WFS and HOA: Simulations and evaluations of planar Higher Order Ambisonic, Wave Field Synthesis and Surround Hybrid Algorithms for lateral spatial reproduction for in theatre

A. Vilkaitis[1], B. Wiggins[2]

[1] *University of Derby, United Kingdom, Email: a.vilkaitis1@derby.ac.uk*

[2] *University of Derby, United Kingdom, Email: b.j.wiggins@derby.ac.uk*

Abstract

Wave Field Synthesis and Higher Order Ambisonics are both spatialisation techniques that could be applied to theatre sound design, but practicalities such as the number of loudspeakers and space required limit their use. Practical setups could consist of a planar array across the stage (for performer localisation) and surround speakers around the auditorium in different configurations (for ambience). This research simulates the use of extrapolated and truncated arrays, with HOA and WFS algorithms in order to create a panned frontal dominant system with potentially increased intelligibility due to source separation and spatial unmasking. Hybrid methods where WFS and ambisonics are used simultaneously will be evaluated to create a system for theatre that is both psychoacoustically sound, homogenous and practicable.

Introduction

Wave Field Synthesis (WFS), and Higher Order Ambisoncis (HOA) are methods of spatial audio reproduction that are capable of a high degree of localisation accuracy. There are drawbacks in their practical implementation, as large numbers of speakers are required for accurate spatial reproduction. Theatrical implementations of spatial audio have mostly been based on *'spot special'* techniques, where single speakers are integrated to the set or in the auditorium to localise a particular sound source. Full scale spatial reproduction methods, though rarely implemented due to practical constraints, would give the sound designer much greater scope and flexibility.

One method of increasing localisation accuracy would be to utilise a planar speaker array situated near the edge of the stage. Ideally, the performers' positions and orientations would be tracked on stage, corresponding with the movement of the performers' voice sources reproduced over the planar array. This would allow for localisation across the plane of the stage, however, any sources would be constrained to the dimensions of the planar array. For atmospheric sounds, this would be insufficient, and therefore an additional surround array distributed around the perimeter of the auditorium would need to be used in conjunction with the planar array. This research aims to simulate both WFS and HOA, and hybrid algorithms over different planar and surround array configurations.

Theory

The development history of HOA and WFS has been discussed many times [1] [2]. Ambisonics and Wave Field Synthesis aim to achieve accurate spatial reproduction of a sound field over a listening area. Both methods have advantages and disadvantages.

Higher Order Ambisonics (HOA) aims to reproduce the sound field by using spherical harmonic decomposition over a given listening area. First order reproduction (the minimum) is based on four channels with figure of 8 polar patterns directed among the axis, with WXYZ denoting mono (pressure only), front/back, left/right and up/down respectively; which is known as B-Format. Encoding and decoding of ambisonics is separate, and therefore it is a very flexible speaker layout agnostic system. First order ambisonics is limited in its localisation acuity, and therefore higher order systems with more channels greatly increase the systems ability to reproduced accurately localised sources due to increased spatial resolution. Ambisonics however has a small *sweet spot* where the spatial reproduction will be reproduced accurately in all directions, which is an issue in live/theatrical environments as the experience will differ greatly depending on listener seating position.

Speaker arrays for ambisonics usually consist of regular polyhedral array speakers of speakers around a small listening area with the listener being in the centre which is impractical for a number of reasons [3]. Firstly, theatre auditoriums are not usually symmetrical in three dimensions, making it impossible for the listeners to be equidistant from all speakers. The ideal regular polyhedral setup would also require speakers be placed well below the listener, i.e. under the floor, which is impractical for a large number of listeners and on the stage which is untenable. It is therefore the scope of this project to investigate other speaker array layouts that are more practicable in a live environment for reproduction over a large audience area.

Wave field synthesis is another method of spatial reproduction. It aims to recreate the waveform of the original acoustic sound source using large speaker arrays. WFS is most often utilised as planar array setups, giving very high localisation accuracy as well as depth/source distance [4]. These high density circular arrays are sometimes difficult to implement in a live situation, and there are now several companies offering live WFS source rendering solutions based on traditional LCR arrays with added fills dependent on the specific event [5]. WFS has difficulty in reproducing full 3D content, especially in the elevation plane however it provides accurate 2D spatial reproduction over a large listening area.

