# Spatial Audio System for Surround Video

[1]Martin Morrell, [2]Chris Baume, [3]Joshua D. Reiss

[1, Corresponding Author]*Queen Mary University of London, Martin.Morrell@eecs.qmul.ac.uk*
[2]*BBC Research & Development, Chris.Baume@bbc.co.uk*
[3]*Queen Mary University of London, Josh.Reiss@eecs.qmul.ac.uk*

## Abstract

*In this paper we present the design processes of a spatial audio system for Surround Video. Surround Video is a method of reproducing two simultaneous video streams captured by two cameras onto a main television screen and onto the walls of a room via a projector. Through the use of distortion software to correctly map the surround image to the geometry of the viewing room, the user experiences 180 degrees of video reproduction, immersing them in the content. The design of a spatial audio system was necessary to give 360 degree coverage of audio so that, like for the video, the viewer is immersed into the programme world. We discuss the design process and decisions made that concluded in using a mixed reproduction system of Vector Base Amplitude Panning with Ambisonics to match the audio localisation precision with the video precision; high localisation around the main monitor image whilst the surrounding audio is immersive, but with less localisation. Attributes associated with objects in the real world are discussed and methods for recreating the effect of distance, in-head panning, sound scene rotations, reverberation and movement that alter the reverberation placement are presented. The end result is an immersive video and audio system that can be used by the BBC Research & Development department to demonstrate the potential of such technologies where the audio system uses 14 loudspeakers, a subwoofer signal and a discrete '4D' type effects channel.*

**Keywords**: *Ambisonics, Vector Base Amplitude Panning, Surround Video, Spatial Audio, 3D*

## 1. Introduction

A three-month collaboration took place between Queen Mary University of London and BBC R&D to create a spatial audio system to accompany a technology called Surround Video [1]. Surround Video is a technology that uses two simultaneous videos where one is replayed on a main television screen directly in front of the viewer and the other video is projected onto a hemispherical mirror at the back of the room that reflects it onto the walls with approximately 180 degrees of coverage all around the viewer. The second video is processed first so that it has a black section in the centre that aligns with the main television screen so no video is projected onto the screen. Secondly the video is processed; warping the aspect so that when it is projected onto the walls the perspective correctly corresponds to as if a person were standing where the cameras were placed. In previous material for Surround Video [1] the content was captured using two coincident cameras where one was a standard camera for the main video and the other camera used a fish eye lens to capture the extended viewing angle needed to recreate the perspective aspect of being there.

For this collaboration the resulting video of a new commission was used. The new material was computer generated, which has the advantage of having the two camera perspectives in the exact same origin. It also does not carry the risk of production equipment being visible due to the extended viewing angle. The produced video is an animatic. Instead of being a continuous video with 25 or more frames per second, it uses only a few frames per minute. There are usually more frames when there is more action occurring, thus the frames are not linearly spaced in time.

The remit of the project was to create a spatial audio system for use with the new surround video animatic that matched the audio localisation attributes with the video attributes. The surround video technology offers high localisation, in terms of high resolution, in the main television screen and lower resolution and therefore localisation in the surround video. The premise for this is to keep the audience's attention to the centre where the *action* should be and for the surround to fill the periphery of the scene without being distracting. The proposed system needed to use only the equipment available; fourteen speakers, a subwoofer and an effects device.

## 2. Spatial Audio Reproduction

In this section two of the main methods for spatial audio reproduction, ambisonics and vector base amplitude panning, are shown and the proposed method for a spatial audio system with enhanced localisation within the frontal section is presented. In the spatial audio reproduction system presented below, sound sources are represented in terms of spherical coordinates. Two angles represent the sound source location; an azimuth angle $\theta$[0 360] in the horizontal plane and an elevation angle $\phi$ [-90 90].

