ISSUES IN PERFORMANCE PREDICTION OF SURROUND SYSTEMS IN SOUND REINFORCEMENT APPLICATIONS

Christian Landone, Mark Sandler

Department of Electronic Engineering,
King's College London
The Strand, London WC2R 2LS, United Kingdom
christian.landone@kcl.ac.uk
mark.sandler@kcl.ac.uk

ABSTRACT.

Multichannel audio is set to change the way we listen to reproduced music, allowing the creation of spatial auditory images that will add, quite literally, more dimensions to the whole listening experience

Current methods for the objective assessment of the imaging imparted by holographic sound systems assume listening conditions that can be met exclusively in household installations, where the audience is usually located in a very restricted area and the acoustical properties of the room can be generally neglected.

In this paper we address the main limitations of these traditional evaluation techniques in the instance of surround sound systems serving large and acoustically non-ideal listening areas.

1. INTRODUCTION.

The directional auditory information contained in a complex sound field can be reproduced by multi-loudspeaker layouts using two different approaches: transaural-based techniques or holographic reconstruction.

Sound reproduction systems employing the transaural method provide the listener with *synthetic* auditory cues corresponding to those generated by the original sound field whereas holographic systems reproduce a two or three-dimensional sound field in a confined area and produce natural auditory cues as a result of the natural diffraction of the combined wavefront, generated by the loudspeakers in the layout, around the listener's head and torso.

Transaural systems are generally unsuitable for applications involving multiple listeners with unknown precise locations in the audience areas [1] and, for this reason, only holographic systems will be considered in this paper.

Sound field reconstruction by means of multiple loudspeakers, in theory, allows more freedom in terms of head movement and listening locations; however, the large majority of two and three-dimensional surround sound schemes have been developed on the basis of conditions that can exist only in a very restricted portion of the audience area (sweet spot).

Holographic reconstruction schemes such as pair-wise amplitude panning [2] and Ambisonic [3], for instance, assume equal time of arrival, at the listening location, of the wavefronts generated by the loudspeakers. This condition in practice can be met exclusively at one point in the listening area.

Furthermore the assumption of plane wave propagation, that constitute one of the basis of Ambisonic theory, is incorrect in the general case and changes in amplitude caused by loudspeaker proximity come into play when the listening location is moved from the sweet spot.

In situations where the listeners are scattered within a relatively large area such as, for instance, concert venues, almost none of the conditions that guarantee the performance of conventional surround schemes are met and the majority of the audience may perceive spatial distortion or no imaging at all.

At present, very little has been done in order to predict the performance of these installations and, consequently, to quantify the loss of spatial auditory imaging in the audience area.

Existing criteria for the assessment of holographic sound field reconstruction methods have also been developed mainly with respect to one listening location and in this paper we attempt to highlight the issues that might impair their accuracy in real world applications.

In the next section an overview of some of the most established objective evaluation techniques will be given and in section 3 we will present and discuss non ideal listening conditions in large audience areas and their influence on traditional assessment methods. Finally, some suggestions for improving the prediction of imaging produced by surround systems installed in auditoria will be given.

2. ASSESSMENT METHODS FOR HOLOGRAPHIC SOUND REPRODUCTION SYSTEMS: AN OVERVIEW.

Since the introduction of stereophony and multichannel sound systems, various ways of quantifying their spatial imaging accuracy have been developed. Broadly speaking these methods can be classified as either *wave-theoretical* or *perceptual*.

The main objective of techniques belonging to the first category is to find a meaningful measure for the error between the original (or notional) and reconstructed acoustical fields over a specific *area* in the listening room.

Perceptual assessment methods, instead, concentrate on the evaluation of the quality and coherence of auditory cues generated by the reproduced sound field.

