Audio Engineering Society

Convention Paper

Presented at the 129th Convention
2010 November 4–7 San Francisco, CA, USA

Design of Ambisonic Decoders for Irregular Arrays of Loudspeakers by Non-Linear Optimization

Aaron J. Heller[1], Eric Benjamin[2], and Richard Lee[3]

[1] Artificial Intelligence Center, SRI International, Menlo Park, CA 94025, US
heller@ai.sri.com

[2] Surround Research, Pacifica, CA, 94044, US
ebenj@pacbell.net

[3] Pandit Littoral, Cooktown, Queensland 4895, AU
ricardo@justnet.com.au

ABSTRACT

In previous papers, the present authors described techniques for design, implementation, and evaluation of Ambisonic decoders for regular loudspeaker arrays. However, irregular arrays are often required to accommodate domestic listening rooms. Because the figures of merit used to predict decoder performance are non-linear functions of speaker positions, non-linear optimization techniques are needed. In this paper we discuss the implementation of an open-source application, based on the NLopt non-linear optimization software library, that derives decoders for arbitrary arrays of loudspeakers, as well as providing a prediction of their performance using psychoacoustic criteria, such as Gerzon's velocity and energy localization vectors. We describe the implementation and optimization criteria, and report on informal listening tests comparing the decoders produced.

1. INTRODUCTION

The main goal of Ambisonic sound is to reproduce localization cues that approximate those experienced in natural hearing, while utilizing a modest number of transmission channels and loudspeakers. The signals carried by the transmission channels define "what it should sound like" and it is the job of the decoder to produce a set of loudspeaker signals that reproduce those cues as accurately as possible over the listening area.

The ability to generate decoders for any given ad hoc array of loudspeakers is needed to accommodate typical domestic listening rooms. Although the ITU-recommended five-loudspeaker configuration is a commonly referenced system, its actual use is relatively uncommon. Real systems may be symmetrical, but not have the ITU shape, or may be completely asymmetrical. If the angular coverage is non-uniform, the derived

decoder may be theoretically suitable but in practice require that the loudspeaker emit unrealistically intense sounds. Likewise, the localization cues may be reproduced correctly at the center of the array (the "sweet-spot") but vary so rapidly with displacement from the center that the localization experienced by the listener is unstable with respect to head movements.

There are numerous numerical techniques for optimizing non-linear objective functions; in particular, modified tabu search [31][24][25], neural networks [30], and genetic search [2] have been applied to the problem of designing Ambisonic decoders.

We have employed the NLopt library for non-linear optimization [20] and other free/open-source software packages to produce an open-source application for the generation of practical Ambisonic decoders for end-users. NLopt provides a common application programmer interface (API) to a number of non-linear optimization techniques, allowing a common framework to be developed to support rapid-turnaround experiments.

In the present work we limit ourselves to first-order decoders because the vast majority of Ambisonic recordings are first order. Furthermore, the listening facilities available to us have horizontal arrays, so our listening tests were limited to those arrays. The present work has been limited to ITU-like arrays initially, because they are a configuration that others have worked on and therefore provide a good benchmark of our approach. However, there is nothing that limits the techniques presented to a particular Ambisonic order or speaker array.[1]

Differences with work in this area published previously include:

- The software is being released as an open-source project[2].
- The system is not limited to ITU 5 loudspeaker arrays. In particular there are no assumptions about left/right symmetry. It will operate with arbitrary arrays.
- The user can impose constraints on the ranges of the individual parameters.

In addition, we present two new results:

- A decoder for the ITU 5 array that makes full use of the center loudspeaker. In our listening tests it was preferred over decoders that did not use the center speaker.[3]
- A decoder optimized for a left/right asymmetric five-speaker array where the front speakers and one of the surround speakers are placed according to the ITU recommendations and the remaining surround loudspeaker is significantly displaced. In our listening tests this was preferred over misapplied decoders.

Because an Ambisonic recording is a definition of "what it should sound like" rather than "what comes out of a speaker," the promise it offers is that results for (at least) the central listener should be independent of the speaker layout. This is subject to certain constraints. One logical expectation might be that the sound is "better" or more "accurate" in a direction with more speakers. Previous decoder designs for irregular arrays have not lived up to this promise. We feel the techniques in this paper, especially what we are calling "Vienna-like decodes" are an important step in achieving this promise.

At the time of this writing it is not known what the precise perceptual tradeoffs are between maximizing the various parameters associated with a given decoder design. As a separate project, the present authors have created a methodology for assessing the effective reproduction of various factors such as ITD and ILD cues and the perception of envelopment.

We do not claim that the present work represents the ultimate, or even best available, solution, but we do hope that by making our tools available on an open-source basis, it will foster further work in this area, as well as provide a useful tool for listeners.

[1] We recognize that as more parameters are added, the more slowly the system will converge and the more likely it will settle in a local minimum rather than the global minimum.
[2] Please email the authors for access information.

[3] Wiggins, Moore, and others have published ITU B-format decoders where the center speaker is effectively shut off. Gerzon's "Vienna decoders" use the center speaker, but he does not show a decoder for the ITU array. Hence we refer to the current five-speaker decoder as "Vienna-like".

2. BACKGROUND

2.1. Classic Decoders

It is likely that, until the last five years, the majority of good experiences with Ambisonic playback have been with the hardware designs by Dr. Geoffrey Barton of the original Ambisonic team, e.g., [23][19]. When we attempted to duplicate these experiences with modern software decoders, we found many would not have been deemed Ambisonic by the original team and did not give as good results as the 'Classic' hardware designs done in accordance with Gerzon's theories. [17]

Our first paper "Localization in Ambisonic Systems" [4] confirmed via controlled listening tests, the importance of each aspect of "Classic" Ambisonic decoder design: a decoding matrix matched to the geometry of the loudspeaker array in use, phase-matched shelf filters, and near-field compensation.[4]

"Ambisonic Localisation – Part 2" [22] compared successful decoders from the first paper with an ITU-R decoder by Wiggins [31]. We investigated how robust decoders were to movement away from the central 'sweet spot' and when used with the wrong speaker layout.

"Is My Decoder Ambisonic" [18] put into the public domain what we had learned about the design of Classic decoders. It provides decoder writers with the necessary knowledge and tools to write decoders with performance equal to the original Classic decoders. These set the standard for first-order Ambisonic playback and are still appropriate for many domestic situations.

2.2. Decoders for Irregular Arrays

It is relatively straightforward to calculate a decoder for loudspeaker arrays for which it can be proved that the velocity and energy localization vectors are parallel for all source angles (i.e., regular polygons or diametric opposites). The decoder that optimizes the low-frequency performance ($\mathbf{r_V}$) is calculated by taking the pseudoinverse of the projections of the loudspeaker directions. The high-frequency performance ($\mathbf{r_E}$) is then optimized by adjusting the pressure-to-velocity ratio of the low-frequency decoder using phase-matched shelf filters. [18]

Decoder design for general irregular arrays, such as ITU-R, is more difficult because the energy localization vector is not guaranteed to point in the same direction as the velocity localization vector, and in general will not point in the same direction as the velocity vector, and is therefore not a linear function of the speaker locations.

The idea of using separate decoders for the different frequency regimes (below and above 400 Hz) was a major advance leading to the so-called "Vienna Decoders." [16] These are also known generically as "dual-band decoders." Gerzon and Barton outline an analytic optimization technique (characterized as "very tedious and messy") and show results for some five-speaker arrays, but do not give enough details to generalize these to other arrays. Furthermore, Wiggins has pointed out that these solutions have flawed localization. [33]

More recent work in this area has made use of numerical optimization techniques. While steady progress on the design of decoders for the ITU 5-channel loudspeaker arrays has been accomplished by Gerzon and Barton [16], Craven [7], Wiggins [31], Moore and Wakefield [25], and others [30] [2], significant work remains to be done in this area, in particular for playback of first-order B-format recordings.

2.3. Ambisonics at Home

We share with the original Ambisonic team, an emphasis on domestic reproduction. While Ambisonic techniques including the methods presented here are useful in studios or large theatres, our main interest is using it to play music at home. But what speaker arrangements are found or are possible at home?

We assume the listener wants to have good sound and will attempt to follow advice if it doesn't drastically disrupt his domestic arrangements. The listener is unlikely to convert his listening room into an anechoic chamber enclosing a regular dodecahedral arrangement of speakers with his armchair suspended in mid air, but would certainly consider a 1970's "Quadraphonic" layout with four speakers, perhaps in the corners of the room.[5] This is still a sensible domestic arrangement for surround sound.

[4] Near Field Compensation (NFC) is called "Distance Compensation" in Gerzon's papers. We prefer the term, NFC [9] because "distance compensation" is easily confused with delay and 1/*r* amplitude compensation. [11] NFC (at 1st order) corrects the curved wave front from a nearby source. [5] [18]

[5] The main problem with this layout is hardly any modern speakers are designed to give good performance when placed in corners. With the advent of digital speaker and room EQ,

Because most rooms are rectangular, the rectangle decoder (and its 3-D extension, the bi-rectangle) is the single most important case for actual applications in reproduction of first-order Ambisonic recordings. Rectangular speaker layouts are handled well by the 'Classic' techniques outlined in [18] and perform well [4] [22]. For these systems, the listener is in the centre of the rectangle, which can sometimes be difficult in smaller rooms, especially as the surround system is probably part of a home theatre system with a large screen TV.

2.4. The Mythical ITU-R System

In 1994, the ITU specified a speaker layout for surround sound with 3 speakers at 0° and ±30° in front and surround speakers at ±110° as ITU-R BS.775 [26].

