ENHANCED 3D SOUND FIELD SYNTHESIS AND REPRODUCTION SYSTEM BY COMPENSATING INTERFERING REFLEXIONS

A. Sontacchi, R. Hoeldrich

Institute of Electronic Music and Acoustics
University of Music and Dramatic Arts Graz
alois.sontacchi@mhsg.ac.at, robert.hoeldrich@mhsg.ac.at

ABSTRACT

The antique stereophonic recording and playback format is going to be replaced by new surround sound formats in the near future. At the moment, various surround techniques are being investigated in many artistic and technical applications. The main concern is to find an appropriate recording and playback format which supports the natural spatial hearing cues. Therefore, surround sound systems should provide a homogeneous and coherent sound field image, both for recorded and synthesized sound fields [1]. In a homogeneous sound reproduction system, no direction is treated preferentially. Coherent sound field image means that the image remains stable under changes of the listener position, though the image may change as a natural sound field does. The Holophony and Ambisonic system described by Nicol and Emerit [2] is the basic approach. This system will be extended by a new approach to compensate the interfering reflections of the reproduction room. Further possibilities to determine higher order Ambisonic signals using the beam forming approach are investigated.

1. INTRODUCTION

This paper deals with a improved implementation of a sound spatialisation system rendering a recorded or synthesized 3D sound field using an arbitrary loudspeaker array in any desired room. First of all, the main spatial hearing cues must be considered. The following mechanisms indicate the position of sound in 3D space to the auditory cortex:

- Time differences between the ears (ITD, interaural time difference)
- Intensity differences between the ears (ILD, interaural level difference)
- Complex distance cues (reverberation, spectral balance modifications)
- Spectral cues caused by the shape of the head and the body (HRTFs Head Related Transfer Functions)
- Sensory augmentation (supplementary visual cues)

Conventional cinema or TV surround systems (e.g. Dolby Prologic, DTS etc.) do not accomplish all these requirements because they only attempt to give a general impression of the space. Due to additional visual cues they work quite well and satisfying. HRTF related systems perform very accurately but depend on the match of the listener's and the dummy's head and torso shape. The periphony systems like Wave Field Synthesis, Holophony and Ambisonic try to replicate the original sound pressure field in a two- or three dimensional region around the listener. Thus the localization cues are accomplished naturally. Ambisonic is a special case of Holophony and both Holophony and Wave Field Synthesis can be regarded as equivalent, because both exploit the Huygens‘ principle [2], [3] and [4] .

In the second section, the advantages and disadvantages of the Ambisonic and the Holophony system will be considered. In the third section the new approach to compensate the reflections of the reproduction room will be presented. Section 4 and 5 give a brief summery and mentions further investigations concerning the beam forming approach.

2. AMBISONIC AND HOLOPHONY

2.1. Ambisonic

Ambisonic offers a suitable general basis for sound field synthesis and sound field reproduction [5]. In the Ambisonic system, the sounds and their directional components are encoded vectorially in a set of spherical harmonics. The advantages of Ambisonics are the simple encoding formulations for arbitrarily sound sources (synthesized sound field), straightforward rotations of the whole sound field and the independency of encoded (recorded) signals and reproducing loudspeaker arrays. The disadvantages are the two basic assumptions: The sources radiate plane waves and the sound fields produced by the loudspeakers are plane waves at the position of the listener, too. These assumptions are almost met if the sources and loudspeakers are far away from the listener position. Nevertheless there are non-negligible problems in small rooms [6]. Below the coding and decoding rules are derived.

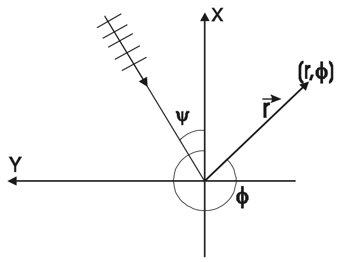

Figure 1. Recording of a single source with direction ψ.