2.1. Wave theoretical techniques: Integrated-D error.

This measure has become almost an "industry standard" for assessing the reconstruction accuracy of holographic-based multichannel sound systems [4].
The D error is based on Huygens' principle, stating that "*Each point on a primary wavefront can be considered to be a new source of a secondary spherical wave and that a secondary wave front can be constructed as the envelope of these secondary waves* " [5], as shown in figure 1.
Therefore, if the wave field reconstruction measured (or calculated) on a closed contour is satisfactory, by induction it will also be satisfactory within the enclosed area.
For the Integrated-D error the contour is taken as a circle of radius *r*, encompassing the centre of the listening area (figure 2) and the reconstruction error is the integral, over this path, of the magnitude of the complex difference between the pressure distributions of an ideal monochromatic plane wave and its holographic reconstruction:

$$D(kr,\psi) = \frac{1}{2\pi|p_\psi|}\int_0^{2\pi}\left|S(kr,\theta) - \tilde{S}(kr,\theta)\right|d\theta \quad (1)$$

where:

$$S(kr,\theta) = p_\psi \exp(j\omega t)\exp\left(jkr\cos(\theta-\psi)\right) \quad (2)$$

is the ideal plane wave with incidence ψ and pressure p_ψ, and

$$\tilde{S}(kr,\theta) = \sum_{i=1}^{N} a_i \exp(j\omega t)\exp\left(jkr\cos(\theta-\phi_i)\right) \quad (3)$$

is the combined wavefront generated by the *N* loudspeakers in the layout located at angles ϕ_i with pressures a_i determined by the holographic reconstruction technique.
The integrated-D error in practice returns an *average* error that relies exclusively on pressure distribution differences but gives no indication regarding the nature of the error or its exact location and magnitude within the listening area.
This error measure, however, can be employed to gauge the relative accuracy of the reconstructed wavefront for various holographic reconstruction schemes [6], as a function of the distance from the centre of the sweet spot, the frequency of the reconstructed plane wave and its incidence angle.
Figures 3 and 4 depict, respectively, the D error relative to Ambisonic reproduction systems, as a function of the distance from the centre of the sweet spot and the angle of the reconstructed wavefront.
In both cases, the plots relative to different loudspeaker layouts and order of the reproduced spherical harmonics [3] are shown. These results indicate that the difference, in terms of pressure distributions, between the original and reconstructed wavefronts diminish as we increase the encoding order and that the largest reconstruction errors occur at locations exactly mid-way between loudspeaker pairs.
However, from a localisation point of view, the question regarding the maximum acceptable error cannot be answered.

