Investigations on control and perception of virtual distance using loudspeakers.

Alois Sontacchi, Robert Höldrich

Institute of Electronic Music and Acoustics
Inffeldgasse 10/3, 8010 Graz, AUSTRIA
sontacchi@iem.at

Abstract - The proposed system enables the control of sound sources in a large auditorium both in direction and distance. Reproduction of 3D sound fields over loudspeakers whereas distance coding is taken into account is a rather difficult goal. Because different simple and complex properties of a sound event contribute to the perception of distance. Similar to the perception of pitch, the perception of distance is bounded in both directions. Small deviations concerning far and very short distances can hardly be distinguished. Furthermore, caused be the associated memory, the impression of distance is always biased and relative. Therefore, similair to the visual sence the auditory sence can be easily confused. To overcome these problemes the reproduced sound field should be identical to the natural one as much as possible. As will be shown below the presented approach is based on the reconstruction of the wavefront curvature in a defined listening area in order to perceive the desired distance. The subject of this paper is to find out and validate objective description which correlates with subjective perception. Based on this correlation, optimization criterias can be evaluated to maximize the reconstruction performance in order to control the desired spatial auditory stimuli of perception.

I. INTRODUCTION

A realistic auditory environment can increase the overall subjective sense of precence in virtual environment applications. The proposed approach to realize efficient distance coding in virtual sound scapes is based on the wave field synthesis approach (WFS) and on the ambisonic approach using higher orders (HOA). The WFS approach is based on the Huygens` Principle which is mathematically described by the Kirchhoff-Helmholtz-Integral. This integral states that the wave field of a source free volume V can be described by the knowledge of the pressure along the enclosure surface and the gradient of the pressure normal to the surface. Therefore an arbitrary sound field inside the volume caused by a primary source anywhere outside can be realized by secondary monopole and/or dipole sources distributed over the bordering surface. This principal even holds if the space of sources and the inspected sound field are exchanged. Expanding the bordering surface between the source space and the listening space (reproduced sound field space) to an infinite large plane will introduce a more feasible mathematical description. This leads to a technique called „holographic audio" [1] or also known as „holophone systems".

In our approach the derivation of the virtual loudspeaker feeds is closely related to the WFS approach. The feeds of the distributed loudspeakers are filtered in order to obtain the specified sound field in a defined area. In [2] and [3] the calculation of this filters for linear and planar loudspeaker arrays are presented. Caused by the finite and discrete arrangement artifacts are introduced which are addressed there, too. In our case a special arrangement of virtual loudspeakers along a segment of a circle is used. The calculation procedure for these loudspeaker weights (filters) is given in section 2.

In general the Ambisonics approach is presented with the well-known B-format with the signals W, X, Y and Z. This system approach is based on the sound field's spatial decomposition in spherical harmonics of 0th and 1st order. In common this approach can be extended to higher order systems (HOA) [4,5,6], resulting in better localisation properties and a wider listening area. However increasing the system order will also increase the required transmission channels and also the amount of necessary loudspeakers. In the three dimensional case the sound field produced by a plane wave can be expanded into a series of spherical harmonics. If the wave field is inspected in the origin the mathematical description reduce to a series of an infinite sum of cosine and sine weighted Legendre-functions. This leads to the common coding and decoding equations which can be found elsewhere in [5] and [6]. Using higher order signals will reduce the reconstruction errors and also energy spread over the loudspeakers. There are compromises necessary to overcome the problem of finite system orders. We suggest the approach of applying spatial filters, which we call window applied decoding. Unwanted artefacts are reduced by weighting higher order signals less than lower orders. Therefore the localisation blur is slightly increased but the perceived direction is much more stable [7], [8]. In the case of recording real sound fields even new complex microphone characteristics are required. A possible solution approach is given in [9].

To realize virtual sound scapes the benefits both using the WFS and the HOA approach are combined. The main goal is to reconstruct the wavefront curvature in a defined optimal listening area. The coding scheme is proposed in section two. Using the properties of the complex velocity vector, introduced in section three, the reconstruced sound fields can

be objectively observed. An derivation of objective quantities is given in section four. Finally the paper is concluded.