Let us consider[1] the sound field in a point $\vec{r} = (r,\phi)$, caused by a planar wave (reference wave) arriving from direction ψ, as given in Eq. 1 (and Fig.1).

$$S_{\text{Referenz}}(r,\phi) = P_\psi \cdot e^{j\vec{k}\cdot\vec{r}} = P_\psi \cdot e^{jk\cdot r\cdot\cos(\phi-\psi)} \tag{1}$$

The vector $\vec{k}$ represents the wavenumber $k = \frac{2\pi}{\lambda}$ and the direction of arrival. Furthermore, the vector $\vec{r}$ describes the point of interest (listening position). During reproduction, N loudspeakers are arranged symmetrically along a circuit (Fig. 2). Each of the loudspeakers (P_n, ϕ_n) radiates a plane wave (Eq. 2).

$$S_n(r,\phi) = P_n \cdot e^{j\vec{k}_n\cdot\vec{r}} = P_n \cdot e^{jkr.\cos(\phi-\varphi_n)} \tag{2}$$

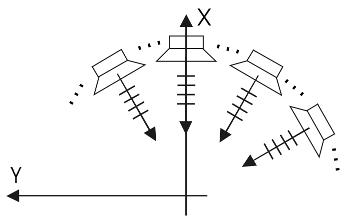

Figure 2. *Reproduction*

The superposition of the individual waves is described by the system wave $S_{\text{Ambisonic}}$.

$$\begin{aligned} S_{\text{Ambisonic}}(r,\phi) &= \sum_{n=1}^{N} S_n(r,\phi) = \sum_{n=1}^{N} P_n \cdot e^{j\vec{k}_n\cdot\vec{r}} \\ &= \sum_{n=1}^{N} P_n \cdot e^{jkr\cdot\cos(\phi-\phi_n)} \end{aligned} \tag{3}$$

In order to reproduce the original sound field, the sound field caused by the Ambisonic system has to meet Eq. 4.

$$S_{\text{Referenz}}(r,\phi) \equiv S_{\text{Ambisonic}}(r,\phi) \tag{4}$$

Planar waves are specified by a constant pressure term and an oscillating phase term (e.g. see Eq.1). The oscillating term can be developed in a Bessel-Fourier series [7]. Using this property, we can rewrite equation 1 and 3:

$$\begin{aligned} S_{\text{Referenz}}(r,\phi) &= P_\psi \cdot e^{jkr\cos(\phi-\psi)} \\ &= P_\psi J_0(kr) + 2P_\psi \sum_{m=1}^{\infty} i^m J_m(kr)\cos[m(\phi-\psi)] \\ &= P_\psi\left(J_0(kr) + 2\sum_{m=1}^{+\infty} i^m J_m(kr)[\cos(m\phi)\cos(m\psi) + \sin(m\phi)\sin(m\psi)]\right) \end{aligned}$$

$$\begin{aligned} S_{\text{Ambisonic}}(r,\phi) &= \sum_{n=1}^{N} S_n(r,\phi) = \sum_{n=1}^{N} P_n \cdot e^{j\overrightarrow{k_n}\cdot\vec{r}} \\ &= \sum_{n=1}^{N} P_n \cdot e^{jkr\cdot\cos(\phi-\phi_n)} \\ &= \sum_{1}^{N} P_n\left(J_0(kr) + 2\sum_{m=1}^{+\infty} i^m J_m(kr)[\cos(m\phi)\cos(m\phi_n) + \sin(m\phi)\sin(m\phi_n)]\right) \end{aligned}$$

Equation 4 is met if the coefficients of the $\cos m\psi$ and $\sin m\psi$ functions are identical. Therefore we get the matching conditions given in equation 5.

$$\begin{aligned} P_\psi &= \sum_{n=1}^{N} P_n \\ P_\psi \cdot \cos(m\psi) &= \sum_{n=1}^{N} P_n \cdot \cos(m\phi_n) \qquad m = 1,\dots,\infty \\ P_\psi \cdot \sin(m\psi) &= \sum_{n=1}^{N} P_n \cdot \sin(m\phi_n) \end{aligned} \tag{5}$$

This system consists of an infinite number of equations. Under practical considerations the number of equation is limited by the chosen system order m. The higher the system order, the more accurate the original sound field will be reproduced. Furthermore, the area of correct reproduction will be enlarged.