### 2.1. Ambisonics

Ambisonics was developed in the 1970s, primarily by Michael Gerzon [2]. Ambisonics has been further developed since then to what is called higher order ambisonics (HOA) [3]. Ambisonics can be represented as spherical harmonics [4,5] calculated from Legendre polynomial functions based on an order, $m$, where a system is defined by the highest order used, $M$, such that $0 \leq m \leq M$. Each Legendre function is also dependent on the degree of the order, $n$, defined by $-m \leq n \leq m$ for each order used and the associated Schmidt semi-normalised version. The spherical harmonics are thus calculated as:
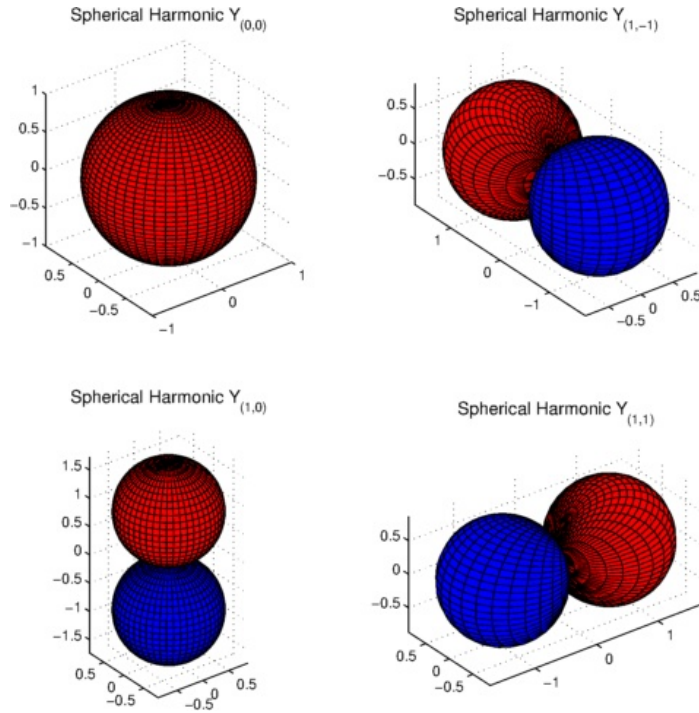
$$Y_{(mn)}^{\sigma(N3D)}(\theta,\phi) = \sqrt{2m+1}\,\hat{P}_{mn}(\sin\phi)\begin{cases} \cos n\theta & n \geq 0 \\ \sin n\theta & n < 0 \end{cases}$$

$$\hat{P}_{mn}(\sin\phi) = \sqrt{(2-\phi_{0,n})\frac{(m-n)!}{(m+n)!}}P_{mn}(\sin\phi) \tag{1}$$

The spherical harmonics are then normalised using the N3D scheme by using the included factor on each component $\sqrt{2m+1}$. These N3D normalised spherical harmonics are shown in figure 1 for first order ambisonics. This collection of spherical harmonics can be used to represent a sound field of a single point in space and is known as B-Format.

The sound field is then reconstructed on playback to a loudspeaker layout using what is known as an ambisonic decoder. In the simplest case the decoder is the pseudo-inverse of a matrix containing spherical harmonics for each speaker position, $(\theta_N, \phi_N)$ given in equation 2. The sound field representation, B-Format, is then multiplied by the decoder to create the speaker signals. The pseudo-inverse is used because the loudspeaker matrix is not square.

$$pinv\begin{pmatrix} Y_{(0,0)}(spk1) & Y_{(1,-1)}(spk1) & Y_{(1,0)}(spk1) & Y_{(1,1)}(spk1) & \cdots & Y_{(M,n)}(spk1) \\ Y_{(0,0)}(spk2) & Y_{(1,-1)}(spk2) & Y_{(1,0)}(spk2) & Y_{(1,1)}(spk2) & \cdots & Y_{(M,n)}(spk2) \\ Y_{(0,0)}(spk3) & Y_{(1,-1)}(spk3) & Y_{(1,0)}(spk3) & Y_{(1,1)}(spk3) & \cdots & Y_{(M,n)}(spk3) \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ Y_{(0,0)}(spkN) & Y_{(1,-1)}(spkN) & Y_{(1,0)}(spkN) & Y_{(1,1)}(spkN) & \cdots & Y_{(M,n)}(spkN) \end{pmatrix} \tag{2}$$

Spherical Harmonic $Y_{(0,0)}$

Spherical Harmonic $Y_{(1,-1)}$

Spherical Harmonic $Y_{(1,0)}$

Spherical Harmonic $Y_{(1,1)}$

**Figure 1.** N3D Spherical Harmonics for orders 0 and 1; $Y_{(0,0)}$ , $Y_{(1,-1)}$ , $Y_{(1,0)}$ and $Y_{(1,1)}$

### 2.2. Vector Base Amplitude Panning

Vector base amplitude panning (VBAP) [6,7] uses any speaker arrangement where the speakers are equidistant from the listener position on the surface of a sphere. The speaker layout is separated into three dimensional triangles where a) the speakers are not in the same plane and b) the triangles are non-overlapping. The position of a virtual source, $p$, is a weighted combination of the same signal coming from all three loudspeakers, $L_{123}$, where the gains, $g$, can be calculated by:

$$g = p^T L_{123}^{-1} = [p_x p_y p_z] \begin{bmatrix} l_{1x} & l_{1y} & l_{1z} \\ l_{2x} & l_{2y} & l_{2z} \\ l_{3x} & l_{3y} & l_{3z} \end{bmatrix}^{-1} \tag{3}$$

