies and vector lengths in figure 4), Gerzon and Barton [1] proposed that two separate decoders be used, one for low frequency (<~700Hz) and another for high frequencies (>~700 Hz).

This can be achieved using a simple cross-over network (preferably using linear phase, FIR, filters) feeding low and high passed versions of the Ambisonic, b-format, signal to the two decoders where totally separate decoding can be achieved, not just a microphone polar pattern adjustment as in a regular speaker array decode.

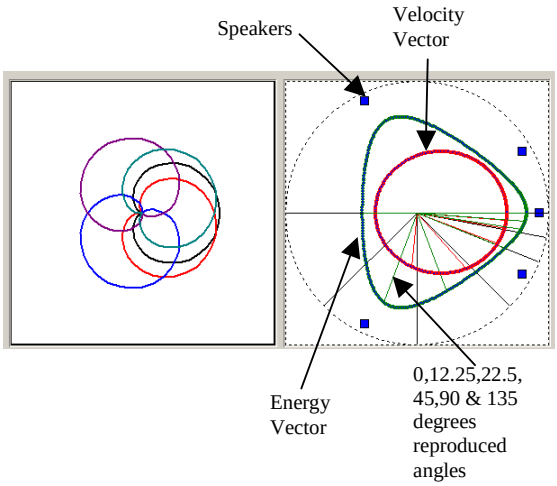

Figure 4 – Energy and velocity vector response of an ITU 5-speaker system, using virtual cardioids.

DECODER SYSTEM

1st order Ambisonics is based on four different signals, as shown in figure 5, an omni-directional pressure signal (W), a front-back figure of eight (X), a left-right figure of eight (Y), and an up-down figure of eight (Z).

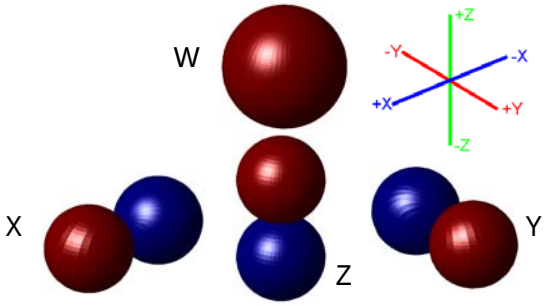

Figure 5 – Polar patterns of the four B-format signals used in 1st order Ambisonics. Red shows an in-phase response, and blue shows an out-of-phase response.

The 5-speaker system shown in figure 1 is a horizontal only system, and hence, only three of the four available b-format signals are required at the input of the decoder (W, X and Y). Also, the speaker array in figure 1 is left/right symmetric such that the decoder coefficients are arranged to work in mid and side pairs (i.e. sum and difference). The Ambisonic encoding equations are given in equation 2.

The incorporation of a 'frontal dominace' control in the decoding system can also be considered, the definition for this is given in equation 3. Although this form of the frontal dominance equation exhibits a non-linear response to the dominance parameter, it is used in this investigation to keep compatibility with Gerzon's previous paper on this subject [1].

$$W = 1/\sqrt{2}$$
$$X = \cos(\theta)$$
$$Y = \sin(\theta)$$

where θ is the encoded angle, taken anti-clockwise from straight ahead.

Equ 2. – Ambisonic Encoding coefficients.

$$W' = 0.5\left(\lambda + \lambda^{-1}\right)W + 8^{-\frac{1}{2}}\left(\lambda - \lambda^{-1}\right)X$$
$$X' = 0.5\left(\lambda + \lambda^{-1}\right)X + 2^{-\frac{1}{2}}\left(\lambda - \lambda^{-1}\right)W$$
$$Y' = Y$$

where λ is the forward dominance parameter (2>λ>1 for front, and 1>λ>0 for rear dominance).

Equ 3. – Forward dominance equation.

The frontal dominance terms are then substituted into the decoding equations to give a numerical value for each speaker output. Equation 4 shows the substitutions used for a 5 channel system.

$$C_F = (kW_C \times W') + (kX_C \times X')$$
$$L_F = (kW_F \times W') + (kX_F \times X') + (kY_F \times Y')$$
$$R_F = (kW_F \times W') + (kX_F \times X') - (kY_F \times Y')$$
$$L_B = (kW_B \times W') + (kX_B \times X') + (kY_B \times Y')$$
$$R_B = (kW_B \times W') + (kX_B \times X') - (kY_B \times Y')$$

where k denotes a decoding coefficient.

