WAVE FIELD SYNTHESIS: FROM RESEARCH TO APPLICATIONS

Karlheinz Brandenburg, Sandra Brix and Thomas Sporer

Fraunhofer IDMT & Ilmenau Technical University
Langewiesener Str. 22, 98693 Ilmenau, Germany
Phone: +49 3677 69-4341, Fax: +49 3677 69-4399, E-mail: bdg@idmt.fraunhofer.de
Web: www.idmt.fraunhofer.de

ABSTRACT

Wave Field Synthesis can dramatically increase the quality of sound reproduction. To create new virtual rooms with realistic ambience within any listening space, new ideas for room simulation and virtual acoustics are needed. The work originally started at Delft Technical University (TU Delft) is now continued at a number of research locations. The paper will introduce the basic techniques, show the requirements for a number of real world applications and give an overview of current research topics. The applications scenarios include cinemas, concert halls and all kinds of music performances (indoor and outdoor).

1. INTRODUCTION

The history of spatial sound reproduction began originally with the stereophonic reproduction by using the concept of the acoustic curtain (many microphones wired 1:1 with many loudspeakers) at the Bell Laboratories. Research conducted by Blumlein resulted in channel reduction down to basically three channels, but due to long-time practical limitations only two channel stereo was applicable. Looking at further evolution from two-channel stereophony over quadrophony to 5.1, there are limitations which have not been overcome since the early days of Blumlein. The reproduction system quality depends strongly on the properties of the reproduced sound field and on psychoacoustic effects (phantom sources). Besides phantom sources and problems that come with them (no precise source positioning, no precise source localization, etc.) the well-known existence of a “sweet spot” represents a limit of best spatial impression and immersion in a reproduction room.

There have been many efforts to solve these problems especially at universities, but all these investigations did not achieve economic impact. In the late 90th the 3D audio profile of the MPEG-4 standard prepared the field for a significant step forward. Since the beginning of 2001 universities, research institutes and companies joined their efforts in the development of 3D audio. The EU-project called CARROUSO [1] has developed key technologies for recording, encoding, transmitting, decoding and rendering a sound field in an efficient way at highest perceived quality. Important key components for these technologies were the Wave-Field-Synthesis (WFS) as a new way of reproducing sound and MPEG-4 codecs. WFS was invented at the TU Delft in Holland and has been demonstrated in academic environments successfully in the past [2, 3]. Due to its high computational complexity it has not found broad application until today. The progress in microelectronics with decreasing costs of computing power made the first application in the professional market possible. In February 2003 Fraunhofer IDMT implemented a large WFS array in the Ilmenau cinema and new applications for WFS technology are around the corner.

2. BASIC THEORY

Wave-Field-Synthesis (WFS) is based on the wave theory concept of Huygens: All points on a wave front serve as individual point sources of spherical secondary wave fronts. This principle is applied in acoustics by using a large number of small and closely spaced loudspeakers (loudspeaker arrays) (Figure 1). Each loudspeaker in the array is fed with a corresponding driving signal calculated by means of algorithms based on the Kirchhoff-Helmholtz integrals and Rayleighs representation theorems [4].

$$P_A = \frac{1}{4\pi}\int_S \left[\left(P\frac{1+jkr}{r}\cos\varphi\,\frac{\exp(-jkr)}{r}\right)+\left(j\omega\rho_0 V_n\frac{\exp(-jkr)}{r}\right)\right]dS$$

The superposition of the sound fields generated by each loudspeaker composes the wave-field. This technique enables an accurate representation of the original wave-field with its natural temporal and spatial properties in the entire listening space.

By means of WFS virtual sound sources (point sources) can be placed anywhere in the room, both behind the loudspeaker arrays as well as inside the room. WFS is also capable of reproducing plane waves. An important property of plain waves is that the sound pressure level approximately does not decrease along the propagation direction.
Natural sound fields in general are composed of the sound fields of each individual sound source and the room impulse response. In WFS individual sound sources are modeled by means of point sources. The acoustical properties of the reproduced sound scene can either be those of the recording room, those of a prerecorded different venue or obtained from artificial room models. It has been shown that these properties can also be reproduced using a small number of plane waves [5] (Figure 2).

