Acoust. Sci. & Tech. **25**, 6 (2004)

INVITED REVIEW

Wave field synthesis: A promising spatial audio rendering concept

Günther Theile* and Helmut Wittek†

Institut für Rundfunktechnik (IRT),
Munich, Germany

Abstract: Modern convolution technologies offer possibilities to overcome principle shortcomings of loudspeaker stereophony by exploiting the Wave Field Synthesis (WFS) concept for rendering virtual spatial characteristics of sound events. Based on the Huygens principle loudspeaker arrays are reproducing a synthetic sound field around the listener, whereby the dry audio signal is combined with measured or modelled information about the room and the source's position to enable the accurate reproduction of the source within its acoustical environment. Not surprisingly, basic and practical constraints of WFS systems limit the rendering accurateness and the perceived spatial audio quality to a certain degree, dependent on characteristic features and technical parameters of the sound field synthesis. However, recent developments have shown already that a number of applications could be possible in the near future. An attractive example is the synthesis of WFS and stereophony offering enhanced freedom in sound design as well as improved quality and more flexibility in practical playback situations for multichannel sound mixes.

Keywords: Wave Field Synthesis, Loudspeaker array, Virtual Panning Spot, CARROUSO

PACS number: 43.38.Hz, 43.38.Wl, 43.60.Sx [DOI: 10.1250/ast.25.393]

1. INTRODUCTION

Three psychoacoustic fundamentally different spatial audio imaging methods should be distinguished:

- (Multichannel) loudspeaker stereophony
- Binaural reconstruction of the ear input signals
- Syntheses of the sound field around the listener

All known spatial sound systems can be traced back to one of these methods or can contain mixed forms thereof, whereby certain advantages of the methods are being exploited, respectively its disadvantages are avoided, dependent on the intended application area.

1.1. Loudspeaker Stereophony

This is in principle based on the characteristics of localization in the superimposed sound field, generated by two loudspeakers [1]. Directional imaging is done in the imaging area between two adjacent loudspeakers [2]. In the case of 3/2 stereophony, with the assistance of surround channels the imaging area between the front loudspeakers can be extended. Therefore possibilities are offered for the reproduction of early lateral sound for imaging of spatial depth as well as reverberation, in order to produce the spatial impression and the envelopment. Details are described in [3].

*e-mail: theile@irt.de
†e-mail: wittek@irt.de

1.2. Binaural Reconstruction of the Ear Input Signals

The original employment of this method is the known dummy head stereophony. It is not intended to reproduce a suitable sound field at the reproduction location. Instead, the effective ear signals in the recording location are recorded with the assistance of a dummy head — and replayed in principle via headphones. Under ideal circumstances, the reproduced binaural signals are identical to the original ear signals that the listener received in the recording location. In practice it is possible to reproduce auditory events with excellent realism regarding spatial characteristics and sound color.

1.3. Synthesis of the Sound Field around the Listener

The third approach, which is the main topic of this paper, has been pursued within the framework of the European Research Project "CARROUSO" [4] It is based on the concept of Wave Field Synthesis (WFS, developed at the Technical University Delft, refer e.g. [5,6].), i.e. the representation of a virtual source and a virtual room is achieved by rendering an acoustically correct sound field. The principle of WFS is based upon the assistance of loudspeaker arrays, when a complete sound field is generated in the listening zone which is identical to an appropriate real sound event (see Chap. 2). This acoustical counterpart to the optical holography is also described as "holophony." The binaural ear input signals that are active for the auditory event thus arise in a natural way within the sound field, contrary to dummy head stereophony. It should be mentioned that also Komiyama and Ono ([7,8]) performed investigations on this reproduction principle.

1.4. Combining Stereophony and WFS

Further developments of spatial audio systems are based on useful combinations of basic methods, using sophisticated real time convolution algorithm. This paper also focuses on overcoming certain practical drawbacks of multichannel sound on the one hand and of WFS on the other. So called "Virtual Panning Spots" (VPS) are introduced to improve the WFS rendering quality of large complex sources (e.g. "choir"), to reduce the number of WFS transmission channels and to ensure compatibility and scalability. Useful combinations of VPS and conventional or sophisticated stereophonic panning and mixing techniques will provide advanced facilities for spatial sound design. A special VPS application allows play back of conventional multichannel mixes in a virtual high quality listening room rendered by means of WFS technologies, offering full backwards compatibility with usual loudspeaker stereophony, optimum multichannel format flexibility, as well as attractive practical benefits in the home, in the cinema, or in other applications.

