ISCAS 2000 - IEEE International Symposium on Circuits and Systems, May 28-31, 2000, Geneva, Switzerland

APPLICATIONS OF BINAURAL PROCESSING TO SURROUND SOUND REPRODUCTION IN LARGE SPACES

C. Landone, M. Sandler
Department of Electronic Engineering, King's College London
The Strand
WC2R 2LS, London, UK

ABSTRACT

This paper presents a method for evaluating the performance of multi-channel sound reproduction systems serving large audience areas.

The technique, based on the synthesis of binaural cues, emulates the pressures at the listener's ears corresponding to those produced by a real sound field thus, in theory, allowing the assessment of the imaging produced by holographic sound systems to be carried out simply by using a pair of headphones.

Preliminary results obtained by applying this method to a simple surround sound system are presented and discussed.

1. INTRODUCTION

Holographic-based multi-loudspeaker systems reproduce a two or three-dimensional sound field in a confined area and generate natural auditory cues as a result of the diffraction of the combined wavefront around the listener's head and torso.

The common two channel stereo system can also be considered holographic since it was designed to create low frequency auditory cues that allow the localisation of a sound image in an arc subtended by the two loudspeakers.
Stereo, however, is unsuitable for large listening areas due to the instability of phantom sources for listening positions not coinciding with the mid-loudspeakers axis.

Multi-channel reproduction systems have been introduced partly to reduce the problem of image stability, but mainly for improving the quality of the auditory experience by extending the sound stage to rear and lateral locations.
The large majority of holographic techniques, however, are based on the assumption that the listener is located in a point in space equidistant from the loudspeakers of the reproduction layout.
Such condition, however, cannot be met in large installations, such as movie theatres or concert venues, where the majority of the listeners are located in non-optimal positions.

Predicting the quality of the perceived image as a function of the listener's position is an extremely difficult task since current analytical methods for assessing the performance of multi-channel systems are unable to quantify the image degradation occurring at non optimal locations.

Wave-theoretical techniques such as the Integrated-D error [1] estimate the mismatch, in terms of instant pressure distribution, between a real and a reconstructed wavefront but neglect the perceptual aspects of the combined wavefront generated by the loudspeaker layout.
Other measures, based on rough approximations of human auditory localisation, such as the velocity and energy vectors [2], on the other hand, assume equal time of arrival of the sound from the loudspeakers and are, therefore, reliable only at the "sweet spot".

One of the major culprits for the degeneration of phantom images in locations away from the sweet spot is an auditory mechanism that allows localisation in presence of reverberation: *the precedence effect*.
In the instance of two loudspeakers emitting the same signal at the same amplitude but with a relative delay of more than 1mS, the perceived auditory image will coincide with the undelayed speaker. This "law of the first wavefront" is only one of many aspects that current assessment methods for holographic sound systems do not take into account and, in practice, the only reliable method for fully evaluating the reproduced auditory images is by means of listening tests at different locations within the audience area.

Other than physically installing a complete loudspeaker layout, which can be extremely costly and time consuming, the only method for subjectively assessing the imaging generated by a holographic sound systems under non optimal conditions is to resort to auralisation.
This technique renders an arbitrary sound field through a pair of headphones, placing the listener within a virtual auditory scene.

The paper describes a method for simulating the auditory cues generated by holographic reproduction systems in a virtual space employing a binaural processor developed by the authors and, currently, this approach is being used to evaluate the Ambisonic surround sound format restricted to the horizontal plane.
In section 2 a brief description of the principles behind Ambisonic will be given while in section 3, after an overview of spatial perception, the technique employed for rendering the surround sound field will be presented.

Finally, the results of some preliminary, and informal, listening tests will be discussed in order to highlight the limitations of this method

2. OVERVIEW OF AMBISONIC THEORY

The primary assumption in Ambisonic is that, for relatively distant sound sources, the generated wavefront can be approximated by a plane wave. With reference to figure 1, a monochromatic plane wave can be fully characterized by its incidence angle ψ and by its pressure at the co-ordinate origin[1], therefore the pressure of the wave observed at the point (r, θ) can be expressed as:

$$p(r,\theta) = A\exp(j\omega t)\exp(jkr\cos(\theta - \psi)) \quad (1)$$

where A is the peak pressure, ω is the angular frequency and k is the wave number $2\pi/\lambda$.

Equation 1 can also be expressed in terms of Bessel functions [3]:

$$p = A\exp(j\omega t)\left\{ J_0(kr) + 2\sum_{n=1}^{\infty} j^n J_n(kr)\left[\cos(n\theta)\cos(n\psi) + \sin(n\theta)\sin(n\psi)\right]\right\} \quad (2)$$

where the term $J_n(kr)$ represents Bessel functions of the first kind.