The main idea discussed in this paper is evaluating both spatial reproduction methods based on planar and semicircular arrays due to their practicability in a theatrical/live performance environment. A Hybrid system will briefly be explored by simulations combining WFS and ambisonics on the same array. The practical limitations of ambisonics and WFS could be reduced by using a hybrid speaker array design, with a high density linear planar array at the proscenium arch for accurate localisation across the stage, and then surrounds placed around the auditorium on a circular basis as in traditional 2D ambisonics for ambient material.

Method

Matlab Simulations

For this current research, the array size will be kept small as future work will include real world evaluation in a hemi-anechoic environment. All simulations will be carried out in octave bands starting at 125Hz, to 4kHz. Both single planar arrays and combined planar and surround systems will be simulated and their performance evaluated while using Ambisonics, WFS or a hybrid of the two. First the array dimensions, number of speakers, and the room dimensions are initialised. For ambisonic simulation, a given maximum order is also specified. The virtual source panning angle is specified, and symmetry is assumed as all arrays are symmetrical at this stage.

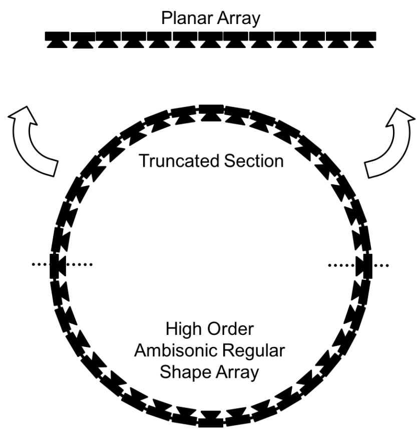

Figure 1: HOA Array diagram showing truncation and extrapolation of half the circular array to planar array

Figure 1 shows the transformation from a regular spaced ambisonic array, to a planar array. The circular array is split in half, and one side is extrapolated to create a planar array. The original horizontal spacing of the array (keeping the x-dimension spacing as if it was still part of the circle, with greater speaker density at the edges and less dense in the centre) will be simulated and evaluated.

Speaker spacing within the array is calculated based on its dimensions and number of speakers. Each speaker location has a relative cartesian coordinate. A virtual source position is then defined as an azimuth pan direction.

For the ambisonic version, the gain coefficients for each speaker, and the virtual source are calculated in a separate function.

$$N_n^{m(SN3D)} = \left(\frac{2-\delta_m}{4\pi} \times \frac{(n-|m|)!}{(n+|m|)!}\right) \text{ where } \delta_m = \begin{cases} 1 \text{ if } m = 0 \\ 0 \text{ if } m \neq 0 \end{cases}$$

(1)

Where:

- N_n^m = SN3D normalisation value used to normalise the gain of the harmonics
- n = ambisonic order
- m = the degree

$$Y_n^m(\theta,\phi) = N_n^{|m|} P_n^{|m|}(\sin(\phi)) \begin{cases} \cos(|m|\theta) \text{ if } m \geq 0 \\ \sin(|m|\theta) \text{ if } m < 0 \end{cases}$$

Where:

- $N_n^{|m|}$ = SN3D normalisation value used to normalise the gain of the harmonics
- $P_n^{|m|}$ = associated Legendre function
- θ = azimuth
- ϕ = elevation

In this very simplified WFS model, the distance from the virtual source to all loudspeakers needs to be calculated first. Further work would include utilising a more complex method of WFS as per the Sound Field Synthesis Toolbox [6].

$$r = \sqrt{(ls_x - source_x)^2 + (ls_y - source_y)^2} \qquad (2)$$

Where:

- ls_x = loudspeaker x coordinate
- $source_x$ = source x coordinate
- ls_y = loudspeaker x coordinate
- $source_y$ = source y coordinate

Then, turn the distance into delay in ms:

$$delay = \frac{r}{c} \qquad (3)$$

The gain of each loudspeaker is then calculated using the distance from the source (with added normalisation):

$$gain = \frac{1}{r+1}$$
$$gain = \frac{gain}{gain_{max}}$$

These gain values are then used, along with the analytic signal, to simulate waveform propagation.