2.1.1 Encoding

On the left hand side of Eq. 5, we already see the necessary information to be recorded. The obtained signals – the so-called Ambisonic signals - contain the information about the source and the directions. If we regard synthesized sound fields, no problems occur in deriving higher order Ambisonic signals. On the other hand, recording higher order signals in real sound fields would require microphone characteristics that do not exist yet. As an example the Ambisonic signals for a source with signal S and direction ψ are given with

m=0 $W = 0.707 \cdot S$
m=1 $X = S \cdot \cos\psi \quad Y = S \cdot \sin\psi$
m=2 $U = S \cdot \cos 2\psi \quad V = S \cdot \sin 2\psi$
and so on.

The notation (W,X,Y,..) of the signals is a common convention.

[1] The following considerations are restricted to the 2-dimensional case.

2.1.2 Decoding

The decoding equations can be obtained by rewriting Equation 5 in matrix form.

$$\mathbf{b} = \mathbf{A} \cdot \mathbf{x} \tag{6}$$

with $\mathbf{b} = [W, X, Y, U, V, ...]^{\mathbf{T}}$

$$\mathbf{x} = \left[P_1, P_2, P_3, ... P_N\right]^{\mathbf{T}}$$

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \cos(\varphi_1) & \cos(\varphi_2) & & \cos(\varphi_N) \\ \sin(\varphi_1) & \sin(\varphi_2) & & \sin(\varphi_N) \\ \vdots & \vdots & & \vdots \\ \sin(m\varphi_1) & \sin(m\varphi_2) & \cdots & \sin(m\varphi_N) \end{bmatrix}$$

Vector **b** represents the Ambsionic signals, and the matrix **A** is defined by the arrangement of the loudspeaker array. The unknown loudspeaker feeds P_n are described by vector **x**. We obtain the loudspeaker feeds by solving the following normal equations

$$\mathbf{x} = \mathbf{A}^{\mathbf{T}} \cdot (\mathbf{A} \cdot \mathbf{A}^{\mathbf{T}})^{-1} \cdot \mathbf{b} \tag{7}$$

The loudspeaker feeds are found by weighting the Ambisonic signals according to the loudspeaker position. If the loudspeaker array is symmetrically arranged along a circuit (2 dimensional case, in the 3 dimensional case along a sphere) this will lead to:

$$P_i = \frac{1}{N}(W + 2X\cos(\varphi_i) + 2Y\sin(\varphi_i) + 2U\cos(2\varphi_i) + 2V\sin(2\varphi_i)\cdots)$$

As a restriction for the decoding array, the number of the loudspeakers must exceed the number of the encoded Ambisonic signals and the loudspeaker array should be as symmetrical as possible, otherwise a parameterization of the loudspeaker feeds is necessary [8], [9]. Using signals of higher order spherical harmonics, the „sweet spot“ can be enlarged resulting in better localization and more stable images. As a drawback the number of required transmission channels increases dramatically. Furthermore no microphone for the recording of higher order Ambisonic signals exists at the moment.

2.2. Holophony

In [2], a new method derives higher order signals from the sensor signals of a microphone array using the basic approach of Holophony. The approach is based on the “Huygens Principle” [1690], which states ’...each point of a wave front can be regarded as a starting point of an elementary wave which interacts with its surroundings...’ and is depicted in figure 3.

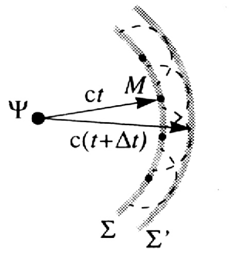

Figure 3. Illustration of the “Huygens Principle”.

In other words, the same principle can be described as follows: The wave field produced by a primary source Ψ can be synthesized by distributed secondary sources M.
The Kirchhoff-Helmholtz Integral is its mathematical formulation.

$$P(\vec{r}_R) = \frac{1}{4\pi}\oint_S \left[P(\vec{r}_S)\cdot\nabla_S G(\vec{r}_R|\vec{r}_S) - G(\vec{r}_R|\vec{r}_S)\cdot\nabla_S P(\vec{r}_S)\right]\cdot \vec{n}dS$$

with $G(\vec{r}_R|\vec{r}_S) = \dfrac{e^{-jk|\vec{r}_R - \vec{r}_S|}}{|\vec{r}_R - \vec{r}_S|}$ ...Green’s function.

