Acoustics'08
Paris
June 29-July 4, 2008
www.acoustics08-paris.org

Simulation of wave field synthesis

F. Völk, J. Konradl and H. Fastl

AG Technische Akustik, MMK, TU München, Arcisstr. 21, 80333 München, Germany
florian.voelk@mytum.de

Wave field synthesis utilizes a large number of loudspeakers to generate a desired wave field. It therefore is necessary to drive each speaker with an independent signal, which requires as many amplifier and soundcard channels as there are loudspeakers. These enormous hardware costs make research and development expensive and time consuming. Additionally, different rooms influence the wave field synthesis arrays in different ways. For this reason, a simulation technique is of advantage that permits the evaluation of the perceived properties of arbitrary wave field synthesis configurations without the need to physically construct them. This paper proposes a simulation system capable to simulate wave field synthesis systems in different rooms based on physical measurements of loudspeakers in each room. It presents this system, which is called *virtual wave field synthesis*, and discusses possibilities as well as limits of this system based on preliminary listening experiments.

1 Introduction

At the moment, the number of different publications on approaches for the creation of virtual acoustical environments increases nearly every month. All these concepts can be grouped with regard to their underlying principles: one group aims for synthesizing a desired wave field in the whole reproduction area; the aim of the second group is to reproduce only those signals that are present at the listener's ears in the recording situation or at least would be present there in a virtual scene that should be synthesized.

A well-known technique belonging to the first group is the so-called wave field synthesis (WFS, cf. [1]), aiming for (re-)constructing a wave field in a certain listening area. A member of the second group is the binaural technique (cf. [2], [3]), which aims to (re-)construct the listeners' ear signals using dummy or human head recordings or a proper synthesis procedure.

Both of the mentioned techniques require audio signal processing in real time and therefore a relatively fast signal processing. In addition, wave field synthesis utilizes a huge number of loudspeaker channels, including one amplifier and one soundcard channel per loudspeaker as well as an array for the loudspeakers' placement. Hardware requirements for the binaural technique on the other hand are comparably small; two output channels, a head tracking system, and a pair of headphones are needed.

Considering the background given above, it is obvious that development and scientific research in the field of wave field synthesis is hardware costly and time consuming. For this reason, a simulation system is desirable that creates the same hearing sensation as a given wave field synthesis system, but reduces the necessary hardware costs. Besides, it would be helpful to be able to compare different wave field synthesis systems directly without the delay necessary for their physical construction. Another interesting point in current research on wave field synthesis would be the computation of the ear signals occurring in a sound scene rendered by wave field synthesis. This ear signal computation may be called a missing link between wave field synthesis and psychoacoustics.

This article presents the basics of a wave field synthesis simulation system based on binaural reproduction, which possesses all the aforementioned advantages. After a short overview on previous work in this field, intention and goals of this study are presented. Afterwards, the concept of this system is presented in detail, followed by a critical technical review leading to some problems to be solved in the future. As will be shown in the current paper, the ear signal computation comes with no additional costs if the proposed wave field synthesis system, we call it *virtual wave field synthesis*, is realized. Finally, a preliminary listening test, first results and a short discussion conclude this article.

2 Previous work

Until now, most psychoacoustic research in the field of wave field synthesis has been carried out in the traditional way by building up the system under consideration and presenting the rendered wave fields to some listeners, mostly under anechoic conditions (cf. [4], [5], and [6]).

Additionally, there was an attempt to reduce the hardware requirements necessary for listening tests in the context of wave field synthesis: Wittek et al. ([6], [7], [8]) used the so called Binaural Room Scanning (BRS, cf. [9]) in combination with offline computed wave field synthesis driving signals for the generation of a virtual wave field synthesis array. This system is restricted to static wave field synthesis because of the offline wave field synthesis computation. They do not describe in detail the used solution and only a short verification experiment is mentioned. This is the only experiment known to the authors, which points in the direction of the work described in the current paper.