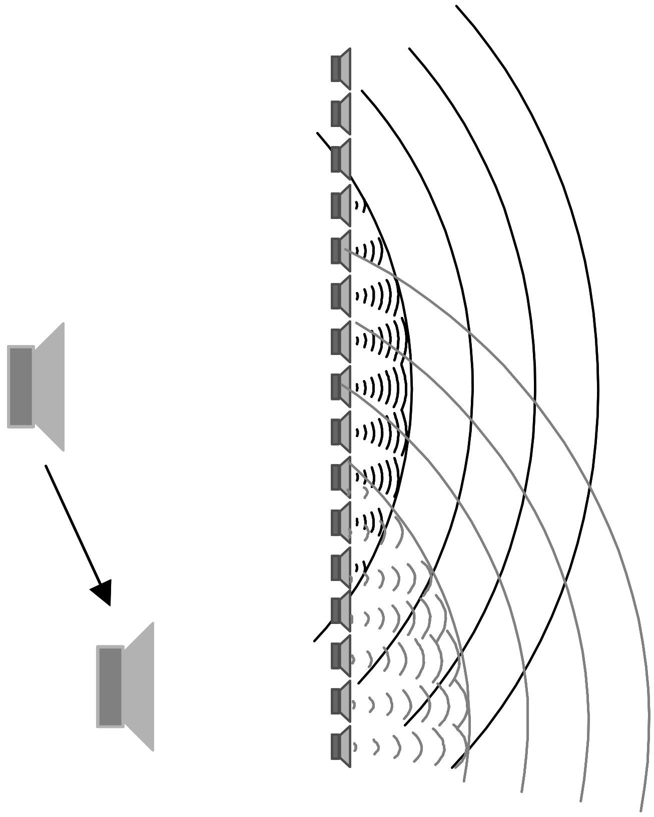

Figure 1: Wave-Field-Synthesis based on the wave theory. Virtual sources can be placed anywhere.

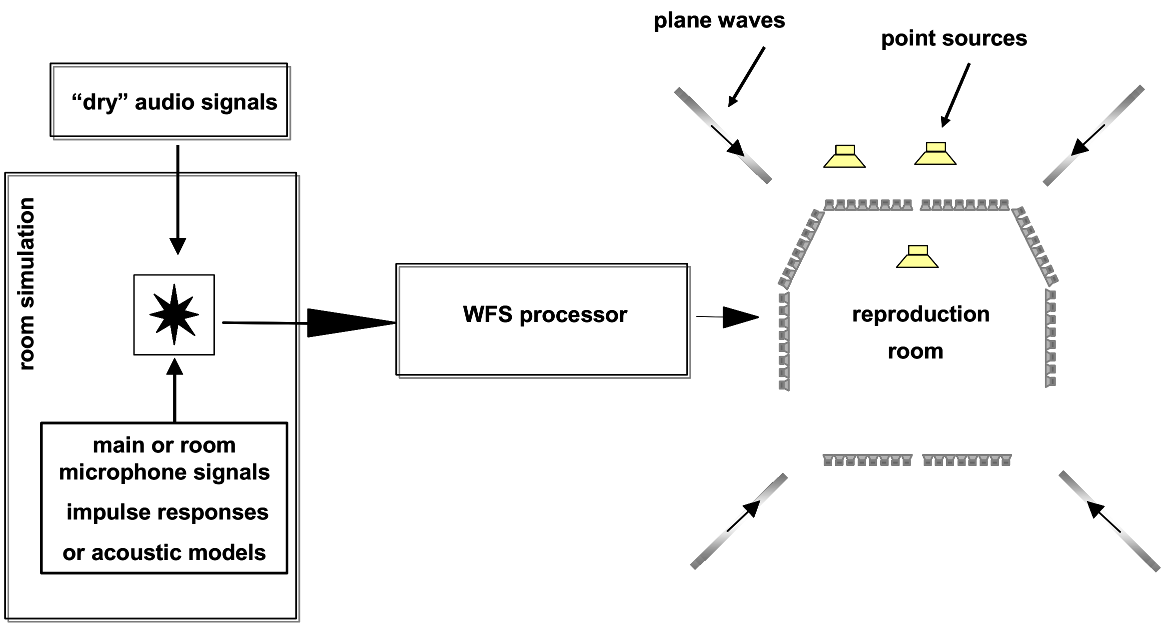

Figure 2: Workflow of creating or recording signals for WFS reproduction.