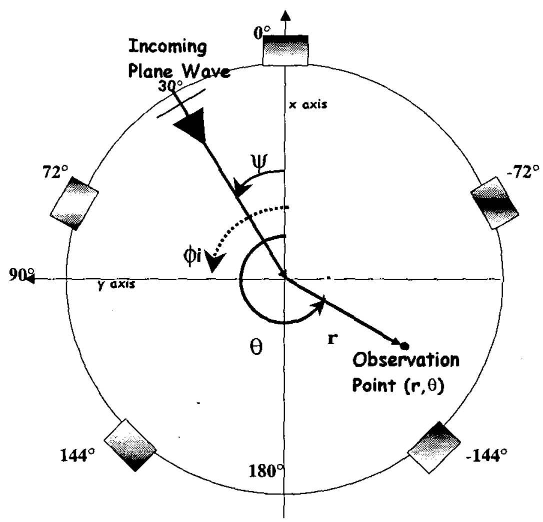

Figure 1: Ambisonic Reproduction System

The Ambisonic system consists of a circular array of N loudspeakers enclosing the listening area, represented by the shaded boxes in figure 1.

Again we assume that the acoustical field, s_i, generated by the i_{th} loudspeaker, located at an angle ϕ_i and with peak pressure a_i is a plane wave and can be written as:

$$s_i = a_i \exp(j\omega t)\exp(jkr\cos(\theta - \phi_i)) \quad (3)$$

[1] Note that in Ambisonic theory, the co-ordinate system is rotated anticlockwise by 90°

Therefore, if we express (3) as a Bessel series, the total sound field at the listening point generated by a layout with N in-phase loudspeakers is:

$$S = \exp(j\omega t)\{ \sum_{i=1}^{n} a_i J_0(kr) + 2\sum_{n=1}^{\infty}[\ j^n J_n(kr)\sum_{i=1}^{N} a_i(\ \cos(n\theta)\times$$
$$\times\cos(n\phi_i) + \sin(n\theta)\sin(n\phi_i)\ )\ ]\ \} \quad (4)$$

Values for the coefficients a_i that allow the loudspeaker layout to reproduce the original wavefront can be found by truncating the Bessel series to an order K and by comparing equations (2) and (4), yielding the set of equations:

$$\begin{aligned} A &= \sum_{i=1}^{N} a_i \\ A\cos(n\psi) &= \sum_{i=1}^{N} a_i \cos(n\phi_i) \\ A\sin(n\psi) &= \sum_{i=1}^{N} a_i \sin(n\phi_i) \end{aligned} \quad (5)$$

$$\text{for } n = 1, 2, \ldots, k$$

Equation (5) can be easily solved in matrix form [4] and, provided that the loudspeakers in the layout have equal angular spacing (as in the case depicted in figure 1), a simple analytical expression for the loudspeaker feeds can be found:

$$a_i = \frac{A}{N}\left[1 + 2\sum_{n=1}^{K}\cos(n(\psi - \phi_i))\right] \quad (6)$$

Some features of the Ambisonic technique must be highlighted following this brief overview.

Theoretically, in order to be able to perfectly reconstruct a plane wave, an infinite number of Bessel functions (and, according to equation (5), also of loudspeakers) are required, as it can be deduced from the upper bound of the sums in eqns.(2) and (4).

The plane wave assumption is also a rough approximation since loudspeakers cannot be regarded as plane wave radiators, therefore in a real Ambisonic system a wavefront reconstruction error will be introduced by the distance dependent sound pressure attenuation.

By evaluating eqn. (6) for different layouts and orders of the Bessel series truncation, some interesting "clues" regarding the behaviour of Ambisonic reproduction systems have also emerged during our investigation.

For a given truncation order K, if the number of loudspeakers N follows the rule $N = 2K + 1$, then perfect reconstruction will be assured everywhere in the listening area for plane waves with an incidence angle ψ coinciding with a loudspeaker location ϕ_i since the feeds of all the other sources will be set to zero. Also, the polarity of the feeds is not restricted to positive values, hence for most incidence angles of the notional plane wave, some loudspeakers will emit phase-reversed signals, causing unpleasant auditory artifacts away from the sweet spot (as well as severely degrading the image).

3. BINAURAL RENDERING OF AMBISONIC SOUND FIELDS

Sound sources can be localised in a three-dimensional space according to a number of auditory cues.
In the low frequency regime (<1KHz) displacements from the frontal position produce a difference in the time of arrival of the sound between the ears; this delay, commonly referred to as interaural time difference (ITD) reaches a maximum of 0.7 mS for sources with a lateral displacement of 90°. At higher frequencies, the acoustical shadowing effect of the head affects the amplitude of smaller wavelengths reaching the ear located farther from the wavefront, creating a frequency dependent interaural level difference (ILD) between the two ears.
The combined effect of ITD and ILD on the auditory system is generally regarded as the dominant localisation cue on the horizontal plane but is not sufficient to explain the discrimination of rear or elevated sources.
According to studies on perception [5], the disambiguation of loci sharing the same combination of interaural time and level differences is allowed by the unique direction dependent filtering of the sound caused by the external ears (pinna cues).