3 Concept

The proposed system should reduce the hardware requirements necessary for psychoacoustic research in the context of wave field synthesis but should nevertheless allow the rendering and modification of virtual scenes in real-time. As an additional advantage, the computation of the ear signals in the considered situation should be an inherent part of this system.

3.1 Computation of ear signals

Wave field synthesis relies on the idea of driving each speaker of a loudspeaker array with a different signal so that the output signals altogether lead to a desired spatial and temporal wave field in the so-called reproduction area (in most cases a theoretical completely correct reproduction is solely possible on a line or on a single point in the three dimensional space, cf. [10], [11]). The loudspeakers' input signals are called driving signals.

If we assume the system under consideration as linear and time invariant (common praxis for binaural technique, cf. Møller ([2]) and holds as first approximation under static circumstances), the following examination is possible: From a system theoretical point of view, the loudspeakers' output signals, each of them convolved with the appropriate propagation paths' impulse responses, called the appropriate head related impulse responses (HRIRs), superimpose at each of the listener's eardrums. For clarity of terms, it should be mentioned that head related impulse responses could be recorded under anechoic or reflective conditions. To distinguish theses two cases explicitly, the latter ones are most often called binaural room impulse responses (BRIRs) to denote the inclusion of room information. In this paper, the recording environment plays a minor role and therefore the term head related impulse response is used in any case to avoid unessential distinctions.

The sound pressure signals at the eardrums are called ear signals. To subsume the preceding paragraph, the ear signals for one static listener position and orientation can be described as superposition of the loudspeakers' input signals (the driving signals), each convolved with the associated loudspeaker's impulse response and with the appropriate propagation paths' impulse responses. Fig. 1 shows this consideration as a schematic block diagram.

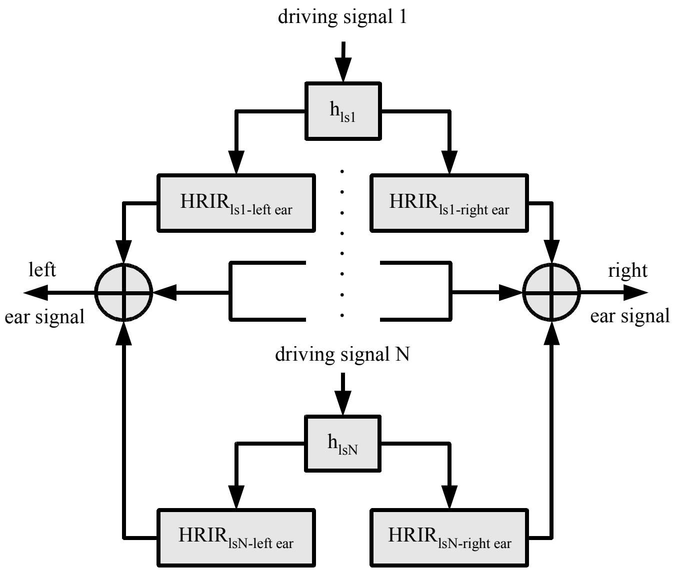

Fig. 1 Wave field synthesis.
Schematic system theoretical consideration of the listening situation in wave field synthesis reproduction. h_{lsn} denotes the impulse response of loudspeaker n, $HRIR_x$ identifies the head related impulse response between a certain loudspeaker and one ear of the listener.

With this background, it becomes clear that the ear signals occurring while listening to a wave field synthesis system at one instant of time could be computed, if the involved head related impulse responses, the loudspeakers' impulse responses and the used driving signals are known.

This holds for all possible geometries of the wave field synthesis array, because spatial information about the locations of the loudspeakers is contained in the corresponding head related impulse responses. For that reason, the reachable quality of synthesis strongly depends on the chosen HRIRs. The latter may be acquired by measurements (the so-called data-driven approach) or with a proper rendering method (model-driven). Therefore, it remains up to the operator to select the HRIR-acquiring method which best fits his necessities. The easiest, practically working but not at all theoretically correct way would be to measure the HRIR-set necessary for the rendering of one WFS-speaker in the room to be synthesized and to control the position of the other speakers simply by adjusting the included delay, depending on the distance between the considered loudspeaker and the listener.