The Kirchhoff-Helmholtz Integral implies that the wave field of an source free volume V can be described by the knowledge of the pressure along the enclosure surface S and the gradient of the pressure normal to the surface S.

In the following the 2 dimensional case is concerned. The continuous sound field is sampled at M points (as depicted in figure 4).

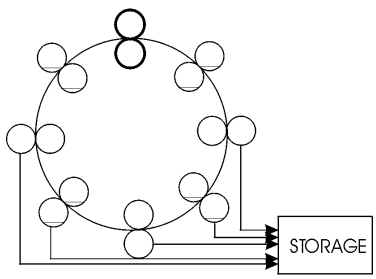

Figure 4. Recording situation with Holophony.

The number M of microphones required to prevent spatial aliasing increases with the size of the enclosed area. During reproduction, the N loudspeaker feeds are adjusted by an adaptive system to obtain the identical signals at the microphone positions.

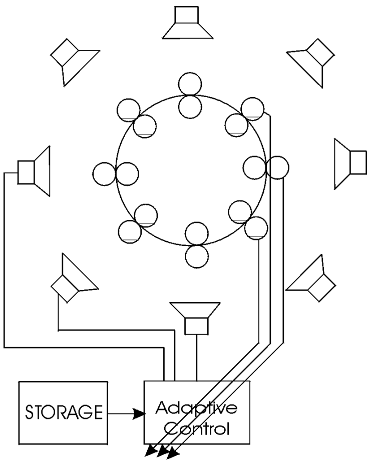

Figure 5. Playback situation with Holophony.

The advantages of the Holophony system are the simple recording procedure and the simultaneous elimination of the reproduction room during playback. On the other hand, a drawback of the playback system is the use of adaptive filters. The number of filters is $N \cdot M$. The length of the filter impulse response is determined by the length of the room reverberation time.

Considering the various advantages and disadvantages, it seems to be useful to combine both systems. Under a few assumptions, it can be proofed that Ambisonic is a special case of Holophony. In [2], it is shown how higher order Ambisonic signals can be derived out of the Holophony array signals. Therefore, the Holophony approach is used for recording. The recorded signals are transformed into Ambsionic signals, and the Ambsonic approach is applied to the reproduction system. Nevertheless the approach does not account for the reflections of the reproduction room. In the following part a new extension for the combined system is proposed that compensate these reflections.

3. COMPENSATION OF REFLECTIONS

In general, we are confronted with interfering room reflections during reproduction. These reflections can be interpreted as additional sources. The new approach presents a possibility to compensate these reflections by using the three dimensional room impulse response in a prefiltering process of the Ambisonic signals. The room impulse response of each loudspeaker is recorded with a microphone array and transformed to an Ambisonic representation.

Consider a 1st order, 2 dimensional Ambisonic system. The room impulse response caused by the jth loudspeaker at angle ϕ_j is described by equation 8.

$$h_j(t) = \delta_{\phi_j}(t - t_0) + \sum_{k=1}^{+\infty} a_k \cdot \delta_{\theta_{j,k}}(t - t_{j,k}) \tag{8}$$

Each reflection can be expressed as a delayed ($t_{j,k}$) source with the gain a_k originating from angle $\theta_{j,k}$. The associated Ambisonic representation of the reflective part of the 3D room impulse response is expressed by equation 9 (discrete time index n).

$$W_{h_{j,r}}(n) = -\sum_{k=1}^{+\infty} a_k \cdot \delta(n - m_{j,k}) \tag{9a}$$

$$X_{h_{j,r}}(n) = -\sum_{k=1}^{+\infty} a_k \cdot \delta(n - m_{j,k}) \cdot \cos(\theta_{j,k}) \tag{9b}$$

$$Y_{h_{j,r}}(n) = -\sum_{k=1}^{+\infty} a_k \cdot \delta(n - m_{j,k}) \cdot \sin(\theta_{j,k}) \tag{9c}$$