The first diagram in Figure 1 shows this as ITU 5.1 including the ±30° and 100° to 120° range that is allowed. It is immediately obvious that few living rooms can accommodate such a layout without considerable domestic disruption. The few surveys of surround systems in real homes (e.g., Fig 16 in [3], Fig 14 in [27]) show that ITU-R systems are found only in research establishments and professional studios; they are rarely present in a domestic setting. In fact, ITU-R BS.775 is an attempt to match the cinema practice of three speakers behind the screen and multiple surround speakers scattered around the auditorium. It makes no concessions to domestic acceptability.

Why is so much effort dedicated to decoders for a mythical speaker layout?

Figure 1 shows seven of the "Vienna" decoders [16], which are all rectangular, or near rectangular layouts with a centre speaker added. They share the classic rectangle layout's listening position in the centre of the room but the speaker layouts are quite sensible from a domestic viewpoint.

Most home theatre guides (e.g., [6][10]) also support and recommend this layout. In the first diagram of Figure 1, below, the "Real World" system (red circles) is superimposed on the ITU recommendation (blue circles).

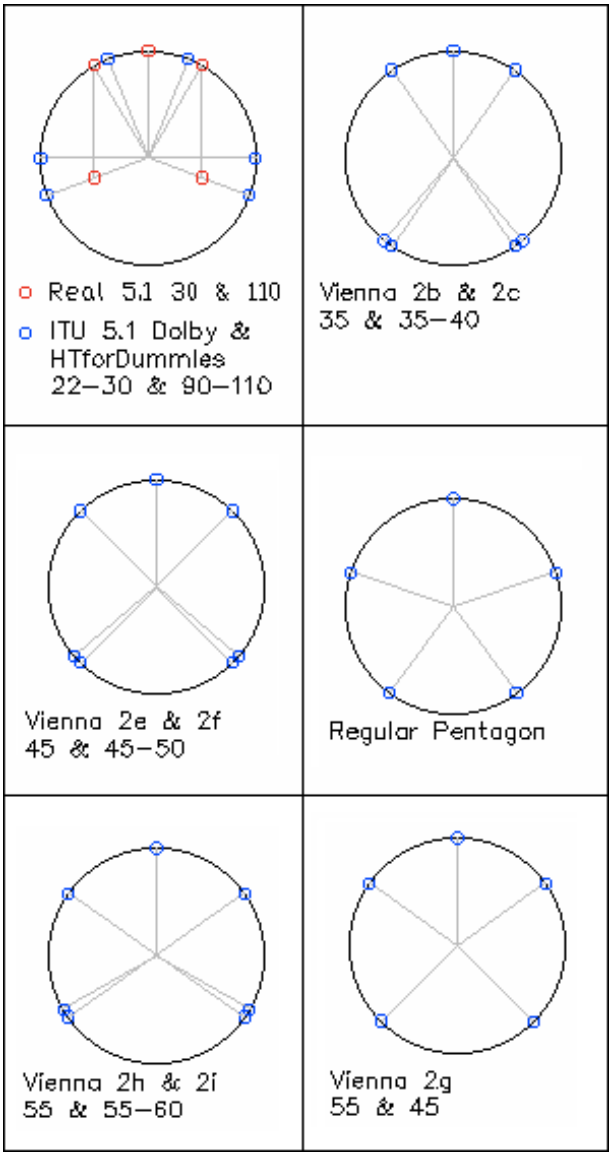

Figure 1: Five-loudspeaker layouts for home use.

If the speaker layout (without center loudspeaker) is a square and the listener moves towards the back boundary, the speakers will now subtend the same angles as ITU-R and *this is why ITU-R decoders are important.* They form the basis of decoders for these 'real world' systems where the front and surround speakers form a (near) rectangle or square but the listener sits near the back of the layout.

There is an important difference between these real world systems and ITU-R BS.775. Although the speakers may subtend the same angles, the rear speakers are

which might conveniently be part of an Ambisonic decoder, this approach may be due for resurrection.

much closer to the listener. If the speakers are in a 1.05 x 1.0 wide 'square' with LF and RF at ±30° and LS and RS at ±110° (probably the closest to a universal arrangement), the rear speaker outputs will be 5.48 dB louder and arrive in 0.532 of the time it takes for the front speakers. This can be corrected in the decoder by applying different delays and '1/r' amplitude compensation as in [11], with individual speaker Near Field Compensation to correct for the different curvature of the wavefronts due to the different distances of the speakers.

Lastly, even the most dedicated audiophile has to deal with doors, windows, furniture, and other uses of the space, so it may be impossible to place one or more speakers in the logical position. A good optimizer should still give the best decoder for such layouts.

The work in this paper is another step towards a domestic Ambisonic decoder that takes all the above into account; perhaps with an inexpensive Ambisonic microphone to detect the position of the speakers and for room and speaker EQ. The Trinnov Optimizer for studio use already incorporates similar techniques. [28]

2.5. Metatheory for dummies

Michael Gerzon's Metatheory of Localization [17] combines information theoretic reasoning with what was known about the psychoacoustics of localization, and from them realizes a set of metrics that describe various aspects of auditory localization.

The Velocity Localization Vector, $\mathbf{r_V}$, and the Energy Localization Vector, $\mathbf{r_E}$ are two particularly useful metrics. The velocity vector is intended to predict low-frequency localization, at frequencies below which the interaural time difference is unambiguous, that is to say below about 800 Hz. The energy vector is intended to predict high-frequency localization above 800 Hz.

The ITD is ambiguous at high frequencies and that fact is central to this discussion. An average human head is about 18 cm in diameter, or about 28 cm in half circumference. Since sound travels at about 343 m/sec, the time required for sound to travel from one side of the head to the other is about 800 μsec. That length of time corresponds to a wavelength of sound at about 1200 Hz. For a system, such as human hearing, that uses phase to determine direction, there is no way to tell if a continuous sound at 1200 Hz, which is heard by the two ears as being in phase, is separated by one cycle, or two cycles, or any other multiple of a wavelength.

The result of this is that the auditory system switches from a form of localization at low frequencies dependant on interaural time differences, to another form of localization, dependant on interaural level differences at high frequencies, and it does so rather abruptly. It is incumbent, then, on a surround sound system to get the ITDs correct at low frequencies and the ILDs correct at high frequencies.

It should be noted that our hearing is also sensitive to ILDs at low frequencies, but that large values of ILD at low frequencies appear only when the source is very near to the listener's head. That fact may be significant if the audio system unintentionally produces large low-frequency ILDs that would not have been present in natural hearing.

2.6. Review of Ambisonic Criteria

As introduced above, Gerzon defined two primitive models, the velocity localization vector ($\mathbf{r_V}$) and energy localization vector ($\mathbf{r_E}$). These models encapsulate the primary Interaural Time Difference (ITD) and Interaural Level Difference (ILD) theories of auditory localization. The direction of each indicates the direction of the expected localization perception, and the magnitude indicates the quality of the localization. In natural hearing from a single source, the magnitude of each vector should be exactly 1, and the direction of the vectors is the direction to the source. It should be noted that, while $\mathbf{r_V}$ is proportional to the physical quantity of the acoustic particle velocity, $\mathbf{r_E}$ is an abstract construct.

For a particular source direction, these are computed as:

$$P = \sum_{i \in S} G_i \tag{1}$$

$$E = \sum_{i \in S} (G_i G_i^*) \tag{2}$$

$$\mathbf{r_V} = \frac{1}{P} \mathrm{Re} \sum_{i \in S} G_i \hat{\mathbf{u}}_i \tag{3}$$

$$\mathbf{r_E} = \frac{1}{E} \sum_{i \in S} (G_i G_i^*) \hat{\mathbf{u}}_i \tag{4}$$

where S is the set of loudspeakers, the G_i are the (possibly complex) gains from the source to the i^{th} loud-

speaker and $\hat{\mathbf{u}}_i$ is the unit vector pointing in the direction of the i^{th} loudspeaker. The "*" indicates the complex conjugate.[6] P and E are the overall reproduced acoustic pressure and energy gains for the given source direction.

Ideally, both types of cue will be recreated accurately by a multispeaker playback environment and their directions will be in agreement with each other. In terms of Gerzon's models this means that $\mathbf{r_V}$ and $\mathbf{r_E}$ should agree in direction up to around 4 kHz; and that below 400 Hz the magnitude of $\mathbf{r_V}$ should be near unity for all reproduced directions; and that between 700 Hz and 4 kHz, $\mathbf{r_E}$ is maximized over as many reproduced directions as is possible. $\mathbf{r_E}$ has a maximum value of 1 for a single source but is always less than 1 for multiple sources. Gerzon observes that a value of the magnitude of $\mathbf{r_E}$ of 0.5 or less "gives rather poor image stability." [15] In general, the magnitude predicts how compact and stable the image is with respect to head movement.

It is important to note that Gerzon's velocity and energy models are not in any way new but simply provide a convenient mathematical description of all known mechanisms of auditory localization (and the body of experimental data they represent) except for high frequency (impulsive) ITD and pinna coloration models. In particular, $\mathbf{r_E}$ describes all the ILD models and $\mathbf{r_V}$, the low frequency (phase) ITD models. They unify the fixed head theories and the moving head theories, both those where the listener is allowed to turn and face the virtual source as well as those which only allow small head movements. Gerzon [17] provides a list of these theories and shows their equivalence with his models using well-known stereo phenomena.

It is clear that these metrics do not explain all of what we hear; however the authors feel that they are necessary conditions for good surround sound reproduction. One example of their failing is they predict that square and hexagonal speaker arrays will have the same performance, yet the difference between the two is discerned easily in a listening test. Their key advantage over other psychoacoustic models is that they are quick to compute and well behaved numerically, making them suitable for non-linear optimization techniques. Without them, optimization of decoder performance would be much more difficult.

