Two-Dimensional Hybrid Spatial Audio Systems with User Variable Controls of Sound Source Attributes

Martin J. Morrell and Joshua D. Reiss

Centre for Digital Music, Queen Mary University of London, London, UK
martin.morrell@eecs.qmul.ac.uk

Abstract. This paper presents two novel hybrid spatial audio systems demonstrated for use in two-dimensional applications with their scalability to three-dimensions. The emphasis of these hybrid systems is to give further creative freedom to a composer, sound engineer or sound designer. The systems are principally based on the end result of Ambisonics spatial audio reproduction systems. Since Ambisonics systems are used primarily for temporary sound installations and exhibits, the use of B-Format can be unnecessary. Therefore these systems revert to producing channel based content rather than sound field content that is later separately decoded. The presented systems use the decoder as a real-time sound manipulation feature on a per sound source basis. A comparison is drawn between the two systems and each method is described as to how it can be used as part of a standard music production workflow.

Keywords: Ambisonics, variable-order, variable-decoder, polar pattern, octagon, spatial audio, surround sound, 2D, 3D.

1 Ambsionics Background

The work in this paper is based on Higher Order Ambisonics systems. Michael Gerzon led the original Ambsionics development team in the 1970s and wrote papers on the subject throughout his career [11–13]. Further work has been done to expand Ambisonics into Higher Order Ambisonics [3–5, 15] and to develop decoders, speaker layouts and evaluation of systems [1, 9, 10, 14, 17, 18, 31]. The basis of Ambisonics is to represent a three-dimensional auditory scene as a sound field representation that can later be reconstructed for any user speaker layout. An Ambisonics representation is based on a fixed order that is linked to the localisation attributes of sound sources. Ambisonics theory is based on spherical harmonics calculated from legendre polynomials.

$$Y_{mn}(N2D)(\theta,\phi) = \sqrt{2}\hat{P}_{mn}(\sin\phi) = \begin{cases} \cos n\theta & n \geq 0 \\ \sin n\theta & n < 0 \end{cases} . \tag{1}$$

$$\hat{P}_{mn}(\sin\phi) = \sqrt{(2-\phi_{0,n})\frac{(m-n)!}{(m+n)!}} P_{mn}(\sin\phi) . \tag{2}$$

M. Aramaki et al. (Eds.): CMMR 2012, LNCS 7900, pp. 58–81, 2013.
© Springer-Verlag Berlin Heidelberg 2013

Where Y_{mn} is the spherical harmonic, n is the degree, m is the order and $\hat{P}_{mn}$ the associated Legendre polynomial.

The above equations use the N2D normalisation scheme. Several schemes exist for Ambisonics and affect the maximum gain of each spherical harmonic. When these are applied to a monaural sound source a sound field representation is created and is known as B-Format. The 2D representation is based only on the angular value θ as $\phi = 0^o$. The spherical harmonic expansion of the sound field is truncated to a finite representation known as the Ambisonic order M and each prior order m is included, $0 \leq m \leq M$. For each included order m the degrees calculated are $n = \mp m$. The total amount of harmonics in the sound field representation is $2M + 1$.

Once encoded, Ambisonics material can be played back over various speaker layouts using a suitable decoder. The minimum number of speakers to correctly reproduce 2D Ambisonics is $2M + 2$ [22]. For a regular layout, i.e. one that has the speakers equally spaced, the angular separation is simply $360^o/L$ where L is the number of speakers for 2D reproduction. For a regular layout the decoder matrix can be calculated by using the Moore-Penrose pseudo-inverse matrix of the spherical harmonics at each speaker position.

$$\begin{pmatrix} Y_{(0,0)}(spk1) & Y_{(1,-1)}(spk1) & Y_{(1,1)}(spk1) & \dots Y_{(M,m)}(spk1) \\ \vdots & \vdots & \vdots & \ddots \vdots \\ Y_{(0,0)}(spkN) & Y_{(1,-1)}(spkN) & Y_{(1,1)}(spkN) & \dots Y_{(M,m)}(spkN) \end{pmatrix}^{\dagger} . \tag{3}$$

Gerzon specified criteria for low and high frequencies reproduction known as rV and rE vectors [11, 12, 14]. The given pseudo-inverse decoder results in the standard, rV, decoder matrix. To create a decoder that maximises the rE vector the decoder is then multiplied with gains g_{rE} based on each component's order and the system order.

$$g_{rE} = P_m(\text{largest root of } P_{M+1}) . \tag{4}$$

Furthermore the decoding can be changed to what is known as In-Phase decoding, using the $g_{In-Phase}$ coefficients, so that there are no negative gains used to create the sound's directionality.

$$g_{In-Phase} = \frac{M!}{(M+m)!(M-m)!} . \tag{5}$$

Ambisonics can be seen as creating a polar pattern of M^{th} order in the direction of the sound source where the polar pattern is sampled by discrete speaker positions. By increasing the amount of speakers the resolution of the polar pattern is increased. In turn, by increasing the order, the directionality is increased and by using different decoders as described above, the rear-lobe is altered.

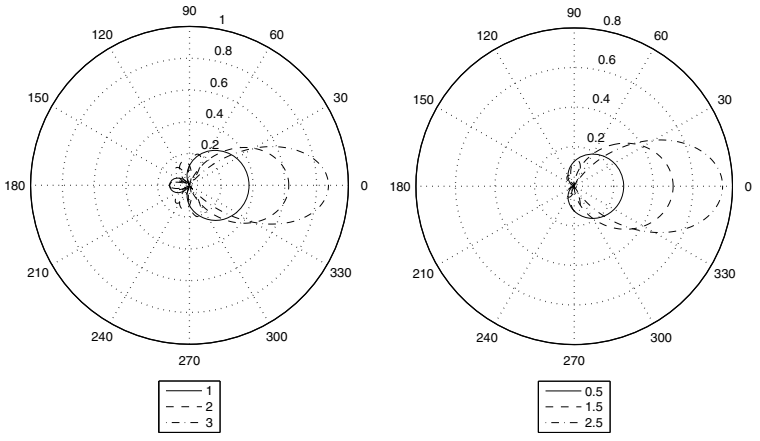
2 Variable-Order, Variable-Decoder Ambisonics

This section presents the novel idea of Variable-Order, Variable-Decoder Ambisonics. The concept allows for varying the reproduced polar pattern, and therefore the sharpness of localisation, by setting the order used to a non-integer value. Further to this, the idea of a variable-decoder is discussed that can alter the amount of rear lobe of the sampled polar pattern. The two variables are linked but not interchangeable. The order alters the width of the main lobe, whilst altering the amount and gain of, the rear lobes. The decoder alters the gain of rear lobes whilst consequently altering the width and gain of the main lobe.

2.1 Variable-Order

The result of encoding a monaural sound source to Ambisonics B-Format and then decoding it for a speaker layout is equivalent to applying a gain to the monaural sound and sending it to each speaker. Therefore in this described approach, the audio signal is not converted to B-Format. Instead, the gains are calculated numerically and applied based on the octagonal layout.

The variable-order is created by calculating the decoders of the identical type, for each order. Since we are dealing with an octagonal layout the orders used are 0 through 3. The spherical harmonic values are calculated for all included orders for the sound source location θ and speaker gains obtained. By using



(a) Orders 1, 2 and 3 **(b)** Orders 0.5, 1.5 and 2.5

Fig. 1. The reproduced polar pattern of a sound source at $\theta = 0^o$ for Ambisonics orders 1 through 3 are shown in (a). The half orders of 0.5, 1.5 and 2.5 are shown in (b).

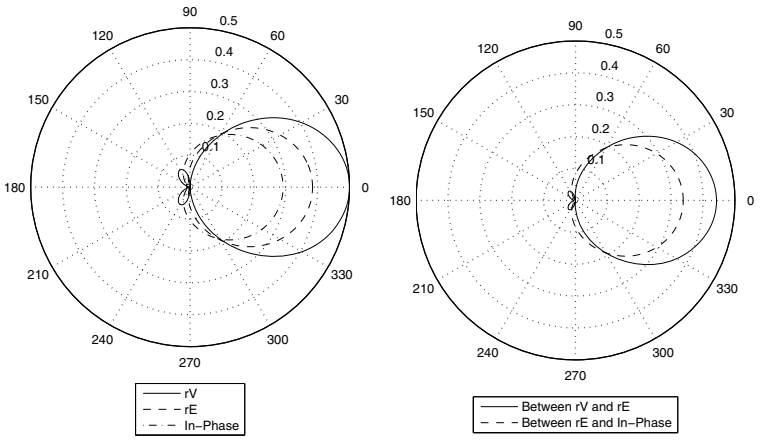
interpolation the variable-order can be created by a mixture of 0^{th} and 1^{st}, 1^{st} and 2^{nd}, and 2^{nd} and 3^{rd} speaker gains. Figure 1 (a) shows the sampled polar pattern for the whole orders. Figure 1 (b) shows the half orders using the variable-order approach. As can be expected, the polar pattern of half orders are directly between the whole orders. The variable-order approach can be used to create the polar pattern of any decimal value order representation. For an Ambisonics representation the gain of all speakers must equal one. This fact is important so that a sound source does not experience an overall gain boost when the variable-order is used as a creative feature.