2. WFS PRINCIPLES AND PROPERTIES

2.1. The "Huygens" Principle

"If from a point S of a homogeneous isotropic medium a spherical wave is emitted, one can imagine the procedure of the individual wave reproduction in that a particle brought into oscillation by external forces, transfers its movements to its neighboring particles. This procedure then continues symmetrically in all directions and in this way gives cause to a spherical wave..." [9].

If a sound source S (the so-called "primary source," Fig. 1a) emits spherical wave fronts, one can imagine in accordance with the "Huygens" Principle each emitted wave front (refer to Fig. 1a), through the addition of all participating "secondary sources" (which also emit spherical waves) on the surface O. Due to the knowledge of the wave front on surface O (Fig. 1a) the state of the oscillation can be determined at an arbitrary point P of the sound field. The wave front through point P is constructed through the summation of all participating secondary source signals.

In principle, in the case of the WFS, one replaces the secondary point sources by loudspeakers and in this way again produces a spherical wave (refer to Fig. 1b). The sound source S is virtual; the listener in point P receives the same wave front which is transmitted by the sound source S.

a)

b)

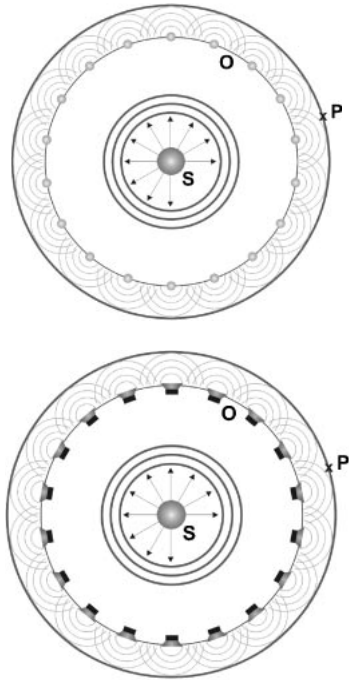

Fig. 1 The "Huygens" principle. a) Theoretical Model, b) Application WFS.

2.2. WFS — The Application of the "Huygens" Principle

This applies correspondingly for a circular arrangement of the loudspeakers on a two dimensional level. Concerning a sound source S, (which emits a sine impulse and is located in an infinitely large plane without demarcation of walls), a wave front results as illustrated in Fig. 2a). If one now places an array of n microphones (M) in this primary sound field and one reproduces the recorded microphone signals via an equally arranged array of n loudspeakers (L) — special equalization has to be included according to the relevant physical basics — in a reproduction room (Fig. 2b)), one obtains the synthesized wave front in the (dotted) listening area. At any place in the listening area the listener perceives a virtual sound source S, as he can move around freely, whilst the virtual sound source remains correctly localized in terms of its direction (see [5] or [10]).

2.3. Special Properties of WFS

Through WFS the sound engineer has a powerful tool to design a sound scene. One of the most important (with respect to conventional techniques) novel properties is its outstanding capability of providing a realistic localization of virtual sources. Typical problems and constraints of a stereophonic image vanish in a WFS sound scene.

In contrast to stereophony WFS is able to:

- produce virtual sources that are localized on the same position throughout the entire listening area, refer Fig. 3: The dashed and dotted arrows indicate the directions of the auditory events when the virtual