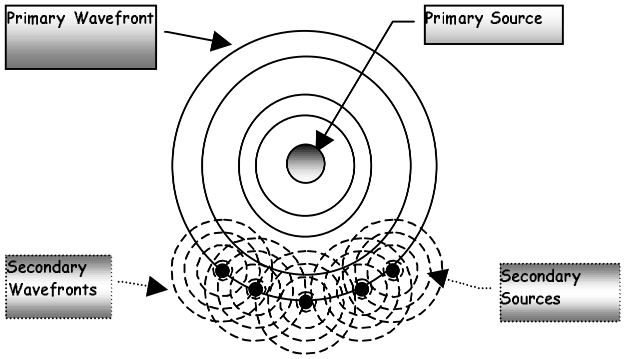

Figure 1. *Huygens' Principle.*

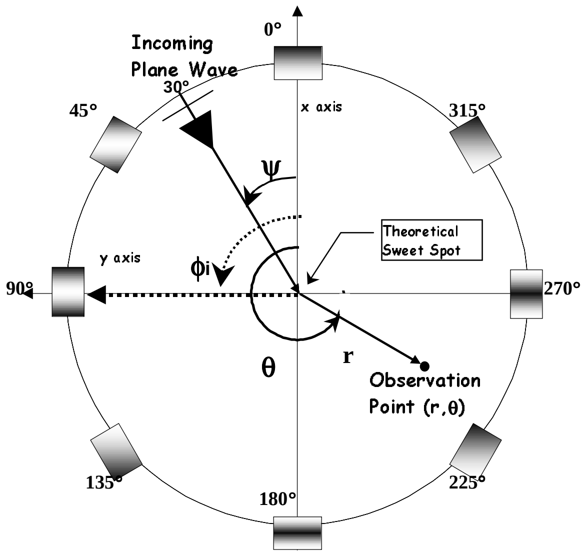

Figure 2. *Horizontal Surround Sound Layout.*

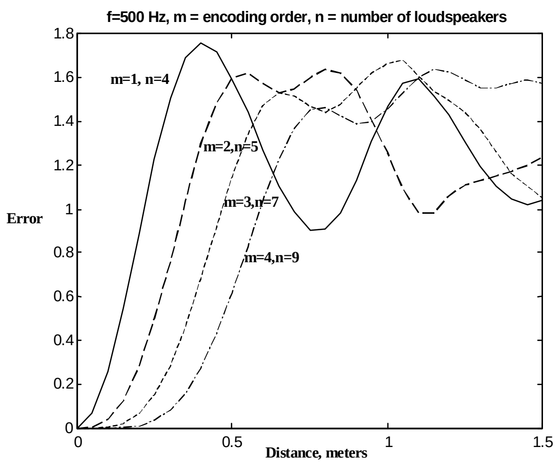

Figure 3: *D-Error Vs. Distance; Ambisonic Systems*

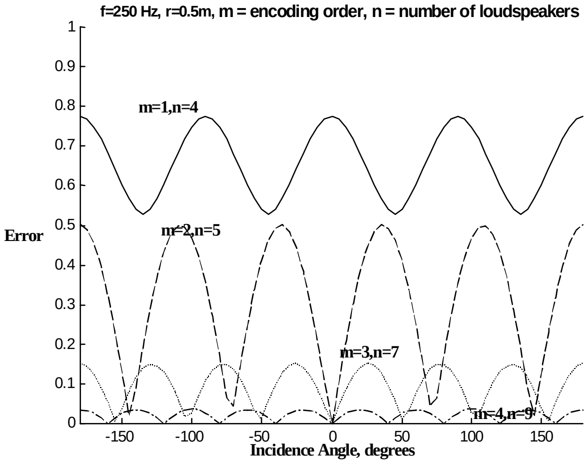

Figure 4: *D-Error vs. Incidence Angle; Ambisonic Systems*

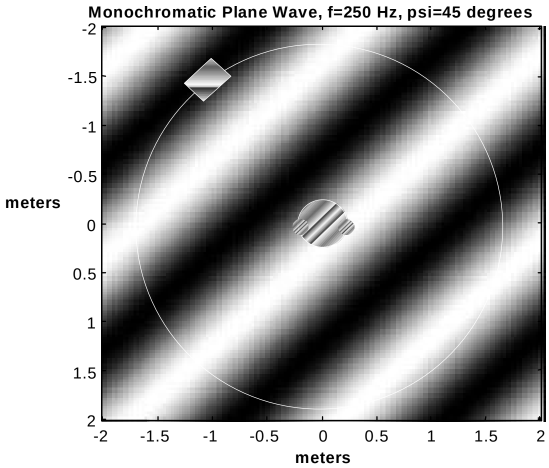

Figure 5: *Pressure Field of a Single Plane Wave*

2.2. Wave theoretical techniques: Interference Patterns.

This simple approach allows the visualisation of the acoustical field generated in the listening space by multiple sources.

The technique, implemented by the authors as a module for the "Ambitools" Ambisonic evaluation toolbox for Matlab [7], treats the loudspeakers in the layout as ideal plane wave sources (equation 3) emitting a monochromatic wavefront at an angle of incidence ϕ and displays the combined instantaneous pressure field generated over the area under investigation. Free-field propagation is assumed in order to keep the complexity at a minimum.

Figure 5 shows the pattern generated over a $16\,m^2$ area by a single sinusoidal 250 Hz plane wave at 45° clockwise from the x-axis[1].

It can be noticed, however, that due to the theoretical nature of the source (plane wave radiator), the magnitude of the acoustical field does not decrease with distance and extends beyond the position of the radiator.