2.2 Variable-Decoder

Three types of Ambisonics decoders have been presented in section 1 where each is used for a specific purpose. However, these decoders offer an aspect of creativity in being able to manipulate the rear lobe of the polar pattern, thus altering the shape of the sound source's polar pattern.

The variable-decoder can be calculated in the same manner as for the variable-order concept. By using a weighted ratio that equals 1 of two types of decoder, a variable pattern can be created. The weightings are calculated between rV and rE decoders and the rE and In-Phase decoders. This is because the rE polar pattern lies between the basic and In-Phase patterns.

Figure 3 shows the three decoders for order 1.5 on the left and the decoders half way between the rV and rE decoders and the rE and In-Phase decoders. The variable-decoder lies at the given ratio between the standard decoders.



(a) rV, rE and In-Phase decoders **(b)** Variable-decoders

Fig. 2. The three standard decoder types for order 1.5 are shown in (a) and the intermediate decoders in (b)

2.3 Observations

The proposed methodology creates a set of variable-order, variable-decoder speaker signals for an octagonal arrangement of speakers. The end result is sampling the sound scene at regular intervals of a third order polar pattern [6]. The resultant gain G_L for the speaker at position θ_L can be calculated by eq. (6). The sum of the gain of the orders must equal one; that is $\sum_{L=1}^{N} G_L = 1$.

$$G_L = a_0 + a_1 \cos(\theta + \theta_L) + a_2 \cos(2(\theta + \theta_L)) + a_3 \cos(3(\theta + \theta_L)) \; . \qquad (6)$$

Therefore the variable-order is equivalent to increasing the next order gain whilst the ratio of the prior orders' gains remain the same. The variable-decoder, is like altering the ratio between the a_0 and a_1 gain coefficients thus changing the base polar pattern, as well as altering the ratio between higher orders.

2.4 Test Case

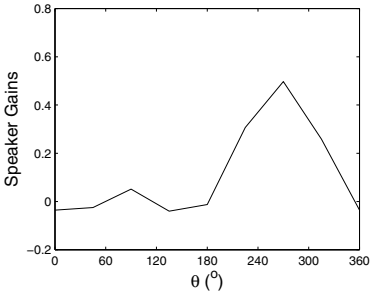
Figure 3 shows the plots for both second and third order using a variable-decoder of 0.8 rV and 0.2 rE for a sound source at $\theta = 93^o$. The lower plot (c) shows 2.2 variable-order. The resultant variable-order has a maximum point between the two whole orders and the other lobes are smoothed out. The secondary lobes become more like an In-Phase decoder. The sum of the speakers for the variable-order, variable-decoder remains 1. Hence no normalisation of the speaker signals is needed. Due to the changing of the secondary lobe gains, the decoder type attributes associated to integer orders are lost.

2.5 Calculating in the Decoder

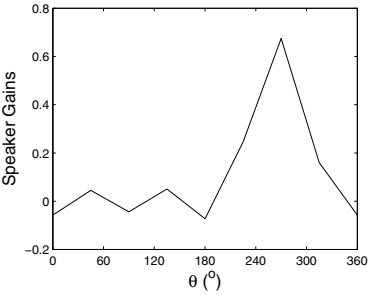
The methodology presented here to calculate the Variable-Order, Variable-Decoder Ambisonics has been to use a lookup table approach. First, all of the values between the two decoders for the lower and higher integer order are interpolated. Then the resultant variable-decoders for both orders are interpolated to produce the final signals. This has involved no creation of B-Format due to the speaker feeds directly being produced by multiplying the sound source by the resultant speaker gains. The same effect can be obtained by means of manipulating the decoder. To calculate a variable-order decoder directly, the $n = \mp m$ components for individual order $m = \lceil M \rceil$ of the variable order need to be multiplied by a factor, ν as shown in eq. (7), that is chosen by the user.

$$\begin{pmatrix} Y_{(0,0)}(spk1) & \dots & Y_{(\lceil M \rceil, -\lceil M \rceil)}(spk1) & Y_{(\lceil M \rceil, \lceil M \rceil)}(spk1) \\ \vdots & \ddots & \vdots & \vdots \\ Y_{(0,0)}(spkN) & \dots & Y_{(\lceil M \rceil, -\lceil M \rceil)}(spk1) & Y_{(\lceil M \rceil, \lceil M \rceil)}(spkN) \end{pmatrix}^{\dagger} \begin{pmatrix} 1 \dots & \nu & \nu \\ \vdots \dots & \ddots & \vdots \\ 1 \dots & \nu & \nu \end{pmatrix} . \qquad (7)$$

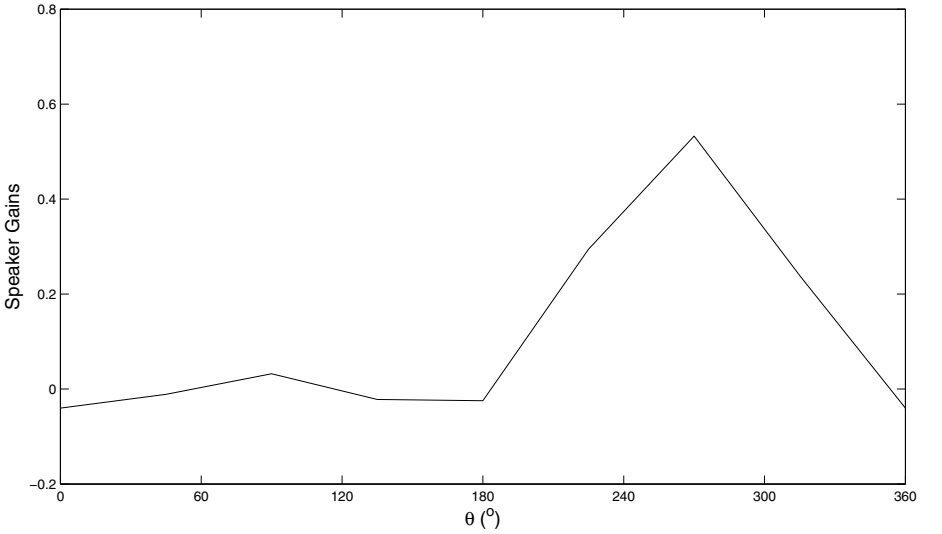
The decoder type gains can then be multiplied to eq. (7) by use of a further matrix. This uses the variable κ as the interpolation factor between the normal



(a) Second Order



(b) Third Order



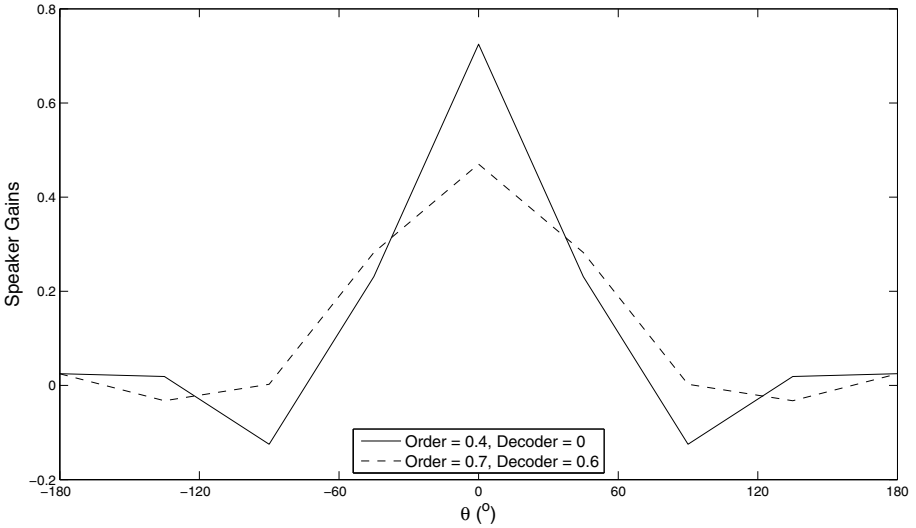
(c) 2.2 Order

Fig. 3. The speaker gains of $\theta = 93^o$ for second (a) and third (b) order and the gains for 2.2 order (c) using a variable-decoder as presented in section 2.2 with rV of 0.8 and rE of 0.2.

