DAFx04 Invited paper

WAVE FIELD SYNTHESIS - GENERATION AND REPRODUCTION OF NATURAL SOUND ENVIRONMENTS

Thomas Sporer

Fraunhofer Institute for Digital Media Technology IDMT
Ilmenau, Germany
spo@idmt.fhg.de

ABSTRACT

Since the early days of stereo good spatial sound impression had been limited to a small region, the so-called sweet spot. About 15 years ago the concept of wave field synthesis (WFS) solving this problem has been invented at TU Delft, but due to its computational complexity it has not been used outside universities and research institutes. Today the progress of microelectronics makes a variety of applications of WFS possible, like themed environments, cinemas, and exhibition spaces. This paper will highlight the basics of WFS and discuss some of the solutions beyond the basics to make it work in applications.

1. INTRODUCTION

In 1931 Blumlein [1] filled a British patent which described the basics of stereo recording and reproduction, and which is up to now the basic of all stereo recording techniques. In 1934 researchers at AT&T Bell labs described two major configurations of spatial audio reproduction: binaural, that is two channels recorded at the ear position (dummy head) and multi-channel [2]. In experiments they proved three loudspeakers (left, center, right) provide superior quality to a larger audience compared to two loudspeakers. About twenty years later, after careful research on a variety of loudspeaker number and configurations, William Small concluded in [3] that " The number of channels will depend upon the size of the stage and listening rooms, and the precision in localization required." " ...for a use such as rendition of music in the home, where economy is required and accurate placement of sources is not of great importance if the feeling of separation of sources is preserved, two-channel reproduction is of real importance." Due to limitations in physical delivery media (LP, CD) and broadcast formats two-channel stereo was and is dominant in most applications. After a commercially non-successful extension to four channels (quadrophony) in the seventies today 5-channel stereo, which adds 2 surround channels and a center channel is used more and more. But two, three and five channel stereo are doomed to preserve some limitations along the time:

- the position of the loudspeakers has to be exactly the position as predefined when producing the content
- good spatial audio quality is limited to a small portion of the reproduction room, the so-called sweet spot
- (virtual) sound sources can be placed at loudspeaker position, between loudspeaker positions or farer apart from the listener, but not in the gap between loudspeaker and listener
- (virtual) sound sources placed between loudspeakers sound differently than sound sources placed on loudspeakers positions
- sound sources placed between front and surround speaker are rather unstable, that is they are even more dependent on the exact position of the listener
- in larger venues, like cinemas, where 5 loudspeakers are not sufficient to provide acceptable sound pressure level for all listeners, additional loudspeakers have to be used. Dependent on the position of the listener some or even most surround channels are no more in the back but in front, disturbing the sound image from front.
- all formats are lacking the 3rd dimension
- all formats store only final mixes of sound scenes making interactive manipulation of content by the end user impossible

Newer multi-channel formats like Tom Holmans 10.2 [4] improve the size of the sweet spot and try to add a third dimension. However they do not bridge the gap between listener and loudspeaker, and mixing in that formats is very difficult. A way to overcome the all limitations listed above is to store all content in an object oriented approach: Each audio source is stored as an audio object comprising certain properties like position, size and (mono) audio track. A scene description describes how the individual objects are composed to build up a scene.

Since 2001 key technologies for recording, encoding, transmitting, decoding and rendering of object-oriented sound fields have been investigated in an European project called CARROUSO [5]. Key components in this project are Wave Field-Synthesis as a new way of reproducing sound and the object-oriented features and codecs of MPEG-4. Wave Field Synthesis (WFS) was invented at the TU Delft in the Netherlands [6]. In the nineties WFS has been demonstrated in academic environments and basic research was conducted. For long time the main reason prohibiting applications outside academic research was its high computational complexity. Decreasing costs of computing power enabled the first application in the professional market. In February 2003 Fraunhofer IDMT has implemented a large WFS array in the Ilmenau cinema and new applications for WFS technology are around the corner.