In order to build a more complete picture of the acoustical field in the listening area, the Matlab script also takes into account the time dependence of equation (3), allowing an animated visualisation of the combined wavefront and, therefore, simplifying the task of recognising patterns on the boundaries of the sweet spot and to roughly estimate their extent.

Figure 6 depicts the reconstruction of the single monochromatic wave by means of a second order Ambisonic system employing five loudspeakers; the loudspeakers are superimposed onto the interference pattern only to provide a directional reference, since we assume that the radiators are placed at infinite distance in order to account for the plane wave nature of propagation.

The ellipse at the centre of the area under investigation roughly describes the extent of a pattern that appears to be travelling in the desired direction (shown by the arrow).

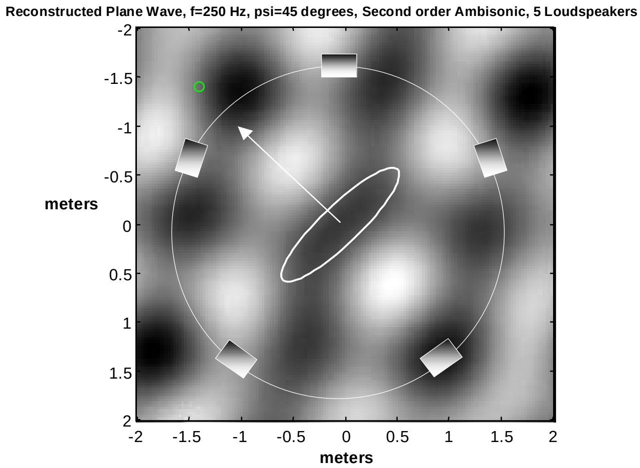

Figure 6: *Pressure Field of the Reconstructed Plane Wave, 2nd Order Ambisonic,5 loudspeakers*

Curiously, by increasing the extent of the area under investigation, it was also noticed how such pattern can be observed at locations away from the sweet spot, suggesting the existence of multiple ideal listening locations.

2.3. Perceptual techniques: Localisation Vectors.

In surround sound theory the Velocity and Energy localisation vectors are often employed for assessing the quality of auditory images.

These measures were originally proposed by Gerzon [8] as approximate psychoacoustic criteria for the development of optimal Ambisonic encoders and decoders but have also been used for assessing other holographic techniques [9].

The velocity vector describes the properties of the image generated by a loudspeaker layout according to the low frequency theory of human auditory localisation (interaural time difference-

[1] The modified co-ordinate system of figure 2 is still being used as the reference.

based) therefore it can be considered valid for frequencies below 700 Hz [8]; the velocity vector can be mathematically defined as:

$$\mathbf{r}_v = \frac{\sum_{i=1}^{N} a_i \mathbf{u}_i}{\sum_{i=1}^{N} a_i} = a_V \mathbf{u}_V \quad (4)$$

where a_i is the amplitude gain of the *ith* loudspeaker in the layout and $\mathbf{u}_i$ is the unitary vector pointing at the loudspeaker angle ϕ_i.

Equation (4) is shown graphically in figure 7; in practice the velocity vector $\mathbf{r}_V$ represents a synthetic plane wave with an apparent direction of propagation:

$$-\mathbf{u}_V = (\cos(\phi_v), \sin(\phi_v)) \quad (5)$$

It can be shown [10] that, near the centre of the listening area, the speed of propagation of the wavefront is directly related to the factor a_V:

$$c' = c / a_V \quad (6)$$

where *c* is the velocity of sound in air (~340 m/s) and *c'* is the propagation speed of the synthetic wavefront.

Therefore, the interaural time difference induced by the synthetic wavefront will roughly correspond to that of a natural plane wave only if $a_V = 1$.

The velocity vector starts losing its usefulness at frequencies above 700Hz, where the main localisation cues are the interaural level differences (ILD).

In order to assess the properties of the synthetic wavefront at these frequencies, a modified version of equation (4), where the loudspeaker gains are replaced by their squared magnitudes, can be used:

$$\mathbf{r}_E = \frac{\sum_{i=1}^{N} g_i^2 \mathbf{u}_i}{\sum_{i=1}^{N} g_i^2} = a_E \mathbf{u}_E \quad (7)$$

The quantity $\mathbf{r}_E$ is commonly referred to as the "energy vector", and describes the direction and stability of the apparent sound source at high frequencies.