However, this does not keep power constant. A further step is needed to calculate the final gains $\hat{g}$:

$$\hat{g} = gC \tag{4}$$

where we calculate $C$ based on the originally computed gains $g$:

$$C = \frac{1}{\sqrt{\Sigma g^2}} \tag{5}$$

resulting in constant power panning within a 3D speaker triplet. If a sound source lies on the line between two adjacent speakers in a triplet the end result is the same as stereo constant power panning between two loudspeakers i.e. $g_l = \cos\theta, g_r = \sin\theta$ for $0° \le \theta \le 90°$.

## 2.3. Prior Real-World Systems

There are many real-world spatial audio systems that have been documented [8-17] which attempt to accurately recreate spatial audio throughout the entire listening area. However, whilst the specification of the system is to reproduce spatial audio from all possible angles, it should also have higher directionality within the frontal section. None of the systems described have such an attribute and as such are not suitable to meet the criteria specified for the use of a spatial audio system for Surround Video.

## 2.4. Authors' Approach

To create the desired spatial audio system a combination of ambisonics and VBAP is used. VBAP reproduces sounds around the main television screen to give high localisation attributes. This is created using an area of triangles from six speakers. Outside of this bounded VBAP area, first order ambisonics is used to reproduce atmospheric sounds and audio that should not distract the attention of the listener.

All audio material used for the system was monaural sound sources. These sources were time aligned with the main video material in the Nuendo digital audio workstation. The sources can be individually processed using equalization, dynamic range control and so forth as any other production within Nuendo. The audio sources were then sent to a custom built Max/MSP application along with control data that was stored as midi messages. Inside the spatial audio application the sound sources are categorized into those that lie within the VBAP reproduction area or ambisonics area and processed with the corresponding spatialisation technique.

The ambisonics decoder used is of a dual band design where high and low frequencies are decoded differently. Low frequencies are decoded using the pseudo-inverse method presented. The high frequency section modifies the pseudo-inverse decoder by matching the gain of the first-order spherical harmonic components and then applying $\max_{rE}$ gain weightings, $g'_m$ [3]:

$$g'_m = P_m(rE)$$
$$g_0 = \sqrt{\frac{N}{E_{g'_m}}} \tag{6}$$
$$rE = g'+1$$

where $rE$ is the largest root of $P_{M+1}$. Both decoder sections are increased by $\sqrt{2}$ to match the perceived level of the VBAP section. A dual band decoder is used to match Gerzon's criteria [18]. The decoder coefficients used are given in table 1.

**Table 1.** Ambisonic dual band decoding matrix coefficients

| Speaker | | $Y_{(0,0)}$ | $Y_{(1,-1)}$ | $Y_{(1,0)}$ | $Y_{(1,1)}$ |
|---|---|---|---|---|---|
| Low Frequencies | Left Front Up | 0.1768 | 0.2041 | 0.1433 | 0.2041 |
| | Left Back Up | 0.1768 | 0.2041 | 0.1433 | -0.2041 |
| | Right Back Up | 0.1768 | -0.2041 | 0.1433 | -0.2041 |
| | Right Front Up | 0.1768 | -0.2041 | 0.1433 | 0.2041 |
| | Left Front Down | 0.1768 | 0.2041 | -0.1433 | 0.2041 |
| | Left Back Down | 0.1768 | 0.2041 | -0.1433 | -0.2041 |
| | Right Back Down | 0.1768 | -0.2041 | -0.1433 | -0.2041 |
| | Right Front Down | 0.1768 | -0.2041 | -0.1433 | 0.2041 |
| High Frequencies | Left Front Up | 0.1768 | 0.0833 | 0.0833 | 0.0833 |
| | Left Back Up | 0.1768 | 0.0833 | 0.0833 | -0.0833 |

| | | | | |
|---|---|---|---|---|
| Right Back Up | 0.1768 | -0.0833 | 0.0833 | -0.0833 |
| Right Front Up | 0.1768 | -0.0833 | 0.0833 | 0.0833 |
| Left Front Down | 0.1768 | 0.0833 | -0.0833 | 0.0833 |
| Left Back Down | 0.1768 | 0.0833 | -0.0833 | -0.0833 |
| Right Back Down | 0.1768 | -0.0833 | -0.0833 | -0.0833 |
| Right Front Down | 0.1768 | -0.0833 | -0.0833 | 0.0833 |