WFS can easily be combined with room equalization: To lower the effect of the actual listening space on the perceived sound, partial cancellation of early reflections can be used.

3. REALIZATION

Using WFS it is possible to treat signals coming from sound sources separately from signals coming from the room. Thus it is possible to manipulate signals from sources and room independent of each other. One example would be choosing between different concerts and playing them in different concert halls.

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 [6]. Besides the microphone array technique, the conventional 5.1 recording technique (spots, main and room microphones) can be also applied.

The audio information (recorded material or synthetic sources and/or room acoustics) and the 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 side.

The necessary storage capacity for a two hour movie can be estimated as followed: In a first version 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. 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 broad 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 the 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 [7] 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 array.

3.1 Channel-oriented versus object-oriented

Current sound mixing is based on the channel or track paradigm, i.e. the coding format defines the reproduction set-up. 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 workflow. 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 (Figure 3). 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 rendering PCs, which calculate the signals for all loudspeakers. This process is explained in more details in chapter 4.

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 versions in one bit stream enables multi-lingual listeners to hear all characters in a movie in their original language, and to replace only unknown languages with the preferred dubbed version.

4. THE WFS SYSTEM

The WFS system is basically divided into authoring, processing and rendering components (Figure 4). The whole system is modularly designed. All components are integrated in a so-called Spatial Audio Workstation (SAW), which consists of an audio server module, a rendering module, a room simulation module and an user interface module.

A control unit which handles the communication between all modules, the system status and the scene description is part of the user interface module. The audio scene description can be stored in hard disk as uncompressed PCM-data or XML-format. Advantages in using XML are easily portability and flexible extensibility. A so-called XML-SAW format – optimized for WFS – was introduced for storing and exchanging of audio scene descriptions.

The task of the audio server module is to record a play-back audio data. All audio connections between these modules are routable by an audio routing matrix. The capability of input channels for virtual sources is 64. This allows the use of conventional mixing consoles for the pre-mixing of signals of virtual sound sources.

The rendering module is a scalable rendering cluster, which calculates WFS-data. This cluster receives virtual source signals including a channel oriented scene scheme and calculates loudspeaker driving signals as mentioned before in chapter 2.

The room simulation module works with artificial rooms as well as with rooms based on real acoustic measurements. Artificial rooms can be achieved by means of perception based parameters.

At the reproduction side (Figure 4) the reproduction system, i.e. loudspeaker arrays (so called loudspeaker panels – each panel is composed of 8 loudspeakers) and subwoofers are used.

