"GETTING MIXED UP WITH WFS, VBAP, HOA, TRM …"
FROM ACRONYMIC CACOPHONY TO A GENERALIZED RENDERING TOOLBOX

Alois Sontacchi and Robert Höldrich

Institute of Electronic Music and Acoustics, University of Music and dramatic Arts
Inffeldgasse 10/3, A-8010, Graz, Austria
phone: + (43) 316 389 3548, fax: + (43) 316 389 3171, email: sontacchi@iem.at
web: iem.at

ABSTRACT

This paper focuses on various application scenarios based on the wave field synthesis (WFS) approach which have been implemented and/or investigated in our laboratories lately. Within the few different selected scenarios, we try to show the possibility to combine different state-of-the-art audio rendering approaches to obtain an efficient solution concerning computation load, hardware request, and audio reproduction quality. The major aspect is related to the fact that each of the existing rendering strategies suffer on different drawbacks or limitations. The presented (hopefully) "best practice" models should depict how we try to overcome those problems.

The paper is organized in five sections. The first one will briefly introduce to the WFS rendering technique. Benefits and drawbacks will be highlighted. Within section 2 and 3, we present examples how to combine WFS with Vector Based Amplitude Panning (VBAP) and Higher-Order-Ambisonic (HOA), respectively. The relation of WFS with the time reversal mirror (TRM) will be figured out in section 4. Finally, section 5 draws a short conclusion .

1. INTRODUCTION TO WFS

Wave Field Synthesis was initially invented in the late 80's by Berkhout and has been further developed at the TU Delft [1, 2]. The basic idea is related to the Huygens' Principle which states, that an arbitrary wave front may be considered as a secondary source distribution. Regarding the propagating wave from the given wave front we cannot differentiate if it was either emitted by the original sound source (the primary source) or by a secondary source distribution along this wave front. As a consequence, the secondary source distribution may be substituted for the primary source, in order to reproduce the primary sound field. Based on this physical background we can state that WFS aims at reproducing sound waves by (distributed) loudspeaker arrays. In Figure 1 the well know illustration of this concept is depicted.

Mathematically the described concept is completely modelled and formulated by the Kirchhoff-Helmholtz-Integral (KHI, see Eq. 1). This integral states that the wave field inside a source free volume V can be described by the knowledge of the pressure $p(\vec{r}_S)$ along the enclosure surface S and the gradient of the pressure normal to the surface $\nabla_S\, p(\vec{r}_S)$. This principal even holds if the space of sources and the inspected sound field are exchanged.

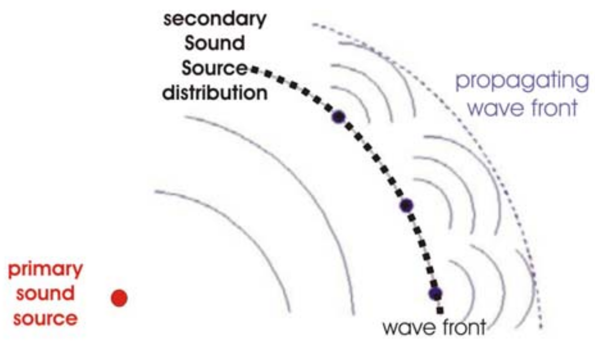

Figure 1 – Illustrated WFS concept based on Huygens' Principal

Therefore, an arbitrary sound field inside the volume caused by a primary source anywhere outside can be realized by secondary monopole (second integrant in Eq. 1) and/or dipole sources (first integrant in Eq. 1) distributed over the bordering surface.

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

Whereby $G(\vec{r}_R|\vec{r}_S) = \dfrac{e^{-jk|\vec{r}_R-\vec{r}_S|}}{|\vec{r}_R-\vec{r}_S|}$ is known as the Green's function, the index S is related to the surface and R denotes the point of interest.

Expanding the bordering surface between the source space and the listening space (reproduced sound field space) to an infinite large plane and regarding solely monopole or dipole distributions will introduce a more feasible mathematical description. Furthermore, if the separating (secondary source distribution) plane between the primary source space and the listening space is reduced to a separating line we will just arrive at the place where the applied WFS technique starts. A detailed derivation of the WFS approach deduced form the KHI can be found in [3].