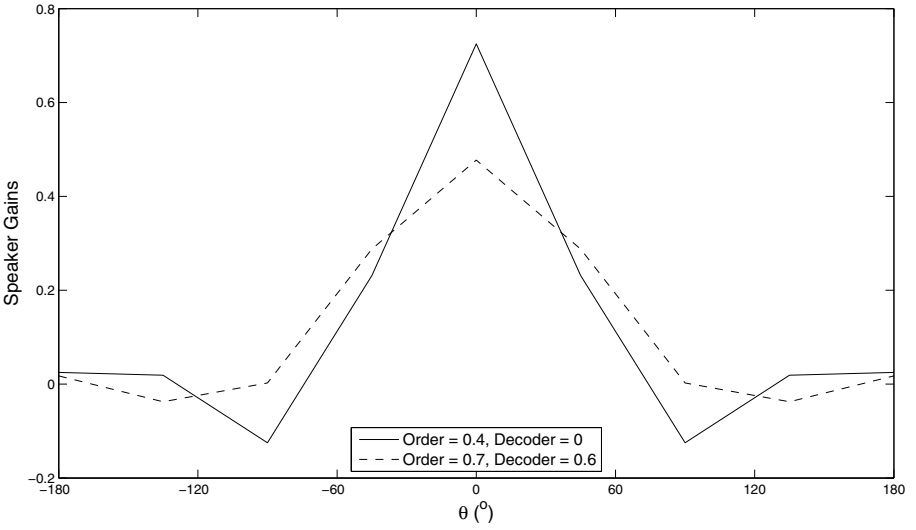
and g_{rE} or g_{rE} and $g_{In-Phase}$ gains as presented in section 1. The matrix to change the decoder type is given as:

$$\begin{pmatrix} 1-\kappa_{(0,0)}(spk1) & \dots & 1-\kappa_{(\lceil M\rceil,-\lceil M\rceil)}(spk1) & 1-\kappa_{(\lceil M\rceil,\lceil M\rceil)}(spk1) \\ \vdots & \ddots & \vdots & \vdots \\ 1-\kappa_{(0,0)}(spkN) & \dots & 1-\kappa_{(\lceil M\rceil,-\lceil M\rceil)}(spkN) & 1-\kappa_{(\lceil M\rceil,\lceil M\rceil)}(spkN) \end{pmatrix} \dots$$
$$\begin{pmatrix} \kappa g'_{(0,0)}(spk1) & \dots & \kappa g'_{(\lceil M\rceil,-\lceil M\rceil)}(spk1) & \kappa g'_{(\lceil M\rceil,\lceil M\rceil)}(spk1) \\ \vdots & \ddots & \vdots & \vdots \\ \kappa g'_{(0,0)}(spkN) & \dots & \kappa g'_{(\lceil M\rceil,\lceil M\rceil)}(spkN) & \kappa g'_{(\lceil M\rceil,\lceil M\rceil)}(spkN) \end{pmatrix}. \quad (8)$$

The use of eqs. (7) and (8) are compared to the initial interpolation, or lookup table method in fig. 4. The original methodology as described in the beginning of this paper is shown in (a) and the direct manipulation of the decoder approach in (b). It can be seen that they do in fact result in the same outcome.



(a) Lookup Table Approach



(b) In Decoder Calculation

Fig. 4. This figure shows the identically produced speakers' gains for a sound source placed at $\theta = 0^o$ using both the lookup table approach and calculating the variable-order and variable-decoder within the decoder.

2.6 rV, rE, Power and Energy of Variable-Ambisonics

widt To further examine this system's behaviour as a result of the variable-order and variabe-decoder, the rV and rE vectors proposed by Gerzon [3, 11, 12, 14, 31] will be evaluated, as well as the power and energy values upon which those metrics are based. The results are displayed in fig. 5. Between zeroth and first order the rV linearly goes between zero and one. This is a somewhat obvious result as zeroth order has no directionality and first order is the minimum for a directional response. The power for which is the denominator for calculating rV, resulting in a constant value of one for each value of θ. This is to be expected since the speakers are regularly spaced and meet the minimum N criterion of $N > 2M + 1$. Although an expected result, it does determine that variable-order is valid for rV cues and that the power is constant irrespective of variable-order. The rE result however is an interesting one. We can see that the rE is maximised at approximately $m + 0.6$ orders, not at the whole integer orders. This can be attributed to the polar pattern produced becoming more like an rE decoder polar pattern than an rV decoder. The response of rE is not linear to the $m+0.6$ points but curved either side. Conversely the energy is not maximised at $m+0.6$ orders. The energy between integer orders shows an exponential growth. These results indicate that the variable-order effect should be examined for the variable-decoders.

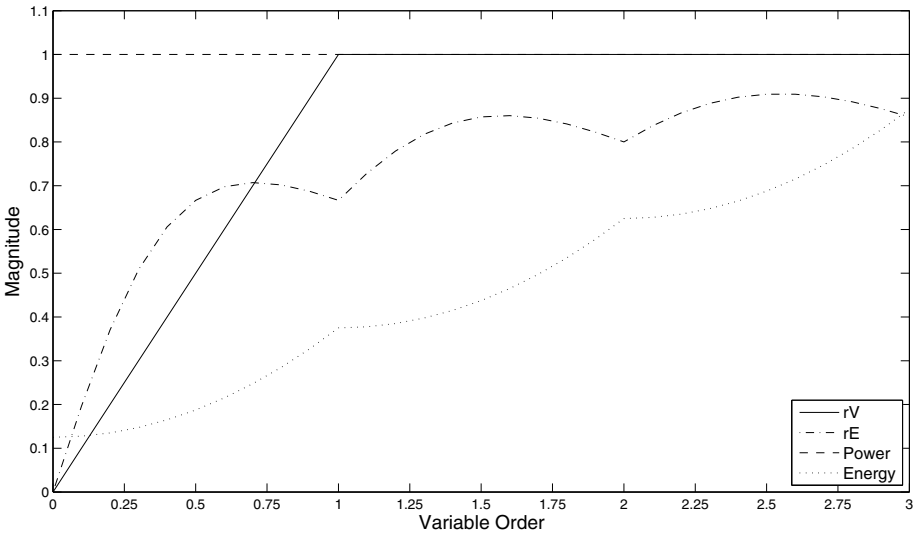
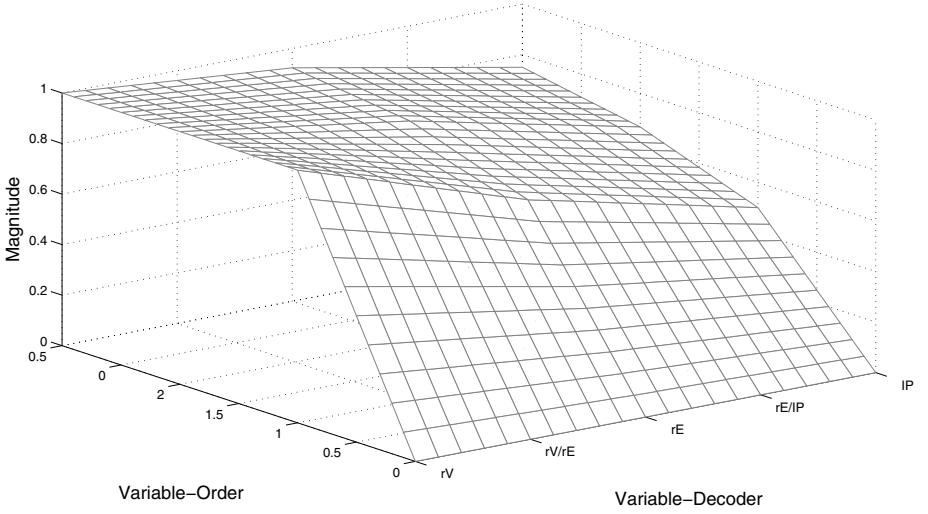
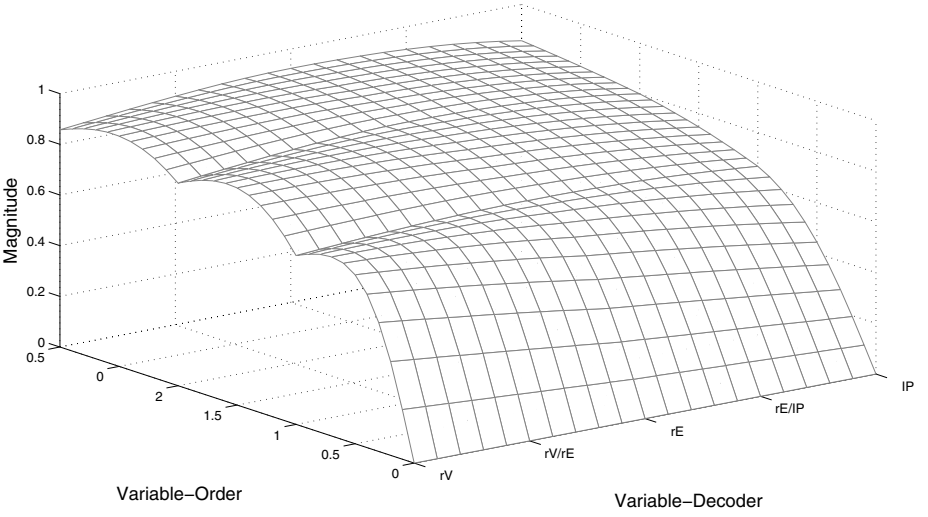


Fig. 5. The rV and rE vector, power and energy values for zeroth through to third order for Variable-Order, Variable-Decoder Ambisonics. The decoder used is the regular rV decoder from the pseudo-inverse function. The values are independent of θ since in a regular speaker array the rV, rE, power and energy are constant for $N > 2M + 1$.