Wave Propagation

The first set of results is image and gif plots of the waveforms resulting from the ambisonic/WFS panned virtual source, and a waveform as if a single source was placed in the same location using the analytic wave equation.

$$r = \sqrt{x^2 + y^2} \tag{4}$$

$$\psi(r,t) = \frac{L_s \times e^{j2\pi ft} \times e^{-jkr}}{r} \tag{5}$$

Where:

- L_s = inverted speaker encoding matrix (ambisonics) or the speaker gains (WFS) convolved with the virtual source coefficients
- $e^{j2\pi ft}$ = analytic signal
- r = coordinate point distance from source

The waveform plots of the virtual source (WFS or Ambisonics) and the single point source in the same source position are normalised so that they can be compared. A reference coordinate point at the centre of the array containing the virtual panned source is chosen as the comparison point at the first frame of the process, so the two are at the same point in time. The same reference point in the array containing the single point source data is compared and a normalisation factor is calculated by division of the two reference points. Every element in the single point source array is then multiplied by this normalisation factor.

Velocity & Energy Vector Analysis

Velocity and energy vector analysis is a useful tool for evaluating the performance of spatial audio algorithms. Both are methods of evaluating the performance of speaker arrays without full-scale perceptual listening tests. At each point of the listening area, the output from each speaker sums to create a pressure and velocity vector that points towards the perceived source location [7]. The velocity vector is particularly suited to evaluating spatial array performance at low frequency, and energy vector at high frequencies. This is due to the psychoacoustics of localisation and the size of the head. Below 700Hz (wavelength approximate to the width of the head), there is very little difference in energy between the ears, as the head is not large enough to cause an obstruction and a resultant reduction in level at one ear, with the only inter-auditory-differences (IAD) being phase related. The difference in the waveforms at both ears is the velocity of the sound field along the ear-axis [8].

Energy vectors are the sum of the unit vectors. In an ideal scenario, that of a single source, the magnitudes and the direction of the energy vectors would reach unity. This will never be the case for more than one source, as the energy is emanating from more than one loudspeaker and thus cannot reach unity [3]. Extended energy vectors, as per Stitt [9] give even greater improvements in localisation performance prediction, however this method is currently beyond the scope of this paper.

The velocity and energy vectors are calculated for each point on the 400x400 coordinate grid. Rather than displaying every 20th vector, a simple averaging function takes a 20x20 block of coordinate points and averages the energy vector and velocity vectors of block, representing an area with a single velocity and energy vector.

Localisation Error

A useful tool for evaluating listening areas using surround sound is localisation error. This compares the predicted localisation angle (based on the energy vector) of a virtual panned source, to the ideal scenario of a single *'real'* source in the same location. This resulting difference in the localisation angle for each coordinate point can then be calculated and mapped onto a colour spectrum.

Results

The simulations have resulted in a very large database of figures, and it is therefore only possible to show a small sample in this paper.

Ambisonic Array Simulation Results

The behaviour of the planar ambisonic array is greatly dependent on the ambisonic order. It is useful to visualise the wave propagation from the different array configurations and algorithms in order to evaluate their ability to produce a coherent waveform for a virtual source compared to a real source placed in the same position.

Figure 2 shows the simulated waveform for an ambisonic virtual source panned to the centre of the array at 0 degrees, as well as the velocity vector (white) and energy vector (pink) analysis. At first order, the audience area would perceive the source as coming from the speaker closest to them, and the constructed waveform is very fragmented and not representative of a single point source waveform. As the order increases, the reproduction of a point source like waveform increases, and the resultant energy and vector analysis demonstrate a greater homogeneity throughout the listening area with the source being perceived as a coming from its panned position as if a point source.