In order to compensate the reflections, they are interpreted as negative sources. In the free field condition, the loudspeaker feeds L_i for i=1 to N are obtained by equation 10.

$$\begin{pmatrix} L_1 \\ \vdots \\ L_N \end{pmatrix} = D \cdot \begin{pmatrix} W_o \\ X_o \\ Y_o \end{pmatrix} \tag{10}$$

W_o, X_o and Y_o are the original Ambisonic signals. The matrix D with elements[1] $d_{i,c}$ describes the weighting of each signal with regard to the geometrical arrangement of the loudspeaker array. The jth loudspeaker feed of the original (uncompensated) system is given in equation 11.

$$L_j(n) = d_{j,1} \cdot W_o(n) + d_{j,2} \cdot X_o(n) + d_{j,3} \cdot Y_o(n) \tag{11}$$

To compensate the interfering reflections caused by the jth loudspeaker its feed L_j (Eq. 10) is convolved with the reflective part of its room impulse response,

$$W_{L_{j,r,comp}}(n) = L_j(n) \otimes W_{h_{j,r}}(n) \tag{12a}$$

$$X_{L_{j,r,comp}}(n) = L_j(n) \otimes X_{h_{j,r}}(n) \tag{12b}$$

$$Y_{L_{j,r,comp}}(n) = L_j(n) \otimes Y_{h_{j,r}}(n) \tag{12c}$$

and we obtain the loudspeaker feeds $L_{i,j}$ that compensate the reflection caused by the jth speaker.

$$\begin{pmatrix} L_{1,j} \\ \vdots \\ L_{N,j} \end{pmatrix} = D \cdot \begin{pmatrix} W_o \\ X_o \\ Y_o \end{pmatrix} + D \cdot \begin{pmatrix} W_{L_{j,r,comp}} \\ X_{L_{j,r,comp}} \\ Y_{L_{j,r,comp}} \end{pmatrix} \tag{13}$$

[1] The index i of the decoder matrix D element $d_{i,c}$ describes the used loudspeaker L_i and the index c depends on the used channel (Ambisonic signal).

Eq. 13 can be rewritten as:

$$\begin{pmatrix} L_{1,\mathrm{j}} \\ \vdots \\ L_{\mathrm{N},\mathrm{j}} \end{pmatrix} = D\cdot \begin{pmatrix} \delta(n)+\mathrm{d}_{\mathrm{j},1}W_{h_{j,r}}(n) & \mathrm{d}_{\mathrm{j},2}W_{h_{j,r}}(n) & \mathrm{d}_{\mathrm{j},3}W_{h_{j,r}}(n) \\ \mathrm{d}_{\mathrm{j},1}X_{h_{j,r}}(n) & \delta(n)+\mathrm{d}_{\mathrm{j},2}X_{h_{j,r}}(n) & \mathrm{d}_{\mathrm{j},3}X_{h_{j,r}}(n) \\ \mathrm{d}_{\mathrm{j},1}Y_{h_{j,r}}(n) & \mathrm{d}_{\mathrm{j},2}Y_{h_{j,r}}(n) & \delta(n)+\mathrm{d}_{\mathrm{j},3}Y_{h_{j,r}}(n) \end{pmatrix} \otimes \begin{pmatrix} W_o \\ X_o \\ Y_o \end{pmatrix}$$

If all impulse responses caused by the N loudspeakers are taken into account, we obtain the compensated loudspeaker feeds given in equation 14.