WFS is a powerful and appealing acoustic rendering technique, but there are advantages and disadvantages. Summing up the pros and cons in a brief characterization we get the following list of attributes:

Pros:

- WFS aims at physically reconstructing the sound field in a global sense.
- The direction of rendered point sound sources is independent of the listeners' position.
- Sound source distance can be handled and controlled.
- Various source types between a point source and a plane wave, and even focused sources can be realized

Cons:

- The proposed technique is restricted to applications that render sound sources within a plane (2½D).
- Amount of required loudspeakers grows with the requested reproduction area.
- Focused sound sources (within the listener space) can be perceived just from the front and suffer from inconsistent spatial hearing cues.

2. WFS AND VBAP

In this chapter the usage of the WFS technique applied to distributed tiny line-arrays around a computer screen is highlighted. The aim of the proposed auditory interface is to acoustically render the "visual" information content at its position on the screen (see Fig. 2). In [4, 5], a detailed description concerning the immersive audio environment for desktop applications and evaluation can be found.

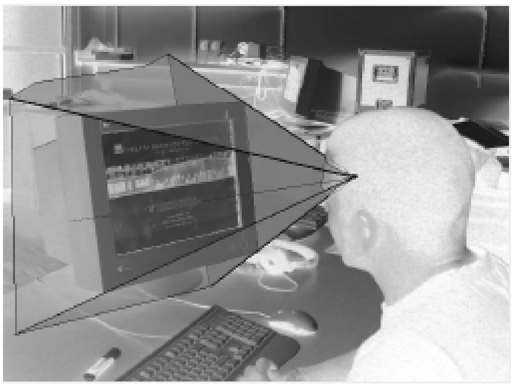

Figure 2 – Region of visual perception.

Several loudspeakers positioned along the screen edges should generate the target sound field. All loudspeakers along one edge are considered as a line array (see Fig. 3). Caused by the fact, that the WFS technique is restricted to the sound field reproduction within a plane, an additional panning law is required in order to position virtual sources arbitrarily. The vector based amplitude panning technique (VBAP), which was introduced by Pulkki [6, 7],came out to be an appropriate candidate. VBAP can render a virtual sound source within an area defined by a loudspeaker triple. The sound source signal is feed to each speaker with different gains which depend on the virtual source position in relation to the three speaker locations. The line array produce WFS-based virtual point sources that are combined using the VBAP approach to simulate virtual sources on arbitrary locations within the screen area..

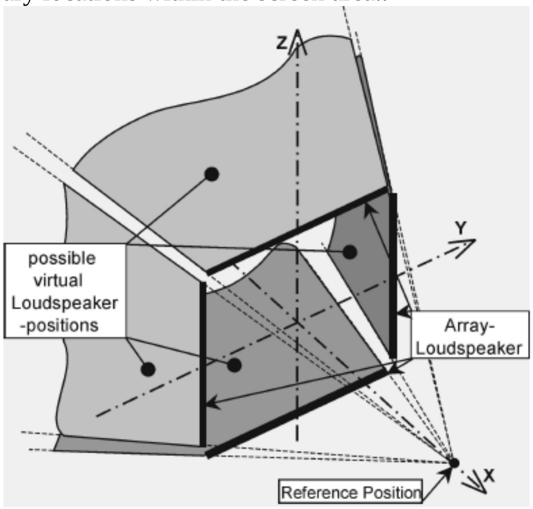

Figure 3 – Region of visual perception.

Since the loudspeakers are placed in front of the listener only sources in front of the listener can be reproduced faithfully. However the position of sources can vary in azimuth, elevation and distance.
Sources behind and in front (see Fig.3) of the desktop can be realized. The size of the reproduction area is restricted to small extend around the listeners head. Consequently the listeners head can move free inside this region without being tracked.

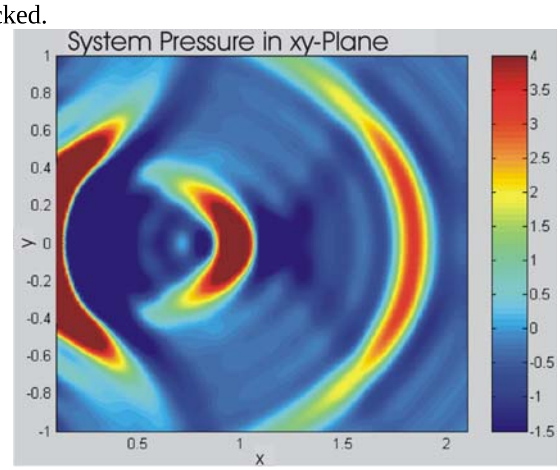

Figure 4 – Pressure distribution of a focused sound source.