[6] If the gains, G_i, are real, as they are in this paper, then GG^* is equivalent to G^2.

3. THE DECODER DESIGN SOFTWARE

While others have discussed how their design programs operate and published the resulting decoders, there were none readily available to support our experiments. Thus we undertook our own open-source implementation.

In the current implementation, the user enters the loudspeaker coordinates, relative weights for the various psychoacoustic figures of merit, and other particular constraints. The application then performs the optimization, producing a summary of predicted performance of the decoder and a set of decoder coefficients. The coefficients are suitable for use a real-time decoder such as Ambdec [1]. Formats suitable for use with other decoders could be added easily.

The application was tested by deriving first-order Ambisonic decoders for a number of regular and irregular arrays. The resulting decoders were compared to those reported in the literature, both numerically and in informal listening tests, and found to be equivalent to or better than the previously reported ones.

3.1. Implementation

Our implementation comprises three components:

- A program called "sph" that takes a speaker array definition as input, projects it on to set of spherical harmonics, and then computes the pseudoinverse to produce the velocity matching decoder that is used below 400 Hz.[7]
- A program called "ambi_opt" that takes the low-frequency solution, a set of constraints, and a set of weights and other parameters as input; and produces an $\mathbf{r_E}$ optimized solution for use above 400 Hz, along with performance metrics and logs.[8]
- A set of utility programs that produce graphs showing the decoder's performance, translate the decoder parameters into presets for software decoders, and so forth.

[7] We sometimes call this the "pinv solution" referring to the use of the pseudoinverse. See Appendix A of [18]. It is a form of Wavefield Synthesis [34] as it recreates the soundfield over an area that is dependent on wavelength and the order of Ambisonic reproduction.

[8] We sometimes refer to this as the "NLopt solution," referring to the use of the NLopt package.

These programs are written in ANSI-standard C and C++ and have been tested on Windows, MacOSX, and Linux platforms. The first utilizes the GNU Scientific Library to compute spherical harmonic projections and carry out the singular value decomposition (SVD) needed for the pseudo-inverse. The second utilizes the NLopt package to perform the constrained non-linear optimization.

Due to the stochastic nature of the optimization algorithm, successive runs may yield slightly different results within the tolerances specified by the stopping criteria. We have observed that occasionally, the optimization process fails to converge and is stopped when it reaches the time limit (200 sec). When this happens, rerunning the optimizer will usually result in finding good parameters. If not, the problem is usually an error in the specification of the speaker configuration or constraints, or it may be that the problem that has been posed does not have a good solution.

A typical run finds decoder parameters to a precision of 1 in 10^5, examining 20,000 to 500,000 sets of parameters each at 180 source directions during the optimization process. Our experience is that the optimization arrives at a solution in less than 20 seconds on a 2.6 GHz Intel Core 2 Duo processor, allowing rapid experiments with the parameters and constraints. More source directions can be evaluated and greater precision can be obtained at the expense of longer runtimes.

An example run is shown in Appendix 1.

3.2. Optimization algorithm

The programmer supplies NLopt with the following:

- Choice of optimization algorithm to be used
- Set of parameters to be adjusted
- Initial value for each parameter
- Constraints on the values each parameter can take on during the optimization process
- Objective function to be minimized
- Stopping criteria (e.g., relative or absolute precision, maximum running time)

In our implementation, the parameters to be adjusted are the elements in the decoder matrix. In the case of first-order horizontal Ambisonics there are three parameters for each loudspeaker signal, namely the gain of the W, X, and Y B-format signals. Thus a four-speaker array has 12 parameters and a five-speaker array has 15.

The optimizer stops when either the relative change in the objective function falls below a certain value (10^{-5} currently) or exceeds the time allotted (200 sec).[9] The user can change these by editing a configuration file.

The objective function is a weighted sum of criteria derived directly from the definition of an Ambisonic reproduction system [17][18]. This is discussed further in the next section.

The specific optimization algorithm used is "Controlled Random Search (CRS) with local mutation" as described in [21]. This algorithm is "derivative free", meaning that it does not require that the objective function provide gradients with respect to the parameters, but that it simply returns a single figure of merit for a given set of parameters. This is well suited to problems, such as the current one, where the objective function is defined algorithmically (as opposed to analytically).

NLopt also allows the user to define constraints on the parameters. The implementation has two options:

- [-2 .. 2] for each parameter (essentially no constraints)
- [0.5 .. 1.5] times the initial value of the parameter

The latter constraint was used in experiments to force the system to find solutions that utilized the center loudspeaker.[10] *NLopt* also has facilites for non-linear and vector constraints, which could be used to impose symmetry constraints on particular subsets of the speakers, if desired.

[9] In our current experiments, we have found that if the optimizer exceeds the time limit, something has gone wrong.

[10] It is also useful to ensure the $\mathbf{r_V}$ and $\mathbf{r_E}$ solutions are not too dissimilar. Wiggins eschews "dual band" solutions as his methods often produce $\mathbf{r_V}$ and $\mathbf{r_E}$ solutions which are very different making it difficult to transition smoothly between his high frequency and low frequency decoders. [32]

3.3. Objective functions

The optimizer requires an objective function that takes as input the parameters being optimized and returns a single figure-of-merit. It then adjusts the parameters to attempt to minimize that figure. In our case, the individual criteria are drawn directly from the definition of Ambisonic reproduction as given by Gerzon and Barton in section 3 of reference [16]. For each source direction (currently 180), the following are calculated:

- P, the overall acoustic pressure gain
- E, the overall acoustic energy gain
- $\mathbf{r_V}$, the velocity localization vector
- $\mathbf{r_E}$, the energy localization vector

Then the following figures of merit are calculated over all directions:

- Root-mean-square deviation (RMSD) of pressure gain from gain at 0º
- RMSD of the magnitude of $\mathbf{r_V}$ from 1
- Average deviation of the magnitude of $\mathbf{r_E}$ from 1. (since $0 < |\mathbf{r_E}| < 1$, this maximizes $|\mathbf{r_E}|$)
- RMSD of the direction of $\mathbf{r_V}$ from the source direction
- RMSD of the direction of $\mathbf{r_E}$ from the source direction
- RMSD of the difference in direction of $\mathbf{r_V}$ and $\mathbf{r_E}$
- Standard deviation of P
- Standard deviation of E

Each individual figure of merit is then multiplied by a weighting factor, to produce an overall figure of merit for a particular set of parameters. This is returned to the optimizer as the value of the objective function. Moore and Wakefield have published methods for determining these weights based on perceptual criteria and the range of the quantities. [24]

A simpler approach has taken in the present work; the system divides the weights on angular measures by 2π to convert to arc lengths, but otherwise leaves the weights to the user. We have found that the exact values are not that important and the convergence behavior is "cuspy," meaning that the optimizer will converge to a single solution over a fairly large range of weights and a different solution over another range, with a small transition region. The weights used are changed by editing a configuration file.

3.4. Examples

In this section, we show examples, for the ITU 5.0 speaker array. These are five-speaker arrays, comprising a center-front (CF), left-front (LF) and right-front (RF) speakers at ±30°, and left and right surround speakers (LS and RS) at ±110°, as defined in ITU-R BS.775.

Once regular arrays are abandoned there is no longer a single best solution where all of the criteria listed in Section 3.3 take on their optimum values. Hence, design of decoders for irregular arrays, such as the ITU array, is an exercise in trading off aspects of the array performance in various ways. Because there is no one 'right' solution, these tradeoffs may cause the perceived performance to depend on the type of program material used, with one decoder preferred for one type of material and another decoder preferred for a second type of program material.

The exact low-frequency solution (< 400Hz) is obtained directly from the pseudoinverse method. The low-frequency solution is shown here. In each case, we use the low-frequency solution as the initial values for the parameters and then run the optimizer to produce the HF solution. The initial parameters are:

```
     W-gain      X-gain      Y-gain
CF: +0.10237800 +0.31154100 +0.00000000
LF: +0.14332900 +0.24084600 +0.22064900
LS: +0.51258900 -0.39661600 +0.41468400
RS: +0.51258900 -0.39661600 -0.41468400
RF: +0.14332900 +0.24084600 -0.22064900
```

The performance is summarized in the following graph.

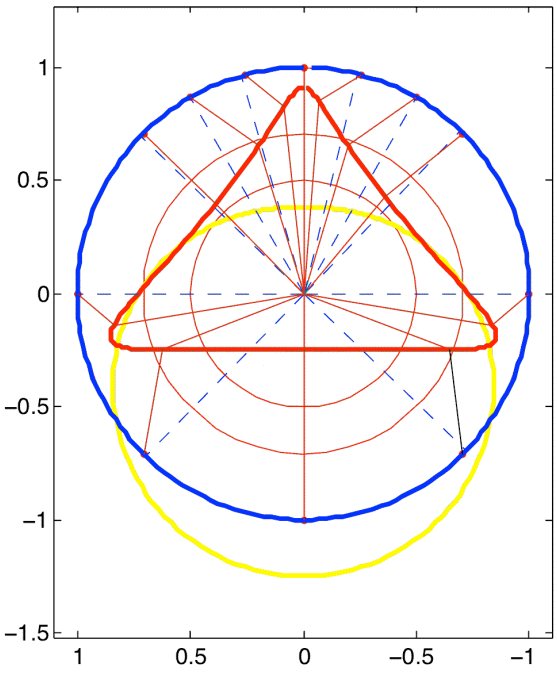

Figure 2: Plot of $\mathbf{r_V}$ and $\mathbf{r_E}$ for the LF solution of the ITU array. $\mathbf{r_V}$ is plotted in blue and is optimal. $\mathbf{r_E}$ is plotted in green and is very poor, sound is drawn to the center and sides. The distribution of E is skewed heavily to the back. The red dots show the source directions. The black lines associate the perceived vs. the true source direction. The red circles are drawn at radii of ½, √½, and 1. (in this instance, the last one is under the blue circle)

This is the common starting point for all the ITU arrays.