3.2 Dynamic virtual wave field synthesis

For the realistic simulation of a wave field synthesis array the static solution described in the previous section is insufficient, especially because of the lack of the important dynamic localisation cues (cf. Wightman and Kistler, [12]). To overcome this limitation, an existing dynamic binaural technique system (cf. Völk et al., [13]) was modified to allow at each moment the computation of the current ear signals resulting from a sufficient number of (virtual) wave field synthesis loudspeakers in real time and to adopt the used filters dependent on the position and orientation of the listener. Fig. 2 shows a block diagram of the resulting system.

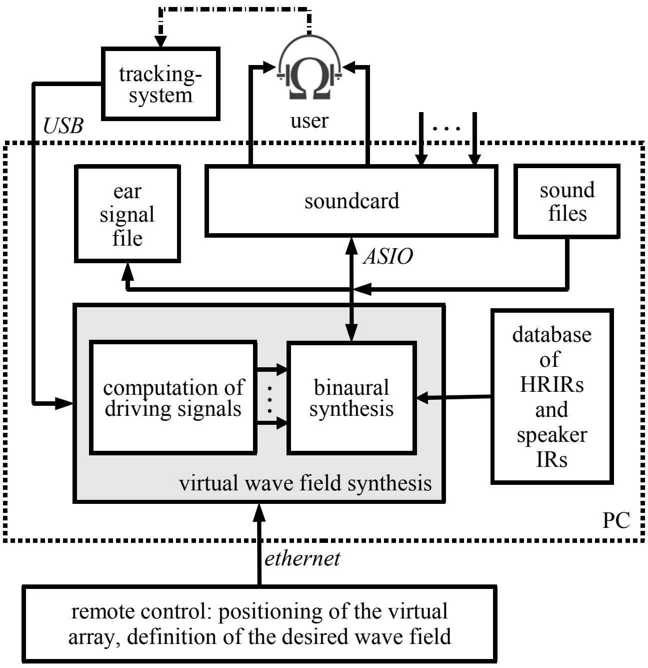

Fig. 2 Virtual wave field synthesis.
Schematic block diagram; the signal processing (computation of driving signals and binaural synthesis) is done in the core of the system (grey box), which is realised as a software tool on a consumer PC, the latter indicated by a dashed line. Additionally, the necessary tracking system, the remote control, and the possible data sources and sinks are shown.

The core of the presented system, the signal processing software, is drawn with a light grey background; it runs on a consumer PC that is indicated with a dashed line. This core is composed of two major engines: the online computation of driving signals necessary for moving primary sources (the sources to be synthesized via the original WFS), and the binaural synthesis, which accounts for movements of the listener.

All parameters can be adjusted at runtime over a network remote control. For that reason, moving primary sources and varying array geometries easily can be realised. The user's orientation and position are tracked and could therefore be accounted for in the computation of the driving signals and in the binaural synthesis. For that reason, with the presented system it is possible to set the point of optimal amplitude adjustment resulting from the wave field synthesis (cf. [11]) dependent on the listener's position. In other words, it is possible to avoid the amplitude errors occurring in real wave field synthesis systems caused by the deviation from the theoretical case due to the finite array size. It is clear that, if the synthesis of a real system is intended, this adoption is not correct and should be avoided.

With the presented system, it is possible to choose any distance between two neighbouring loudspeakers and therefore to decrease their spacing until the spatial aliasing frequency becomes higher than the upper frequency limit of the hearing area. Another advantage of this procedure would be the possibility to select any loudspeaker for the HRIR measurement, independent of its physical size, what would make transfer functions possible that are nearly frequency independent with no spatial aliasing.

At the moment, no psychoacoustic experiments have been carried out to support these assumptions; they are only based on a theoretical consideration. The subjective verification of the constraints is a current project at our laboratories.