The accuracy of the reproduced sound direction can be compared with the intended direction by calculating the rV vector at low frequencies and the rE vector at high frequencies [18]. Figure 2 shows rV at low frequencies below 700Hz and rE at high frequencies above 700Hz for the proposed speaker layout for all angles of $\theta$ in the horizontal plane, where $\phi = 0°$. The proposed speaker layout has fourteen speakers; eight arranged as a cube at azimuths $\theta = \pm45°$ and $\theta = \pm135°$ with $\phi = \pm45°$, two loudspeakers at $\theta = \pm30°, \phi = 0°$ and four loudspeakers in a square at $\theta = \pm15°, \phi = \pm15°$. It can be seen in figure 2 that VBAP reproduces the high frequency content more accurately in regards to localisation, where the largest error is the centre point between two adjacent speakers. Low frequency reproduction is slightly compromised for VBAP when compared to the ambisonic reproduction area, however this is not such a problem since a subwoofer is used for the lowest frequencies, with a second-order crossover at 120Hz. Thus the full frequency range is utilised. The subwoofer derives its feed from the origin sound source audio for the VBAP bounded area and from the $Y_{(0,0)}$ spherical harmonic in the ambisonic reproduced area. There is also a dedicated subwoofer send for each source within the authors' built system. All filters in the system use a phase-matched design as described in [18].

The system also uses a so-called 'ButtKicker' device to give (4D) effects. The device is driven like a speaker using a low frequency signal, but instead of creating acoustic waves, the device is attached to a chair or sofa to provide the vibration effect to the listener. This device is driven by a discrete channel and not derived from any other signal. The speaker layout along with the effects devices can be seen in figure 3.
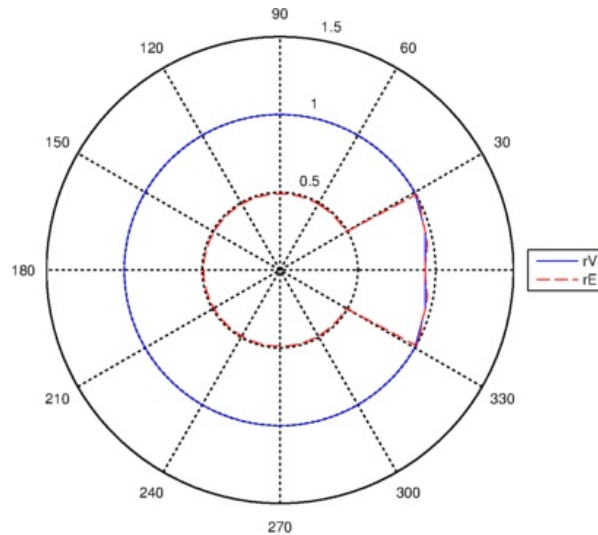


**Figure 2.** The rV and rE vectors representing the accuracy of low and high frequency content respectively within the horizontal plane
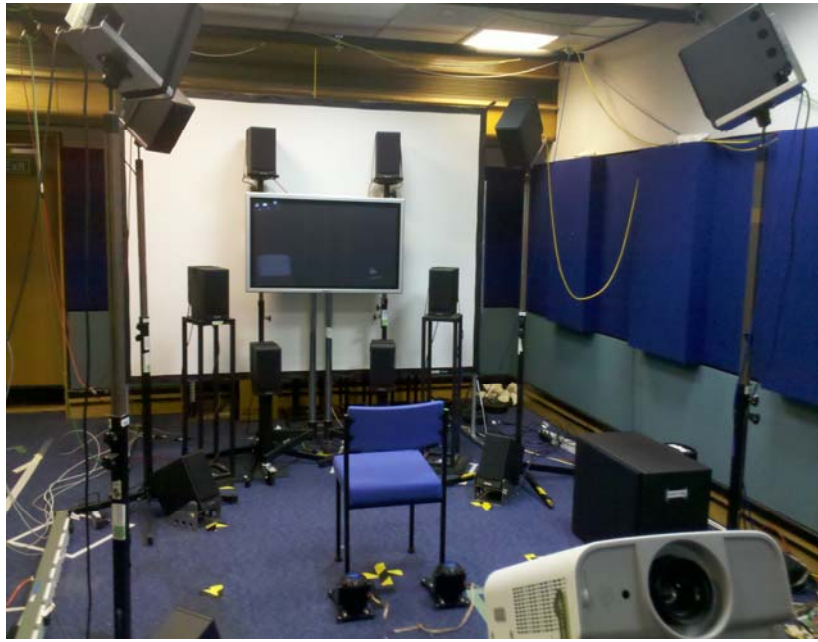
**Figure 3.** Speaker layout of the authors proposed 4D reproduction system