3.4.1. ITU 5.0 Array without constraints

For this run, high weights were placed on the average $\mathbf{r_E}$ and angular accuracy. The optimizer ran for 2.48 sec. and considered 60274 configurations, producing the following:

```
     W-gain      X-gain      Y-gain
CF: -0.04881270 -0.02756928 +0.00001724
LF:  0.28205165 +0.24760232 +0.18790454
LS:  0.44947336 -0.23381746 +0.31911519
RS: +0.44945895 -0.23380219 -0.31911386
RF: +0.28204229 +0.24758662 -0.18792311
```

It can be seen from the matrix elements that the CF speaker is essentially shut down. Wiggins and others have noted this behavior. [33]

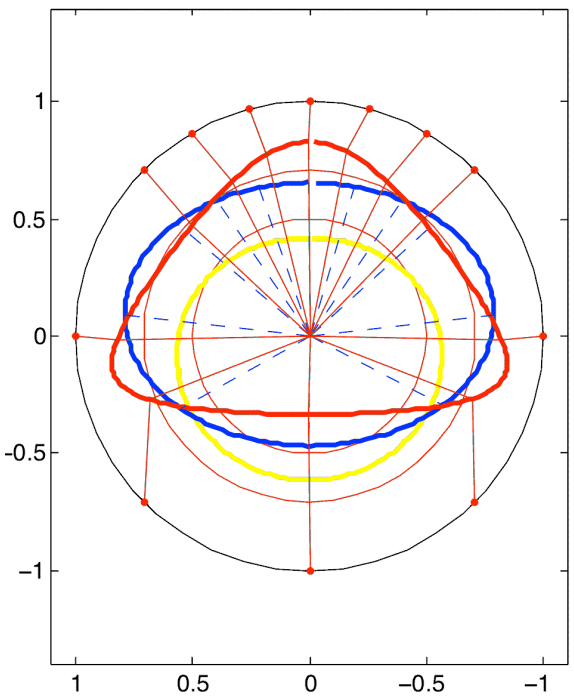

Figure 3: ITU array performance after optimization. Note that the average value of $\mathbf{r_E}$ is increased, the angular accuracy in the front is improved and the distribution of E is more uniform.

In listening tests, this array produced a "detent" effect in the center, meaning that there is a compact buildup of sound coming from directly ahead of the listener.

3.4.2. ITU 5.0 Array with constraints

In 1992, Gerzon and Barton published decoders that utilized the center loudspeaker [16]. To explore the performance of such decoders, constraints were added to the optimizer that kept the parameters in the range of -50% to +100% of their initial values. The following decoder was produced in 1.48 sec, after evaluating 43,901 configurations.

```
     W-gain      X-gain      Y-gain
CF: +0.09993309 +0.15577050 +0.00000000
LF: +0.21426224 +0.19218459 +0.20409261
LS: +0.44287748 -0.27006948 +0.30405695
RS: +0.44287676 -0.27006941 -0.30405595
RF: +0.21426400 +0.19218379 -0.20409362
```

Note that the center speaker is now active.

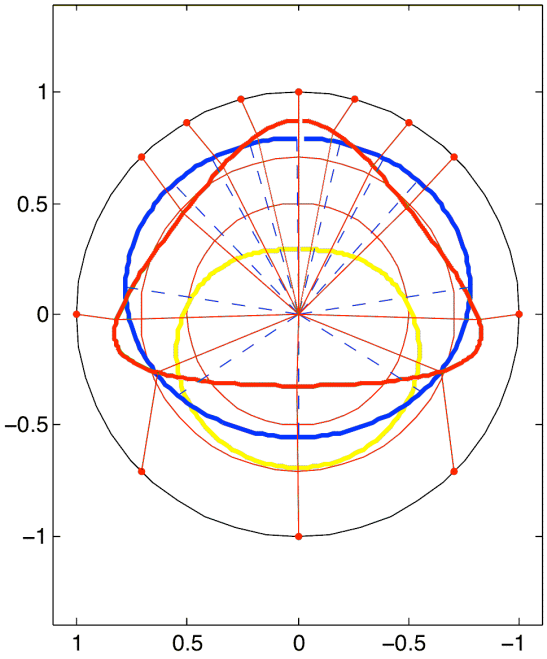

Figure 4: ITU array performance after optimization with constraints to keep the CF loudspeaker active. Note that the performance is nearly identical, with an increase in $\mathbf{r}_E$ in the front.

This decoder was preferred in listening tests to the one shown in Section 3.4.1 for music with sources in front and reverberation to the rear, due to more uniform source placement across the front. It did not perform as well for environmental recording where sounds originated from all directions.

3.4.3. ITU 4.0 Array with uniform *E* distribution

The total energy in the two previous decoders is skewed to the rear of the listener. This may cause a change in tonal balance for sounds originating from the rear. Greater weight was placed on the uniformity of the overall energy gain, *E*, and lowered angular accuracy weights. The CF speaker was also removed from the optimization process. The optimizer ran for 67.04 seconds, considering 1,680,755 configurations.

```
     W-gain      X-gain      Y-gain
LF: +0.35355493 +0.24825490 +0.22602911
LS: +0.35355209 -0.24826414 +0.26865556
RS: +0.35355474 -0.24825447 -0.26866481
RF: +0.35355179 +0.24826371 -0.22601987
```

Except for small variation in Y-gain, this is identical to the $\mathbf{r}_E$-max decoder for a square array.

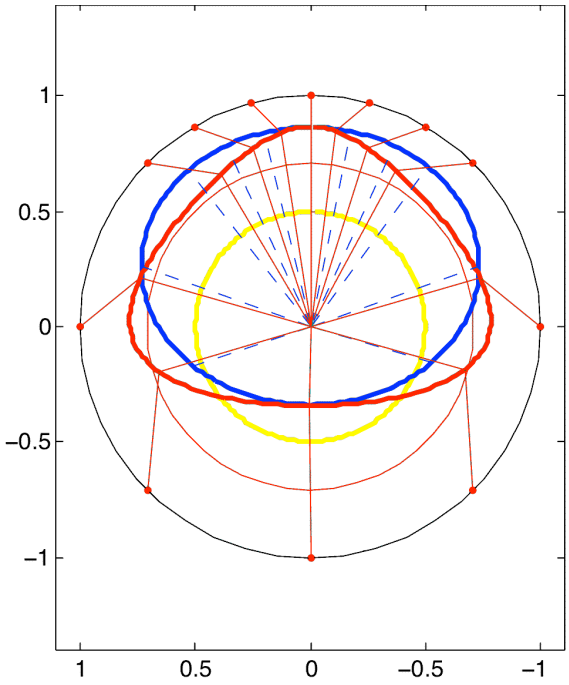

Figure 5: ITU 4.0 array (no CF) where uniform distribution of total energy E is weighted heavily.

While the energy distribution is uniform, all sources are drawn to the center front, producing a 'detent' effect.

3.4.4. Tradeoffs

These examples illustrate some of the tradeoffs involved with designing first-order decoders for the ITU array. In general, it is not possible to have both correct directions and even energy gain for all directions. These tradeoffs are a matter of preference on the part of the listener and the type of source program being reproduced. This highlights the utility of the listener having a decoder design tool and flexible decoder.

4. ANALYSIS AND EXPERIMENTS

Ambisonics has historically been deployed using regular arrays of loudspeakers. Previous publications in the area of decoder optimization have used these techniques to develop decoders for ITU-R arrays with speakers at 0°, ±30° and ±110°. While this is a worthwhile addition, real-world systems typically have different angles or are

even asymmetrical from left to right. Having a tool which easily provides optimized decoder solutions for haphazard loudspeaker arrays allows us to delve into some interesting practical questions:

- How well do various arrays perform relative to each other?
- What is the effect of imprecision in loudspeaker location?
- What is the sensitivity to off-center positioning of the listener?
- What is the effect of decoder/array mismatch?

The first question to be addressed is one of how well this optimization software duplicates previous solutions. Given that heuristic methods do not generate exact solutions but rather approximations, and that furthermore the result depends on the constraints applied to the optimizer, it is expected that the solutions obtained here will only approximately match what has been published previously. First order energy-optimized decoders for the ITU-R arrays have been published by Gerzon and Barton (the 'Vienna' decoders) [16], Wiggins [31], Moore and Wakefield [24][25], and Tsang, *et al.* [29][30]. The result from Gerzon & Barton [16] is for loudspeakers located at 0°, ±45°, and ±130°, an array shape which is very different than the 0°, ±30°, ±110° that is recommended in BS.775.

The value of $|\mathbf{r_E}|$ can be calculated as a function of angle for these various decoders and that has been done for the following figure:

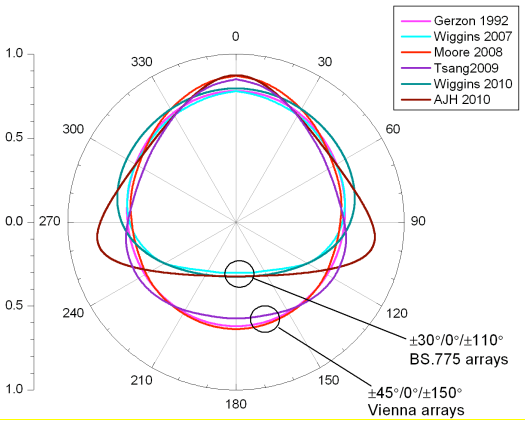

Figure 6: $|\mathbf{r_E}|$ for various decoder solutions.