In Fig. 4 the resulting pressure distribution of a focused sound source at 0.7m is depicted for the xy-plane. If the listener is positioned next to the x-Axis, at a distance greater than 0.7m in the x-direction, than the reproduction error of the sound field can be almost neglected. Further results can be found in [5].

3. WFS & AMBISONIC

A realistic auditory environment can increase the overall subjective sense of presence in virtual environment applications. Within this section we propose an approach (see [8,9]) to realize efficient distance coding in virtual 3D "sound scapes" based on the wave field synthesis approach (WFS) and on the ambisonic approach using higher orders (HOA). Both WFS and HOA aim at physically reconstructing the sound field. Though they derive from distinct theoretical fundamentals, they have already been shown as equivalent under given assumptions [10].

However, as already mentioned above, WFS is restricted to the reconstruction within a plane. Fortunately, HOA is able to do audio render even in the third dimension [11] and convinces with a compact notation and properties concerning the handling of a decomposed sound field (see [12]). Nevertheless, HOA is not able to render sound sources at arbitrary distant locations. Therefore an appropriate combination of both techniques can overcome both insufficiencies. In [8] we proposed a two stage model depicted in Fig. 5.

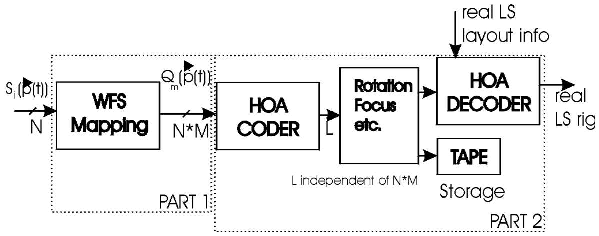

Figure 5 – Two stage model to provide 3D distance coding.

In the first stage the distance of each sound source is coded based on the WFS approach. Related to the geometrical layout of the loudspeaker arrangement and the desired source position (within, at, or outside the bordering loudspeaker distribution plane) so called driving functions (a derivation can be found in [3]) are used to decode the source distance. Caused by the fact, that in this stage we are just interested at the distance of a sound source, we can neglect the source direction and therefore we reduce the 3D problem to a 2D scenario which is sketched in figure 6. The resulting driving functions of several loudspeakers along the defined circular arc (the ends are defined by the enclosed angle between $\vec{r}$ and $\vec{n}$ e.g. outlying sources: $180° - \varphi_{inc} \geq 90°$) can be interpreted as a defined source distribution along the bordering spherical segment by simple rotation around the conduit of the sound source and the centre of the loudspeaker arrangement.

In the second stage the obtained source distribution is encoded in the HOA domain and transformed (rotated) related to the desired source direction. Afterwards each HOA representation of several sound sources at various positions can be superposed. The highest resolution of the encoded source distribution is directly related to the greatest order M of the used HOA system (which depends on the actual loudspeaker layout, i.e. in 2D case: $\varphi_{\min} = 360/(2M+1)$).

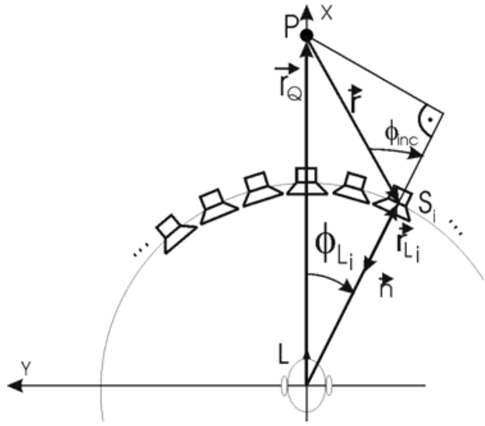

Figure 6 – General 2D distance coding scenario.

The calculation of the driving functions in the first stage lead to complex modification (gain & delay) of the source signal. In figure 7 the resulting pressure field 1x1m (left) and the error (right, relative deviation of the synthesised sound field from the reference sound field) of a virtual source (band-limited impulse <2kHz) at a distance of 2m, 74° direction synthesized with 5 real sources at 5m, apex angle of 10° is depicted. To prevent further degradations in the synthesised sound field an HOA system of order 18 is required.

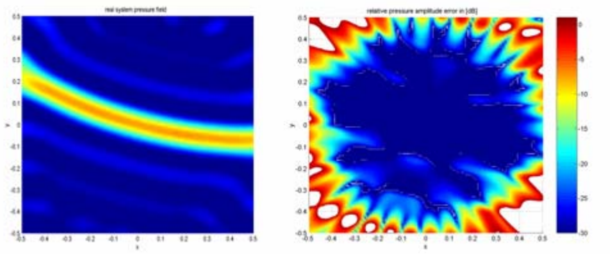

Figure7 – Results of the proposed two stage model.