Equ 4. – Decoding Equations for each of the five speakers.

The λ and 'k' values are chosen so as to optimise the decoded output, with λ having possible values

between 0 and 2, and 'k' values having a nominal range between 0 and 1.

Equation 5 shows the conditions which are used to assess the performance of a given solution. The conditions that must be met are:

- Radius of the localisation vector lengths (R_V and R_E) should be as close to 1 as possible for all values of θ.
- $\theta = \theta_V = \theta_E$ for all values of θ (where θ is the encoded source angle).
- $P_V = P_E$ and must be constant for all values of θ.

$$Vx = \sum_{i=1}^{N} g_i \times \cos(SPos_i) / P_V$$

$$Vy = \sum_{i=1}^{N} g_i \times \sin(SPos_i) / P_V$$

$$Ex = \sum_{i=1}^{N} g_i^2 \times \cos(SPos_i) / P_E$$

$$Ey = \sum_{i=1}^{N} g_i^2 \times \sin(SPos_i) / P_E$$

$$R_E = \sqrt{E_x^2 + E_y^2} \qquad \theta_E = \tan^{-1}\left(E_y / E_x\right)$$

$$R_V = \sqrt{V_x^2 + V_y^2} \qquad \theta_V = \tan^{-1}\left(V_y / V_x\right)$$

$$P_V = \sum_{i=1}^{n} g_i \qquad P_E = \sum_{i=1}^{n} g_i^2$$

where:

g_i = Gain of the i^{th} speaker

$SPos_i$ = Angular position of the I^{th} speaker.

Equ 5. – Equations used to measure the performance of a decoder design.

In practice, the conditions defined in equation 5 are difficult to solve because the best result must be found over the whole 360^0 of encoded source positions. It is known that these equations are laborious to solve for five speaker systems [1]. Furthermore, an increase in the number of speakers will result in a disproportionate increase in the complexity of the decoding optimisation problem. Also, more than one valid solution for each decoder design exists at low and high frequencies. This means that a group of solutions need to be found, followed by subjective listening tests in order to find the best performing coefficient set.

Due to the laborious and time-consuming nature of decoder optimisation, a method is needed that can automate this process, so that the onus in designing Ambisonic decoders is shifted from the calculating of the decoding coefficients to listening to the different decoders so the optimal system can be decided upon.

THE HEURISTIC SEARCH METHODS

The word heuristic can be used to mean 'using trial and error', and mathematical searches using this technique can lead to the solutions of complex numerical problems. Heuristic search methods work on the simple principle that any result that is found (using, say, random starting values) can be tested on its 'correctness', with this then being used to decide on values to try next following some rule depending on the type of search method used.

As a result of the fact that each parameter has a value from a well defined range, 0 to 1 or 0 to 2, a search method seemed to be a very viable solution. However, if we wish to determine the settings to two decimal places there are 2×10^{18} possible solutions (given that there are 9 search parameters) and an exhaustive search is not feasible. The first avenue of research taken was that of a Genetic Algorithm approach. However, Genetic Algorithms are well suited to problems that have large search spaces (i.e. large parameter ranges), and this is not the case here. Also, a Genetic Algorithm approach is very good at getting reasonably close to an accurate solution, but will then need optimising further [2]. This was seen to be overly complicated for our needs, and a method based on the Tabu Search (memory based search) was developed, which is a method that can achieve accurate results, and is a viable option as long as the parameters that are to be altered have defined limits [2].