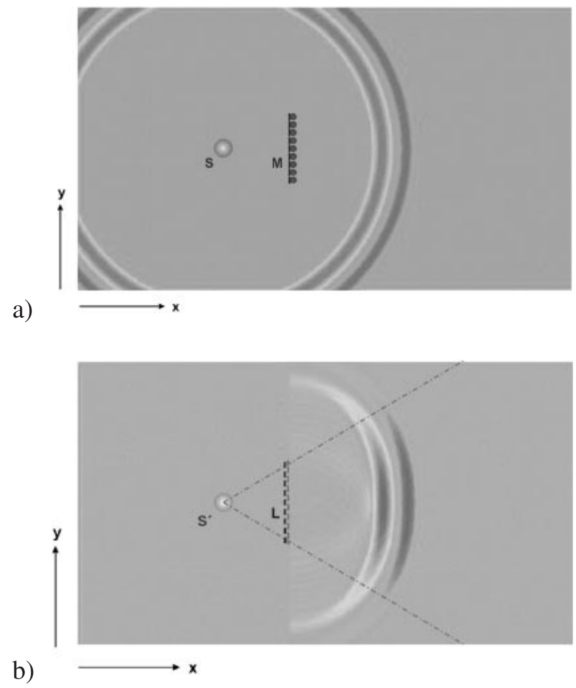

Fig. 2 Principle of WFS. a) ideal source response, b) typical output of a finite WFS array.

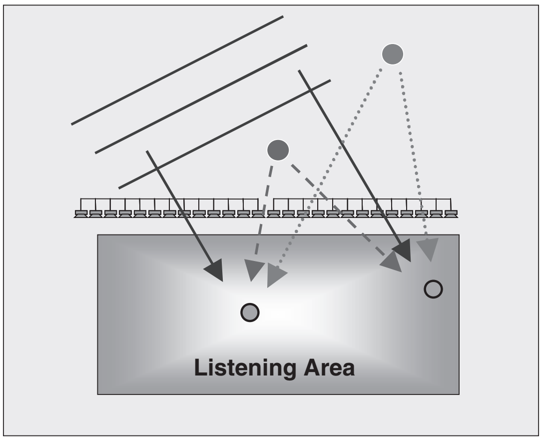

Fig. 3 WFS is capable of reproducing both the stable positions of point sources (dashed and dotted) and the stable direction of a plane wave (solid).

point sources are reproduced.

- ○ produce plane waves that are localized in the same direction throughout the entire listening area, refer Fig. 3: The solid arrows indicate the direction of the auditory event when the plane wave is reproduced.
- ○ enhance the localization of virtual sources and the sense of presence and envelopment through a realistic reproduction of the amplitude distribution of a virtual source. In other words, when the listener is approaching the location of a virtual source the amplitude increases in a realistic way. Accordingly, the amplitude of a plane wave — which can be seen as a source in infinite distance — changes least on different listener positions.

These properties enable the synthesis of complex sound scenes which can be experienced by the listener while moving around within the listening area. Figure 3 illustrates the way in which the sound image changes at different listening positions. This feature can be made use of deliberately by the sound engineer to realize new spatial sound design ideas.

Moreover, it has been shown that the enhanced resolution of the localization compared with stereophony [11] enables the listener to easily distinguish between different virtual sources which makes the sound scene significantly more transparent.

3. WFS PRACTICAL CONSTRAINTS

Not surprisingly, in practice it is not possible to match all theoretical requirements for a perfect result. The rendered WFS sound field differs from the desired sound field to some degree for a number of reasons (for details see [12]):

3.1. Discreteness of the Array (Spatial Aliasing)

This effect produces spatial and spectral errors of the synthesized sound field due to the discretisation of a continuous secondary source distribution. Above the spatial aliasing frequency f_{alias} the time difference between two successive loudspeaker signals interferes at the listener's position, depending on the spatial sampling interval, i.e. the loudspeaker/microphone inter-spacing.

3.2. Reflections of the Reproduction Room (Spatial Interference)

A WFS array can not render the desired sound field perfectly if reflections of the reproduction room produce interference in spatial perception. In particular, perception of distance, depth and spatial impression are affected, because fragile distance cues of synthesised sources can be dominated by the stronger distance cues generated by the array speakers. They interfere with the desired reflection pattern of the synthesised source. Special room compensation algorithms being under investigation [13,14] will perhaps be able to minimize this effect.