In [13] the distance perception of virtual sound sources produced by planar loudspeaker arrays has been investigated. The show, that in anechoic rooms the distance perception only depends on the reproduction level, i.e. the louder the nearer. Under normal listening conditions in normal rooms, there exists a relation of the listener position to the array and the amount of active loudspeakers. In other words distance perception is than largely controlled by the ratio between direct to diffuse sound energy. Furthermore the early reflections patterns are of great importance, which can be additionally applied to improve the distance impression (see [13, 14]).

An alternative approach concerning distance coding directly in the HOA domain can be found in [15], called near-field compensated higher order ambisonics (NFC-HOA).

4. TRM OR HO-WFS

Applying the basic concept of Wave Field Synthesis to reflective sound in addition to the direct sound will lead to the approach of the Time Reversal Mirror (TRM). The basic idea was first investigated and applied in medicine and inhomogeneous medium [16]. Recently, the TRM has also been applied to acoustic applications [17, 18]. Within this chapter the extension of the WFS concept – the TRM - will be presented. The TRM achieves an improved stability concerning focussed / real projected sound sources within the listening area. Room adaptations concerning absorption and/or special loudspeaker arrangements are not required.

In figure 8 the principal of the TRM technique is shown. Starting at the left side of Fig. 8 proceeding to the right, an impulse is emitted from loudspeaker and after transmission through the room (for the moment just considering the direct path) received at various microphones (here arranged along a line). Depending on the microphone position we will obtain different delayed versions of the emitted impulse (impulse responses, respectively). If the received impulse responses are reverses (inverse time shift: the beginning is shifted to the end and the end vice versa) we will get a mirrored set of delayed versions of the emitted impulse. After exchanging the microphones with loudspeakers at the same positions (assumed similar acoustic properties concerning directivity etc.) and feeding these loudspeakers with the mirrored recorded impulse responses will cause a spatial focused impulse at the same position where the initial impulse was emitted.

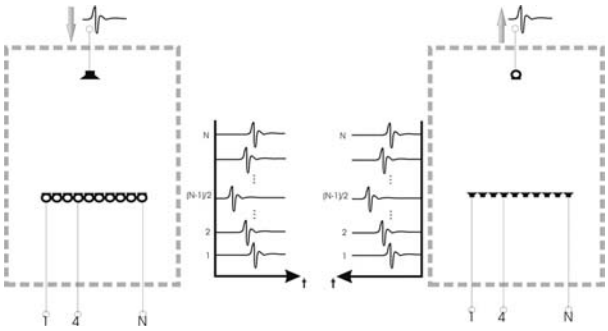

Figure 8 – Pressure distribution of a focused sound source.

Based on the evidence that the solution of the wave equations exhibit both positive and negative time solutions (incoming and outgoing waves) the above mentioned principal can be mathematically tracked. Furthermore within the Greens' function (see sec. 1), which describes the propagation of a point source, the source point and the receiving point can be exchanged without altering the function. Therefore, provided similar acoustic properties concerning directivity, frequency response, dynamic etc., the transmission paths between the loudspeaker and the microphones in Fig.8 on the left are identical to the measured transmission paths in Fig.8 on the right. Hence we can proceed measuring the transmission paths from a loudspeaker array to any arbitrary point in the room consecutively, do the inverse time shift and apply each reversed impulse response as a filter to each loudspeaker feed. Providing a mono signal to the loudspeaker feeds (filtered with the corresponding reversed impulse responses) will cause a focused sound source at the measured point playing the mono signal.

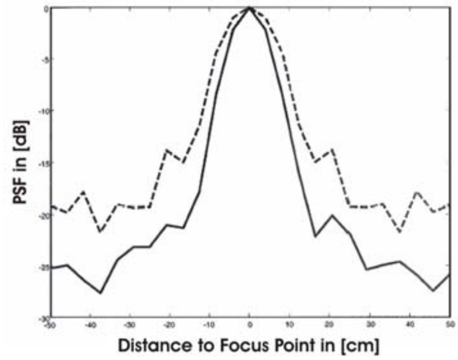

Figure 9 – Point Spread function at focus point (cf. [17]).