This figure shows that these solutions are largely similar, with $|\mathbf{r_E}|$ being greatest in the front and least in the rear.

A tool that allows decoders for arbitrary arrays to be simply and quickly generated gives the opportunity to investigate some interesting questions. As we know, typical surround sound loudspeaker arrays in domestic situations almost never conform precisely to ITU-R BS.775. The positions of the front three loudspeakers are likely to be relatively close to ±30° as recommended, but the surround loudspeakers are likely to be either directly to the sides of the listener or much further behind the listener than ±110°. This being the case, it is important to know what may happen if the wrong decoder is used for a particular loudspeaker arrangement, as might happen if Ambisonic program material were distributed in a 5-channel format decoded directly for the recommended BS.775 layout, or if the loudspeakers are in an asymmetrical arrangement as may be necessitated by the geometry of real rooms.

4.1. Decoder for a square

A square, being a regular array, is easy to derive a dual band decoder which gives $|\mathbf{r_V}|$ equal to 1 in all horizontal directions, $|\mathbf{r_E}|$ = 0.7071, which is the maximum amount possible, and has the angles of $\mathbf{r_V}$ and $\mathbf{r_E}$ pointing exactly in the intended directions.

The first application of the optimization software is a decoder for a square loudspeaker array. The square array is the simplest possible Ambisonic system, any fewer loudspeakers not being sufficient to properly reproduce the velocity and energy vectors. The output of the first program, which computes the pseudoinverse (pinv) of the projections of the loudspeaker coordinates, is the following decoder:

	W Gain	X Gain	Y Gain
LF	+0.35355	+0.35355	+0.35355
RF	+0.35355	+0.35355	-0.35355
RR	+0.35355	-0.35355	-0.35355
LR	+0.35355	-0.35355	+0.35355

Which gives the following result for $\mathbf{r_V}$ and $\mathbf{r_E}$:

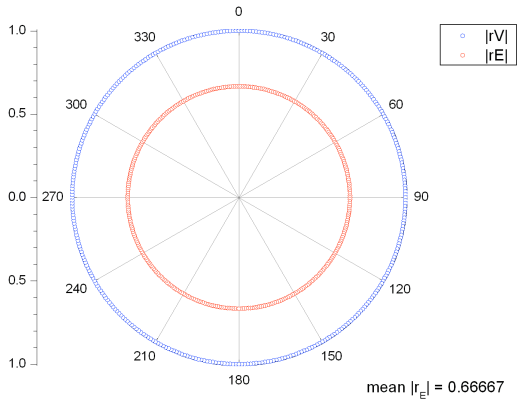

Figure 7: $\mathbf{r_V}$ and $\mathbf{r_E}$ analysis of velocity matching decoder for a square array.

The blue circles represent the magnitude of $\mathbf{r_V}$ in the direction of $\mathbf{r_V}$, the red circles represent the magnitude of $\mathbf{r_E}$ in the direction of $\mathbf{r_E}$. $\mathbf{r_V}$ is the correct magnitude and direction everywhere but $\mathbf{r_E}$ points in the correct direction but has a magnitude of only 0.667 instead of 1.

If the optimizer program is used with the output of the PINV program, the following result is obtained for $\mathbf{r_E}$:

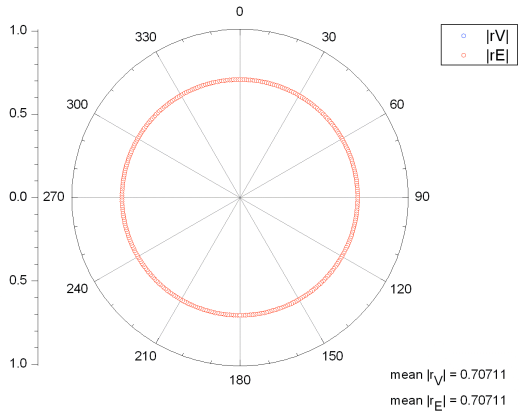

Figure 8: $\mathbf{r_V}$ and $\mathbf{r_E}$ analysis of NLopt derived decoder for a square array ($\mathbf{r_V}$ not visible behind $\mathbf{r_E}$)

$\mathbf{r_E}$ has been increased from 0.667 to 0.707, but $\mathbf{r_V}$ has been reduced from 1 to 0.707. As discussed previously, the velocity matching (PINV) solution can be used at low frequencies and the NLopt solution can be used at high frequencies with the result that:

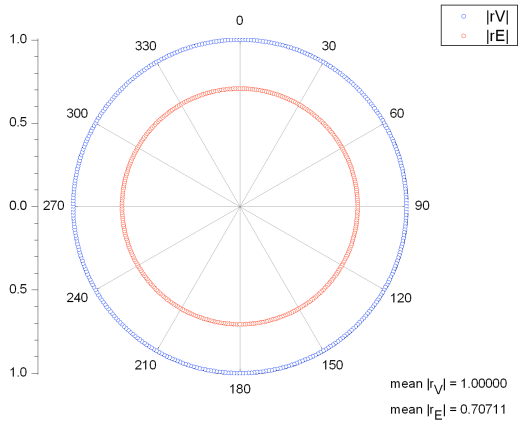

Figure 9: Dual band decoder for a square array using the velocity matching solution at low frequencies and the NLopt solution at high frequencies.

In this particular case the dual-band decoder is only slightly better than the PINV solution alone. For non-regular arrays the problems are much more complicated because the PINV solution will have the velocity vector and energy vectors pointing in different directions.

4.2. Evolution from square to ITU array

If the decoder for a square array is applied to the four speakers of an ITU array (no CF loudspeaker), the following results:

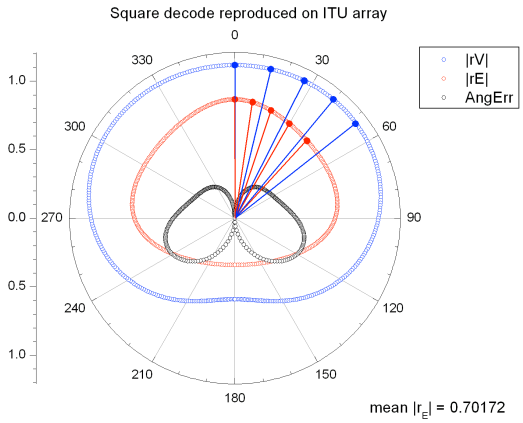

Figure 10: Gerzon vector analysis of square Ambisonic decoder on ITU BS.775 array.

The shape of the array and the relatively narrow spacing at the front has caused the magnitude of the velocity vector to increase beyond what occurs with natural sources (> 1), and the energy vector is concentrated in the front and reduced in the rear, relative to the square array. The black kidney shapes show the energy vectors pointing in the wrong directions, only being exactly correct for sources directly in front or directly behind. The five locations at 0°, 15°, 30°, 45° and 60° are crowded toward the front. The mean value of $\mathbf{r_E}$ is only negligibly decreased, and the velocity and energy localization cues are in conflict.

This in fact, describes the situation in [22] which led the present authors to cautiously endorse the Nimbus use of Classic Square decode to generate 4.0 (5.0 with center channel muted) DVD-A and DTS recordings. It results in four channel "one channel/speaker" recordings which are robust to variations from the intended speaker layout and listener position compared to the dedicated ITU-R decoder we tested at the time. It has been reported that when such recordings are played on domestic systems with roughly square speaker layout, listeners instinctively move forward to the correct centre position. [12] Perhaps the narrowing of the front sound stage as the listener moves back helps to maintain the illusion as that would be what happens in real life.

Optimization of the decoder for the specific target loudspeaker array can improve this performance. The optimization parameters can be chosen such that the magnitude of $\mathbf{r_E}$ is globally maximized while still constraining the directions of $\mathbf{r_E}$ to be close to the desired direction gives the following results:

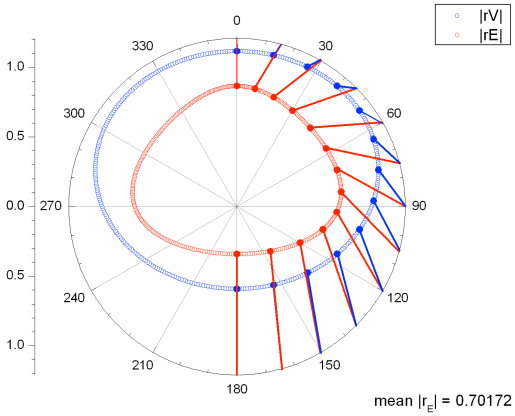

Figure 11: ITU loudspeaker array with ITU4 decoder.

The reproduction now has the magnitudes and directions of $\mathbf{r_V}$ exactly correct at low frequencies, and the directions of $\mathbf{r_E}$ substantially correct at high frequencies. The mean value of $\mathbf{r_E}$ is nearly $\sqrt{2}/2$, which is as high as it can be. The directions of the five vectors at 0°, 15°, 30°, 45° and 60° are almost exactly correct. The magnitude of $\mathbf{r_E}$ still peaks in the direction of the surround speakers and there is significant directional error in the rear. Compared to the square array, the magnitude of $\mathbf{r_E}$ is greater over most of the front half and the angle error is acceptably good, but the magnitude of $\mathbf{r_E}$ is too small in the rear and the angle error is large around +130° and -130°. Compared to a square decoder used on an ITU array, the magnitude of $\mathbf{r_V}$ is substantially increased and both the direction and the magnitude of $\mathbf{r_E}$ are improved.