This, slightly adapted, form of a Tabu search works by having the decoder coefficients initialised at random values (or values of a previous decoder, if these values are to be optimised further). Then the Tabu search program changes each of the tweakable values in turn, plus or minus the step size. The result that is deemed to be most correct is then kept and the parameter changed is then restricted to only move in the successful direction for a set number of iterations (which, of course, will only happen if this parameter, again, is the best one to move). It must be noted that the random start position is of great importance, as it helps in the search for a wide range of solutions as, if the Tabu search starts in exactly the same place each time, exactly the same results will be found (as there is no randomness in the search process itself, unlike a Genetic Algorithm). The most important part of the Tabu search algorithm is the equations used to measure the fitness (or correctness) of the coefficients used, as it is this one figure that will determine the course that the Tabu search takes. As

mentioned above, three parameters must be used in an equation that represents the overall fitness of the decoder coefficients presented. These are:

- Localisation measure (vector lengths, R_V & R_E).
- Localisation Angle (vector angles, θ_V & θ_E).
- Volume (Sound pressure gain, P_V & energy gain, P_E) of each encoded direction.

As each of these results must be as good a fit as possible for the whole 360^0 sound stage, the three parameters must be evaluated for a number of different encoded source positions. Gerzon evaluated these parameters at 14 points around the unit circle (7 around a semi-circle assuming left/right symmetry), but as computers can calculate these results extremely quickly, it was decided that encoded sources at 4^0 intervals would be used (90 points around the unit circle). Due to the large number of results for each of the fitness values an average was taken for each fitness parameter using a route mean square approach. If we take the example of the fitness of the vector lengths (localisation quality parameter), then if a mean average is taken, then a less than one vector length in one part of the circle could be compensated for by a greater than one vector length else where. However, if we take a good fit to be zero, and use a route mean square approach then a non-perfect fit around the circle will always give a positive error value, meaning it is a true measure of the fitness. The equations used for each of the fitness parameters are shown in equation 6.

$$VFit = \sqrt{\sum_{i=0}^{n} \frac{\left(1 - P_0 / P_i\right)^2}{n}}$$

$$MFit = \sqrt{\sum_{i=0}^{n} \frac{\left(1 - R_i\right)^2}{n}}$$

$$AFit = \sqrt{\sum_{i=0}^{n} \frac{\left(\theta_i^{Enc} - \theta_i\right)^2}{n}}$$

where:
P_0 is the pressure at an encoded direction of 0^0.

n is the number of points taken around the unit circle.

θ_{Enc} is the encoded source angle and θ is the localisation angle.

Equ 6. – Equations of fitness used to evaluate the decoder coefficients.

Given the three measures of fitness in equation 6, the overall fitness for the high and low frequency versions of the decoder are actually calculated slightly differently. The low frequency decoder can achieve a near perfect fit, but the best fit that the high frequency decoder can expect to achieve is shown in figure 6. The best results were obtained from the Tabu search algorithm if the overall fitness was weighted more towards the angle fitness (*AFit* from Equ 6.) as shown in equation 7.

$$LFFitness = AFit + MFit + VFit$$

$$HFFitness = AFit + \left(MFit + VFit\right)/2$$

Equ 7. – Fitness equations for low and high frequency models.

A block diagram of the Tabu search algorithm described in this paper is shown in figure 6.

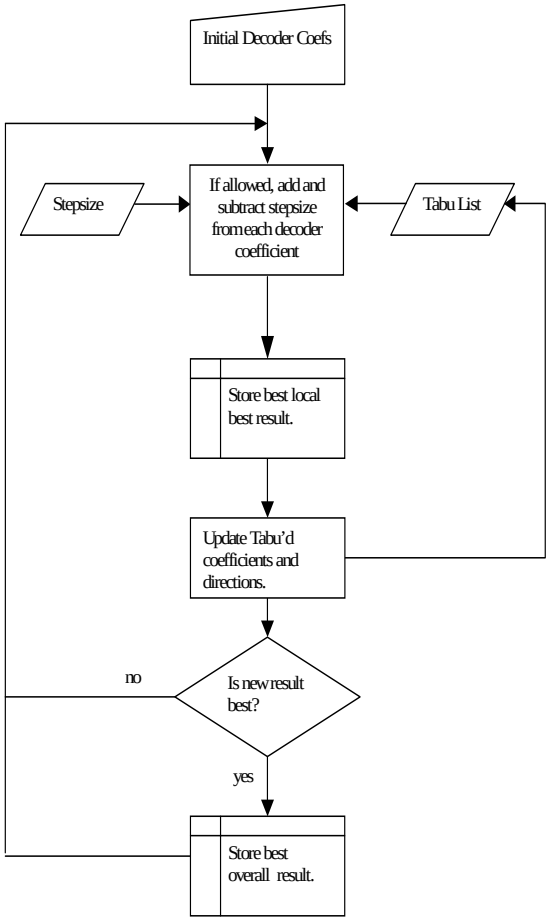

Figure 6 – Block diagram of a simple Tabu Search application.