3.3. Restriction to the Horizontal Plane

Theory does not restrict WFS to the horizontal plane. However, the reduction of the array dimension to the horizontal plane is the practical approach, having a number of consequences. First, virtual sources can be synthesized only within the horizontal plane. This includes virtual reflections affecting the completeness of a natural reflection pattern and thus possibly resulting in impairments of perception of distance, depth, spatial impression and

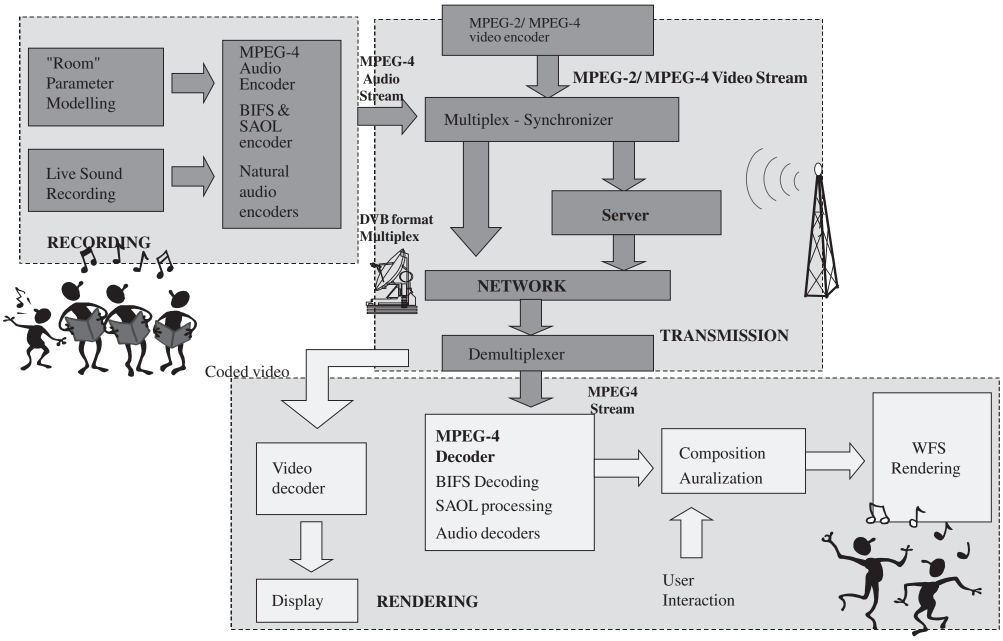

Fig. 4 Priciple block diagram of the CARROUSO demonstrator.

envelopment.

Furthermore, horizontal arrays do not generate real spherical waves, but cylindrical waves. In the case of imaging a plane wave for example there results an error with respect to the level roll-off (3 dB/doubling of distance), in comparison with the ideal plane wave (no roll-off) [12,15].

3.4. Limitation of Array Dimensions (Diffraction)

In practical applications the loudspeaker array will have a finite length. Due to a finite array so-called diffraction waves originate from the edges of the loudspeaker array [12,15]. These contributions appear as after-echoes, and — depending on their level and time-offset at the receiver's location — may give rise to colouration.

However, methods to reduce these truncation effects are known, e.g. by applying a tapering window to the array signals. This means that a decreasing weight is given to the loudspeakers near the edges of the array. In this way the amount of diffraction effects can substantially be reduced at the cost of a limitation of the listening area [15].

3.5. Effects on Perception

Although a number of authors have suggested methods to deal with the practical limits of rendering accurateness or to minimize their effects, there is still a lack of knowledge (some details can be found e.g. in [5,6,12,13,16]). Several effects of the constraints on specific perceptual attributes are not known yet in detail. However, this knowledge is important for further developments of WFS systems in view of future applications.

Current psychoacoustic studies are concentrating on the subjective evaluation of principle characteristics of WFS systems in comparison with stereophonic or binaural systems. They are necessary to evaluate the resulting impacts on attributes of spatial perception not only with respect to the development of WFS systems for different applications but also in view of scientific knowledge. Particular attention should be turned to the perception of direction, distance, spatial depth, spatial perspective, spatial impression, reverberance, and envelopment, as well as sound colour.

4. WFS APPLICATIONS

4.1. The European CARROUSO Project[1]

The European CARROUSO Project ("Creating, Assessing and Rendering in Real Time of High Quality Audio-Visual Environments in MPEG-4 Context") has intended to break several limitations of these current commercial systems by merging the new WFS rendering technique with the flexible new coding technology MPEG-4 standard, allowing object-oriented and interactive sound manipulation.