The first investigation is for the case where the loudspeakers are arranged with bilateral symmetry but with a different angle between the surround loudspeakers than 110°.

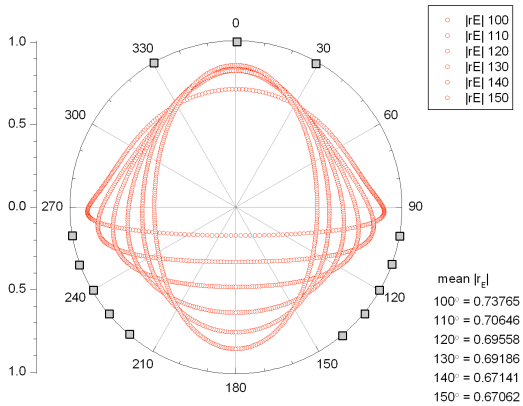

Figure 12: $|\mathbf{r_E}|$ for variations in the location of surround loudspeakers.

As the left back and right back loudspeakers are moved backwards from the initial position, the average value of $\mathbf{r_E}$ stays relatively constant because $\mathbf{r_E}$ is reduced to the sides and increased in the front and the back. The optimizer only minimally utilizes the front loudspeaker, and for the case where the angle is 150°, the front loudspeaker is shut down entirely.

The second investigation is for the case where a single surround loudspeaker is moved backwards from the recommended 110° displacement in 10° increments. The distribution of $\mathbf{r_E}$ was calculated for the ITU rec-

ommendation and for the case where the right back loudspeaker is moved back to 120°, 130°, 140°, and 150°, but the decoder remains the one derived for the original case with the loudspeakers at 110°.

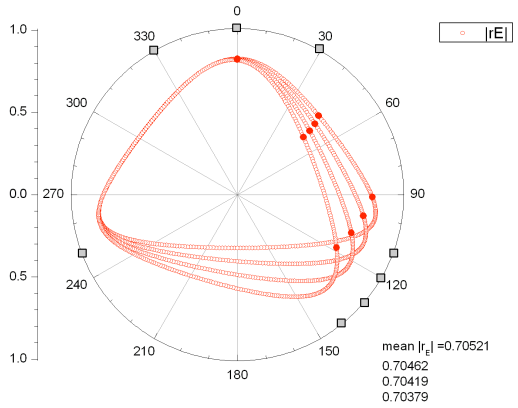

Figure 13: Distribution of $\mathbf{r_E}$ with various aberrations in the position of the left back loudspeaker, incorrect decoder used.

This analysis shows that, unsurprisingly, the magnitude of $\mathbf{r_E}$ is increase in the direction of the displaced loudspeaker, but that the locations of sounds are also displaced backwards along with the loudspeaker.

If the loudspeaker array is varied but the correct decoder is still used, the following results:

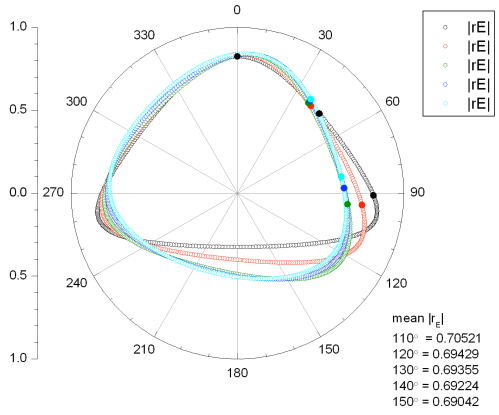

Figure 14: $|\mathbf{r_E}|$ for displacement of one of the surround loudspeakers backwards, with a new decoder derived for each new configuration.

It should be noted that the mean value of $|\mathbf{r_E}|$ varies relatively little when the position of one loudspeaker is altered. The overall performance of the system is maintained, including correct high frequency localization.

4.3. Decoders for Loudspeakers at ±90°

The optimization tool was used to derive decoders for the case where an ITU-R array is modified to have the surround loudspeakers at ±90°. Such an array cannot produce a result that truly surrounds the listener, since all of the loudspeakers are in the front hemisphere. It is useful, however, to derive a decoder because many domestic installations have such a setup.

A decoder derived using the pseudoinverse gives the remarkable result ensues that the velocity vector can be reconstructed exactly (correct magnitude and direction). The energy vector is not too good; $\mathbf{r_E}$ is more concentrated in some areas than others and the direction of $\mathbf{r_E}$ is very far off in some areas, especially in the back. That result is shown in the following figure:

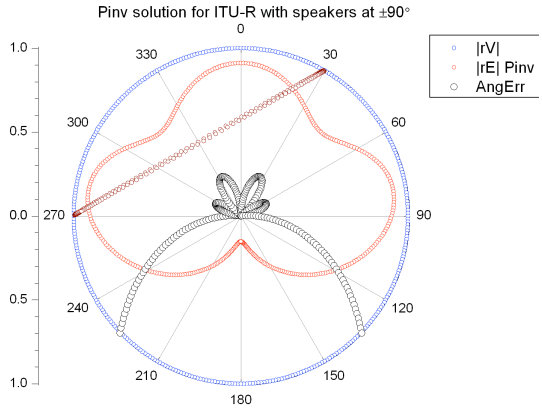

Figure 15: Distribution of $\mathbf{r_E}$ with surround loudspeakers at ±90°.

This result shows the correct value and distribution of the velocity vector and moderately large values of $\mathbf{r_E}$ in the front half, but greatly reduced in the rear half. A decoder optimized for an extreme layout is very different from one for a slightly less extreme speaker layout like ITU-R. A decoder for a less extreme layout may be more robust with listener position and provide 'better' if not 'accurate' results for the extreme layout.

In the next section the performance of these various decoders will be compared in listening tests.

5. LISTENING TESTS

The preceding analysis of decoder performance using the Gerzon-defined quantities of $\mathbf{r_V}$ and $\mathbf{r_E}$ is intended to quantify the reproduction performance, but it is not expected that that analysis will describe every aspect of the performance. For that reason listening tests were performed on several of the decoder solutions described above.

The listening tests are intended to answer the following questions. What is the relative performance of:

- the ITU-R array vs. a 1.732:1 rectangle?
- variations on the ITU-R array?
- decoders that concentrate energy at the front?

The listening tests were done by having both loudspeaker arrays present at the same time, using previously decoded test materials. A software application allowed the listener to choose between any of the predecoded files with seamless switching. The tests were performed in the multichannel listening room CMAP described below.

5.1. CMAP listening room

The listening room used in these tests has been designed as a compromise between the heavily treated listening rooms designed to meet the requirements of BS.1116 and the acoustics of normal domestic rooms.

- Dimensions 6.7 m × 4.5 m × 2.44 m
- RT60 ~0.32 sec in octave bands
- Loudspeakers JBL LSR25p (self powered)
- Frequency response 70 Hz to 20 kHz, +1/-2 dB,

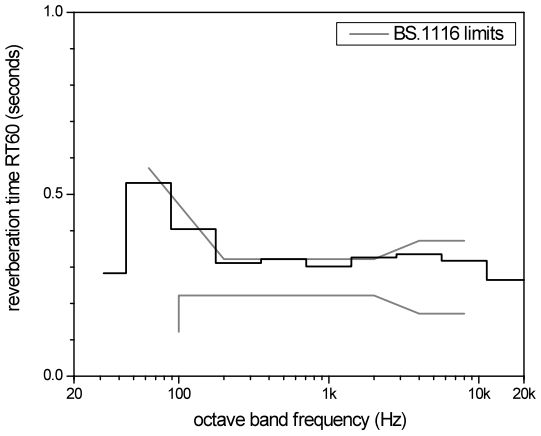

Figure 16: Reverberation times in octave bands.

Up to 16 channels of audio can be reproduced from sources in the horizontal plane and at elevations from +30° to -30° with respect to horizontal.

The loudspeakers were installed on stands such that the geometrical center of the loudspeakers was at 1.1 m height, approximately ear height for the listeners. The horizontal location was set using a laser pointing device and the loudspeakers were placed on a 4-meter diameter circle around the principal listening position. System performance was verified by measuring the impulse response from each of the loudspeakers to an upwards-pointing ¼" measurement microphone. The grazing incidence orientation results in about 3 dB of attenuation in the measurements at 20 kHz, but guarantees the same response from each of the loudspeaker positions. Playback of the previously decoded B-format test sources was from a nearly silent Dell Optiplex server via two Echo Layla digital audio interfaces, only the first of which was used in these tests. Switching between program sources was done in software.

The loudspeakers were arranged on a 4-meter diameter circle as shown in the following figure:

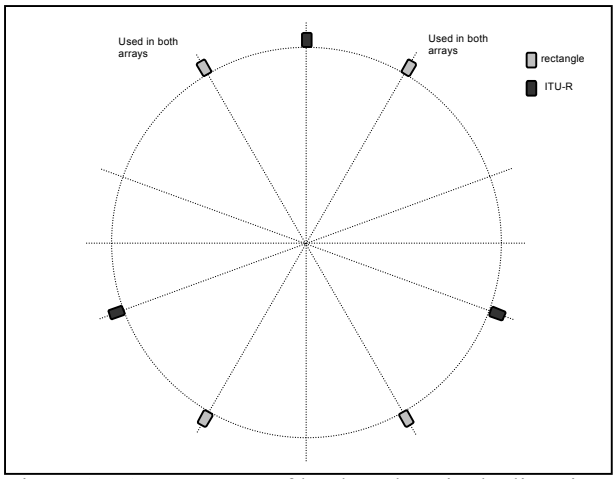

Figure 17: Arrangement of loudspeakers in the listening room for ITU-R vs. rectangle comparison.