II. DISTANCE CODING SCHEME

A. Proposed System Model for 3D-Rendering.

Using an arbitrary finite amount of discrete distributed loudspeakers (virtual loudspeaker setup) a specified sound field within a defined area can be synthesised. In order to code the virtual source distances the driving functions for a virtual loudspeaker setup using a derivative of the WFS approach are primarily calculated (see fig. 1, part 1). In the second step the apparent solid angle of the sources are coded using the HOA approach (see fig. 1, part 2).
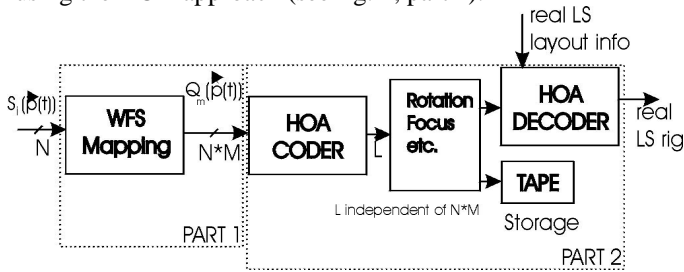

Figure 1. System model.

B. Calculation of the loudspeaker feeds.

The virtual loudspeakers are positioned along a sector around the ideal listening position at a defined fixed distance. The number of the adjacent virtual loudspeakers and their apex angle are important design parameters. Their proper choice will be discussed in section 3. Within the defined area the field caused by the virtual source and the sound field radiated by the loudspeakers are compared. In order to minimize the overall error concerning the wavefront curvatures the loudspeaker feeds must be adjusted properly. The solution is found by solving an over determined equation system using the least mean squares approach. According to the distance of the reproduced virtual source different filter sets will be obtained. With regard to the computation load these filters can be approximated with fixed transfer functions and adjustable frequency independent gains and delays.
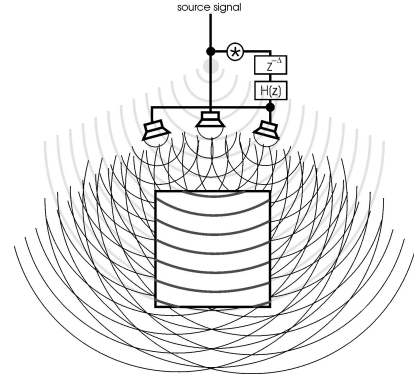

Figure 2. Scheme of virtual loudspeaker placement and restricted reproduction area.

C. Channel coding and audio rendering

Each virtual loudspeaker feed is encoded to the ambisonic domain according to its real position. Subsequently these signals are summed up and decoded to the existing real loudspeaker rig. The choice of the decoder is important to the system performance and will be investigated in section 3. The order of the ambisonic coding and decoding system has to be ajusted according to the spacing of the choosen virtual loudspeakers to prevent spatial aliasing.

III. COMPLEX VELOCITY VECTOR

Objective sound field indicators describing sound fields should have physically useful interpretations. Schiffrer and Stanzial [10] have considered the average velocity of the energy transport as an useful objective indicator of sound fields. The average velocity is obtained by dividing the time averaged intensity $\mathbf{I}$ by the time averaged energy density w.

$$\mathbf{u} = \frac{\langle \mathbf{I} \rangle}{\langle w \rangle} \tag{1}$$

The energy density of a real pressure field is given in Eq.2.

$$w = \frac{\rho}{2} \cdot \left[\|\mathbf{v}\|^2 + \frac{|p|^2}{(\rho c)^2} \right] \tag{2}$$

and the intensity is obtained by

$$\mathbf{I} = p \cdot \mathbf{v} \tag{3}$$

whereby p indicates the sound pressure and $\mathbf{v}$ the sound velocity.