The quality of the focal spot width (see Fig. 9) depends on the number of primary sources (loudspeakers), the shape and the acoustic properties of the room, the provided length of the reversed impulse response, the used bandwidth and the stability of temperature (cf. [17]). Fig.9 depicts the point spread functions (PSF) at focal point for free-field simulation (dashed line) and a real test setup (solid line) (cf. [17]). The PSF is defined as the maximum of the temporal impulse response at a defined position.

$$PSF(\mathbf{x}) = \max_t \{h_{\mathbf{x}}(t)\} \qquad (2)$$

We have investigated the properties of various loudspeaker arrangements concerning the focal point spread and the localisation quality of focused sound sources (cf. [19]). The different arrays are figured below.

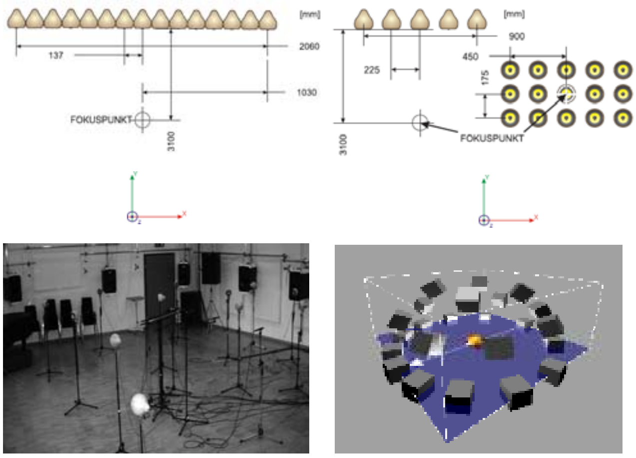

Figure 10 – Different investigated loudspeaker arrays (starting from the left row1: Array 1&2; row 2: Array 3, 4&5).

The array 1 is a 2.06m long line array consisting of 15 similar loudspeakers spaced 137 mm. Array 2 consists of the same 15 loudspeakers arranged in planar array (3x5) and Array 3 is a chaotic arrangement in the xy-plane. Array 4 is the IEM CUBE consisting of 24 loudspeakers evenly distributed over the upper hemisphere and Array 5 is just the lower rig of the IEM CUBE consisting of a ring of 12 loudspeakers. In figure 11, the measured PSF along all three axes are depicted for Array 1-3 and the simulated PSF for both IEM CUBE arrangements. The origin of the coordinate system is placed at the focus point and the three axes are orientated as follows: the x-axis is oriented parallel to the line array axis and to the ground (pointing to the right); y-axis is oriented orthogonal to the line array axis, parallel to ground (pointing forward); and the z-axis is bound upward.

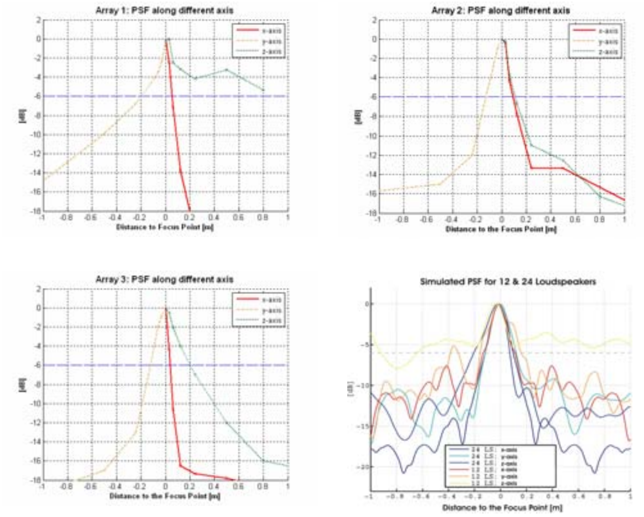

Figure 11 – Results of focused sound source.

It can be seen, that the quality of the focused sound source depends on the number of loudspeakers (see Array 4&5 for the x and y axes) and on the arrangement, too. If loudspeaker are just arranged e.g. in the xy-plane, the sharpness of the focus quality in the z direction is reduced (Array 1 and Array 5). In the case of the chaotic arrangement in the xy-plane the results are even better than expected, particularly for the z-direction.

In general there is a good agreement of the simulated data with the measurements in vicinity of the focus point. Interestingliy, the simulated results show a slightly reduced focusing quality. This is caused by the fact that each reflection contributes to the focusing process in TRM, but for the room simulation only mirror sources up to the 2nd order have been considered.

However, the temporal structures measured near the focal point exhibit several severe pre-echoes (cf. [20]). These phenomena can be reduced by the truncation with sliding windows at the end of the reversed impulse responses, whereby the psychoacoustic findings known as the "precedence-effect" can be used as a design guide.