Figure 6 shows the rV vector (a), rE vector (b) and energy (c) values for variable-orders of variable-decoder. The power is not shown in this figure as for all cases it has a value of one. Thus for power the Variable-Order, Variable-Decoder Ambisonics satisfies the constant power condition. The rV plot shows that the

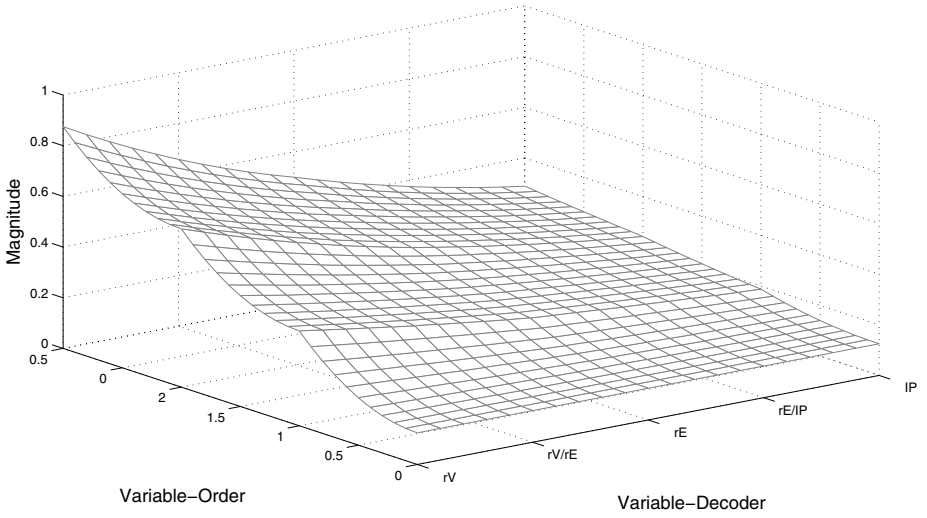


(a) rV Vector



(b) rE Vector

Fig. 6. rV, rE vector (figures a and b) and Energy attributes (figure c) of Variable-Order, Variable-Decoder Ambisonics. Under various conditions the intermediate order values are not linear between the normal integer orders.



(c) Energy

Fig. 6. (*Continued*)

rV vector is reduced as the variable-decoder increases from zero. For all variable-decoder values the rV linearly increases between integer order values. For the same variable-order the rV value is linear between rV and rE decoders, and then the rE and In-Phase decoders. The rE values show that the curved nature of the rV decoder for variable-order is smoothed out as the variable-decoder increases above one. The rE decoder shows a slight curve between the integer orders in rE value, but for the In-Phase decoder the increase is linear between integer orders. The results show that for the $m+0.6$ order the rV decoder now performs similarly to the In-Phase decoder for rE value. Finally, the energy plot shows the reduction of the curvature of energy value as the variable-decoder is increased from zero order. It shows that the energy is always greatest for variable-decoder of zero, rV decoder, and falls in value as the variable-decoder is increased. The maximum energy values for a variable-order are for the rV decoder. It can be determined that the greater energy value does not result in the best rE vector value.

To conclude, using variable-order can result in a heightened rE vector value for Variable-Order, Variable-Decoder Ambisonics and is better than the next highest integer order. This is where the rE value is maximised around the $m+0.6$ point.

2.7 Composition/Production Tool Implementation

The tool to use the variable-order and variable-decoder methodology has been implemented in the Max/MSP 5 software environment for Mac OSX. The tool is designed to receive audio signals from digital audio workstations (DAW), e.g. via Jack or Soundflower, for a total of 16 monaural and 4 stereo signals. The controls

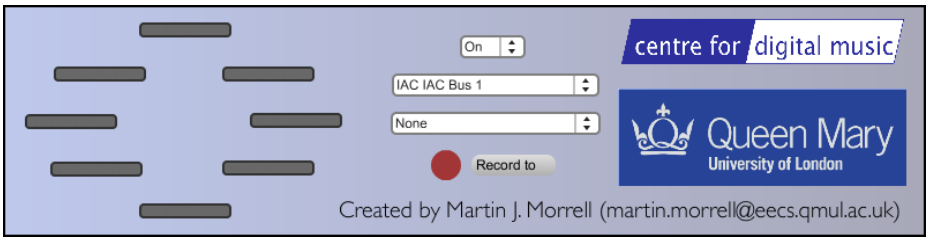


Fig. 7. The user interface for the Variable-Order, Variable-Decoder Ambisonics spatialisation tool

for each channel are sent via midi commands which are stored in a digital audio workstation project. User Control Panels were built for this function for the Cubase/Nuendo environment, but VSTs, AUs or other midi capable software can be used to control the settings for each sound source. The premise for this is that no extra saved data is needed that cannot be stored in a common DAW project.

Figure 7 shows the user interface for the tool. The user definable parameters on the interface are On/Off, midi driver, audio driver and where to save a recorded file. The interface has eight LED style meters for monitoring the signal level going to each speaker so that distortion can be avoided. Since users may not always have an eight speaker layout available, a binaural (over headphones) mix is simultaneously available.

2.8 Distance

Distance is a user definable parameter and is accomplished by gain attenuation only. No delay has been included since for music purposes pitch shifting of sound sources will affect the overall tonal effect and harmonicity of the work, alter the speed and therefore ensemble timing of the music and finally can include zipper noise. The $1/r$ inverse law is used to implement the gain change at sources greater than 1.0 where the maximum value is 10. Since the roll off of $1/r$ simulates anechoic conditions, the feature is given as a creative and not real-world application. For sources that are placed inside the speaker layout the distance calculation changes to $1 + \cos(90^o r)$ so that infinite gain is not reached. The maximum gain at the central position is 2.0, or approximately +6dB.

2.9 Inside Panning

Sound sources that have a distance of less than 1.0 are placed inside the speaker array. This is done by altering the reproduced polar pattern. If the order of reproduction is 1 then this is the same as cancelling out the 1st order spherical harmonics and doubling the zeroth order spherical harmonics. This methodology was first presented in [19]. For the case of third order two-dimensional Ambisonics, the maximum allowed in this tool, the inside-panning function is expanded.

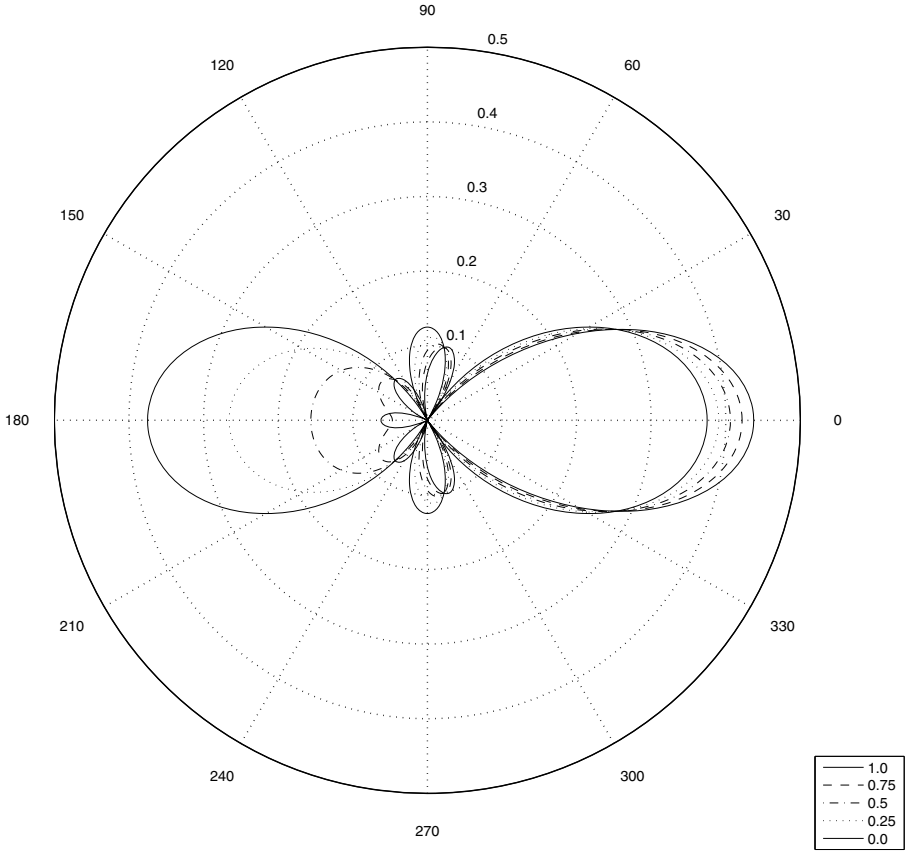


Fig. 8. The change in polar pattern exerted by a third order sound source as it is moved from a distance of 1.0 to 0.0 to be placed in the middle of the speaker array