The key objective of the project CARROUSO was to provide a new technology that enables to transfer a sound field, generated at a certain real or virtual space, to another usually remote located space, in a bit efficient way at highest perceived quality. The principle block diagram is illustrated in Fig. 4, it shows three functional components:

Capturing

For recording of the sound field a microphone array technology is applied. Signal processing calculates the

[1]EU-Project IST–1999-20993 (Jan. 2001–June 2003): [4].

position of the sound sources which could be fixed or moving. The microphone array as well as the video cameras is used to gather relevant information on the acoustical conditions in the recording room. Acoustic models can be obtained to parameterise the acoustic data set, thus making it suitable for transmission.

Transmission

Encoding of audio objects is operated by the MPEG-4 standard and encapsulated into specific data streams. For broadcasting applications the transmission adopts digital video broadcasting (DVB) streams. The coding uses a subset of MPEG-4 components (no predefined profile for the given application in the standard).

Rendering

In the rendering process the transmitted data are demultiplexed, decoded and processed by a compositor. It enables the reproduction of a recorded or simulated sound field via WFS loudspeaker arrays, ensuring immersive sound perception in a wide listening area. The original acoustics of the reproduction room, which may negatively influence the obtained result, is optimally corrected for.

All components were combined to a demonstrator, as a basis for validation. This was done using perceptual experiments and field tests. The results of the project are a first step towards a new quality of high quality spatial audio imaging. CARROUSO has shown the possibility to capture, transmit and render sound sources and their related acoustic environment with more realism, compared to existing stereophonic methods.

This novel spatial audio technology was developed for applications in conjunction with moving pictures, using the recently introduced MPEG-4 standard. It is considered as a major milestone for immersive audio representation at public places and in private households. Two applications have been targeted within this project. The first one concerns high quality spatial audio with associated video for broadcasting. The second application is related to cooperative and interactive work on immersive audio objects. CARROUSO is expected to contribute to information, communication and media technology.

4.2. Synthesis of WFS and Stereophony

This paper also describes the synthesis of WFS and stereophony. Figure 5 illustrates the basic concept by means of music recording. Step one is always the room response measurement in the music hall, done e.g. with a stepwise rotating microphone. This measured spatial information is stored in the WFS processor.

For recording of orchestra and soloist closely spaced spot microphones are used. The stereophonic orchestra mix should be composed in a way that it contains as little room information (reverb, reflections, etc.) as possible; but it should contain the adequate spatial distribution of elements. This three channel stereophonic mix signals and the soloist signal are being convolved with the appropriate spatial impulse responses.

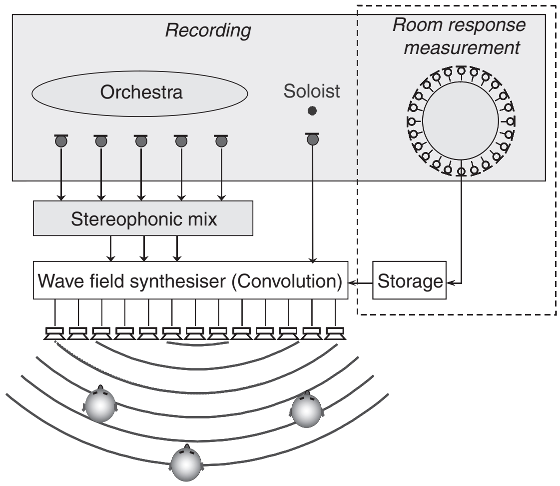

Fig. 5 WFS: Separate handling of sources and spatial information.

As a result, the rendered WFS sound field represents stable virtual sources located in the concert hall. Listeners within the listening area perceive a three-channel stereophonic image of the orchestra and a point source image of the soloist, whereby the reproduced characteristics of the concert hall give a new sense of realism.

On this basis apparent advantages of established conventional stereophonic recording techniques on the one hand and of WFS technologies on the other can, in principle, be utilized through a purposeful combination.

4.3. Virtual Panning Spots (VPS)

The key tool is use of so-called Virtual Panning Spots (VPS) [17], virtual point sources to be applied for panning across any stereophonic imaging plane in the virtual WFS imaging area. VPS can be understood as virtual "loudspeakers" which reproduce the stereophonic sound image of a spacious sound source (e.g. a choir) in the recording room (see also [17–19]). The suitable room impulse responses have to be measured in the original room or to be created artificially in a suitable way. In the example in terms of Fig. 7, the orchestra is imaged with the assistance of six VPS, which are reproduced via WFS and are relatively freely configurable with regard to localization, expansion and distance.