As already mentioned, the reciprocity is only valid if transmitter and receiver have the same acoustical properties. Neglecting the radiation properties, the frequency response of the transmission path "loudspeaker – room –microphone" has a serious impact on the sound quality of the focused sound, because these paths are run through twice. In [18] an appropriate inverse filter solution to this equalisation problem has already been proposed.

5. CONCLUSSION

WFS is a very powerful and appealing acoustic rendering technique. However in some application scenarios it might be necessary to combine the WFS technique with some other existing rendering approaches to overcome existing drawbacks. We argue for mixed rendering approaches that can lead to convincing solutions.

We have revisited the time reversal mirror which can be regarded as the generalization of the WFS technique. It has been posed that beside the number of loudspeakers the array arrangement has a direct effect on the sharpness of the focus. Improvements of the TRM approach concerning the temporal structure at points near the focal and the coloration of the focused sound signal have been considered.

REFERENCES

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

[2] A. J. Berkhout, D. de Vries, and P. Vogel, Acoustic Control by Wave Field Synthesis, J. Acoust. Soc. Am., vol. 93, pp. 2764-2778, 1993.

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

[4] M. Strauß, "Simulation und Implementation eines Audio Interface in Wellenfeldsynthese", Diploma Thesis at the IEM, University of Music and dramatic Arts Graz, Austria, 2002.

[5] A. Sontacchi, M. Strauß, and R. Höldrich, "A Spatial Audio Interface for Desktop Applications", AES 24th International Conference on Multichannel Audio, St. Petersburg, Russia, 2003.

[6] V. Pulkki, "Spatial Sound Generation and Perception by Amplitude Panning Techniques", PhD Thesis, Helsinki University of Technology, Espoo, Finland, 2001.

[7] V. Pulkki, „Virtual Sound Source Positioning Using Vector Base Amplitude Panning," J. Audio Eng. Soc., Vol. 45, No. 6, pp. 944-951, 1997.

[8] A. Sontacchi and R.Höldrich, "Further investigations on 3D sound fields using distance coding", DAFx'01 Proceedings, Limerick, Ireland, 2001.

[9] A. Sontacchi and R.Höldrich, "Investigations on control and perception of virtual distance using loudspeakers.", VECIMS 2003, Lugano, Switzerland, 27-29 July 2003.

[10] R. Nicol and M. Emerit, "3D-Sound Reproduction over an Extensive Listening Area: A Hybrid Method Derived from Holophony and Ambisonic", presented at the AES 16th Int. Conference on Spatial Sound Reproduction, Rovaniemi, Finland, 1999.

[11] M. A. Gerzon, "Periphony: With-Height Sound Reproduction", J. Audio Eng. Soc., vol. 21(1), pp. 2-10, 1973.

[12] A. Sontacchi, "Dreidimensionale Schallfeldreproduktion für Lautsprecher- und Kopfhöreranwendungen", PhD Thesis, TU Graz, Austria, 2003.

[13] S. Komiyama et al., "A Loudspeaker-Array to Control Sound Image Distance", Acoust. Sci. & Tech. Vol. 24, No.5, 2003.

[14] D. Y. Jang, et al., "Room Impulse Response Shaping for Enhancement of Perceived Spaciousness and Auditory Distance", DAFx'04 Proceedings, Naples, Italy, 2004.

[15] J. Daniel, R. Nicol, and S. Moreai, "Further Investigations of High Order Ambisonics and Wavefield Synthesis for Holophonic Sound Imaging", 114th AES Convention, Amsterdam, Netherlands, 2003.

[16] M. Fink, "Time reversal of ultrasonic fields - part I: Basic principles." IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency, Vol.39, pp.555-566, September, 1992.

[17] S. Yon, M. Tanter, and M. Fink, “Sound focussing in rooms I: The time-reversal approach.” Acoustical Society of America, 113(3):1533-1543, March, 2003.
[18] S. Yon, M. Tanter, and M. Fink, “Sound focussing in rooms II: The spatio-temporal inverse Filter.”, Acoustical Society of America, 114(6):3044 - 3052, December 2003.
[19] I. Csonka, „Schallquellenpositionierung mit Hilfe des Zeitumkehrspiegels“, Diploma Thesis at the IEM, University of Music and dramatic Arts Graz, Austria, 2006.
[20] R. Höldrich and A. Sontacchi, „Wellenfeldsynthese – Erweiterungen und Alternativen“, DAGA´05, Munich, Germany, 2005.