$$\begin{pmatrix} \mathrm{L}_1^* \\ \vdots \\ \mathrm{L}_\mathrm{N}^* \end{pmatrix} = D\cdot \begin{pmatrix} \delta(n)+\sum_{i=1}^{N}\mathrm{d}_{\mathrm{i},1}W_{h_{i,r}} & \sum_{i=1}^{N}\mathrm{d}_{\mathrm{i},2}W_{h_{i,r}}(n) & \sum_{i=1}^{N}\mathrm{d}_{\mathrm{i},3}W_{h_{i,r}}(n) \\ \sum_{i=1}^{N}\mathrm{d}_{\mathrm{i},1}X_{h_{i,r}}(n) & \delta(n)+\sum_{i=1}^{N}\mathrm{d}_{\mathrm{i},2}X_{h_{i,r}}(n) & \sum_{i=1}^{N}\mathrm{d}_{\mathrm{i},3}X_{h_{i,r}}(n) \\ \sum_{i=1}^{N}\mathrm{d}_{\mathrm{i},1}Y_{h_{i,r}}(n) & \sum_{i=1}^{N}\mathrm{d}_{\mathrm{i},2}Y_{h_{i,r}}(n) & \delta(n)+\sum_{i=1}^{N}\mathrm{d}_{\mathrm{i},3}Y_{h_{i,r}}(n) \end{pmatrix} \otimes \begin{pmatrix} W_o \\ X_o \\ Y_o \end{pmatrix}$$

(14)

The proposed procedure requires C^2 FIR filters depending on the number of channels C. ($C=2m+1$ for two dimensional and $C=(m+1)^2$ for three dimensional Ambisonic systems of order m). However, the amount of filters is independent of the number of loudspeakers.

4. THE BEAM FORMING APPROACH

In the chapter 2, the approach of Emerit and Nicol [2] has been proposed to obtain higher order Ambisonic signals. The transformation is mathematical very elegant but a drawback arises concerning the required number and the arrangement of the microphones. Therefore, a new approach using beam forming to obtain the Ambisonic signals is considered. At the moment, investigations are directed to find the proper microphone array. In figure 6, the signal flow chart is depicted for one target function.

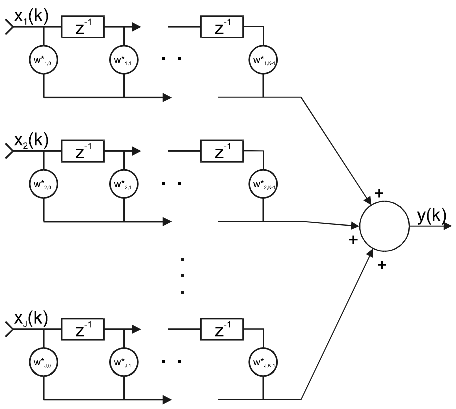

Figure 6. Beam forming signal flow chart for broad band signals.

The target functions are defined by the different microphone characteristics.

5. CONCLUSION

The basic approaches of Ambisonic and Holophony have been introduced. Both systems can be merged to combine the advantages of each approach. A new extension of the Ambisonic system has been proposed which offers the possibility to compensate the interfering room reflections of the playback room. Further investigations should yield arbitrarily higher order Ambisonic signals using the beam forming approach.

The proposed extension concerning the interfering room reflections can be used to generate artificial 3D reverb and seems to be very promising for virtual reality applications.

REFERENCES

[1] Malham, D., ”Homogeneus and Nonhomogeneous Surround Sound Systems”, presented at AES Conference ‘Second Century of Audio’, UK, London, 7-8th June, 1999.

[2] Nicol, R., and Emerit, M., “3D-sound reproduction over an extensive listening area. A hybrid method derived from Holophony and Ambisonic.” Proc. 16th International AES Conference on Spatial Sound Reproduction, Helsinki, pp. 436-453, 1998.

[3] E. Verheijen, “Sound Reproduction by Wave Field Synthesis.”, Thesis, TU Delft, 1998.

[4] E. W. Start, ”Direct sound enhancement by wave field synthesis.”, Thesis, TU Delft, 1997.

[5] Gerzon, M., “With-height Sound Reproduction”, J. Audio Eng. Soc., Vol. 21, No.1, Jan/Feb 1972, pp 2-10.

[6] Farina, A., Email to the sursound mailing list, 29/01/2000

[7] P. M. Morse and K. U. Ingard. Theoretical Acoustics. Mc Graw-Hill, New York City, 1968.

[8] Bamford, J., ”An Analysis of Ambisonic Sound Systems of First and Second Order ”, Thesis presented to the University of Waterloo, Waterloo, Ontario, Canada, 1995.

[9] Sontacchi, A., and Hoeldrich, R.,”Ambisonic System mit unsymmetrischem Lautsprecher Layout”, IEM-Report 09/99, 1999.