Unsurprisingly, as order increases, the localisability of the source increases and the localisation error decreases. In real terms, the higher the order, the larger the audience

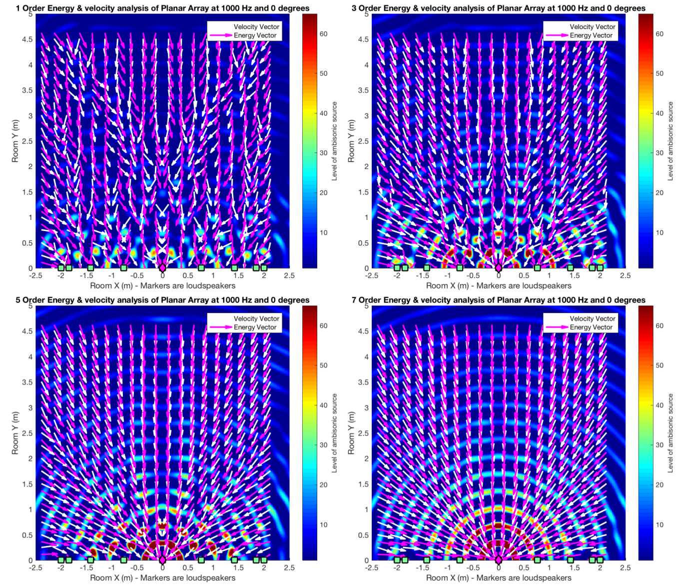

Figure 2: Simulation results for a virtual ambisonic source panned to the centre of the planar array at (left to right) 1st, 3rd, 5th and 7th order

area is where the virtual source is perceived to be emanating from its panned position, and not from the speakers closest to them.

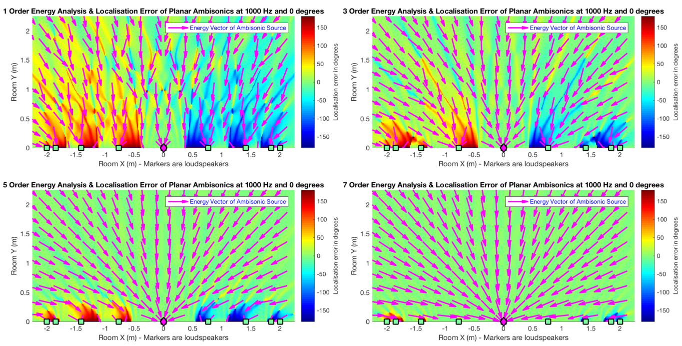

Figure 3: Energy vector and localisation error analysis of a virtual ambisonic source panned to the centre of the planar array at (left to right) 1st, 3rd, 5th and 7th order

Figure 3 shows the energy vector analysis of a 4m long planar array, as well as the localisation error at each point with a virtual source panned to 0 degrees. As the ambisonic order increases, the energy vector shows that a greater portion of the simulation area perceives the sound source as coming from the correct location, instead of a speaker closer to that point. At 7th order, only the areas near the end of the array have any discernable degree of localisation error.

Figure 4 shows the simulation of a source panned to 60 degrees. Once again, an increase in order reduces the localisation error, but only to a point, with the first 0.5m in-front of the array experiencing a large degree of localisation error, with the source being perceived as originating from the speaker with the closest proximity.

As the distance between the speakers increases, the aliasing frequency decreases. A higher density speaker array (with a resultant increase in ambisonic order)

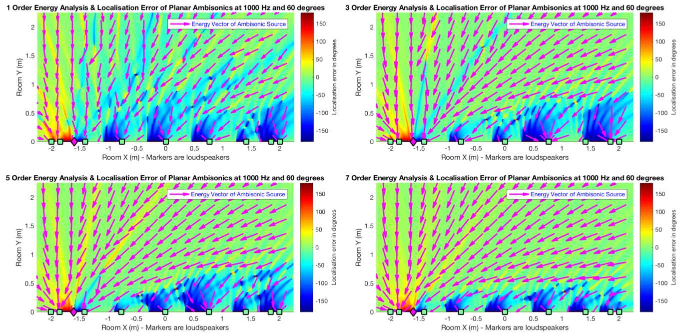

Figure 4: Energy vector and localisation error analysis of a virtual ambisonic source panned to 60 degrees on the planar array (left to right) 1st, 3rd, 5th and 7th orders

performs better at high frequencies than a sparse, low order array, however there is a practicable trade off in terms of production budget and the number of speakers required. So far, a 16 channel array has been truncated to a 9 speaker planar array. In order to create a full surround system, the remaining 7 speakers can be used as surrounds, creating a semi-circular array. This would allow for sound sources to be panned around the auditorium to create immersive effects. The surround spatialisation does not have to be as accurately reproduced as the front planar array, as it will be used to create a sense of ambience, which would be enriched by a diverse sonic experience throughout the audience area. This is why this semi-circular system is front weighted with a greater speaker density in-front of the proscenium arch, in order to accurately reproduce the localisation of the actors' voices and instrument/sound effect positions that are directly pertinent to the drama.