Besides the mentioned constraints, Fig. 2 shows the necessary database of impulse responses and the input and output paths. As inputs (primary source signals), sound files or soundcard inputs might be used. In the synthesis case, the output signals are sent over the soundcard to a pair of headphones. If the computation of ear signals in the dynamic case (this is possible at discrete instants of time for moving primary sources and/or listeners) is desired, then the output signals might be streamed to a data file.

3.3 Known problems

The most obvious problem at the moment is insufficient computation power. For that reason, only a relatively small number of secondary sources can be rendered at the same instant of time.

Another shortcoming is the restriction of the wave field synthesis reproduction to only one listener. Therefore, the intended goal of WFS, the reproduction of the whole wave field, which would allow supplying multiple listeners with the correct ear signals, is not reachable with this system.

4 Subjective verification

A listening experiment was carried out to assess the properties of the auditory events created by a simple implementation of the introduced system. 14 subjects (one female and 13 male) aged between 22 and 34 years (mean value: 26.4 years) had to judge some properties of the auditory events created by a (virtual) primary point source.

Three participants were experienced listeners; two of them had previous experience with listening in virtual auditory displays and with localization experiments. They also knew about the presentation system. All other subjects were naive regarding the presentation system and had no experience with listening tests.

4.1 Stimuli and Procedure

A (virtual) primary point source was placed at nine different positions in the horizontal plane (cf. Blauert, [14]). Stimuli were presented to each person in an individual random order from different spatial source locations; each stimulus occurred four times. Subjects could enter the direction and the distance at which they perceived the auditory events via a graphical interface. The whole trial was automated using a software-program running on a Tablet-PC that also served as input device for the subjects. They could give their answers by pencil-click directly on the touch screen of the Tablet-PC. For this purpose, they saw first a top down view and afterwards a side view of the shape of a head, as shown in Fig. 3.

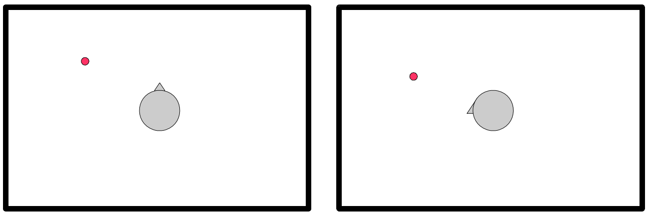

Fig. 3 Input screen for the localisation trial.
The frame on the left shows the sketch used for input of horizontal direction and distance, the frame on the right was used to ask for the vertical position of the auditory events. In the middle of each frame, the shape of a head is depicted. Both frames contain a response sample as might be set as indicator of the auditory event by the subjects.

The subjects had to mark up the position corresponding to the auditory event's location on a completely black sketch, only showing the shape of a head in the respective representation. The persons had the possibility to correct a given answer as many times as they wanted.

A virtual wave field synthesis array consisting of 80 loudspeakers with a distance of 10 cm grouped as circle, leading to a diameter of 255 cm, was used. The position of optimal amplitude adjustment was chosen in the center of the array. The middle of the array was positioned at a chair. The subjects were sitting on this chair and could move their heads freely but were not allowed to walk around. That led to the situation that the position of the virtual array was independent of head movements like a hardware implementation would be.

The system was operated at 48 kHz; all involved digital signals were originally captured or synthesized at this sample rate. All filters were realized as FIR filters. A HRIR-set, recorded with a dummy head (Neumann KU 100) in a reflective environment, was used (filter-length: 8192 samples). Only HRIRs measured in the horizontal plane were used in the binaural synthesis process. For that reason, vertical head movements and rotations had no impact on the rendered ear signals. The length of the wave field synthesis filter used for the computation of the driving signal was selected to 2048 samples. All calculations in the audio-processing chain were performed at a block-size of 512 samples, which made partitioning of the driving signal as well as of the impulse responses necessary. This was realized by an adaptive overlap and save procedure.