When used together, equations (4) and (7) should give a complete picture of the quality of the image generated by a surround sound system.

According to Gerzon, in fact, if the magnitude of both vectors is equal to 1, and their direction of propagation coincides, the perceived phantom image should be virtually indistinguishable from a real acoustical source.

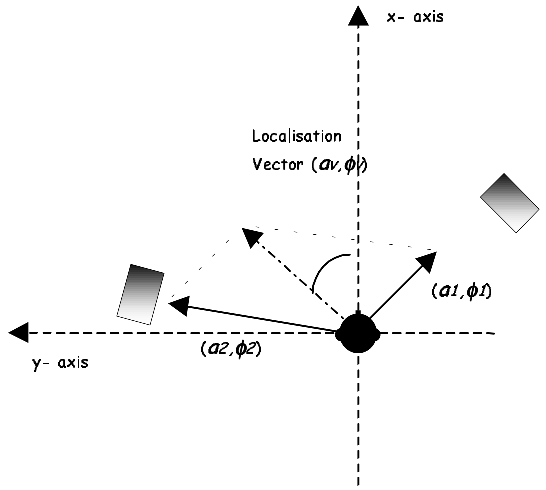

Figure 7: *Velocity Vector*

3. NON IDEAL LISTENING CONDITIONS IN LARGE AUDIENCE AREAS.

As discussed in the introduction, the ideal listening conditions assumed by the large majority of the assessment methods described in the previous section do not apply in the instance of holographic reproduction systems serving large audience areas and, in this section we attempt to describe how "real world" issues might affect the accuracy of these assessment techniques.

3.1. Incoherence of the Combined Wavefront.

In the general case a listener will be located at a point in the audience area where the times of arrival of the wavefronts from each of the loudspeakers in the reproduction layout will be different and, as a consequence, the precedence effect, or law of the first wavefront [11], might degrade the perceived auditory image.

This perceptual effect, responsible for aiding the localisation of sources in the presence of reverberation, is widely regarded as the biggest culprit for the collapse of images at off-axis locations in two-channel stereo systems.

In the instance of two loudspeakers emitting the same signal at the same amplitude but with a relative delay of more than 1mS, in fact, the perceived auditory image will be pulled towards the undelayed speaker.

In a large surround system installation, however, the direction of the perceived image might not be as easy to predict since multiple wavefronts, each with a different delay and magnitude, will reach the listener located far away from the sweet spot.

Furthermore in extreme cases, such as in concert arenas or stadia, the loudspeaker-listener distances can be such that a wavefront coming from the farthest clusters will be perceived as a separate echo (relative delays > 80-100mS).

Wave theoretical techniques, such as the integrated-D error and interference patterns only assess the reconstructed wavefronts in

terms of their pressure distribution and do not take into account any perceptual aspect.
This could ultimately lead to overoptimistic estimates of the quality of the reconstructed image. In the instance of the D error for a high order Ambisonic system, in fact, we may find that while the predicted reconstruction error is very small within a radius of 1.5 meters, a listener standing at the boundary of the area under investigation could, instead, perceive the image as coming from the nearest loudspeaker.
A similar misjudgment could be also made by exclusively relying on the observation of the interference patterns generated by surround layouts. In section 2.2 it was mentioned, in fact, that the pattern generated at the sweet spot location can also be reproduced elsewhere within, or even outside, the listening area.