2. BASICS OF WAVE FIELD SYNTHESIS

Wave Field Synthesis (WFS) is based on the wave theory concept of Huygens [7]: Each point on a wave front can be regarded as the origin of a point source. The superposition of all these secondary sources form a wave front which is physical indistinguishable from the shape of the original wave front. This principle, originally described for water waves and optics, was first applied to acoustics at TU Delft. A large number of small and closely spaced loudspeakers form a so called loudspeaker array (Figure 1). Each loudspeaker in the array is fed with corresponding driving signal cal-

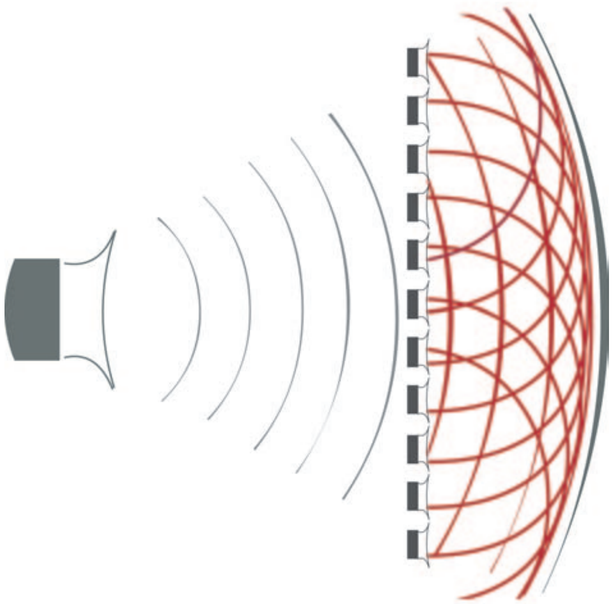

Figure 1: *Principle of WFS: Superposition of secondary sound source recreates sound-field*

culated by means of algorithms based on the Kirchhoff-Helmholtz integrals and Rayleigh's representation theorems [8].

$$P_A = \frac{1}{4\pi} \oint_S \left[\left(P \frac{1+jk\Delta r}{\Delta r} cos\phi \frac{exp(-jk\Delta r)}{\Delta r} \right) + \right. \\ \left. + \left(j\omega\rho_0 V_n \frac{exp(-jk\Delta r)}{\Delta r} \right) \right] dS \quad (1)$$

The Kirchhoff-Helmholtz integral implies that an infinite number of monopoles and dipoles encircling the reproduction space is necessary to achieve perfect results. "Perfect results" includes the property that the reproduced sound field outside the listening space (behind speakers) is zero. Taking either monopoles or dipoles instead of both the sound field inside is the same and only the sound field outside is non-zero. Today most implementations of WFS are based on monopoles only.

A second step to simplify WFS is to reduce the sound-field from 3D to 2D, therefore all loudspeakers are located in one plane.

Reducing the number of loudspeakers to a finite number limits the frequency up to which WFS provides perfect reproduction. Above the alias frequency spatial alias terms occurs. In practice it proved to be sufficient to locate a loudspeaker every 17 cm, giving an alias frequency of about 1 kHz. From an acoustic point of view this seems to be insufficient, but due to psychoacoustical effects a decrease of distance between loudspeakers has only marginal effects on audio quality.

Without room reflection of the reproduction room the sound field reproduced is perfect in nearly the whole space between the loudspeakers. However the superposition of the individual loudspeaker signal does not work properly if the distance between listener and loudspeaker is smaller or similar to the distance between loudspeakers. Such close to the loudspeaker additional effects occur due to the near field of cone loudspeakers.

Three types of sound sources can be reproduced (Figure 2):

Point Sources are sound sources which are between or behind the loudspeakers. At each listening position the position of such sound sources is perceived to be the same.

Focused Sources are point sources in front of the loudspeaker array. While having similar properties as "normal" point sources for most of the listening positions, for positions between source and closest loudspeaker the sound field is inverted and there is no precise location.

Plane Waves behave similar to point sources which are in infinite distance: All listeners perceive the same direction of the source. For this type of signals the effect of distance dependent reduction of sound pressure level is negligible.