The result is that even orders are doubled and odd orders are cancelled out. This again is all done as numerical and not audio calculations. Figure 8 shows the polar pattern change going from 1.0 to 0.0. The result is strong lobes from opposite poles giving the psychoacoustic illusion of the sound source being at the centre of the array.

2.10 Reverberation

Reverberation is produced in the tool by transforming the sound source into B-Format and processing it through either the Wigware VST (Virtual Studio Technology) reverberation plugin [31] based on the freeverb algorithm or by using a convolution plugin using B-Format impulse responses, such as those available [20, 30].

2.11 Original Composition

The authors commissioned a composer to create a multimedia piece that used the creative aspects of the Variable-Order, Variable-Decoder tool and technology. The piece was originally written for speakers and video projection. The work has subsequently been shown at various events and is available as a binaural version online. The following describes the work in the composer's words:

> *The composition was written with the intention of being realised through the use of the Variable-Order, Variable-Decoder Ambisonics two-dimensional tool for an octagonal speaker arrangement (ed. as described in section 2.7). This allowed for a greater creativity in conveying layers within the musical scene at varying distances, widths and positions across a horizontal plane. The tool helped create a greater sense of foreground and background. Placing the 'mechanical' elements in the distance and wide when in city surroundings for example, but brought to the centre and narrowed when portraying a specific man-made character/element such as the record player, horse, dog and heart. The cello and violin parts were generally placed in the stereo field in accordance with their on-screen presence but the distance feature was employed to convey the strength of the character's emotions, getting closer at climax points to create intensity for the listener. The tool was used most creatively in trying to achieve a sense of movement through swirling musical layers around the full range of the eight speakers for example in the rapid bustling city and spiral staircase scene, equally in the slow panning of the opening and pier scenes.*

When working with the composer on the mixing stage of the work, the effect of the source width could be clearly heard, as well as the distance change. The aspect of width helped enhance the use of the space surrounding the listener and the use of distance emphasised the busy nature of the world being portrayed by objects coming and going. The use of rotating the sound field was successfully used to indicate character movement and disorientation. The use of the variable-decoder was sparse, partly due to overlapping of the variable-order control where both alter the sound source width, but where the variable-order is far more intuitive to the user, in this case, the composer.

3 Variable-Polar Pattern Reproduction

Since the final output of an Ambisonics reproduction to the speakers is the same as sampling a polar pattern [3] exhibited by the audio source material, a new method is created whereby the intermediary B-Format and decoding is omitted.

Eargle [6, 21] gives two formulations for calculating higher order polar patterns. The first is given for calculating cardioid patterns in the form of $G = (0.5 + 0.5\cos(\theta))cos^{(M-1)}(\theta)$ for the M^{th} order, which is expanded for any base polar pattern in eq. (9) below. We define a base polar pattern as that created as

a mixture of zeroth, A, and first, B, order components to calculate a gain G at horizontal angular position θ. Where $A + B = 1$ is constant.

$$G = (A + B\cos(\theta))\cos^{(M-1)}(\theta) \,. \tag{9}$$

The second equation for a higher order pattern is given as the product of two or more first order microphone patterns:

$$G = (A_1 + B_1\cos(\theta))(A_2 + B_2\cos(\theta))\ldots(A_M + B_M\cos(\theta)) \,. \tag{10}$$

where $A_{1\cdots M}$ and $B_{1\cdots M}$ are the zeroth and first order terms for each order. To keep controls to a minimum we can limit the possible polar patterns so that $[A_1, B_1] = [A_2, B_2] = [A_M, B_M]$. By using this identity we can use a variable order for M below:

$$G = \begin{cases} (A + B\cos(\theta))^M & M \text{ is odd} \\ -(|A + B\cos(\theta)|^M) & M \text{ is even} \end{cases} . \tag{11}$$

Figure 9 shows the differences for calculating omni-directional, cardioid and figure-of-eight polar patterns using eq. (9) method and eq. (11) method. It can be seen for method A that for the omni-directional above first order the pattern changes to a figure-of-eight pattern of order $M - 1$. When looking at higher order cardioid for method A, we see that rear lobes are formed on the cardioid pattern. Finally the figure-of-eight pattern for method A behaves as expected and so are not shown; as the order increases the angular distance between the -3dB points decreases, giving a narrower polar pattern around the maxima and minima points. In the results of eq. (11) method, the omni-directional pattern remains omni-directional at all orders. The cardioid pattern for eq. (11) method does not develop rear lobes, but becomes a beam like pattern. Finally the figure-of-eight pattern for eq. (11) method behaves like that of method A, as we expect; a tighter figure-of-eight with greater side rejection. From these findings eq. (11) method will be used as it produces the most useful higher order polar patterns.

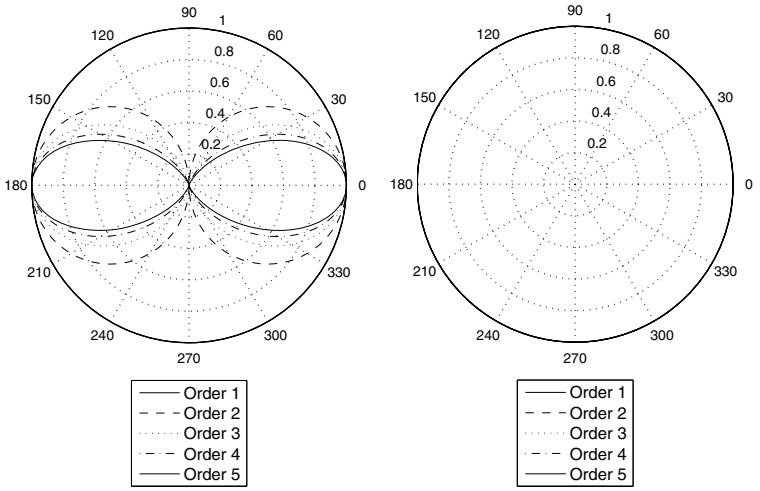
The gain applied to the L^{th} speaker is given as:

$$G_L = \begin{cases} (A + B\cos(\theta - \theta_L))^M & M \text{ is odd} \\ -(|A + B\cos(\theta - \theta_L)|^M) & M \text{ is even} \end{cases} . \tag{12}$$

To maintain a constant level whilst varying the order and/or polar pattern, like in the variable-Ambisonics method, a factor C is needed to scale the speaker gains:

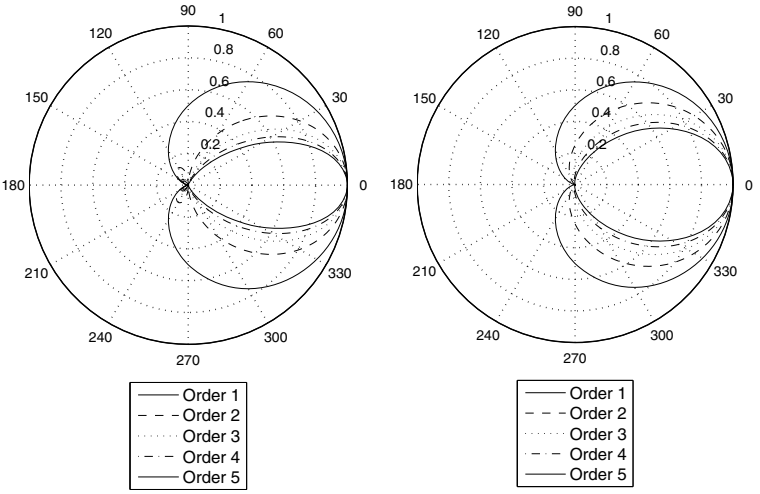
$$C = \frac{1}{\sum_{L=1}^{N} G_L} \,. \tag{13}$$

where the maximum order is based on N number of speakers being used, $M = (N - 2)/2$. Note that this will give a variable order and using the $\lfloor\rfloor$ function will give the highest integer order available.