The sound design advantage of this concept is self-explanatory: The stereophonic recording of the orchestra according to Figure 6 produces a spacious sound image of the sources as there is an image between the VPS in accordance with the principles of phantom sources localization. The "loudspeakers" are virtual sources, generated through WFS and provided with the room characteristics of the recording room. The locations of the VPS behave

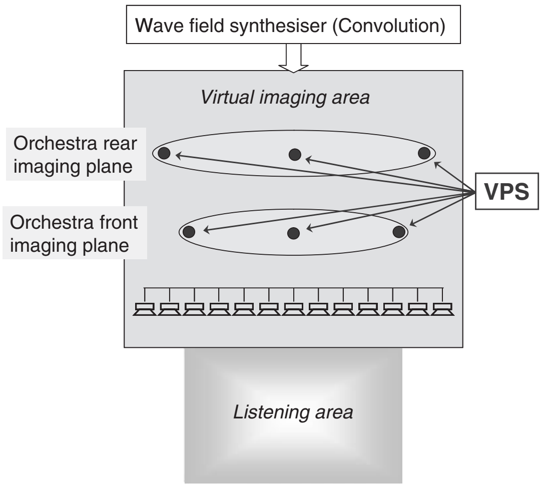

Fig. 6 Use of virtual panning spots (VPS).

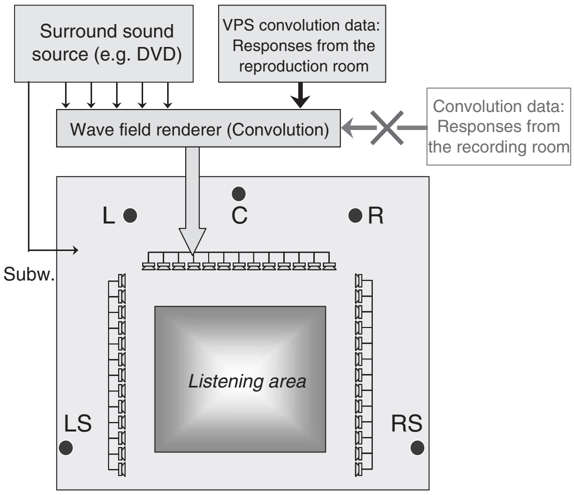

Fig. 7 VPS configuration for rendering virtual multi-channel loudspeakers.

directionally stable in the listening area. The known disadvantages of phantom source localization, especially the low directional stability can be easily avoided by employing a sufficient number of VPS. For further details, see also [20].

4.4. Virtual Loudspeaker Reproduction

An important application of the VPS technique is a special preset of the VPS setup on the reproduction side, which enables the reproduction of conventional multi-channel recordings in a virtual listening room [17]. For this purpose, two modifications are suggested for the WFS decoder, which can be activated in the event of need for application, see Fig. 7:

1. The configuration of the VPS with regard to room impulse responses and spatial arrangement is done in accordance with the preset setup of virtual loudspeakers in a virtual listening room. Arbitrary arrangements of the virtual loudspeakers can be preset and be activated dependent on the stereophonic format to be reproduced.
2. The virtual source signal is not received via the transmission channel, but from the multichannel decoder on the reproduction side (e.g. that of a DVD player).

The WPS reproduction unit operates completely detached from WFS transmission, and can principally offer three attractive advantages:

(1) Diverse stereophonic multichannel formats can be easily reproduced optimally through the selection of a VPS preset, without having to appropriately adjust the loudspeaker arrangement within the living-room.

(2) The virtual loudspeakers can also be placed outside the living-room, i.e. also in a confined area situation, the listening area for multi-channel stereophony is sufficiently large.

(3) A future high quality WFS reproduction unit will allow for an electronic compensation of diverse defects in the reproduction room [13], especially the reduction of the effect of the early reflections and the balancing of asymmetrical arrangements of the speaker array.