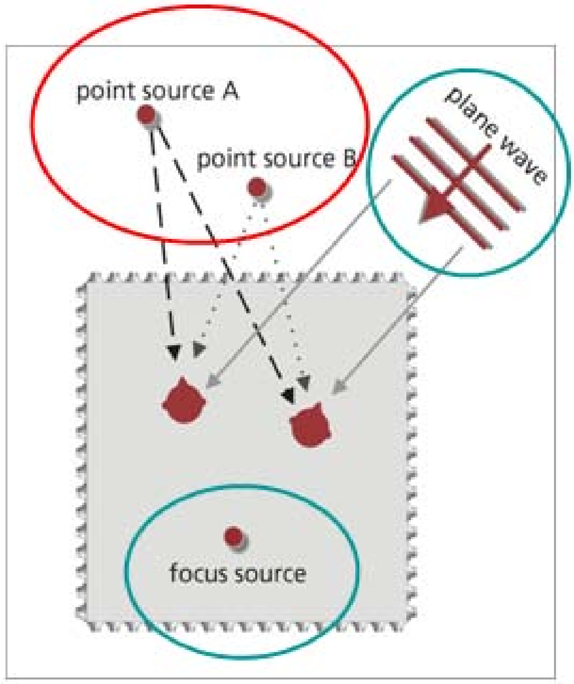

Figure 2: *Reproduction of point source, focused sources and plane waves with WFS*

3. PROPERTIES AND REQUIREMENTS OF WFS

Reproduction of low Frequencies Due to the limited number of loudspeakers (sampling in the spatial domain) frequencies below the alias frequency are amplified by 3 dB per octave [6]. To compensate for this effect the input signals have to be pre-filtered. However the value 3 dB is only valid if the virtual sound source if far apart from the loudspeaker array: if the virtual source is placed exactly on a loudspeaker position there is only the frequency response of a single loudspeaker. The filtering therefore has to be position dependent. In practical applications prototype filters for regions are precomputed and stored. The size of the regions with the same correction filters is smallest near the loudspeaker array, at large distance the approximation of 3 dB per octave is sufficient. That way the computational complexity is reduced, which makes larger numbers of moving virtual sources feasible.

Source Position and Level The simplification of the Kirchhoff-Helmholtz integral as derived in [6] provide correct approximation for the horizontal plane for static (non-moving) sound sources. Due to the fact that these simplifications are done with the assumption of an infinite length of a (linear) loudspeaker array, the variation of the synthesized sound pressure level with position of the virtual source is not correct for smaller systems. This effect is clearly perceptible when audio objects are visible. Similar problems occur if the loudspeaker array is not complete due to limitations of reproduction room (windows, doors, curtains). In practical implementations the level has to be corrected. A simple way to achieve the correction value is to simulate the sound pressure level at a reference point in the listening area under the assumption of an ideal array and with the actual loudspeaker configuration. If the sound pressure level of the actual array would be too small the volume of all active loudspeakers is raised. For a given loudspeaker configuration the correction value only depends on the position of a virtual source. To reduce computational complexity and simplify parallel processing of rendering for different loudspeakers the correction value is precomputed and stored.

Doppler Shift For each position of a virtual sound source WFS creates a natural sound field. If sound sources are moved from one position to another a natural Doppler effect is created automatically. Being a benefit when applying to sound effects in movies this Doppler shift might be unwanted when moving musical instruments. A simple way to overcome this is to use panning between virtual objects as described in [9].

Loudspeaker Characteristics The basic theory of WFS starts with a combination of ideal loudspeakers being either pure monopoles or pure dipoles. Most installations done so far are using cone loudspeaker, which can be approximated as being monopoles up to the alias frequency of WFS. In [10] algorithms are shown to compensate for loudspeakers with directivity. Up to now these algorithms are hardly been used due to complexity.

Characteristics of the Reproduction Room Wave field synthesis is able to reproduce complete sound scenes including the room reflections of a natural or simulated (recording) environment. If the reproduction room has his own reflection pattern the reflection pattern of recording and reproduction room are combined, which often is regarded to sound unnatural. The ideal reproduction room therefore would be an anechoic room. Due to the fact that all walls of the reproduction room are equipped with loudspeakers electro-acoustic cancellation of the first room reflections are possible [11, 12]. Experience from installations show that some diffuse reflections of the reproduction room improve the naturalness of simple simulations of room acoustics.