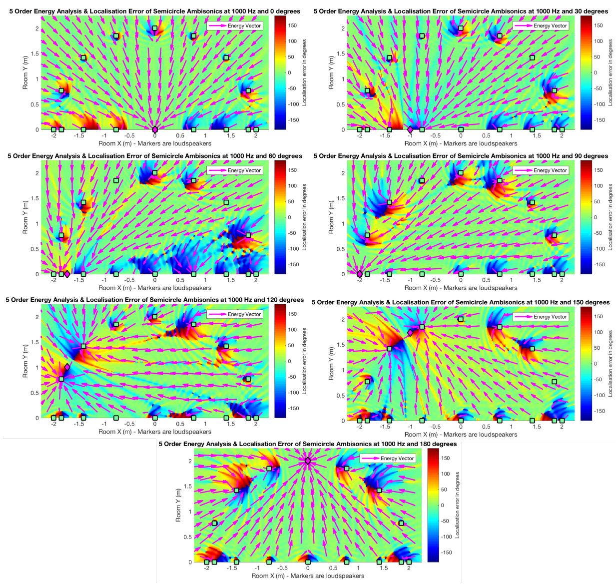

Figure 5: Energy vector and localisation error analysis of a virtual ambisonic source panned to 0 through 180 degrees on the semi-circular array at 5th order

Array performance is also frequency dependent due to the dimensions of the array. Using the same test parameters as the previous graphs, plots for octave bands for 125Hz to 4kHz are compared in Figure 6e. The vector analysis is changed from velocity to energy based at 700Hz, as outlined by Gerzon [8].

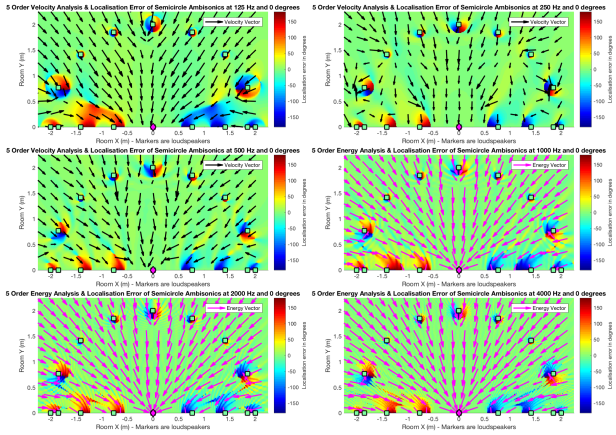

Figure 6: Energy and velocity vector and localisation error analysis of a virtual ambisonic source panned to 0 degrees through octave bands on the semi-circular array at 5th order

WFS Array Simulation Results

The wave field synthesis simulations were undertaken with the same parameters as the ambisonic tests, in order to allow for direct comparison.

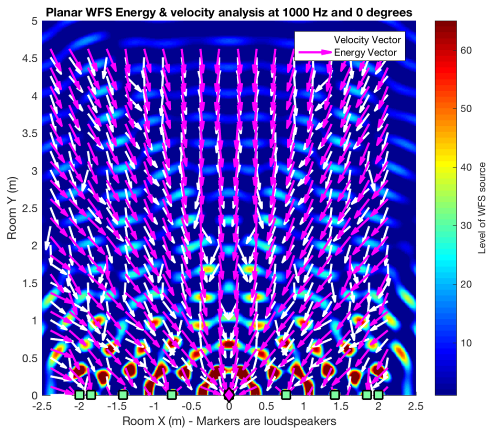

Figure 7: Waveform, Energy and velocity vector analysis of a WFS source at 0 degrees

When compared to a 7th order ambisonic source panned to the same position on a planar array (as per Figure 2, the waveform of the WFS simulation is less coherent (Figure 7). This is likely due to the planar array being relatively sparse in comparison to a traditional, high density WFS array.