The main benefit of the Tabu search method is that all three of the conditions to be met can be optimised for simultaneously, which had not been accomplished in Gerzon's paper [1]. For example if we take the speaker layout used in the Vienna paper, which isn't the ITU standard but is very similar, then the coefficients derived by Gerzon and Barton would give an energy and velocity vector response as shown in figure 7. Several observations can be made from this figure. There is a high/low localisation angle

mismatch due the forward dominance being applied to the high frequency decoders input *after* the localisation parameters were used to calculate the values of the coefficients. Or, if the frontal dominance is applied to both the high and low frequency decoders, a perceived volume mis-match occurs with the low frequency decoder replaying sounds that are louder in the frontal hemisphere than in the rear. Also, even if these mismatches were not present every set of results presented in the Vienna paper showed a distortion of the decoders reproduced angles. Figure 8 shows a set of coefficients calculated using the Tabu search algorithm described in figure 6 and shows that if all three criteria are optimised simultaneously a decoder can be designed that has no angle or volume mis-matches, and should reproduce a recording more faithfully than previously possible.

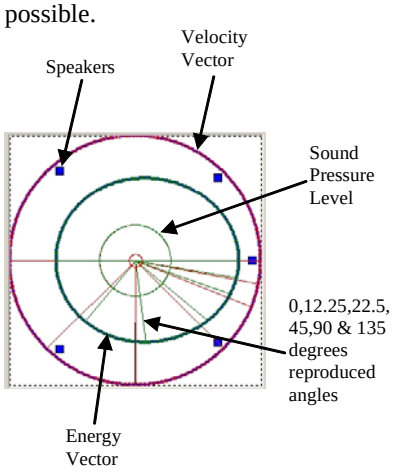

Figure 7 – Graphical plot of the Gerzon/Barton coefficients published in the Vienna paper. Encoded/decoded directions angles shown are 0^0, 12.25^0, 22.5^0, 45^0, 90^0, 135^0 and 180^0.

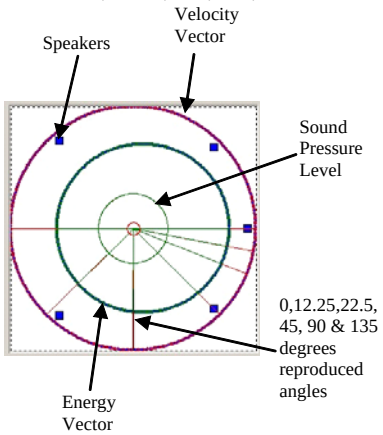

Figure 8 – Graphical plot of the coefficients generated using a tabu search algorithm. Encoded/decoded directions angles shown are 0^0, 12.25^0, 22.5^0, 45^0, 90^0, 135^0 and 180^0.

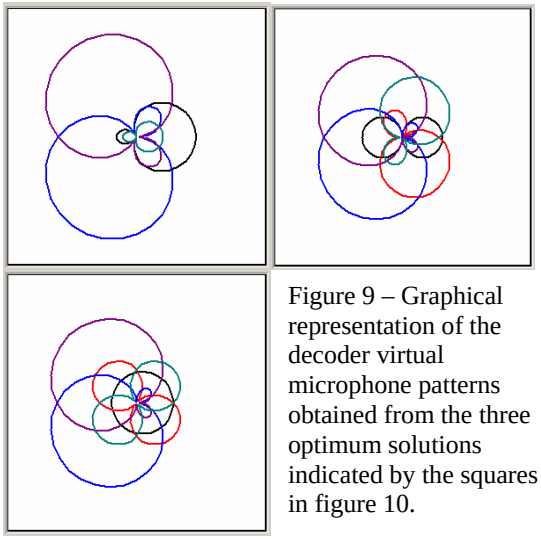

Figure 9 – Graphical representation of the decoder virtual microphone patterns obtained from the three optimum solutions indicated by the squares in figure 10.