(a) Equation (9) Method Omnidirectional

(b) Equation (11) Method Omnidirectional

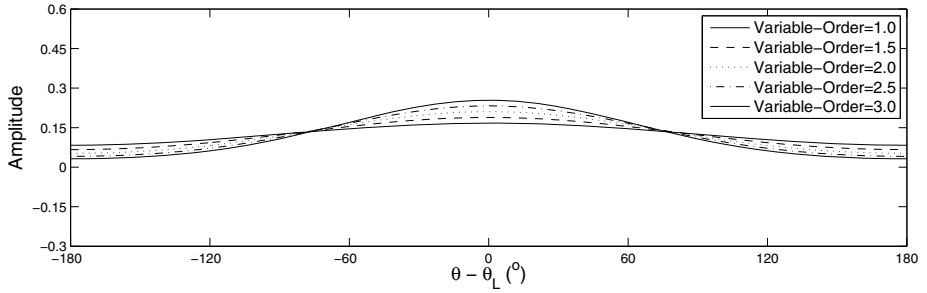


(c) Equation (9) Method Cardioid

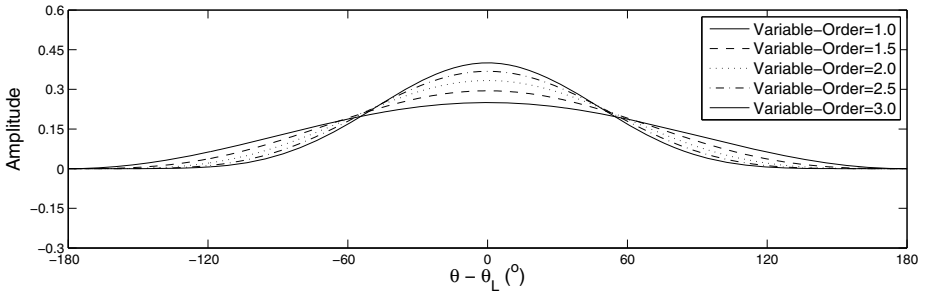
(d) Equation (11) Method Cardioid

Fig. 9. Comparison of pattern methods A from eq. (9) and B from eq. (11) for omni-directional (a and b) and cardioid (c and d) as discussed in section 3

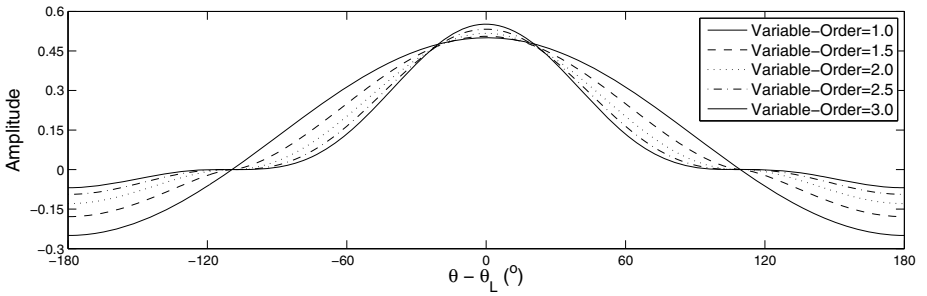
The produced gains for $\theta - \theta_L$ are shown in fig. 10. The sub-cardioid reproduction increases directivity with variable-order whilst the rear and side of the pattern is reduced in gain. The cardioid pattern has a constant zero point at the anti-pole, where the directivity and gain of the single positive lobe increases



(a) Sub-cardioid



(b) Cardioid



(c) Hyper-Cardioid

Fig. 10. Variable orders 1.0 to 3.0 for sub-cardioid (a), cardioid (b) and hyper-cardioid (c) polar patterns

with order. The hyper-cardioid pattern has a single negative lobe at the anti-pole of the main positive lobe. With an increase in order the directivity and gain of the main lobe increases whilst the negative lobe decreases in gain. Since this method uses a base polar pattern, of which the order can be changed variably, a user of a system can see the change in polar pattern easily. The figure-of-eight polar pattern poses a problem. Due to the equal gain of opposite polarities at anti-poles, the gains tend to infinity because of the cancellation when calculating

C in eq. (13). This also creates a problem since a speaker signal would have a maximum above unity gain. For this reason the base polar pattern should be limited so that $0 \leq A \lesssim 0.75$ in eq. (12).

The calculation of speaker gains in this way is similar to 3DVMS [2, 7, 16, 29], three dimensional virtual microphone synthesis, where higher order cardioid patterns are used. This has been compared to Ambisonics by Manola et. al. [16], although in their findings there were errors of localisation circa 180^o which questions the methods used and the validity of the results.

3.1 Real-Time Application

A demonstration application was built using Max/MSP that is controlled and fed audio by a digital audio workstation. Audio is sent from each track using outputs via Jack OS X audio router as monaural sound sources. Control data is sent from a VST audio plugin on each audio track using the OSC (Open Sound Control) protocol [32] using a similar, but reverse, idea to that described in [8]. The audio plug-in does not process the audio in any way as its only use is to communicate OSC commands in this system environment.

The VST presents controls to the user; Azimuth, Pattern, Order and Speakers. The Azimuth control is ranged [-180 180]^o anti-clockwise. The Pattern control varies the base polar pattern. The Order control alters the variable order of the sound source. This control's range is altered by the Speakers control as described in section 3. Therefore it can be set as a relative maximum order, especially if the audio mixture is going to be played back over different speaker configurations. The Speaker control has the range [4 12] in whole integers to represent the amount of speakers in the reproduction array. Finally the VST has 20 programs. These programs are presets to change the audio track that the VST is altering in the application. When changing program the other controls remain the same.

The Max/MSP application presents the user with minimal controls since they are for the most part received from the VSTs within the DAW project. The user can turn audio processing on/off, select the sound source's graph to be plotted, view the number of speakers being used and see output meters for the 12 possible speakers. Of most interest to a user are the graphs that are plotted. This is a plot of the polar pattern being used by the chosen sound source. The positive lobe is shown in red and the negative in blue within the applications display window. This is plotted on top of up to twelve black circles representing the speaker positions. This gives the user visual feedback of how the controls of the VST are affecting the sound source reproduction. The graph to make things clear is normalised, meaning eq. (13) is ignored for plotting purposes to avoid confusion to the user.

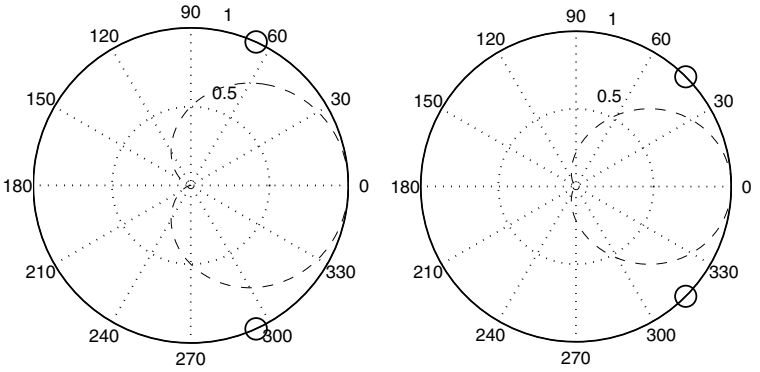
3.2 Vector Driven Variable-Polar Pattern Reproduction

Vector Base Amplitude Panning [23–28] in two dimensions gives the same result as the cosine/sine power panning law resulting between the two speakers neighbouring the sound source. This fact can be exploited to calculate the highest

order reproducible by the neighbouring speakers. If the sound source is assumed to be directly inbetween the speakers and that the polar pattern reproduced is that of a cardioid, where the most highly directional sound source is desired, then the highest order can be calculated. The differing cardioid patterns are shown for different speaker separation in fig. 11. Equation (14) shows the calculated highest order, M_V, under these assumptions.