Poletti has introduced in [9] the complex form of the instantaneous velocity for complex monochromatic fields.

$$\mathbf{u}_c = \frac{\mathbf{I}_c}{w_c} \tag{4}$$

where the energy density of the complex monochromatic sound field will be defined as

$$w_C = \frac{\rho}{4}\left[\|\mathbf{v}\|^2 + \frac{|p|^2}{(\rho c)^2}\right] \tag{5}$$

and the complex instantaneous intensity due to Heyser [11] is defined as

$$\mathbf{I} = \frac{1}{2}p_a \cdot \mathbf{v}_a^* \tag{6}$$

where the subindex a indicates the analytic function and the asterix the conjugate of the function.

The real part of the complex velocity vector in Eq.4 is equal to the average velocity in Eq.1 in the case of monochromatic fields. This quantity can be used to define different objective criteria concerning the wave field curvature and therefore the reproduced apperent source distance.

A. The properties of the complex velocity vector

The real part of the complex velocity vector points in the direction of the energy flow and is time invariant. It has the same direction as the active intensity, and is therefore called active velocity. It is proportional to the gradient of the phase of the sound field, and it is normal to the wavefronts. The calculated vectors can be used to evaluate the curvatures at all positions in the sound field.

IV. OBJECTIVE QUANTITIES

A. Parameters and Arrangements

In order to minimize the computation load, the number of virtual loudspeakers should be choosen as small as possible. The optimal apex angle depends on the investigated frequency. For broadband signals different solutions should be combined. The minimal apex angle defines the required ambisonic order. The choice of the ambisonic decoder can prevent additional artifacts.

B. Objective Indicators and Simulation Results

Below the simulation results will be depicted and observed for a real loudspeaker array (consisting of 37 loudspeakers) positioned along a circle with a radius of 5m reproducing a virtual source at distance of 2m. For the distance coding (part 1 in fig.1) 5 virtual loudspeakers each spaced by an apex angle of 10 degrees are used. The azimuth positioning is realized by the 2 dimensional ambisonic coder and basic decoder of 18th order (part 2 in fig.1). The azimuth dircetion of the source is set to 34 degrees, where no real loudspeaker is postioned (worst reproduction case).
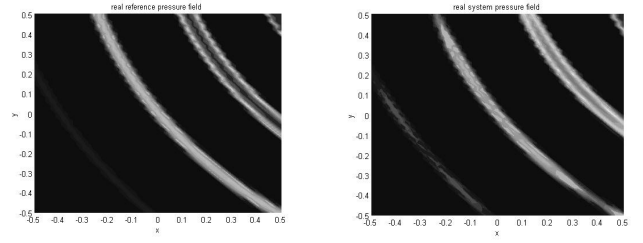

Figure 3. (above) Reference real sound pressure field of a monochromatic source with 1kHz at 2m.
(below) Synthesised real sound pressure field with 5 virtual sources (each apex angle equal 10°) using ambisonic system 18th order.
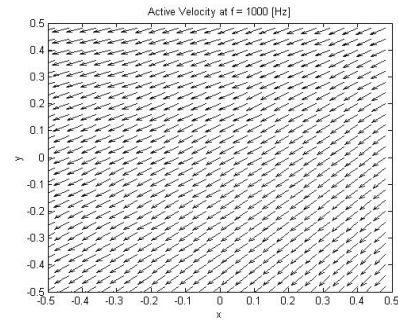

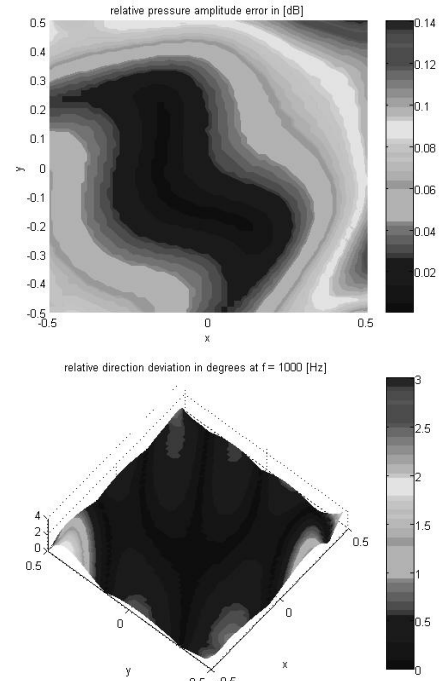Figure 4. Active velocity of the synthesised monochromatic sound field depicted in Fig.3.