The localisation measures presented in section 2.3 also do not seem to consider the precedence effect.
Although both the Velocity and Energy localisation criteria can be expressed as complex vectors [10], the perceptual effect of their imaginary part (the "phasiness" component) is usually described as that of impairing correct localisation by "*producing listener fatigue and poor localization quality as well as affecting tone color*"[9].
The Velocity vector criterion accounts for broadband phase shifts caused, for instance, by phase reversals in the reproduction layout and can be therefore used as a quantitative measure for localization blurs and unpleasantness of the perceived image at locations very close to the centre of the listening area. However, it cannot be considered a reliable predictor for the perceived direction of the image at locations removed from the sweet spot.
The Energy vector, on the other hand, is a more robust measure for off-centre locations since, at high frequencies, the phase relationships between the pressures at the ears have a negligible effect in terms of localisation.
It is claimed [12] that this criterion can also be applied to low frequency localisation in phase-incoherent case; nevertheless, we suspect that this claim can be considered correct only at locations where the precedence effect is not too strong.

3.2. Non-Planar Wavefront Propagation.

All the techniques presented in section 2 assume plane wave propagation from the loudspeakers in the reproduction layouts.
This approximation can be considered valid for radiators placed very far from the audience area; in the general case, however, especially if a listener is in close proximity to a loudspeaker, spherical propagation (attenuation with distance) should be used to model the wavefront properties.
This assumption does not invalidate any of the assessment methods discussed so far since only an amplitude scaling factor as a function of the listener-loudspeaker distance would be introduced.
Discarding the influence of the precedence effect, in fact, proximity of the listening location to a loudspeaker would simply steer, in a predictable way, the Velocity and Energy vectors towards the closest loudspeaker.
Also, the integrated-D error should remain conceptually valid, given that this measure is simply a comparison between two pressure distributions.

By replacing the theoretical sources with monopoles in the interference plots, it is also possible to see how the reconstruction of wavefronts is affected.
Figure 8 depicts the pressure field generated by the same Ambisonic system shown previously in figure 6 with sources placed at a finite distance (4 meters) from the centre of the listening area.
The spherical propagation of wavefronts is evident along with the distance dependent attenuation and the sweet spot also appears to be decreased in size, indicating that the wave field reconstruction produced by Ambisonic systems is degraded when real sources are employed.

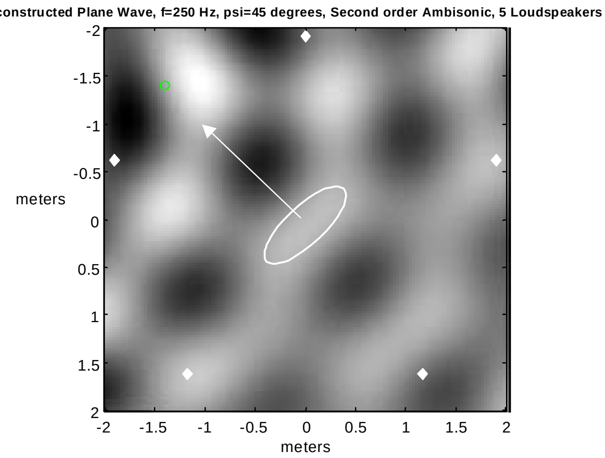

Figure 8: *Ambisonic Reconstruction Using Monopoles*

3.3. Acoustical Properties of the Space.

Free field propagation is generally assumed in holographic reconstruction theory.
In practice, however, the presence of an acoustical source in an enclosed space will generate a reverberant field that can affect the reconstructed wavefront and degrade the performance of the surround reproduction system.
The main cause for localisation inaccuracy can be identified in strong reflections occurring within the first 80 mS from the direct sound, since later reflections will be perceived as separate echoes. From the point of view of image assessment, these reflections can be interpreted as "phantom" loudspeakers present in the room; therefore their influence on the perceived image will depend on the time of arrival, direction and intensity.
Figure 9 shows the reflection pattern [13] in the instance of a four loudspeaker system installed in a rectangular room.
In the general case the absorption coefficients of the surrounding surfaces will reduce the intensity of these reflection to a such degree that the only perceived effect will be an increase in phasiness and localisation blur.
In the author's experience, however, the strength of reflections bouncing from untreated surfaces sometimes can be high enough to steer the image away from the desired direction, therefore a computational tool for the prediction of the quality of phantom images should also take reverberation effects into account.