The reverberant fields, caused by reflection of the wavefront on the surfaces surrounding the listener also appear to be determinant for localisation. The intensity ratio between the direct and reverberant sound fields is, in fact, widely regarded as the main cue for the judgement of auditory distance [6].
In theory, provided that all the localisation cues discussed above are faithfully recreated, it should be possible to produce a realistic three-dimensional auditory experience.

The authors have developed a software-based processor, shown in figure 2 and discussed in more detail elsewhere [7], that allows the placement of a sound source in a three-dimensional virtual auditory space.
The core of the procedure consists in convolving a monoaural audio stream with a pair of filters (directional filters in figure 2) representing the transfer function, between a source located at a specific point in space and the artificial ear canal of an anthropomorphic mannequin used in hearing research. Since these transfer functions are usually measured in an anechoic space, they contain all the directional information discussed above, with the exclusion of distance cues. The signal is then presented via a pair of headphones, effectively by-passing the listener's own auditory cues.

The directional filters were obtained by converting the original Binaural head related impulse responses (B-HRIR) of the MIT database [8] into minimum phase (delay) representations and by estimating the ITD through threshold detection. After diffuse-field equalisation, the length of each FIR filter was truncated to 66 coefficients.
The distance cues were implemented using a reverberator (distance module in fig. 2) based on a modified Schroeder topology [9] and simulating the acoustical response of a rectangular room. The simulated room impulse response consists of an early reflections pattern, produced by two tapped delay lines which, subsequently, feed a network of all pass filters for the generation of the decaying part of the reverberation.

The early reflections are therefore modeled in terms of their time of arrival and attenuation and can provide rough but effective localisation cues in the horizontal plane by setting the proper delay and tap gain values in the delay lines.

The structure of the algorithm employed for generating the virtual Ambisonic sound field is shown in figure 3, where each binaural processor represents a loudspeaker placed in the virtual auditory space. The set of parameters describing the truncation order and number of transducers employed in the Ambisonic system are used in equation (5) to calculate the gain factors to be assigned to each virtual loudspeaker.
For a given observation point (θ,r) in the virtual space, the angular location and relative distance, with respect to the listener, of each transducer is calculated.
The coefficients of the filter pairs from the B-HRIR database are subsequently loaded into the appropriate binaural processor, according to the new angle values, and the calculated distance is employed to set the propagation delay on each virtual loudspeaker.
Since the radiators have been modeled as point sources, the distance between each virtual loudspeaker and the observation point is also used to modify the gain values to take spherical propagation into account.

The reverberation generated by the distance modules within each virtual loudspeaker is mainly employed by this rendering algorithm to obtain source externalization in headphones reproduction and no attempt has been made to simulate the room impulse response relative to a specific reproduction space.
Future improvements of this scheme will include a more complete model of room acoustical response, allowing the inclusion of the effect of early reflections on image degradation in our studies.