As acoustic stimulus, pulsed uniform exciting noise (UEN, cf. Fastl and Zwicker, [15]) was used. This stimulus has equal intensity in each critical band, thus providing all spectral cues contained in the HRIRs to the listener with the same perceptual weight. Therefore, all possible spectral information is available to the hearing system, but no influence of the sound stimulus on the auditory event should be present. To add temporal information besides the random temporal structure of the noise, the UEN was pulsed with 700 ms pulse duration and 300 ms pause duration in between the impulses. Following Blauert and Braasch ([16]), 200 ms is the minimal duration allowing dynamic localization cues, so dynamic localization should also be possible. The pulses were modulated with 20 ms Gaussian gating signals to prevent audible clicks.

To exclude a possible influence of any visual stimuli besides darkness (which may also have an influence on auditory perception), the experiments were conducted under very dark conditions and the listeners did not see the listening room at any instant of time (to ensure this, they were blindfolded before entering and leaving the room).

4.2 Results

For statistical analysis, the individual median over the four denoted auditory event positions per primary source was utilized. In the following graphics, the results for direction, distance, and deviation from the intended horizontal plane are displayed as medians and inter-quartile ranges of the individual medians (blue circles and ranges). Azimuth angles are measured mathematically positive from the frontal plane.

Fig. 4 shows the azimuth results (blue) over the presented primary source azimuths. The red stars indicate the intended directions with uncertainty ranges for real sources at the respective directions after Blauert ([14]).

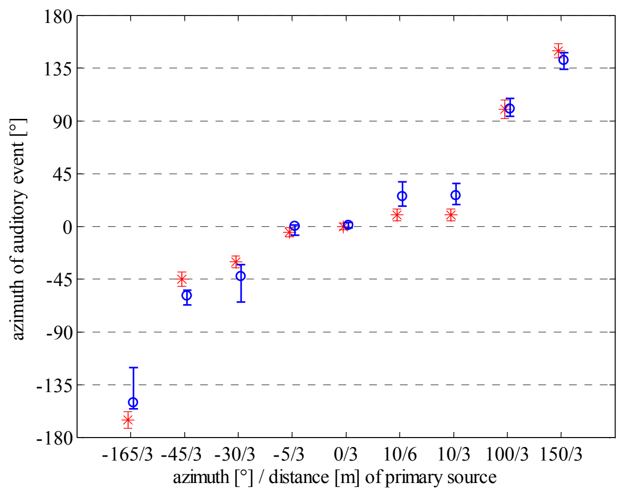

Fig. 4 Denoted azimuths of auditory events.
Blue circles indicate medians of individual medians with inter-quartile ranges. Red stars show the intended virtual source positions with inter-quartile ranges for real sources at these positions (after Blauert, [14]).

The intended vertical position of all primary sound sources is the horizontal plane, because all used HRIRs were recorded in this plane. For their response of the vertical position of the auditory event, the persons saw a shape of a head in the middle of the screen and were instructed to mark the position of the auditory event (see 4.1).

For the computation of the results, rather the distance to the horizontal plane was used than the elevation angle, because this angle depends additionally on the selected distance. Fig. 5 shows the denoted deviation of the auditory events from the horizontal plane. On the ordinate, figure units relative to the head radius, which is set to one, are drawn. The horizontal red line indicates the radius of the head shape that was visible to the subjects during their judgments.