As the source is panned towards the end of the array, the localisation error on the opposite side increases, and extends further into the *'audience area'* than the planar ambisonic implementation (Figure 8).

Adding the surround speakers, the performance of the array is still second to the ambisonic implementation, with a larger degree of localisation error near each of the speakers (Figure9).

The performance of the WFS implementation is also more frequency dependent, with greater variation in degrees of localisation error over a larger portion of the audience area between each octave band than the ambisonic version (Figure10).

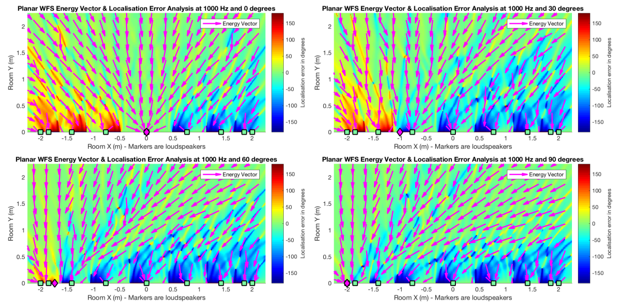

Figure 8: Energy/Velocity vector analysis of a WFS source at (0, 30, 60, 90) degrees on a planar array

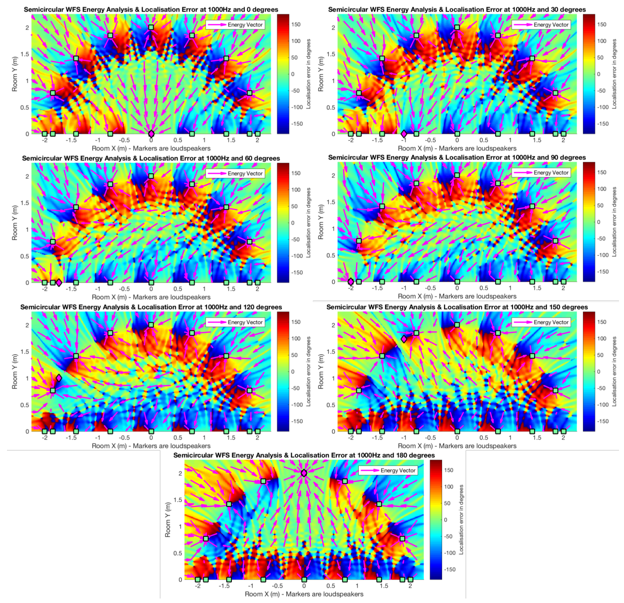

Figure 9: Energy/Velocity vector analysis of a WFS source at (0, 30, 60, 90, 120, 150 180) degrees and 1kHz on a semicircular array

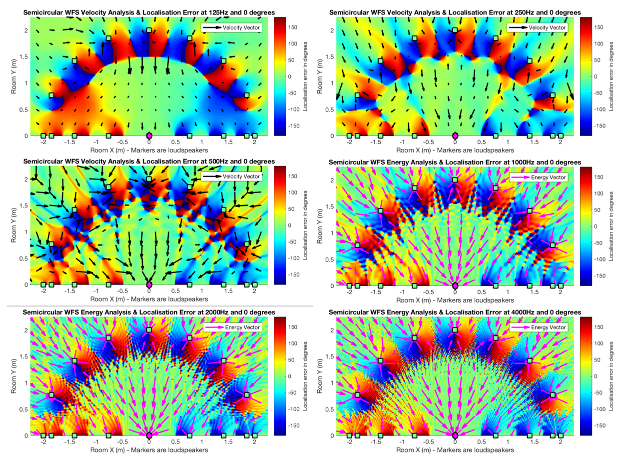

Figure 10: Energy/Velocity vector analysis of a WFS source at 0 degrees and octave bands from 125Hz - 4kHz on a semicircular array

WFS/Ambisonic Hybrid Array Simulation Results

The original aim of this paper was to investigate the simultaneous use of ambisonic and WFS algorithms in order to utilise their individual merits whilst mitigating their disadvantages. As WFS is a system generally implemented in planar arrays, including source distance and depth calculations, this was chosen as the driving method for the planar array, whilst ambisonics is particularly suited to sparser arrays, hence its utilisation in the surround speakers.