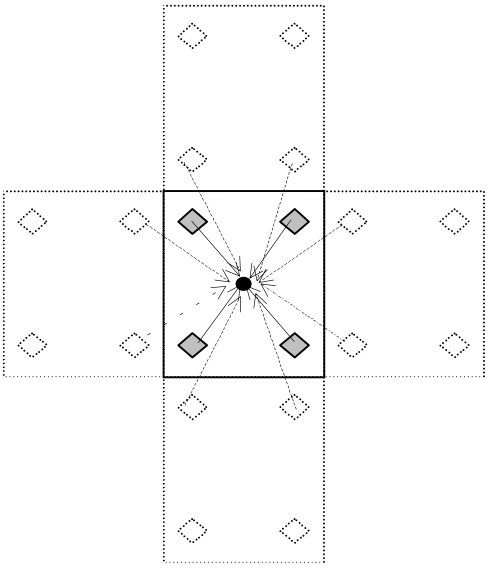

Figure 9: *First Order Reflections*

4. CONCLUSIONS

This paper presented a general overview of objective evaluation methods for the image reconstruction accuracy of holographic sound reproduction systems and a discussion of how their accuracy might be affected under general listening conditions.
The single major source for a possible failure of these techniques has been identified with the precedence effect in human localisation. This feature of the auditory system, in fact, tends to pull phantom images towards the nearest loudspeaker, effectively voiding most of the results returned by wave-theoretical and approximated perceptual methods.

Although countless factors, some of which have been listed in this paper, can affect the localisation of phantom images, it is the authors' belief that the development of a strong computational model for evaluating the precedence effect could lead to a good prediction of the impact of surround sound systems on large audiences.

References:

[1] Begault, D. R.; "3-D Sound for Virtual Reality and Multimedia"; Cambridge, MA, Academic Press Professional (1994)

[2] Pulkki, V.; "Virtual Source Positioning Using Vector Base Amplitude Panning"; J. Audio Eng. Soc., Vol. 45, June 1997

[3] Landone, C.: "Introduction to Ambisonic"; available online at http://www.eee.kcl.ac.uk/~chrisl

[4] Bamford, J.S.; "An Analysis of Ambisonic Sound Systems of First and Second Order" MSc. Thesis, University of Waterloo, Canada, 1995.

[5] Krauss, J.D.; "Antennas"; McGraw-Hill, New York, 1988

[6] Vanderkooy, J.; Lipshitz, P.L.: "Anomalies of Wavefront Reconstruction in Stereo and Surround-Sound Reproduction"; presented at the 83rd Convention of the Audio Engineering Society, New York, October 1987, Preprint 2554,

[7] "Ambitools: Ambisonic Toolbox for Matlab": available online at http://www.eee.kcl.ac.uk/~chrisl

[8] Gerzon, M.A.: "General Metatheory of Auditory Localisation"; presented at the 92nd Convention of the Audio Engineering Society, Vienna, March 1992, Preprint 3306.

[9] Dickins, G.; Flax, M.; McKeag, A.; McGrath, D.; "Optimal 3D Speaker Panning"; Presented at the Audio Engineering Society 16th International Conference, April 1999, Rovaniemi, Finland.

[10] Daniel, J.; Rault, J.;Polack, J: "Acoustic Properties and Perceptive Implications of Stereophonic Phenomena"; Presented at the Audio Engineering Society 16th International Conference, April 1999, Rovaniemi, Finland.

[11] Theiss, B.; Hawksford, M.O.J.; "Binaural Model-Based Measurements of Phantom Sources"; presented at the 107th Convention of the Audio Engineering Society, New York, September 1998, Preprint 5048.

[12] Gerzon, M.A.: "Panpot Laws for Multispeaker Stereo"; presented at the 92nd Convention of the Audio Engineering Society, Vienna, March 1992, Preprint 3309.

[13] Allen,J.B.; Berkley, D.A.; "Image Method for Efficiently Simulating Small-Room Acoustics"; Journal of the Acoustical Society of America, Volume 65 (1979)