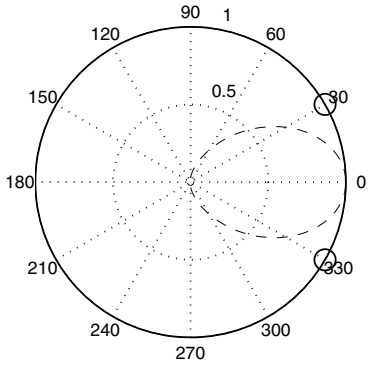
$$M_V = \frac{\log\left(\frac{1}{\sqrt{2}}\right)}{\log\left(0.5 + 0.5\cos\left(\theta_{ML}\right)\right)} \,. \tag{14}$$

Where θ_{ML} is the angle between the mid-point of the neighbouring speakers and one of the speakers, or put another way; the angular separation between the neighbouring speakers divided by two.

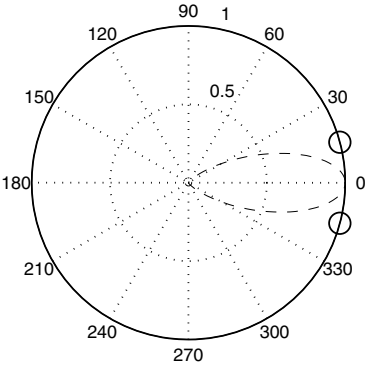


(a) Speaker separation θ_{ML} of 65.53^o.

(b) Speaker separation θ_{ML} of 45^o.



(c) Speaker separation θ_{ML} of 30^o.



(d) Speaker separation θ_{ML} of 15^o.

Fig. 11. The cardioid patterns calculated for four different speaker separations; 65.53^o, 45^o, 30^o and 15^o. The respective M_V values are 1.00, 2.19, 5.00 and 20.17.

This could be expanded for any polar pattern by taking the logarithm of eq. (11) in the denominator of eq. (14). Caution has to be taken when using polar patterns with negative lobes. It is not guaranteed, due to the speaker placement, that the negative gains will be reproduced at all. This could result in the exhibited polar pattern varying with the source position as M_V is changing.

It is, however, not guaranteed that the sound source will be in the middle of the two neighbouring speakers. When at a speaker's location the sound source is only reproduced by that speaker alone and therefore the calculated order would be infinite as there is no directional truncation. To overcome this problem the two neighbouring midpoints between the closest three speaker are found. The highest order of both midpoints is found using eq. (14) and then interpolated based on the angular distance between the sound source and the speaker midpoints. Finally the interpolated M_V is used in eq. (12) to find the initial speaker gains and the power kept constant by eq. (13). This procedure is shown in the flow diagram in fig. 12.

3.3 Expanding to Three-Dimensions

Expansion from the two-dimensional formulae used so far to three-dimensions is trivial. Expanding this theory to the three-dimensional case requires replacing the $\cos(\theta)$ terms to $\cos(\theta)\sin(\phi)$ in eqs. (10), (11) and (12). The resulting three-dimensional polar pattern is thus given as:

$$G_L = \begin{cases} (A + B\cos(\theta - \theta_L)\sin(\phi - \phi_L))^M & M \text{ is odd} \\ -\left(|A + B\cos(\theta - \theta_L)\sin(\phi - \phi_L)|^M\right) & M \text{ is even} \end{cases} . \quad (15)$$

where ϕ is the elevation angle of the sound source and ϕ_L the elevation angle of the speaker gain being calculated. The speaker gains will need normalising after calculation using eq. (13).

4 Comparison

In this section we present a comparison between the Variable-Order, Variable-Decoder Ambisonics system and the Variable-Polar Pattern Reproduction system.

It can be seen in the difference between the speaker gains shown in figs. 1, 3 and 10 that the former system gives a higher degree of directionality due to higher gain at the main lobe position. However, it does also introduce more rear lobes of both negative and positive gain, whereas the latter system retains the amount of rear lobes throughout the change of variable-order. The controls of altering a base polar pattern and variable-order are intuitive to an end user and have a clear distinction when looking at the plots of altering one or another of the parameters. With the first system this is not the case and the two variable controls both alter the same attributes of the polar pattern, although in different ways.

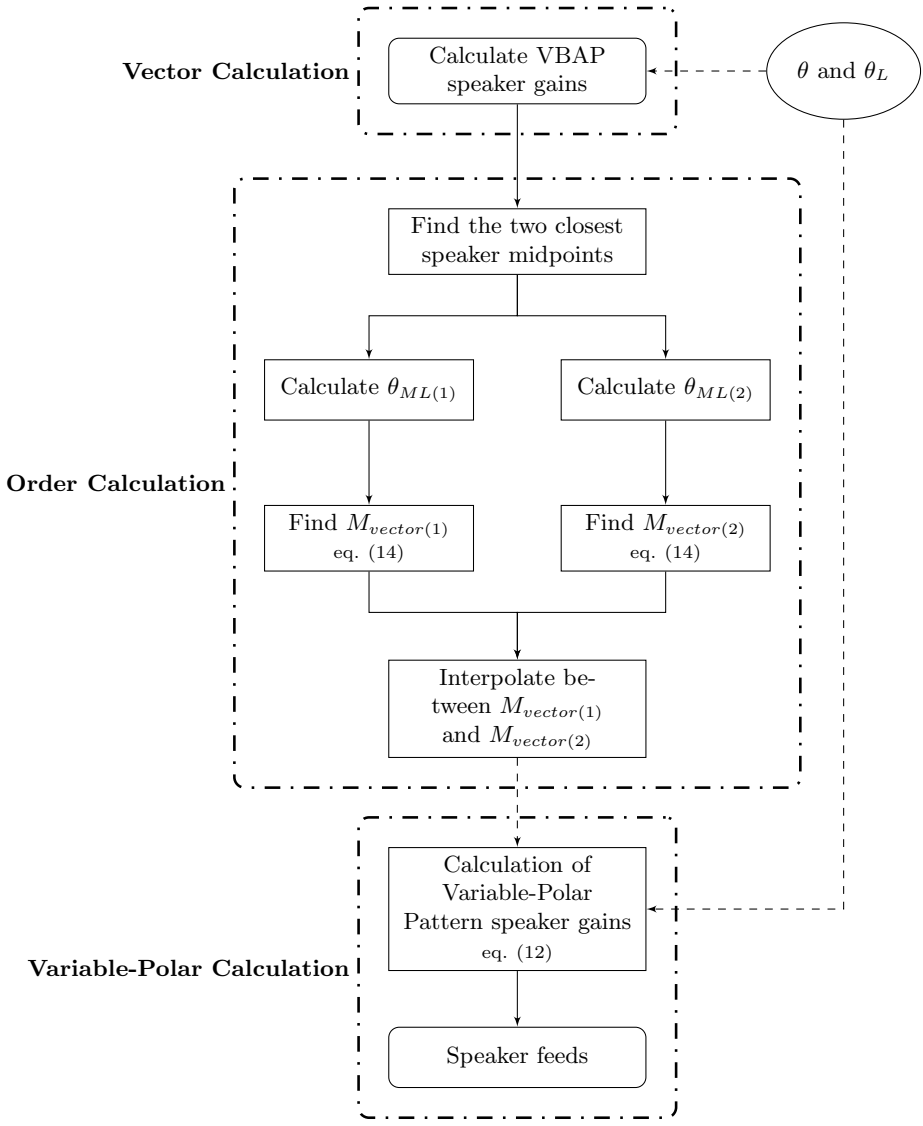
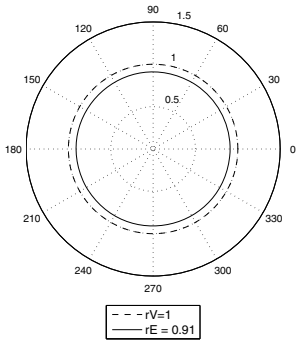
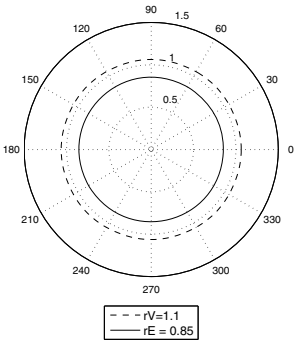


Fig. 12. Block diagram of the calculation of the Vector Driven Variable-Polar Pattern Reproduction. The Variable-Polar Pattern block could be exchanged for Variable-Order, Variable-Decoder Ambisonics or other spatialisation method.