CONCLUSIONS

The Tabu search algorithm has provided an efficient and effective methodology to optimise surround sound decoders. This methodology providing an improvement over the alternative approach [1], allowing for the Vienna equations [1] to be easily solved for virtually any arrangement of speakers and thus simplifying the design process for Ambisonic decoders. Although the software used to generate the results presented here concentrates on a typical five speaker, horizontal arrangement the methodology is applicable to any configuration. This approach has the advantage of generating multiple sets of good solutions (alternative decoders) in a single execution of the Tabu search program, the existing method generates a single solution, thus greatly increasing the number decoders that can be realised and tested in a very short time.

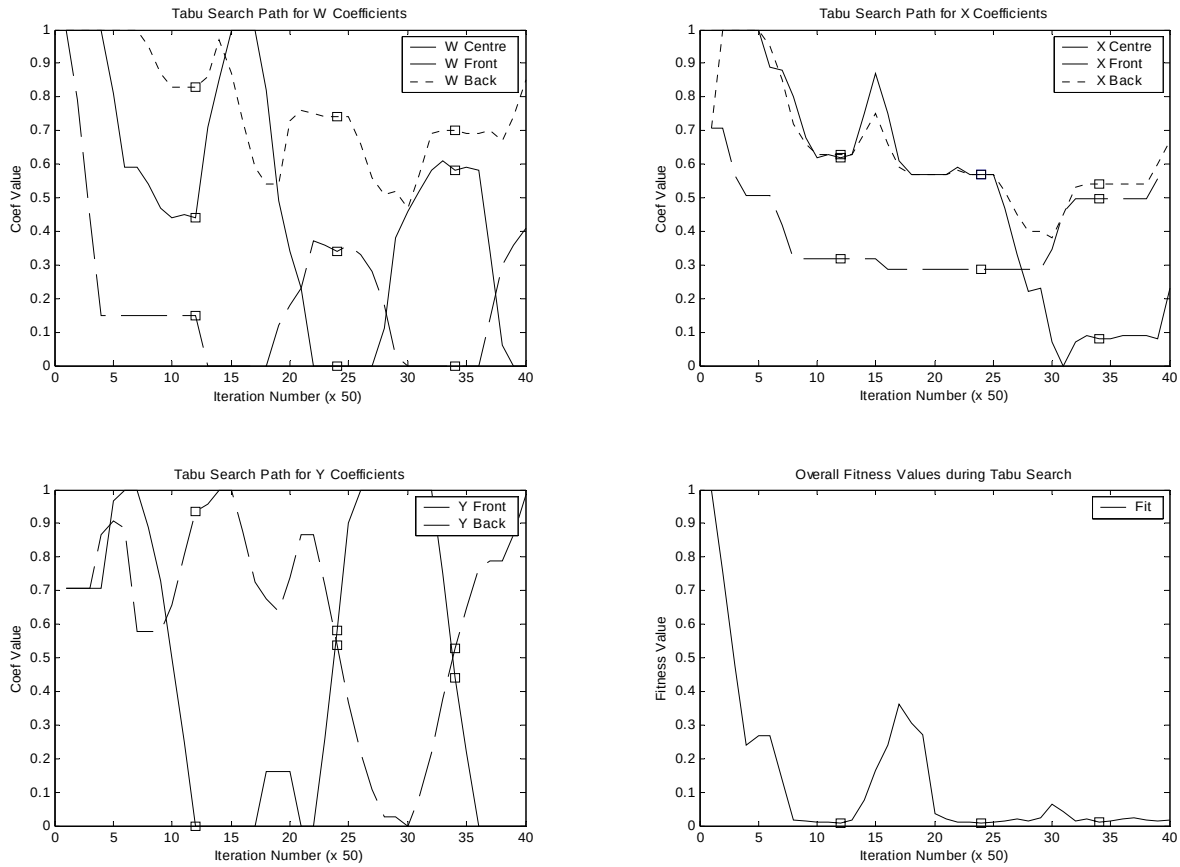

Figure 10 – A graph showing the transition of the eight coefficients in a typical low frequency Tabu search run (2000 iterations). The square markers indicate the three most accurate sets of decoder coefficients (low fitness).