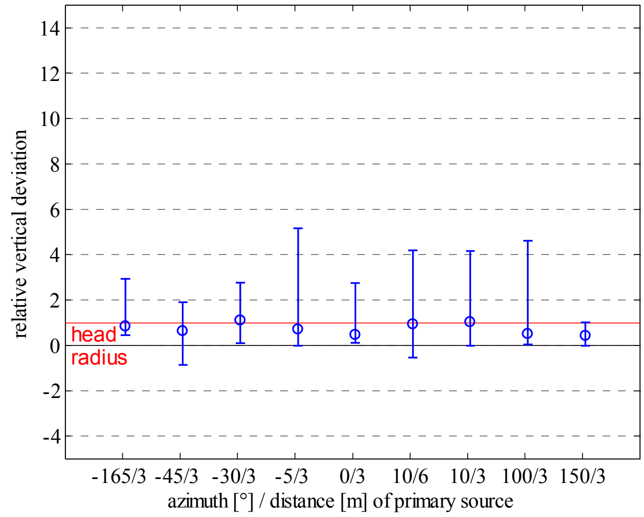

Fig. 5 Stated vertical deviation from the horizontal plane. Blue circles show medians and inter-quartile ranges of individual medians of the perceived deviation of the auditory events from the (intended) horizontal plane.
On the ordinate, figure units relative to the radius of the head shape visible to the listeners during the judgements are drawn.

Like the elevation, the distance of the auditory events is depicted in units relative to the radius of the head shape on the input screen. Fig. 6 shows the stated distances of the auditory events and the radius of the head shape as well as the maximum possible selectable value (due to the physical size of the touch-screen used for input).

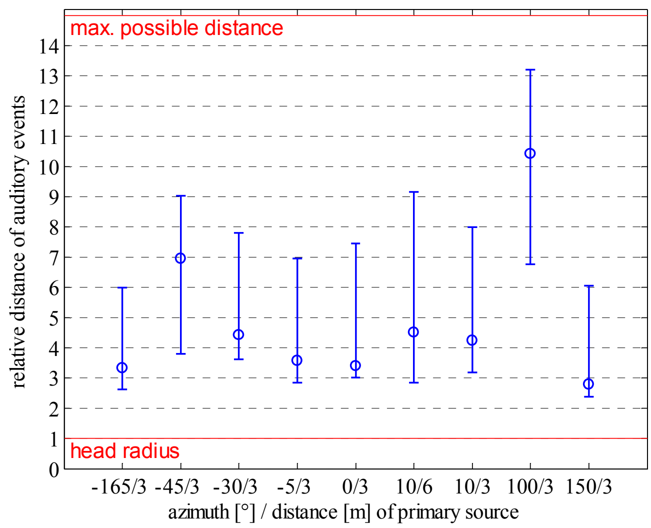

Fig. 6 Denoted distance of the auditory events.
Blue circles represent medians of individual medians with inter-quartile ranges. On the ordinate, figure units relative to the radius of the head shape visible to the listeners during their judgments are shown. Additionally, the maximum selectable distance (due to the screen size) is depicted.

5 Discussion

The results displayed in Fig. 4 show good agreement between the azimuth values of the presented primary sources and the denoted directions of the auditory events. This becomes even more evident if one considers the finite accuracy of the input procedure. The subjects have to map their perceptive space on to the screen of the tablet PC. Additionally, the orientation of each subject, which is important to match the coordinate systems of the input screen and the virtual array, could be adjusted only within certain limits because of the darkness and the procedure that allowed movements of the persons on the chair during the trial. The small inter-quartile ranges suggest that the used procedure leads to repeatable answers and is valid at least for this pilot-study. There is a tendency visible for the auditory events corresponding to primary sources out of the median plane to deviate from the intended position. In all cases, this deviation tends towards 90° or -90° respectively. The auditory events tend to be more laterally than the corresponding primary sources. According to results plotted in Fig. 5, the median values of the denoted elevations of the auditory events lie all between the horizontal plane and the upper limit of the head and there is no tendency for the frontal auditory events to be perceived higher than the lateral or dorsal ones. However, the inter-quartiles indicate significant elevations for some subjects. The results displayed in Fig. 6 illustrate that all auditory events are externalized, i.e. all virtual sources are perceived outside the head. For the time being, a more detailed discussion of the distance results is not feasible, because in the used circular array, for different primary source positions, a different number of loudspeakers are active, which leads to different reproduced sound pressure levels. For that reason, the level, an important cue for distance perception, varies independent of the source distance. This problem has to be solved before a reasonable discussion of the perceived distances will be possible.