The use of a rectangular loudspeaker array in conjunction with the ITU-R array allows the left and right front loudspeakers to be used in both arrays.

Given the open-ended nature of these listening tests, a category-scaling test was not used. Instead, the listeners were allowed to investigate the differences between two (or more) presentations and observe whatever differences were found without regards to categories. But they were encouraged to observe at least the following attributes:

- Timbre of sources
- Soundstage width
- Localization of sources, direction and focus
- Quality of reverberation (if present)
- Sensitivity to listener head movement

The test programs used included natural Soundfield recordings, a studio recording of a female voice, and Gaussian-windowed cosine bursts panned to various directions. The test programs were decoded using an offline, file-to-file version of Adriaensen's AmbDec decoder with the coefficients given by the optimization program.[11] This allowed sample accurate synchronization among the various decodes of the same material to support rapid comparison of the different decodes.

[11] Contact the first author of the present paper to obtain a copy of the offline version of Ambdec.

- Eight Directions; female alto voice announcing directions
- Beethoven *Sym. No. 4*; Symphonic music, Soundfield MkIV recording in a concert hall with excellent acoustics.
- Stravinsky *Pulcinella Suite*; Soundfield MkIV recording of a small orchestra. The sections with interplay among the woodwinds are good for judging the spatial qualities of the frontal image. The brass sections are good for judging the envelopment of the reverberation. Long section of applause at the end.
- Glazunov *Spanish Dance*; large orchestra recorded with a Soundfield MkIV. The same hall, but a considerably more distant perspective than the previous two recordings. The castanets were good for judging the spatial qualities of the frontal image and the more distance perspective is good for judging envelopment.
- *Aran music*; field recording made by John Leonard, using a Soundfield ST250. Outdoors with sound sources primarily in front.
- *Hampi Bazaar*; field recording made by Paul Doornbusch. Outdoors in on busy street with people walking by, and speech and metallic sounds coming from all around the listener. When reproduced correctly, a very good sense of envelopment is achieved.

These recordings are available for download at www.ambisonia.com .

The following additional recordings were also used:

- Beethoven Piano Sonata No. 23 *"Appassionata";* Soundfield MkIV recording of a piano recital in a fairly reverberant hall, with a slight slap echo off the balcony. The second movement with its slow, sustained chords was particularly revealing of the differences among the decoders.
- A recording of shaped noise bursts and voice announcements to identify the individual speaker feeds. This was used to verify the integrity of the playback system before and after each listening session.

Our emphasis is on real life recordings rather than synthesized sounds. The over-riding sensation of good Ambisonic playback has always been "I am there," rather than pin-point localization and the choice of material reflects this. Listening tests that concentrate on strict localization ignore this factor, which is of particular importance to film and video. An important feature is the lack of 'speaker detent' where a speaker draws attention to itself and destroys the illusion. One of the authors claims the Square has less detent effect than a 16-channel horizontal ring with pair-wise panning

Presently, we are limited to first order recordings by available Ambisonic microphones. Higher order microphones are still not quite available but this situation is changing rapidly. [14][8]

Most decoding involved the use of a dual-band decoder. The solution that gives maximum $\mathbf{r_V}$ is different than the solution that gives maximum $\mathbf{r_E}$, and both solutions are used by crossing over between the two using a phase-matched filter.

The use of a dual-band decoder involves the potential problem of normalizing the loudness between the low-frequency and high-frequency decoders. In the traditional Gerzon-inspired decoders this was done by using making the RMS of the low- and high-frequency weights be equal. For instance, if the ratio of P/V needed to be increase by $\sqrt{2}$ (3 dB) to achieve maximum $\mathbf{r_E}$, then some of the increase in the ratio was achieved by increasing P and some by decreasing V.

Remembering that the purpose of the listening tests was to determine the suitability of the various decoders and loudspeaker arrays discussed in section 4, it should be noted that preference was determined, rather than rating. A large number of observations were also recorded, having to do with the quality of the audio reproduction.

The principal systems compared were rectangular arrays vs. ITU arrays, ITU arrays vs. similar arrays with one or both surround loudspeakers moved, and decoders which increased frontal $\mathbf{r_E}$ vs. decoders which maximized $\mathbf{r_E}$ globally.

It was found that the rectangular array produced a frontal soundstage that was substantially different than what was heard with the ITU arrays as a group.

We compared six decodes of each test file:

- 1.71:1 Rectangle. This is the array that was used extensively in our previous listening tests.
- ITU-4 An optimized decode for the ITU-R array that shuts down the CF speaker.
- ITU-5 Hybrid. This uses the 4-speaker optimized solution for HF and the 5-speaker solution for LF.
- ITU-5. This uses all five speakers.
- Square over ITU. A dual-band decode for a square layout, but played over speakers in an ITU-R array (CF is not driven)
- ITU-5-Asymmetric. An ITU-R array with one surround speaker at 150° instead of 110°

The directional announcements were correctly rendered by all decoders, however Left and Left-Back were difficult to distinguish. In general all of the ITU decodes sounded similar when compared to the rectangle decode and square over ITU. The ITU decodes had a better sense of envelopment than the square over ITU. The differences among the ITU decodes were subtle at first, but became more obvious when particular sections of pieces were identified. The ITU-4 and ITU-5 Hybrid arrays had a distinct "center detent" effect that became annoying over time. This effect was completely absent in the ITU-5 decode. The ITU-5 decode was also the most robust when moving out of the sweet spot. The localization on the ITU-4 decode moved "in head"[12] when the listener moved back. However, with an ITU speaker layout, the listener is likely to be already near the back of the room and unlikely to move further back.

Environmental recordings like *Hampi Bazaar* and *Aran Music* did not work well in the ITU decodes. These had important sources in all directions around the listener and the directional distortions in the rear half of the ITU arrays did not work well with these recordings. The classical recordings where the primary sources are in the front and the hall reverberation in the back were handled much better.

The square over ITU did not work well for the musical samples, whereas it worked the best among the ITU

[12] We use the term "in head" to refer to localization where the source sounds close to or inside one's head. This usually arises when the localization cues conflict dramatically.

decodes for the environmental samples. The frontal stage was compressed toward the center and the separation in back was exaggerated and "in head" when the listener moved back. Applause was occasionally "in head".

6. DISCUSSION

6.1. Future Work

- Extend to 3-D and higher-order Ambisonics.
- Add directional-dependant weighting for the objective functions, allowing, for example, different weighting of the objective functions for front and back directions or that sound from azimuth=0 (directly ahead) should give zero output from the surround speakers.
- Add facilities for adding symmetry constraints, and automated detection of symmetries in speaker locations.
- Add facilities evaluation of the objective function at multiple listener positions.
- Add a graphical user interface (GUI)
- Experiment with other spatial hearing models.

7. CONCLUSIONS

A new software application has been presented which allows the generation of optimized Ambisonic decoders for irregular loudspeaker arrays. The software utilizes open source routines for solution of the exact matching decoder (PINV) and for optimization (NLOPT) of the matching decoder to maximize the value of the Gerzon energy vector, $\mathbf{r}_E$. The optimization is under the control of a group of parameters that are selected prior to the beginning of the process. This software in both source form and compiled for Windows, MacOSX, and Linux is available for download for use by any individual who has need to generate Ambisonic decoders.

As is well known, it is a relatively simple process to generate Ambisonic decoders for regular polygonal or polyhedral loudspeaker arrays, but it becomes increasingly difficult to generate decoder solutions that have large and smooth energy vector values when the arrays are irregular. This is because the Energy vector is a non-linear function of the loudspeaker signals. In some cases it is possible to achieve an explicit solution when the array has certain types of symmetry. For the general case, where loudspeakers may be placed somewhat arbitrarily due to the exigencies of domestic environments, solutions are better arrived at by search methods.

This software reaches solutions which are essentially the same as those achieved by other optimization efforts. In this paper it has been applied to certain problems in order to study what may happen when an effort is made to reproduce Ambisonic sources on available surround sound systems, which may follow the ITU-R BS.775 recommendation, or which are more likely to be completely haphazard.

Listening tests were performed using some of the decoders derived using the optimization software. The listening tests had principally to do with comparisons between various decoders on either a rectangular array or on an ITU array. The ITU array was also perturbed to represent various real-world listening situations, such as what happens when one of the surround loudspeakers must be mis-located either in front of or behind its intended position, or when both of the surround loudspeakers are located considerably further rearward than the ITU recommendation.

Different loudspeaker setups were preferred for different types of program material. Generally, a decoder that had been optimized for a particular loudspeaker arrays performed better than one that had not been optimized.

8. ACKNOWLEDGEMENTS

The authors thank Steven Johnson, creator of the NLopt library; Gregory Maxwell for making us aware of it and providing an example program; Don Drewecki, John Leonard, and Paul Doornbusch for providing some of the recordings used in the listening tests; and our families for tolerating rooms with many, many loudspeakers and endless "Front… Right Front… Right… Right Back……"

9. REFERENCES

[1] Adriaensen, Fons , "Ambdec – 0.4.3 User Manual," http://www.kokkinizita.net/linuxaudio/ (accessed 9/1/2010)

[2] Adriaensen, Fons, "One more Ambisonic to 5.0 Decoder," email to Sursound discussion list. (11/13/2006)

[3] Benjamin, Eric and Gannon, Benjamin, "The Effect of Room Acoustics on Subwoofer Performance and Level Setting," 109th AES Convention, Preprint 5232. Los Angeles (2000)

[4] Benjamin, Eric, Richard Lee, and Aaron Heller, "Localization in Horizontal-Only Ambisonic Systems," 123rd AES Convention, Preprint 6967. San Francisco (2006)

[5] Beranek, L, *Acoustic Measurements*. John Wiley, New York (1949) pp 56 – 64.