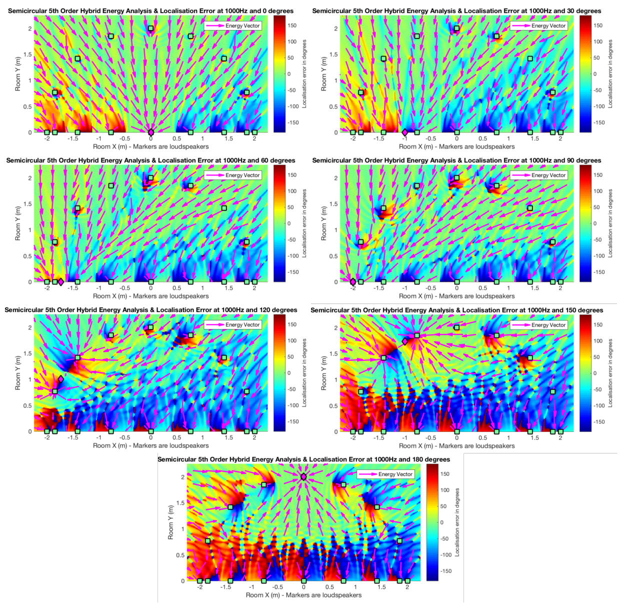

Figure 11: Energy/Velocity vector analysis of a hybrid WFS(planar)/Ambisonic(surround) array with a virtual source at 0 degrees (0, 30, 60, 90, 120, 150 180) degrees and 1kHz on a semicircular array

Again, the hybrid simulation shows a greater degree of localisation error over a much larger portion of the audience area than the ambisonic system, with a particularly interesting result when the virtual source is panned towards the rear of the array. This may be due to the lack of normalisation between the ambisonic and WFS functions, which future work would seek to mitigate.

Conclusion

The results show that ambisonics can indeed be utilised on a planar array to some success. The vector analysis shows that over the audience area, ambisonics produces the least localisation error for this particular array, which is surprising due to its *'sweet spot'* nature. Both the hybrid algorithm and WFS were shown to be poorly suited for this array configuration, possibly due to its relatively sparsity in terms of numbers of speakers. Future work includes simulating multiple array configurations to see whether a different array can be used to improve the results. It must be noted, however, that the WFS implementation used here is a simple one with all secondary sources driven to recreate the primary source. Further work will refine this implementation.

Real world tests, first in a hemi-anechoic environment and then a theatre also need to be undertaken to assess the relative merits of each spatial audio reproduction method over both planar and semicircular speaker arrays.

The search for a hybrid algorithm solution that may mitigate the disadvantages of each system while maintaining their relative advantages will continue.

References

[1] Audio Engineering Society, *Ambisonic Decoders; Is Historical Hardware the Future?*, 2010.

[2] M. M. Boone, E. N. G. Verheijen, and P. F. V. Tol, "Spatial sound-field reproduction by wave-field synthesis," vol. Vol. 43, Audio Engineering Society, 1995.

[3] E. Benjamin, "Ambisonic loudspeaker arrays," *Audio Engineering Society*, 2008.

[4] S. Spors, R. Rabenstein, and J. Ahrens, "The theory of wave field synthesis revisited," *Audio Engineering Society*, 2008.

[5] E. Corteel, R. Foulon, and F. Changenet, "A hybrid approach to live spatial sound mixing," *Audio Engineering Society*, 2016.

[6] S. T. Developers, "Sound field synthesis toolbox - wave field synthesis." http://matlab.sfstoolbox.org/en/2.3.0/frequency-domain/, August 2017.

[7] E. Benjamin, R. Lee, and A. Heller, "Localization in horizontal-only ambisonic systems," *Audio Engineering Society*, 2006.

[8] M. Gerzon, "Surround-sound psychoacoustics," *Wireless World*, 1974.

[9] "Extended energy vector prediction of ambisonically reproduced image direction at off-center listening positions," *Audio Engineering Society*, vol. 64, no. 5, 2016.