6 Summary

This paper presents a system capable of producing a virtual dynamic wave field synthesis environment in real time. It possesses great potential for current research on wave field synthesis because the necessary hardware requirements remain independent of the number of simulated speakers and independent of the array geometry under consideration. Additionally, the computation of ear signals of almost every possible listening situation (including dynamic scenarios) in the context of wave field synthesis is easily possible. This computation of ear signals can be regarded as a missing link between the current research on wave field synthesis and the well-known psychoacoustic results on spatial hearing. First results from localisation experiments show to a certain extent that the intended aim is reached. To ensure that the proposed virtual wave field synthesis system behaves like its real counterpart, further research is necessary.

Acknowledgments

The authors thank Dr. Helmut Wittek and Dr.-Ing. Günter Theile for meaningful ideas, as well as Dipl.-Ing. Daniel Menzel, who contributed a lot of inspiration in many fruitful discussions. Part of this work was supported by grant FA 140/4 of the Deutsche Forschungsgemeinschaft (DFG).

References

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

[2] H. Møller, "Fundamentals of Binaural Technology", *Appl. Acoustics* 36, 171-218 (1992)

[3] D. Hammershøi, H. Møller, "Methods for Binaural Recording and Reproduction", *ACUSTICA - acta acustica* 88, 303-311 (2002)

[4] P. Vogel, "Application of Wavefield Synthesis in Room Acoustics", *PhD Thesis, Technische Universiteit Delft* (1993)

[5] E. Verheijen, "Sound Reproduction by Wave Field Synthesis", *PhD Thesis, Technische Universiteit Delft* (1997)

[6] H. Wittek, F. Rumsey, G. Theile, "Perceptual Enhancement of Wavefield Synthesis by Stereophonic Means", *J. Audio Eng. Soc.* 55, 723-751 (2007)

[7] D. Wegmann, G. Theile, H. Wittek, "Zu Unterschieden in der spektralen Verarbeitung der Ohrsignale bei Stereofonie und Wellenfeldsynthese", *Fortschritte der Akustik, DAGA '06*, DEGA e. V., Berlin (2006)

[8] H. Wittek, "Perceptual differences between wavefield synthesis and stereophony", *PhD Thesis, Department of Music and Sound Recording, School of Arts, Communication and Humanities, University of Surrey* (2007)

[9] P. Mackensen, U. Felderhoff, G. Theile, U. Horbach, R. Pellegrini, "Binaural Room Scanning – A new Tool for Acoustic and Psychoacoustic Research", *ACUSTICA – acta acustica* 85, 417 (1999)

[10] S. Spors, "Active Listening Room Compensation for Spatial Sound Reproduction Systems", *PhD Thesis, Universität Erlangen-Nürnberg* (2005)

[11] J. Sonke, J. Labeeuw, D. de Vries, "Variable Acoustics by Wavefield Synthesis: A Closer Look at Amplitude Effects", *104th AES Convention* (1998)

[12] F. L. Wightman and D. J. Kistler, "Resolution of front-back ambiguity in spatial hearing by listener and source movement", *J. Acoust. Soc. Am.* 105, 2841-2853 (1999)

[13] F. Völk, S. Kerber, H. Fastl, S. Reifinger, "Design und Realisierung von virtueller Akustik für ein Augmented-Reality-Labor", *Fortschritte der Akustik, DAGA '07*, DEGA e. V., Berlin (2007)

[14] J. Blauert, "Spatial Hearing – The Psychophysics of Human Sound Localization", *The MIT Press*, Cambridge, Massachusetts, London, England, Revised Edition (1997)

[15] H. Fastl, E. Zwicker, "Psychoacoustics – Facts and Models", *Springer Verlag, Berlin, Heidelberg*, 3rd Edition (2007)

[16] J. Blauert, J. Braasch, "Räumliches Hören", Contribution for the *Handbuch der Audiotechnik (Chapter 3., Stefan Weinzierl, Ed..), Springer Verlag*, Berlin, Heidelberg (2007)