[6] Briere, D. and Hurley, P., *Home Theater for Dummies.* 2nd Edition, 2003. Wiley, New York.

[7] Craven, P., "Continuous Surround Panning for 5-speaker Reproduction", AES 24th International Conference (2003)

[8] Craven, Peter, G., Malcolm Law, Chris Travis, "Microphone Array." International Patent Application No. WO 2008/040991 A2. (2008)

[9] Daniel, Jerome, "Spatial Sound Encoding Including Near Field Effect: Introducing Distance Coding Filters and a Viable, New Ambisonic Format" 23rd AES International Conference, Copenhagen (2003).

[10] Dolby Laboratories, "Home Theater Speaker Guide," http://www.dolby.com/consumer/setup/speaker-setup-guide/index.html (accessed 9/3/2010)

[11] Eclectic Electronics, "Automatic speaker calibration technologies (MCACC, YPAO, Audyssey MultEQ)" http://www.eclecticelectronics.net/ht-home-theater/automatic-speaker-calibration-technologies-mcacc-ypao-audyssey-multeq/ (accessed 9/3/2010)

[12] Elen, Richard, "Getting Ambisonics Around" http://ambisonic.net/pdf/ambisonics_around.pdf (accessed 9/1/2010).

[13] Farina, Angelo, "Reply to: G-format," email to Sursound discussion list. (11/13/2005)

[14] Farina, Angelo, Andrea Capra, Lorenzo Chiesi, and Leonardo Scopece, "A Spherical Microphone Array for Synthesizing Virtual Directive Microphones in Live Broadcasting and in Post Production," 40th AES Conference. (Toyko 2010)

[15] Gerzon, M., "Practical Periphony: The Reproduction of Full-Sphere Sound", 65th AES Convention, Preprint 1571. London (1980).

[16] Gerzon, Michael A. and Geoffrey J. Barton, "Ambisonic Decoders for HDTV," 92nd AES Convention, Preprint 3345. Vienna (1992).

[17] Gerzon, Michael A., "General Metatheory of Auditory Localization," 92nd AES Convention, Preprint 3306. Vienna (1992).

[18] Heller, Aaron J., Richard Lee, and Eric M. Benjamin, "Is My Decoder Ambisonic?" 125th AES Convention, Preprint 7553. San Francisco (2008).

[19] Integrex Ambisonics Decoder http://ambisonics.dreamhosters.com/Integrex.pdf (accessed 9/3/2010)

[20] Johnson, Steven G., The NLopt nonlinear-optimization package, http://ab-initio.mit.edu/nlopt (accessed 9/1/2010)

[21] Kaelo, P. and M. M. Ali, "Some variants of the controlled random search algorithm for global optimization," J. Optim. Theory Appl. 130 (2), 253-264 (2006)

[22] Lee, Richard and Aaron Heller, "Ambisonic Localization, Part 2," 14th International Conference on Sound and Vibration. (2007)

[23] Leese, Martin, "Minim Decoders," http://sites.google.com/site/minimdecoders/ (accessed 9/3/2010)

[24] Moore, D., Wakefield, J.; "Exploiting human spatial resolution in surround sound decoder design" 125th AES Convention, (Oct 2008).

[25] Moore, D., Wakefield, J.; "The design of ambisonic decoders for the ITU 5.1 layout with even performance characteristics" 124th AES Convention, (May 2008)

[26] Rec. ITU-R BS.775-1, "Multichannel Stereophonic Sound System with and without Accompanying Picture" ITU (1992-1994)

[27] Toole, Floyd E., "Loudspeakers and Rooms for Stereophonic Sound Reproduction" 8th AES Inter-

national Conference: The Sound of Audio, Paper 8-011 (May 1990)

[28] Trinnov Audio, "Broadcast" web page. http://www.trinnov-audio.com/about-us/customers/broadcast(accessed 9/4/2010)

[29] Tsang, P.W.M. and K.W.K. Cheung, "Development of a re-configurable ambisonic decoder for irregular loudspeaker configuration," IET Circuits Devices Syst., 2009, Vol. 3, (4), pp. 197–203.

[30] Tsang, Peter Wai-Ming, Wai Keung Cheung, Chi Sing Leung, "Decoding Ambisonic Signals to Irregular Loudspeaker Configuration Based on Artificial Neural Networks", ICONIP 2009, Part II, LNCS 5864.

[31] Wiggins, B., "The Generation of Panning Laws for Irregular Speaker Arrays using Heuristic Methods", AES 31st International Conference (2007)

[32] Wiggins, Bruce, "An Investigation into the Real-Time Manipulation and Control of Tree-Dimensional Sound Fields." PhD. Thesis, University of Derby. 2004.

[33] Wiggins, Bruce, "Reply to: One more Ambisonic to 5.0 Decoder," email to Sursound discussion list. (11/15/2006)

[34] Wikipedia contributors, "Wave field synthesis," *Wikipedia, The Free Encyclopedia,* http://en.wikipedia.org/wiki/Wave_field_synthesis (accessed September 1, 2010).

1. APPENDIX – EXAMPLE RUN

```
bash-3.2$ sph/sph 1 <configs/itu-110.txt |ambi_NLOpt/ambi_opt itu5-110sd

ambi_opt ver. 0.42
Using NLopt version 2.2.0.
num_speakers = 5

Input Array Projection:
  0    0.00    0.00   0.707107   1.000000   0.000000   0.000000
  1   30.00    0.00   0.707107   0.866025   0.500000   0.000000
  2  110.00    0.00   0.707107  -0.342020   0.939693   0.000000
  3 -110.00    0.00   0.707107  -0.342020  -0.939693   0.000000
  4  -30.00    0.00   0.707107   0.866025  -0.500000   0.000000

Singular Values of A:
[       2.017389       1.505339       1.078934       0.000000 ]

Decoder Matrix:
  0    0.00    0.00   0.102378   0.311541   0.000000   0.000000
  1   30.00    0.00   0.143329   0.240846   0.220649   0.000000
  2  110.00    0.00   0.512589  -0.396616   0.414684   0.000000
  3 -110.00    0.00   0.512589  -0.396616  -0.414684   0.000000
  4  -30.00    0.00   0.143329   0.240846  -0.220649   0.000000

n_spkr = 5

Constraints:
  #0 [+0.051189 +0.204756] [+0.155771 +0.623082] [+0.000000 +0.000000]
  #1 [+0.071665 +0.286658] [+0.120423 +0.481692] [+0.110325 +0.441298]
  #2 [+0.256294 +1.025178] [-0.793232 -0.198308] [+0.207342 +0.829368]
  #3 [+0.256294 +1.025178] [-0.793232 -0.198308] [-0.829368 -0.207342]
  #4 [+0.071665 +0.286658] [+0.120423 +0.481692] [-0.441298 -0.110325]


Initial fit...
#W             X              Y              #Speaker              VM
#+0.10237800   +0.31154100    +0.00000000    #00:   +0.0 deg         +0.0 deg
#+0.14332900   +0.24084600    +0.22064900    #01:  +30.0 deg        +42.5 deg
#+0.51258900   -0.39661600    +0.41468400    #02: +110.0 deg       +133.7 deg
#+0.51258900   -0.39661600    -0.41468400    #03: -110.0 deg       -133.7 deg
#+0.14332900   +0.24084600    -0.22064900    #04: - 30.0 deg        -42.5 deg
gain at 0 deg: 1.000001
vfit:   0.000001 (0.500000)
mfit:   0.000001 (0.000000)
efit:   0.328122 (1.000000)
avfit:  0.000000 (0.159155)
aefit:  0.289429 (0.159155)
avefit: 0.289429 (0.318310)
afit:   0.578858
esd:    0.308045 (1.000000)
Overall metric: 0.77436015
eval() called 0 times.
Searching (tmax=200.000000 sec., tol=0.000010, dirs=180)...
Max XTOL Reached, 1e-05
cpu_time =   4.75 sec.

Optimized fit...
#W             X              Y              #Speaker              VM
#+0.20475327   +0.15577050    +0.00000000    #00:   +0.0 deg         +0.0 deg
#+0.28665690   +0.17671753    +0.24883984    #01:  +30.0 deg        +54.6 deg
#+0.31815325   -0.25530467    +0.23280710    #02: +110.0 deg       +137.6 deg
#+0.31798849   -0.25517030    -0.23327622    #03: -110.0 deg       -137.6 deg
#+0.28665685   +0.17799036    -0.24836979    #04:  -30.0 deg        -54.4 deg
```

```
gain at 0 deg: 1.000000
vfit:   0.000004 (0.500000)
mfit:   0.373626 (0.000000)
efit:   0.304545 (1.000000)
avfit:  0.388068 (0.159155)
aefit:  0.390203 (0.159155)
avefit: 0.057793 (0.318310)
afit:   0.836063
esd:    0.029344 (1.000000)
Overall metric: 0.47615238
eval() called 111296 times.
5
   0.000000   0.000000   0.20475327   0.15577050   0.00000000
  30.000000   0.000000   0.28665690   0.17671753   0.24883984
 110.000000   0.000000   0.31815325  -0.25530467   0.23280710
-110.000000   0.000000   0.31798849  -0.25517030  -0.23327622
 -30.000000   0.000000   0.28665685   0.17799036  -0.24836979
bash-3.2$
```