Recorded and Simulated Rooms Using WFS it is possible to process signals coming from the (primary) sound sources separately from the (secondary) signals coming from the recording room. Thus it is possible to manipulate the two signal classes independent of each other. If done properly it is even possible to change the (virtual) recording venue at the reproduction site. The information of the recording room inherent in a WFS file can be either the recording of the room signature of a real room or the simulated room signature of a room which even does not exist.

4. APPLICATIONS

WFS and mathematically related sound rendering methods like higher-order ambisonics [13] are feasible to all sound reproduction systems where ever it is possible to use more than just one or two loudspeakers. One significant advantage of WFS compared to classical multi-channel reproduction is the possibility to switch from reproduction based storage of audio (the format is defined by the number of loudspeaker channels) to a source based storage (each audio object is stored separately and can be rendered for the best possible audio quality given any reproduction setup). This paradigm shift will revolutionize both production and distribution of audio content, and might be even more important than the improvements of audio quality,

4.1. Application areas

Concert halls: A critical parameter for any kind of life performance is the input-output latency of the signal processing involved. Wave field synthesis has an intrinsic delay which is short enough. If the acoustics of the concert hall is sufficiently good, what should be a property of any concert hall, no room equalization filters, which might cause additional delay, are necessary. WFS can adapt the acoustical behavior of multifunctional venues to the requirements of any kind of music reproduction, and for other purposes, too, for instance for sport events or conferences. In contrast to the systems used today, where electro-acoustical amplification often destroys the sound scene, WFS can make it sound much more natural. Due to the fact that WFS preserves spatial angular and distance precision a much improved audio-visual coherence is achieved.

Open air events: An important requirement for open air concerts is to achieve a sufficiently high sound pressure level for the whole audience without creating dangerously high sound pressure levels near the stage. At the same time the audio quality should be as high as possible and there should be a good spatial coherence of sound and visual scene on stage. While line arrays of loudspeakers are satisfying the requirements concerning sound pressure level WFS additionally preserves the whole auditory scene. For very large venues line-arrays can only control the sound pressure level within (sub-)regions of the reproduction space with problems at the cross-sections of neighboring regions. WFS is based on continuous sound fields. Therefor such problems can not occur with WFS. With WFS it is even possible to create an artificial room around the listening space with acoustical properties like indoors which is especially useful for classical concerts. Focused sound sources enable new creative possibilities for sound effects.

Cinema: For long time cinema sound systems have been screen centric: Most dialogs are just mixed to the center channel, because directors believed that sound which is not localized on screen distracts from the movie. Today almost all cinemas are equipped with sound systems with 5.1 rendering. The surround channels are used to create more immersion to the scene. Due to the fact that there are just two surround channels which are reproduced on the side via several loudspeakers the sound mixer has to fight several restrictions:

- at the back seats in the cinema surround channels sometimes are perceived as coming from the front side and might decrease the intelligibility of dialogs
- all sound sources mixed to surrounds are rather smeared over a large range
- moving sources from front (on screen) via side to the back change their sound color due to different frequency response of loudspeakers and phantom sources
- perception of sound very close to listeners is only possible by using very high sound pressure levels

Newer formats try to overcome some of these limitations by using additional channels, especially in the back.