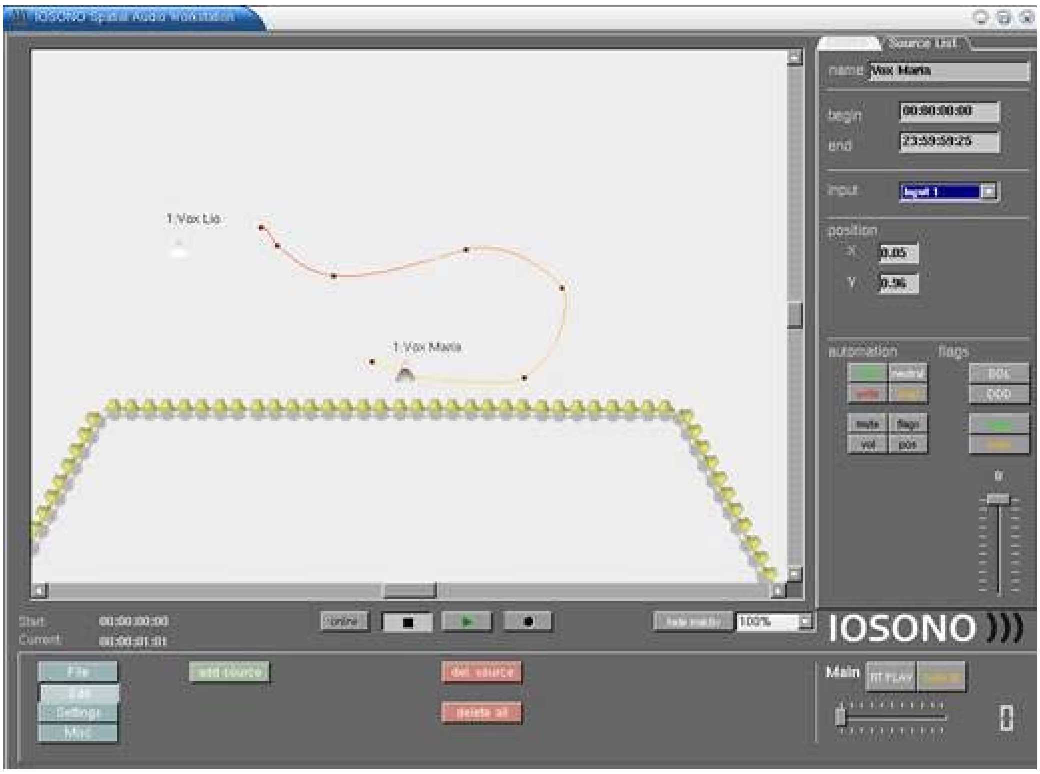

Figure 3: WFS authoring tool: Trajectory of only one sound object shown.

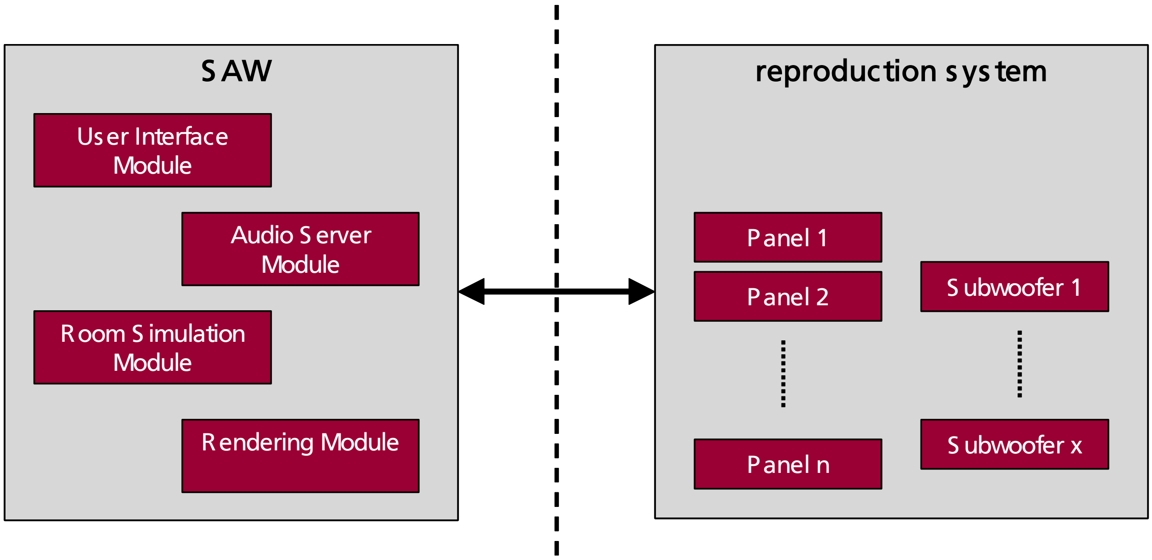

Figure 4: WFS system – authoring, processing and reproduction.

5. APPLICATIONS

Over the long run, WFS and mathematically related sound rendering methods like higher-order ambiphonic reproduction will find its way to all sound reproduction systems where ever it is possible to use more than just one or two loudspeakers. The biggest single advantage of going from classical multi-channel to WFS, beyond the improvements in audio quality, is the paradigm shift from reproduction based audio storage (the format is defined by the number of loudspeaker channels) to source based storage (each audio object is stored separately and can be rendered for the best possible audio quality given any reproduction setup).

5.1 Application areas

Concert halls: The WFS algorithms exhibit intrinsic delay times short enough for life performances. If the acoustics of the concert hall are good enough, so that no room equalization filters are necessary, it is easy to accomplish. With WFS in multifunctional venues the optimum acoustics for each kind of music and other purposes like sports can be adapted. In contrast to the systems used today WFS can provide spatial angular and distance resolution of the acoustic scenes on stage. WFS can make electronically amplified audio sound much more natural.

Open air events: Key requirements for open air concerts are equal distribution of the sound pressure level across the whole listening area and spatial coherence of sound and visual scene on stage. While line arrays of loudspeakers can only satisfy the first requirement, WFS can do both. Optionally it is possible to create an artificial room around the listening area with acoustical properties like in-doors (esp. useful for classical concerts) and to place sound effects even inside the listening space. While line-arrays control the sound pressure level in specific regions this is generating problems at the cross-sections of neighbouring regions. Such problems can not occur with WFS because it is based on continuous sound fields.