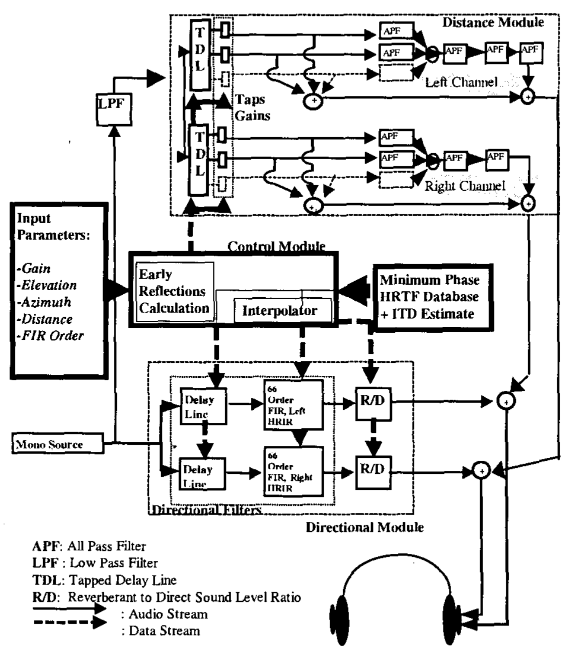

Figure 2: Binaural Processor

4. PRELIMINARY RESULTS

In this section some significant results obtained through informal listening tests are presented and discussed.
A second order Ambisonic system consisting of five equally spaced loudspeakers (figure 1) has been simulated using the algorithm described in the previous section.
The first experiment consisted in generating a virtual image at 72° anticlockwise, the observation point had been initially set at the co-ordinate origin, with the virtual listener facing forward (0°), and the test signal consisted in a music track with a high transient content.
As expected, following the discussion in section 2, since this incidence angle coincides with a loudspeaker location, the rendered image was well focused and, by moving the observation point away from the axis origin, towards the central loudspeaker (0°), the perceived image shifted towards the rear quadrant. In addition, as the distance from the only active loudspeaker increased, the virtual source externalisation was more noticeable.
A further experiment consisted in generating a wavefront at 108° anticlockwise. This angle coincides with a location exactly mid-way between two loudspeaker pairs and, although the perceived image appeared to be located somewhere in a lateral position, it lacked the sharpness of the single loudspeaker case.
Moreover, as the observation point was advanced toward the central loudspeaker, the image first appeared to "jump", with a sudden increase in focus, into the virtual loudspeaker at 72 ° and subsequently, as the virtual listening position was shifted more towards the central speaker, it became "phasey".
By examining the feed values returned by eqn. 5, this result is hardly surprising. For that particular incidence angle, in fact, the gain values are:

(-0.2472 0.6472 0.6472 -0.2472 0.2000)

hence the loudspeakers at 72° and 144° emit a signal with the same intensity and the source at 0° has a phase reversal.
Therefore, moving towards the center speaker first introduced a delay between the time of arrival of sound from the speakers at 72° and 144°, causing the onset of the precedence effect and subsequently, when the intensity between the 0° and 72° sources became comparable, phase reversal artifacts were noticeable.

Although these results are consistent with listening tests carried out in real Ambisonic systems, the technique described in this paper has a number of shortcomings.
The distance cues generated by the processors are quite crude, hence some initial empirical adjustments were required in order to obtain an unambiguous source externalisation. Furthermore, it was generally observed that frontal images are considerably more difficult to localize compared to sources placed in the rear quadrant or in lateralised locations.
This last problem can be attributed to the lack of freedom in terms of head movement and to the mismatch between the listener's own pinna cues and those employed in the simulation
The computational complexity of this procedure is fairly high, since each virtual loudspeaker in the layout requires 180 multiplications/sample; however, considering how inexpensive high end DSP boards have become, this issue should be regarded as the least worrying.

5. CONCLUSIONS

A method for assessing surround sound systems at non-optimal listening locations has been presented.
This approach, based on simultaneous binaural processing of multiple sources, renders the acoustical field generated by multi-loudspeaker layouts, allowing "virtual listening tests" simply by means of headphones.
The results obtained by simulating the combined wavefront of an Ambisonic system are generally consistent with both theory and actual listening tests and some shortcomings of this assessment method have also been highlighted.

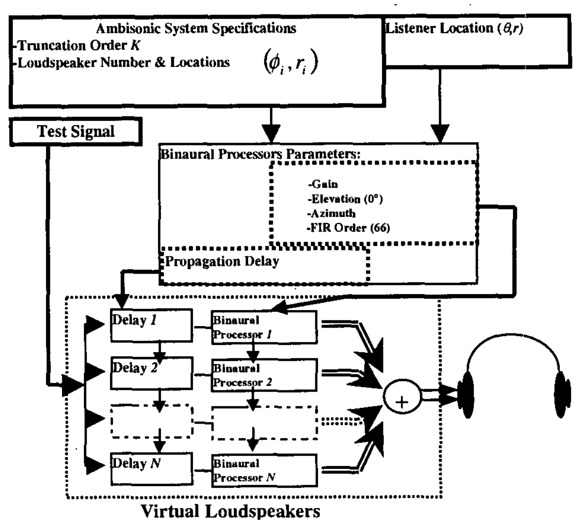

Figure 3: Rendering Algorithm

[1] 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,
[2] Gerzon, M.A.: "General Metatheory of Auditory Localisation"; presented at the 92nd Convention of the Audio Engineering Society, Vienna, March 1992, Preprint 3309.
[3] Morse and Ingard: "Theoretical Acoustics"; McGraw-Hill, New York, 1968
[4] Landone, C.: "Introduction to Ambisonic"; available online at http://www.eee.kcl.ac.uk/~chrisl
[5] Blauert J.:"Spatial Hearing. The Psychophysics of Human Sound Localization" MIT Press,
[6] Mershon D.H., Bowers J.N.: "Absolute and Relative Cues for the Auditory Perception of Egocentric Distance" , Perception, No 8 1979.
[7] Landone, C, Sandler, M..: A Scalable Binaural Processor"; available online at http://www.eee.kcl.ac.uk/~chrisl
[8] Gardner, W; Martin, K.; "HRTF Measurements of a KEMAR Dummy-Head Microphone" MIT Media Lab Perceptual Computing - Technical Report #280 May 1994
[9] Schroeder, M.R.; "Natural Sounding Artificial Reverberation"; Journal of the Audio Engineering Society, No 3 Volume 10, March 1962