REFERENCES

1. Gerzon, M. A. & Barton, G. J. 1992, "Ambisonic Decoders for HDTV" – *92nd AES Convention*, Vienna. Preprint 3345.
2. Berry, S. & Lowndes V. 2001, "Deriving a Memetic Algorithm to Solve Heat Flow Problems" – *University of Derby Technical Report*.
3. Borwick, J. 1981, "Could 'Surround Sound' Bounce Back" – *Gramophone*, February.
4. Malham, D., "Second and Third Order Ambisonics." http://www.york.ac.uk/inst/mustech/3d_audio/secondor.html.
5. Leese, M., "Ambisonic Surround Sound." - http://members.tripod.com/martin_leese/Ambisonic/
6. Gerzon, M. A., 1977 "Multi-system Ambisonic Decoder, parts 1 & 2" – *Wireless World* July & August 1977.
7. Gerzon, M. A. 1992, "General Methatheory of Auditory Localisation" – *92nd AES Convention*, Vienna. Preprint 3306.
8. Farina, A. & Ugolotti 1998, E., "Software Implementation of B-Format Encoding and Decoding" – *104th AES Convention*, Amsterdam. Preprint 4691.
9. Gardner B., Martin K., "HRTF Measurements of a KEMAR Dummy-Head Microphone", 1994. http://sound.media.mit.edu/KEMAR.html
10. Wiggins, B., Paterson-Stephens, I. & Schillebeeckx, P. 2001, "The analysis of multi-channel sound reproduction algorithms using HRTF data" – *19th AES Surround Sound Conention*, Schoss Elmau.
11. Farrar, K. 1979, "Soundfield Microphone. Parts 1 & 2" – *Wireless World*, October & November 1979.
12. Soundfield - www.soundfield.com

BIOGRAPHY

Bruce Wiggins – Lecturer. Bruce began studying at the University of Derby in 1996, and gained an honours degree in Music Technology and Audio System Design. Deciding to stay in education Bruce is a founding member of the Signal Processing Applications Research Group at Derby University (http://sparg.derby.ac.uk), and specialises in signal processing applications in the field of three-dimensional audio, which is the basis for his on-going PhD thesis and a collaboration between the University of Derby and the microphone manufacturing company, Soundfield, via a Teaching Company Scheme. Bruce has now joined the University of Derby as a lecturer in Digital Electronic systems teaching at both undergraduate and post graduate levels.

Iain Paterson-Stephens – Senior Lecturer. After obtaining his BEng(Hons) Electrical and Electronic Engineering from Nottingham Trent University in 1988, Iain worked as an R+D engineer for AEG Telefunken's specialist technologies department in Essen, Germany. At AEG, Iain worked on the design of a DSP based automatic test and evaluation system for aircraft turbines and power supply supervisory equipment.

In 1989 he joined the British Broadcasting Corporation and worked on a wide range of digital audio projects at Broadcasting House and the Maida Vale recording studios in London.

In 1993, Iain joined the Magnetic Resonance Centre at the University of Nottingham. The Centre is primarily involved in the development of high speed medical imaging systems and associated techniques.

In 1996 Iain joined the University of Derby as a Lecturer in Digital Signal Processing. In addition to lecturing, authoring and research activities, Iain is the Programme Leader for the MSc Digital Audio Systems programme at the University of Derby. Iain is also the Director of the Signal Processing Applications Research Group, SPARG.

Val Lowndes – Senior Lecturer University Principal Tutor. BSc Mathematics and MSc Mathematics for Computing, and University Principal Tutor. Interests in the use of Heuristic Methodologies in the solution of practical problems and in the use of Fuzzy Logic in decision making. Heuristic Methodologies: derived solution methodologies based around Genetic Algorithms for use in a range of problems, based around Tabu Search and Simulated Annealing in each case producing efficient and effective solution methodologies by combining elements of each approach. Fuzzy Logic: investigated the use of fuzzy logic in job and flow shop scheduling.

Stuart Berry – Lecturer in Mathematics. Interests in the use of Heuristic Methodologies in the solution of practical problems. Previously derived solution methodologies based around Genetic Algorithms, Tabu Search and Simulated Annealing in each case producing efficient and effective solution methodologies by combining elements of each approach.