WFS gives the possibility to render sound sources in their true spatial depth and direction for any seat in the cinema. In February 2003 the first cinema equipped with WFS system started daily service in Ilmenau, Germany (Fig. 3). Being a regular cinema in daily operation this cinema is also used for research, especially for the perceptual evaluation of loudspeaker configurations [14] algorithms in larger rooms. This cinema system uses 192 loudspeakers. In Figure 3 on top of the loudspeaker array the old 5.1 system can be seen. The theater is an amphitheater style cinema. To provide best audio experience for all seats the loudspeaker array is not in one plane but in two planes which are slightly folded. A trailer, produced in a WFS compliant format, exhibits the potential of the new technology: It starts within an aquarium, with air bubbles modeled as focused sources within the cinema hall, shows slowly moving sound objects leaving the screen, moving around the hall and appearing on screen exactly in audiovisual coherence, virtual two channel stereo (loudspeakers visible and moving on screen) and WFS music reproduction where each instrument is place on each own spatial position. In contrast to trailers for 5.1 formats this trailer does not need to be reproduced at a high sound pressure levels to create the sensation of immersion. The trailer is shown before every film show. All legacy format films benefit from the increased sweet spot offered by the compatible reproduction via the WFS system via virtual loudspeakers placed outside the cinema hall. The reproduction of the surround channels using plane waves improves the spatial sound quality and the intelligibility especially in the back rows. Perceptual experiments to study the performance of wave field synthesis audio reproduction combined with flat video reproductions have been conducted [15]. The results of these experiments already have been considered in the design of authoring tools. In July 2004 the first installation of a WFS system was done in Hollywood. This system, using 304 loudspeakers and a 8 channel WFS-subwoofer configuration, is integrated in a sound stage at Todd AO, Studio City. First experiences with remixing films and trailers show that the positioning of virtual sources anywhere at anytime brings more advantages than expected [16]:

Figure 3: *Cinema of Ilmenau*

- point sources are far less distracting than “old” surround channels
- dialogs moving slowly to the side of the screen are perceived as very natural, and do not draw the attention away from the screen
- no sound colorations appear with moving of sound sources
- Sound designers and mixers expect that the time and costs of mixing films using WFS are the same or lower compared to 5.1.

Home theater systems: Today WFS is regarded as too expensive and too space-consuming for the use in every home. However the installation in some higher rated home cinemas is already reasonable. The possibility to reproduce any content using virtual loudspeakers placed (far) outside the home theater enables to overcome the feeling of being in a small room, which is still inherent in home theaters today [17]. Problems for WFS in normal living rooms are the placement of the loudspeaker arrays and the acoustics of the room. For the later, a combination of acoustic treatment (e.g. curtains) and the application of room equalization techniques (e.g. compensation of a few early reflections) is probably the best solution. To solve the problem of loudspeaker placement new loudspeaker designs are necessary. Flat panel loudspeaker systems like DML panels [18] might play an important role in the home reproduction. Another possibility is to integrate them into furniture or even (far into the future) into wall paper.

Video Conferencing Today video conferencing is using mono or stereo reproduction of audio signals. If rooms with many persons are linked in a conference a lot of discipline is necessary to avoid interference between different persons talking, and to avoid background noise which reduces speech intelligibility. Multi-channel sound systems have not been used because of the problem of acoustic echo cancellation. It has been proven that perfect echo cancellation is impossible for the general case of two or loudspeakers/microphones. However in [19] a method has been presented which takes into account that the loudspeaker array used for WFS turns out to be of benefit for acoustic echo cancellation, too.

4.2. Some remarks on large listening area sound reinforcement

The effect of distance dependent reduction of loudness is more expressed near a sound source. An even distribution of loudness across the whole listening area can be achieved by positioning sound sources far behind the loudspeaker array [8]. Infinite distance of sound sources relates to plain waves which do not have any distance dependent reduction (besides the damping in air).

Due to the expectations of the visitors for the applications listed above a high sound pressure level is essential. In average all loudspeakers contribute for that level, but in the worst case, where a sound object is placed close to the loudspeaker array only a few loudspeakers have to provide the whole power. For most applications it is possible to overcome this by just avoiding such positions of sound objects. For the compatible reproduction in the cinema the worst case are the front speakers: To avoid incoherence of visual and auditory image it is essential to place virtual loudspeakers for left, center and right channels rather close behind the screen.

The possibilities of placing and moving sound sources anywhere in the sound scenes provide new artistic possibilities like musicians acoustically flying thru the audience before appearing on the stage.

5. RECORDING AND DISTRIBUTION OF WAVE FIELD SYNTHESIS CONTENT

The best sound experience using WFS can be achieved when using specially prepared material. Such material consists of dry recordings of separate sound sources, their position in the room and information about the desired room acoustics (e.g. recording room). Using microphone array technique recording of sound sources requires subsequent signal processing. By means of signal processing, sound source signals can be separated and unwanted signals can be suppressed. In addition, information about the position of possibly moving signal sources is extracted [20]. Besides the microphone array technique conventional 5.1 recording technique (spots, main and room microphones) can be also applied.