## 3. Realising The Programme World

In order to put the listener into the programme world, real world sound attributes have to be reproduced. Sound attributes not only need to match visual cues seen in the animatic material but also need to match the effects in a natural sound scene. These natural attributes include, but are not limited to; reverberation, movement, distance, Doppler effect, sound field transformations and in-head panning.

### 3.1. Reverberation

Reverberation is present everywhere in nature and is used to make anechoic or 'dry' recordings sound natural. There are two commonly used types of reverberation in audio production: convolution and algorithmic types. Convolution uses the impulse response of a real room or environment to make sounds appear as though they had been recorded there. Although this reverb type sounds very natural and realistic there are two major drawbacks. The first is that the impulse response is based upon a single combination of source and receiver position; therefore as either the receiver point of view or sound sources move away from the impulse measurement positions the accuracy of being in that location decreases. The second drawback is that convolution can be CPU intensive. Algorithmic reverberation on the other hand is very CPU efficient, being generally based on an algorithm involving delay lines and equalisation. The drawback to this reverberation type is that it does not sound as natural.

The reverberation within the described system is accomplished in the ambisonics domain. There are good reasons to do this. Ambisonics is a sound field representation and so the speaker layout does not need to be taken into consideration. This is ideal when the placement of the speakers for the frontal section may be changed. Impulse responses are available in ambisonics format [19,20,21] and there is an algorithmic reverb available based on the FreeVerb algorithm [22].

To create the reverberation signal(s) a user defined amount of signal, Reverb Send, was taken from the monaural sound sources pre-fader. This was then encoded into an ambisonic first order B-Format signal that was then passed through one of two VSTs that created a reverberated signal by either the FreeVerb algorithm or via convolution with B-Format impulse responses. By computing the reverberation in this way the derived signals are not loudspeaker layout dependent. The reverberation is then sent through the designed ambisonic decoder and sent only to the cube array of surrounding speakers. With less information being sent to the frontal loudspeakers, masking of the highly localisable sources is less likely to occur.
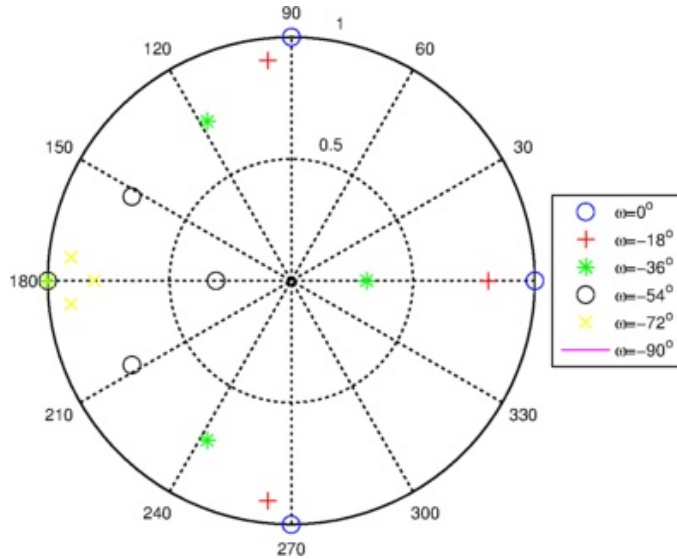
## 3.2. Movement

The animatic content used to demonstrate this system is shot from a 'point-of-view' perspective. As the character moves around, the audio system should be able to accurately reflect this movement. As described above, with conventional reverberation types the source/receiver positions cannot be moved. In order to simulate the movement the reverberation should be altered.