From the technical and practical point of view the application of WFS for multichannel stereo reproduction could be the first step towards acceptance in the market place. In this regard, the development of the so-called MAP technology (see e.g. [21]) is important. The flat panels, e.g. fed with glass fiber cables, can often be better integrated into the living-room and are more attractive than conventional loudspeakers. Thereby not only the application of virtual loudspeakers within the home is envisaged, but also the employment in cinemas, theaters as well as for high quality sound reinforcement.

REFERENCES

[1] G. Theile, "On the localisation in the superimposed sound-field," *PhD Thesis, Techn. University Berlin* (1980). Online at http://www.irt.de/wittek/hauptmikrofon/theile/ON_THE_LOCALISATION_english.pdf

[2] G. Theile, "On the naturalness of two-channel stereo sound," *J. Audio Eng. Soc.*, **39**, 761–767 (1991).

[3] G. Theile, "Multichannel natural music recording based on psychoacoustic principles," *Proc. AES 19th Int. Conf.*, pp. 201–229 (2001). Updated version see http://www.irt.de/wittek/hauptmikrofon/theile/Multich_Recording_30.Oct.2001_.PDF

[4] S. Brix, T. Sporer and J. Plogsties, "CARROUSO — An European approach to 3D-audio," *110th AES Conv.*, Preprint 5314 (2001).

[5] A. J. Berkhout, D. de Vries and P. Vogel, "Acoustic control by wave field synthesis," *J. Acoust. Soc. Am.*, **93**, 2764–2778 (1993).

[6] M. M. Boone, E. N. G. Verheijen and P. F. van Tol, "Spatial sound field reproduction by wave field synthesis," *J. Audio*

Eng. Soc., **43**, 1003–1012 (1995).
[7] S. Komiyama, A. Morita, K. Kurozumi and K. Nakabayashi, “Distance control system for a sound image,” *Proc. AES 9th Int. Conf.*, pp. 233–239 (1991).
[8] K. Ono, S. Komiyama and K. Nakabayashi, “A method of reproducing concert hall sounds by “Loudspeaker Walls.” *102nd AES Conv.*, Preprint 4490 (1997).
[9] H. Geiger and K. Scheel, *Handbuch der Physik*, Band VIII (Verlag Julius Springer, Berlin, 1926).
[10] U. Horbach and M. Boone, “Practical implementation of data-based wave field reproduction system,” *108th AES Conv.*, Preprint 5098 (2000).
[11] M. Boone and W. de Bruijn, “Improving speech intelligibility in teleconferencing by using wave field synthesis,” *114th AES Conv.*, Preprint 5800 (2003).
[12] H. Wittek, “Perception of spatially synthesized sound fields — Literature review about WFS,” Available online at www.hauptmikrofon.de/wfs.htm
[13] S. Spors, A. Kuntz and R. Rabenstein, “An approach to listening room compensation with wave field synthesis,” *Proc. AES 24th Int. Conf.*, pp. 70–82 (2003).
[14] E. Corteel and R. Nicol, “Listening room compensation for wave field synthesis. What can be done?,” *AES 23rd Int. Conf.*, Paper 7 (2003).
[15] J.-J. Sonke, “Variable acoustics by wave field synthesis,” Thela Thesis, Amsterdam, Netherlands, ISBN 90-9014138-3 (2000).
[16] E. Verheijen, “Sound reproduction by wave field synthesis,” Delft University of Technology (1997).
[17] G. Theile, H. Wittek and M Reisinger, “Potential wavefield synthesis applications in the multichannel stereophonic world,” *Proc. AES 24th Int. Conf.*, pp. 43–57 (2003).
[18] M. Boone, U. Horbach, U. de Bruijn and P. J. Werner, “Virtual surround speakers with wave field synthesis,” *108th AES Conv.*, Preprint 4928 (1999).
[19] C. Kuhn, R. Pellegrini, D. Leckschat and E. Corteel,” An approach to miking and mixing of music ensembles using wave field synthesis,” *115th AES Conv.*, Preprint 5929 (2003).
[20] M. Boone, W. de Bruijn and T. Piccolo, “Sound recording techniques for wave field synthesis and other multichannel systems,” *104th AES Conv.*, Preprint 4690 (1998).
[21] M. Boone, “Multi-Actuator Panels (MAPs) as loudspeaker arrays for Wave Field Synthesis,” *J. Audio Eng. Soc.*, **52**, 712–723 (2004).