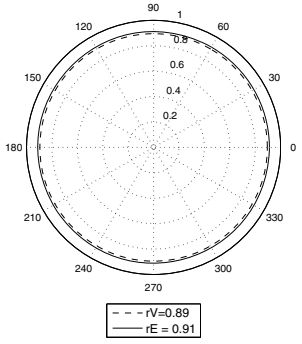
A comparison between the two methods can be drawn using the rV and rE vectors. Figure 13 shows this comparison using 2.5 order. The left column shows the results for the Variable-Order, Variable-Decoder Ambisonics system and the right column the results for the Variable-Polar Pattern Reproduction system. The In-Phase decoder and cardioid pattern both produce identical polar patterns



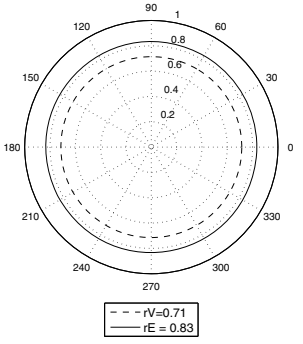
(a) Variable-Ambisonics rV Decoder



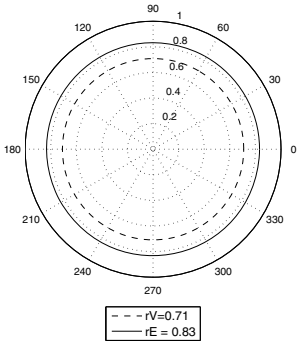
(b) Hyper-Cardioid Reproduction



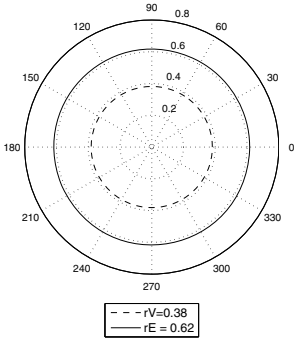
(c) Variable-Ambisonics rE Decoder



(d) Cardioid Reproduction



(e) Variable-Ambisonics In-Phase Decoder



(f) Sub-Cardioid Reproduction

Fig. 13. Comparison of variable-ambisonics and variable source pattern using rV and rE vectors to represent low and high frequency directional cues presented in section 4 for order 2.5

and so have the same rV and rE values. High frequencies are localised better than lower frequencies. The rV decoder and hyper-cardioid are similar in that they both have rear negative lobe(s). The hyper-cardioid has an rV above 1.1 which is unseen in Ambisonics unless the decoding is done for an order that the speakers cannot be replayed on and in that case is an error. The rV decoder however has better high frequency localisation than the hyper cardioid. The sub-cardioid, as one might expect, has poor localisation for both high and low frequencies.

5 Conclusion

Two novel spatial audio systems have been presented and used in real applications that give end users, such as composers, musicians, sound engineers or sound designers, further creative freedom of spatial audio reproduction other than angular position and distance attenuation. The systems have been based on the theoretical underpinnings of Higher Order Ambisonics, however, by eliminating the use of B-Format as a sound scene representation results in a channel based approach like that of stereo, 5.1 and 7.1. It could be argued that by removing the sound field representation format that it has lost one of the best traits of Ambisonics, although in many situations such spatial audio systems are designed for a particular exhibition or speaker reproduction environment where the B-Format signal is not published or shared.

The first system presented here has been used to produce an animation sound track exploring the creative use of the system. From this experience and the composer's feedback the second system was developed. Overall the second system offers intuitive user controls and a wider degree of freedom.

Examples of both these systems are available as binaural sound tracks for playback over headphones and can be seen at:
www.elec.qmul.ac.uk/digitalmusic/audioengineering/spatialaudio/index.html.

Acknowledgments. This research was supported by the Engineering and Physical Sciences Research Council [grant number EP/P503426/1].

References

1. Benjamin, E.: Ambisonic loudspeaker arrays. Audio Engineering Society Convention 125 (October 2008)
2. Capra, A., Chiesi, L., Farina, A., Scopece, L.: A spherical microphone array for synthesizing virtual directive microphones in live broadcasting and in postproduction. In: Proceedings of 40th AES International Conference, Spatial Audio: Sense of the Sound of Space, Tokyo, Japan, October 8-10 (2010)
3. Daniel, J.: Reprèsentation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimèdia. Ph.D. thesis, l'Universite Paris 6 (2000)
4. Daniel, J.: Spatial sound encoding including near field effect: Introducing distance coding filters and a viable, new ambisonic format. In: Audio Engineering Society Conference: 23rd International Conference: Signal Processing in Audio Recording and Reproduction (May 2003)

5. Daniel, J., Moreau, S.: Further study of sound field coding with higher order ambisonics. Audio Engineering Society Convention 116 (May 2004)
6. Eargle, J.: The Microphone Book. Focal Press (2001)
7. Farina, A., Binelli, M., Capra, A., Armelloni, E., Campanini, S., Amendola, A.: Recording, simulation and reproduction of spatial soundfields by spatial pcm sampling (sps). In: International Seminar on Virtual Acoustics, Valencia, Spain, November 24-25 (2011)
8. Freed, A., Zbyszynski, M.: Osc control of vst plug-ins. In: Open Sound Control Conference, July 30 (2004)
9. Furness, R.K.: Ambisonics-an overview. In: Audio Engineering Society Conference: 8th International Conference: The Sound of Audio (May 1990)
10. Furse, R.W.: Building an openal implementation using ambisonics. In: Audio Engineering Society Conference: 35th International Conference: Audio for Games (February 2009)
11. Gerzon, M.A.: Periphony: With-height sound reproduction. J. Audio Eng. Soc. 21(1), 2–10 (1973)
12. Gerzon, M.A.: Practical periphony: The reproduction of full-sphere sound. Audio Engineering Society Convention 65 (February 1980)
13. Gerzon, M.A., Barton, G.J.: Ambisonic decoders for hdtv. Audio Engineering Society Convention 92 (March 1992)
14. Heller, A., Lee, R., Benjamin, E.: Is my decoder ambisonic? Audio Engineering Society Convention 125 (October 2008)
15. Käsbach, J., Favrot, S.: Evaluation of a mixed-order planar and periphonic ambisonics playback implementation. Forum Acusticum 2011 (2011)
16. Manola, F., Genovese, A., Farina, A.: A comparison of different surround sound recording and reproduction techniques based on the use of a 32 capsules microphone array, including the influence of panoramic video. In: AES 25th UK Conference: Spatial Audio in Today's 3D World, York, UK, March 25-27 (2012)
17. Menzies, D.: W-panning and o-format, tools for object spatialization. In: 22nd International Conference on Virtual, Synthetic and Entertainment Audio (June 2002)
18. Moore, D., Wakefield, J.: The design of ambisonic decoders for the itu 5.1 layout with even performance characteristics. Audio Engineering Society Convention 124 (May 2008)
19. Morrell, M.J., Baume, C., Reiss, J.D.: Vambu sound: A mixed-technique 4-d reproduction system with a heightened frontal localisation area. In: AES 25th UK Conference: Spatial Audio in Today's 3D World, York, UK, March 25-27 (2012)
20. Murphy, D.T., Shelley, S.: Openair: An interactive auralization web resource and database. Audio Engineering Society Convention 129 (November 2010)
21. Olson, H.F.: Elements of Acoustical Engineering, 2nd edn. D. Van Nostrand Comapny, Inc. (1957)
22. Poletti, M.A.: Three-dimensional surround sound systems based on spherical harmonics. J. Audio Eng. Soc. 53(11), 1004–1025 (2005)
23. Pulkki, V.: Virtual sound source positioning using vector base amplitude panning. J. Audio Eng. Soc. 45(6), 456–466 (1997)
24. Pulkki, V.: Creating generic soundscapes in multichannel loudspeaker systems using vector base amplitude panning in csound synthesis software. Organised Sound 3(2), 129–134 (1998)
25. Pulkki, V.: Uniform spreading of amplitude panned virtual sources. In: Proceedings of the 1999 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, pp. 187–190. Mohonk Mountain House, New Paltz (1999)

26. Pulkki, V.: Generic panning tools for max/msp. In: The International Computer Music Conference, Berlin, Germany, pp. 304–307 (August 2000)
27. Pulkki, V.: Spatial Sound Generation and Perception By Amplitude Panning Techniques. Ph.D. thesis, Helsinki University of Technology (2001)
28. Pulkki, V., Karjalainen, M.: Directional quality of 3-d amplitude-panned virtual sources. In: The 7th International Conference on Auditory Display, Espoo, Finland, pp. 239–244 (July 2001)
29. Scopece, L., Farina, A., Capra, A.: 360 degrees video and audio recording and broadcasting employing a parabolic mirror camera and a spherical 32-capsules microphone array. In: IBC 2011, Amsterdam, September 8-11 (2011)
30. Stewart, R., Sandler, M.: Database of omnidirectional and b-format room impulse responses. In: ICASSP 2010 (2010)
31. Wiggins, B.: Has ambisonics come of age? Proceedings of the Institute of Acoustics 30(pt. 6) (2008)
32. Wright, M., Freed, A., Momeni, A.: Opensound control: State of the art 2003. In: Proceedings of the 2003 Conference on New Interfaces for Musical Expression (NIME 2003), Montreal, Canada (2003)