There are various ambisonics transforms that can be used to alter the sound field representation. One proposed by Gerzon [23] is known as dominance. Dominance is used as a zoom effect to move sounds around the unit sphere as though zooming in or out of the sound scene. This effect is not ideal since the increase in gain of the sources requires additional headroom to be taken into account within a digital system.

Anderson has shown some different ambisonics transforms [24] called focus, push and press. The transform used here is press. The reverb sound is moved towards the opposite direction of movement through the sound reproduction. The term, $\omega$, in equation 7 alters the sound source positions in the front/back direction as an angle [-90 90], where negative is forward movement.

This equation is for forth/back movement but it is trivial to expand it to all three directions of movement. Figure 3 shows the effect on the position and gain of four audio sources, $\pm 45^o$ and $\pm 135^o$, when the press transform is applied.

$$
\begin{bmatrix}
Y_{(0,0)}{}' \\
Y_{(1,-1)}{}' \\
Y_{(1,0)}{}' \\
Y_{(1,1)}{}'
\end{bmatrix}
=
\begin{bmatrix}
1 & 0 & 0 & 0 \\
\sqrt{3}\,|\sin\omega|\sin\omega & \cos^2\omega & 0 & 0 \\
0 & 0 & \cos\omega & 0 \\
0 & 0 & 0 & \cos\omega
\end{bmatrix}
\begin{bmatrix}
Y_{(0,0)} \\
Y_{(1,-1)} \\
Y_{(1,0)} \\
Y_{(1,1)}
\end{bmatrix}
\tag{7}
$$



**Figure 3.** Press transform altering gain and position of four sound sources to simulate a forward movement

## 3.3 Distance

In the animatic, one of the characters is at various distances from the main character. Therefore the effect of distance needs to be recreated in order for the visual and auditory cues to match properly.

Distance gain can be simulated in the general case by $1/r$, where $r$ is the source to listener distance. The problem with implementing $1/r$ is that it can lead to infinite gain at a distance of 0 and a gain of 20dB at $r = 0.1$. To overcome this problem a different equation, equation 8, is used at a distance of less than 1.

$$g = 1 + \xi \cos(90r) \tag{8}$$

where $\xi$ [0 1] can be used to alter the gain between 0dB and 6dB at $r = 0$ as shown in Figure 4 for $\xi = 0.25, 0.50, 0.75, 1.0$.

The second attribute needed to successfully reproduce distance is the inclusion of a time delay related to $r$. The time delay, $t$, is given as:

$$t = \frac{r}{c} \tag{9}$$

where $c$ is the speed of sound in air (or any other substance, dependent on what the visual cues dictate). With the use of a time delay the Doppler shift is included for a moving sound source [25]. Therefore the frequency heard by the listener $\omega_l$ is not the same as the frequency at the source $\omega_s$ where the shift in frequency is based upon the source velocity component in the direction of the listener $v_{sl}$ and the listener velocity component in regard to the source, $v_{ls}$:

$$\omega_l = \omega_s \frac{1 + \dfrac{v_{ls}}{c}}{1 - \dfrac{v_{sl}}{c}} \tag{10}$$
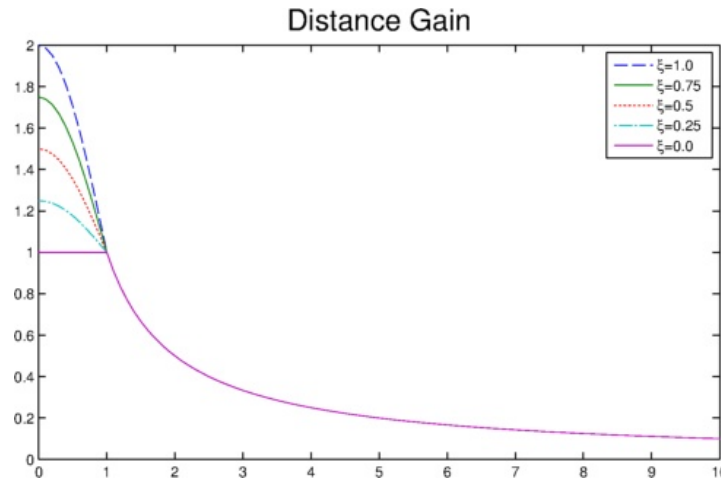


**Figure 4.** The gain applied to sound sources dependant on their distance

However, when including distance in this way there are drawbacks. When a source is at distance of $r = 10$ the gain applied is -20dB which makes the sound difficult to hear, resulting in a conflict between natural and creative sound scene reproduction. The addition of Doppler makes sounds more realistic but several problems are introduced, such as a) the effect is not perceived for low velocities b) sound effects (where it is perceivable) have it pre-added or are recorded with it naturally occurring and c) 'zipper noise' can be noticeable in digital delays.