Figure 5. (above) Relative amplitude error of the synthesised sound field in [dB].
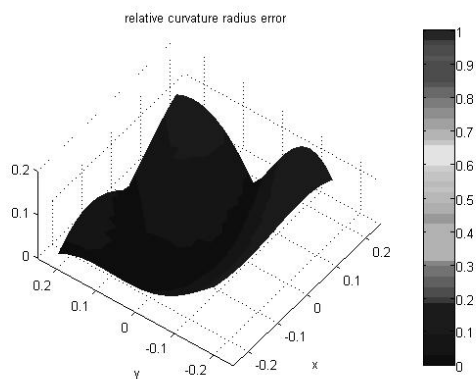(below) Relative direction deviation in degrees.

Figure 6. Relative deviation of the curvature radius
in [m] of the synthesised sound field
around the ideal listening postion.

SUMMARY

Herein a 3D rendering system is investigated. The proposed system enables the control of sound sources in a large auditorium both in direction and distance. The aim is to realize a 3D audio system with a minium effort and maximum efficiency. The quality of the system performance can be resachered in objective and subjective manner. The quality of reproduced sound is most reliably evaluated by formal listening tests. The listening tests are, however, expensive and take a long time to conduct. To avoid them, objective measures have to be developed in order to assist in product development or to characterize known degradations from the targed sound. Using the complex velocity vector any arbitrary local reproduction area can be examined concerning the reproduced direction of a sound source and the curvature of the superimpossed wavefront. Auditory models (e.g. [12]) can be applied to investigate the reproduction of binaural cues. Based on informal listening tests the objective criterias should be validated by subjective decisions. If these objective criterias correlate well with the subjective decisions, they can be used to optimize the distance coding procedure.

REFERENCES

[1] Berkhout, A.J., "A Holographic Approach to Acoustic Control", J. Audio Eng. Soc., 36(12), pp. 977-995, 1988.

[2] Verheijen, E., "Sound Reproduction by Wave Field Synthesis", Thesis, TU Delft, 1998.

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

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

[5] Daniel, Jérôme, "Acoustic field representation, application to the transmission and the reproduction of complex sound environments in the multimedia context" (Engl. translation), PhD Thesis, 2000.

[6] Malham, D.G. "Higher order Ambisonic systems for the spatialisation of sound" Proceedings, ICMC99, Beijing, October 1999.

[7] Sontacchi, A. Noisternig, M. Majdak, P., Höldrich, R., "An Objective Model of Localisation in Binaural Sound Reproduction Systems", 21st AES Conference, St. Petersburg, 2002.

[8] Sontacchi, A. Noisternig, M. Majdak, P., Höldrich, R., "Subjective Validation of an Objective Localisation Model", 21st AES Conference, St. Petersburg, 2002.

[9] Poletti, M.A.: "A Unified Theory of Horizontal Holographic Sound Systems ", J. Audio Eng. Soc., Vol. 48, No. 12, pp. 1155-1182, December, 2000.

[10] Schiffrer, G.; Stanzial, D.: "Energetic Properties of Acoustic Fields ", J. Acoust. Soc. Am. , Vol. 96, pp. 3645-3653, 1994.

[11] Heyser, R.C.: "Instantaneous Intensity," presented at the 81st Convention of the Audio Engineering Society, Preprint Nr. 2399, J. Audio Eng. Soc. (Abstracts), Vol. 34, p.1038, December, 1986.

[12] Pulkki, V.: "Analyzing Virtual Sound Source Attributes Using a Binaural Auditory Model", *J. Audio Eng. Soc., Vol.47, No.4, April, 1999.*