Audio information (recorded or synthetic sources and/or room acoustics) and scene description are treated inside the WFS system on the reproduction side. The number of transmitted audio tracks (either point sources or plane waves) is related to the scene and independent from the number of loudspeakers at the reproduction site.

The necessary storage capacity for a 2 hours movie can be estimated as follows: During the mixing process all sound tracks are stored as PCM using 24 bit at 48 kHz resolution. A reasonable film might be composed of 130 sound tracks in the final mix[1]. This results in a total storage requirement of 125.5 GByte, an amount which easily can be stored on state of the art PC hard drives, but which is beyond the current capacity of cheap magneto-optical storage media (like DVD-ROM). For end-user applications a reduction is necessary. As a first attempt perceptual audio coding can be used: MPEG-4 AAC (Advanced Audio Coding) at comparably high bit-rates (2 bit/sample per channel) achieves a reduction of the combined audio data to about 10.5 GByte (data rate of 12 Mbit/s). By using just some more compression or a slightly lower number of independent sound tracks, current DVD-ROM technology is adequate to provide the audio and metadata (source position information) to control WFS rendering.

For the audio scene description the MPEG-4 standard is very suitable. MPEG-4 is actually the only standardized format that provides a high-level structured coding support to efficiently convey advanced 3D descriptions [21] as those required by WFS. Together with wide band transmission channels, like DVB or the upcoming wide-band Internet, the MPEG-4 3D Audio Profile permits a commercially feasible realization of WFS.

After decoding the final auralization processing is left to the WFS loudspeaker arrays.

5.1. Channel-oriented versus object-oriented

Current sound mixing is based on the channel or track paradigm, i.e. the coding format defines the reproduction setup. Any changes would mean doing the complete mix again. In the current mixing process there is a certain way of arranging tracks for a mix, following the requirements of a mixing desk, routing system and the format (like 5.1 or 7.1) in order to accelerate the work-flow. Looking at the example of mixing a helicopter flying around the listeners head, it would be necessary panning and placing each element from the track system individually or at least copying and pasting the related settings on the desk. This is a time-consuming task, but due to the 5.1 or 7.1 mixes this hardly happens because movements are quite rare in those formats.

The mixing process of the Wave Field Synthesis occurs in a sound object-oriented way. For this the sound source positions are needed. Tracks and channels, which are indirectly considered in the process, form an object and this can be moved in a Wave Field Synthesis authoring system. The position data can either be imported from a tracking system (virtual studio), rendering data (special effects) or manually imported using a pencil. The final WFS mix does not contain loudspeaker related material. Audio signals of all sound sources are transmitted from the final mix to the WFS render PCs, which calculate the signals for all loudspeakers.

The object oriented approach inherent in WFS and MPEG-4 enables additional functionalities: It is able to group sound objects and give (limited) access to some of the mixing parameters to the end user. Putting all dialogs in one group and all remaining objects in a second group enables hearing impaired people to improve the intelligibility by increasing the level of the dialogs only. Keeping the room information separately from the musical objects enables the listener to put an orchestra into another concert hall. Putting sound tracks of different dubbed version in one bit stream enables multilingual listeners to hear all characters in a movie in their original language, and to replace only unknown languages with the preferred dubbed version.

[1] The number of raw tracks is much higher. The normal work flow of film mixing involves the reduction to so-called “stems”. The number 130 mentioned above is equivalent to the number of all channels in all stems used for the final mix. However the work flow for reduction of tracks for WFS is different.

6. CONCLUSIONS

The next generation of audio formats will be object oriented, and Wave Field Synthesis will be used for the reproduction replacing stereo and multi-channel systems. It will find its way into many applications, like concerts, cinemas, theme parks and, eventually, into the home. After long years of research, computational complexity is no longer an obstacle for wide spread adoption of WFS. First professional installations have been already made, and more will come in the next year.

7. ACKNOWLEDGMENTS

The author would like to thank all the many contributors to this work within Fraunhofer, at Delft Technical University and the other CARROUSO partners. The work at Fraunhofer IDMT was partly funded by the Thuringian Minister of Science, Research and Culture TMWFK, the German Minister of Education and Research BMBF and by the European Commission in the context of the CARROUSO project.