### 3.4 In-head Panning

There are several parts of the video material that lead to the requirement to be able to place a sound source within the listeners head, or rather speaker array. This need arises on two occasions. The first is for the voice of the main character, where it should come from slightly in front of the listener but without a strong sense of directionality. The second is when warning messages are displayed on the screen as coming from the HUD (headsup display) that has been built into the character's eyes. These messages should use an in-head voice that is presumed not to be heard by any other characters in the represented world.

When the in-head panning is used the sound source is only reproduced by ambisonics. This initially decreases the localisation accuracy needed to give the in-head effect. The sound source has an opposing sound source that is created, $\theta_{opp} = \theta + 180°, \phi_{opp} = -\phi$ [26]. The two sources are then weighted by $\alpha$ [1 0] so that:

$$source' = \alpha source + (1 - \alpha) source_{opp} \tag{11}$$

The resultant effect is actually a doubling of the $Y_{(0,0)}$ spherical harmonic as $\alpha$ approaches 0 and a cancellation of the $Y_{(1,n)}$ harmonics. With this method of inside panning the type of ambisonic decoder used alters the reproduced gain between, but not including the $\alpha = 0$ and 1 points.

### 3.5. Transformations of The Sound Field

In ambisonics it is possible to apply transformations to the B-Format signal about the x, y and z axis [27]. This makes the transformations very efficient since they only have to be applied once and not to the individual sound sources. The transforms are useful for this project as it easily allows the turning of the audio perspective to match the camera perspective. However since this system used both ambisonics (sound field representation) and VBAP (object based representation) this is not possible to rotate all sound sources in this manner. Instead transforms have to be applied to each sound source. The first step involved is to change the spherical coordinates to cartesian coordinates so that a transform about a given axis can be made by altering the combination of the other two axial components:

$$Rotation = \begin{bmatrix} \cos\beta & -\sin\beta \\ \sin\beta & \cos\beta \end{bmatrix} \tag{12}$$

where $\beta$ represents the rotation around an axis in degrees. Once the transform(s) have been applied the new cartesian coordinates are converted back to spherical coordinates for use within the system.

### 3.6. Haptic Effects

As presented in section 2 there is the ability to drive a ButtKicker device. This is useful for giving haptic feedback to the listener as the video material is of the main characters perspective and therefore the listener should feel that they are themselves the character. There are no equations or rules of when this effect should be used, and it is a totally creative process left to the mix engineer, although it is often used for bangs and loud low frequency effects. In the case of the animatic used for the demonstration, the effect acted as the character's heartbeat, which increased and intensified during stressful moments.

## 4. Conclusion

A system has been presented that has increased localisation around a main television screen so it is suitable to be used with and match the properties of surround video. It has been shown

through the use of rV and rE vectors that the method works as intended. The use of ambisonics, VBAP and a subwoofer can all be combined into a real world system to offer a full range audio system. The authors presented various cues that are needed to be reproduced to match visual aspects shown in the demonstration material used and have presented solutions to reproduce these cues to be both natural and creative. A demonstration of the combined technologies was given where lots of positive feedback was given about the designed audio system.

## 5. Further Work

Natural reproduction of sound scenes can often be difficult to achieve with the creative options currently available for mixing a sound scene. It is in this area where more research can be done to provide further tools for spatial audio sound scene rendering. The use of the visceral feedback to the listener proved popular within the demonstration and further ways of adding to the senses can only increase immersion within a programme world.

## 6. Acknowledgement

## 7. References

[1] P. Mills, A. Sheikh, G. Thomas, P. Debenham, "Surround video", NEM 2011, (Torino, Italy), September 27-29 2011.
[2] M. A. Gerzon, "Periphony: With-height sound reproduction", J. Audio Eng. Soc, vol. 21, no. 1, pp. 2–10, 1973.
[3] J. Daniel, Representation de champs acoustiques, application a la transmission et a la reproduction de scenes sonores complexes dans un contexte multimedia. PhD thesis, l'Universite Paris 6, 2000.
[4] R. Chen, P. Tend, Y. Yang, "Beam Pattern Design of Spherical Arrays for Audio Signal with Particle Swarm Optimization", Advances in information Sciences and Service Sciences(AISS), Volume 4, Number 3, February 2012.
[5] Y. Fang, B. Wu, Z. Yang, "A Novel Haptic Deformation Modelling Nethod using Spherical Harmonic Representation and Radial Basis Function Neural Network", International Journal of Digital Content Technology and its Applications (JDCTA), Volume 5, Number 10, October 2011.
[6] V. Pulkki, "Virtual sound source positioning using vector base amplitude panning", Audio Engineering Society, vol. 45, pp. 456–466, June 1997.
[7] V. Pulkki, Spatial sound generation and perception by amplitude panning techniques. PhD thesis, Helsinki University of Technology, 2001.
[8] J. Zmoling, A. Sontacchi, W. Ritsch, "The iem-cube, a periphonic re-/production system", Audio Engineering Society Conference: 24th International Conference: Multichannel Audio, The New Reality, June 2003.
[9] M. Strauss, P. Gleim, "Application of wave field synthesis and analysis on automotive audio", Audio Engineering Society Convention 131, October 2011.
[10] V. Pulkki, "Spatial Sound Reproduction with Directional Audio Coding", Audio Engineering Society, vol. 55, pp. 503-515, June 2007.
[11] F. Rumsey, D. Murphy, "A scalable sound rendering system", Audio Engineering Society Convention 110, May 2001.
[12] J. Batke, J. Spille, H. Kropp, S. Abeling, B.Shirley, R. Oldfield, "Spatial audio processing for interactive tv services", Audio Engineering Society Convention 130, May 2011.
[13] D. Malham, T. Myatt, O. Larkin, P. Worth, M. Paradis, "Audio spatialization for the morning line", Audio Engineering Society Convention 128, May 2010.
[14] G. Dickins, M. Flax, A. McKeag, D. McGrath, "Optimal 3D speaker panning", Audio Engineering Society Conference: 16th International Conference: Spatial Sound Reproduction, March 1999.

[15] D. Zotkin, R. Duraiswami, L. Davis, "Rendering localized spatial audio in virtual auditory space", Proceedings of the IEEE Transactions on Multimedia, vol. 6, issue 4, pp. 553-564, August 2004.

[16] S. Innami, H. Kasai, "On-demand soundscape generation using spatial audio mixing", Proceedings of the IEEE International Conference on Consumer Electronics (ICCE), January 2011.

[17] D. Kostadinov, J. D. Reiss, V. Mlandenov, "Evaluation of distance based amplitude panning for spatial audio", Proceedings of the IEEE International Conference on Acoustics, Speech and Signal (ICASSP), Dallas, March 2010.

[18] A. Heller, R. Lee, E. Benjamin, "Is my decoder ambisonic?", Audio Engineering Society Convention 125, 10 2008.

[19] D. T. Murphy, S. Shelley, "Openair: An interactive auralization web resource and database", Audio Engineering Society Convention 129, November 2010.

[20] S. Shelley, A. Foteinou, D. T. Murphy, "Openair: An online auralization resource with applications for game audio development", Audio Engineering Society Conference: 41st International Conference: Audio for Games, February 2011.

[21] R. Stewart, M. Sandler, "Database of omnidirectional and b-format room impulse responses", Proceedings of the IEEE International Conference on Acoustics, Speech and Signal (ICASSP), Dallas, March 2010.

[22] B.Wiggins, "Has ambisonics come of age?", Proceedings of the Institute of Acoustics, vol. 30. Pt. 6, 2008.

[23] M. A. Gerzon, G. J. Barton, "Ambisonic decoders for HDTV", Audio Engineering Society Convention 92, 3 1992.

[24] J. Anderson, "Introducing... the ambisonic toolkit", Ambisonics Symposium 2009, (Graz, Austria), 2009.

[25] M. J. Morrell, J. D. Reiss, "Inherent doppler properties of spatial audio", Audio Engineering Society Convention 129, 11 2010.

[26] M. J. Morrell, J. D. Reiss, "A comparative approach to sound localization within a 3-d sound field", Audio Engineering Society Convention 126, 5 2009.

[27] M. Chapman, P. Cotterell, "Towards a comprehensive account of valid ambisonic transformations", Ambisonics Symposium 2009, (Graz, Austria), 2009.