8. REFERENCES

[1] Alan Blumlein, "Improvements in and relating to Sound-transmission, sound-recording and Sound-reprocuction Systems," in *Patent No. 394325*, December 14th, 1931.

[2] J. C. Steinberg and W. B. Snow, "Auditory Percepective - Physical Factors," *Electrical Engineering*, vol. 53, pp. 12–17, 1934.

[3] W. B. Snow, "Basic Principle of Stereophonic Sound," *Journal of SMPTE*, vol. 61, pp. 567–589, 1953.

[4] Tomlison Holman, "The number of loudspeaker channels," in *AES 19th Conference*, Elmau, 2001.

[5] S. Brix, T. Sporer, and J. Plogsties, "CARROUSO - An European Approach to 3D Audio," in *110th AES Convention*, 2001, preprint 5314.

[6] A. J. Berkhout and D. de Vries, "Acoustic holography for sound control," in *86th AES Convention*, 1989, preprint 2801.

[7] Christan Huygens, *Traité de la lumière. Où sont expliquées les causes de ce qui lui arrive dans la réflexion, et dans la réfraction. Et particulièrement dans l'étrange réfraction du cristal d'Islande. Par C.H.D.Z. Avec un discours de la cause de la pesanteur*, Pierre van der Aa, Leiden, 1690.

[8] D. de Vries and P. Vogel, "Experience with a sound enhancement system based on wavefront synthesis," in *105th AES Convention*, 1993 October, preprint 3748.

[9] G. Theile, H. Wittek, and M. Reisinger, "Potential wave-field synthesis applications in the multichannel stereophonic world," in *AES 24th Conference*, Banff, Canada, 2003.

[10] D. de Vries, "Sound Reinforcement by Wave-field Synthesis: Adaptation of the Synthesis Operator to the Loudspeaker Directivity Characteristics," *Journal of the AES*, vol. 44, no. 12, 1996.

[11] Sascha Spors, A. Kuntz, and Rudolf Rabenstein, "An Approach to Listening Room Compensation with Wave Field Synthesis," in *AES 24th Conference*, Banff, Canada, 2003.

[12] Herbert Buchner, Sascha Spors, and Rudolf Rabenstein, "Efficient Active Listening Room Compensation for Wave Field Synthesis," in *AES Convention*, 2004, preprint 6119.

[13] Jérôme Daniel, Sebastian Moreau, and Rozenn Nicol, "Further Investigations of High-Order Ambisonics and Wavefield Synthesis for Holophonic Sound Imaging," in *AES Convention*, 2003, preprint 5788.

[14] T. Sporer and B. Klehs, "Wave Field Systhesis in the real World - Part2: in the movie theater," in *AES Convention*, 2004, preprint 6055.

[15] F. Melchior et al., "Wave-field-synthesis in combination with 2d video projection," in *AES 24th Conference*, Banff, Canada, 2003.

[16] Scott Martin Gershin and Stanley Johnston, "Question and answer session," Wave Field Synthesis Presentation, July 22nd, 2004, Studio City.

[17] B. Klehs and T. Sporer, "Wave Field Synthesis in the real World - Part1: in the living room," in *AES Convention*, 2003, preprint 5727.

[18] Ulrich Horbach, Diemer de Vries, and Etienne Corteel, "Spatial audio reproduction using distributed mode loudspeaker arrays," in *AES 21st Conference*, St. Peterburg, 2002.

[19] Herbert Buchner, Sascha Spors, and Walter Kellermann, "Full-Duplex Systems for Sound Field Recording and Auralization Based on Wave Field Synthesis," in *AES Convention*, 2004, preprint 6120.

[20] N. Strobel, S. Spors, and R. Rabenstein, "Joint audio-video signal processing for object localization and tracking," in *Microphone Arrays: Techniques and Applications*, M. Brandstein and D. Ward, Eds., pp. 197–219. Springer, Berlin, 2001.

[21] J. Plogsties, O. Baum, and B. Grill, "Conveying spatial sound using MPEG-4," in *AES 24th Conference*, Banff, Canada, 2003.