Cinema: In addition to an accurate representation of the original wave field in the listening room, WFS gives the possibility to render sound sources in their true spatial depth and therefore shows enormous potential to be used for creation of audio in combination with motion pictures. On February 19th, 2003 the first cinema equipped with a WFS system started daily service in Ilmenau, Germany (Figure 5). A trailer, produced in a WFS compliant format shows the potential of the new technology (Figure 6). This trailer plays excessively with the new possibilities of the media: air bubbles from within the aquarium modelled as point sources inside the cinema hall, slowly moving sound leaving the screen, moving around the hall and appearing on screen exactly in audio-visual coherence and music changing from two channel stereo to an exact positioning of each instrument. In contrast to trailers for 5.1 formats this trailer does not need to be reproduced at a high sound pressure level to create the sensation of immersion. The trailer is shown before every movie. All legacy format films benefit from the increased sweet spot when reproduced via the WFS system. The five channels are rendered by 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 [8]. The results of these experiments already have been considered in the design of authoring tools.

Figure 5: WFS system applied for cinema

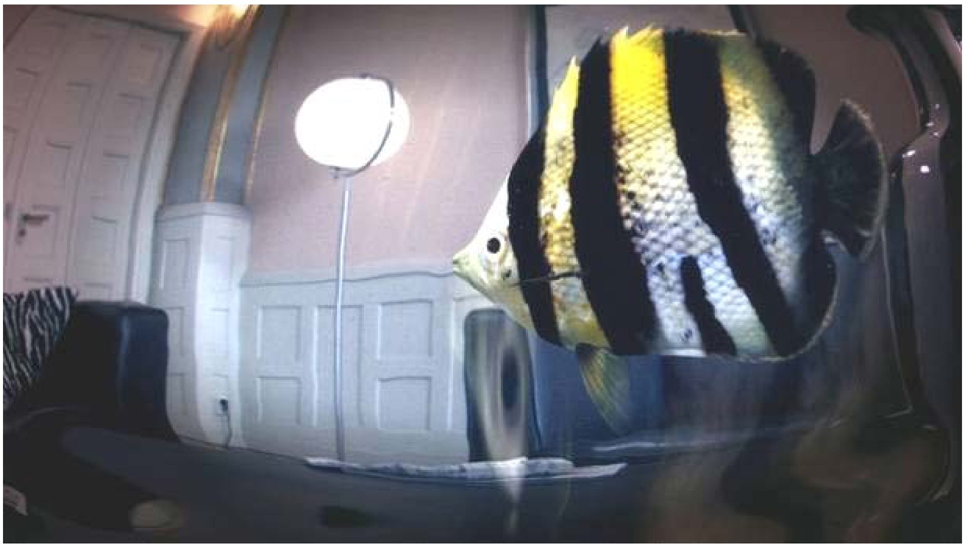

Figure 6: Cinema trailer mixed using WFS technologies

Home theatre systems: While a very expensive solution today, WFS will eventually make its way into home theatre systems. WFS can provide a very natural sound experience to smaller rooms, too. Problems regarding WFS in home-installation are the placement of the loudspeaker arrays and the acoustics of the room. For the latter, 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. For the problem of loudspeaker placement, the proposed solutions all will show up in the market place only some years from now. Flat panel loudspeaker systems like DML panels [9] might be part of the WFS equipped home theatre of the future. Another solution to hide loudspeaker arrays from visibility is to integrate them into furniture or even (far into the future) into wall paper.

5.2 Specifics of large listening area sound reinforcement using WFS

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 arrays. 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, centre and right channels rather close behind the screen.

The potential of placing and moving sound sources anywhere in the sound scenes provides new artistic possibilities like musicians acoustically flying thru the audience before appearing on the stage. WFS automatically generates Doppler effects which might cause unwanted effects when used during music performances.

6. CURRENT RESEARCH TOPICS

While WFS systems are ready for widespread applications, a number of research topics still remains. They cover the whole chain of signal processing, from microphone arrays for recording dry sources suitable as audio objects to novel room simulation algorithms to room equalization and changes to the rendering algorithms. While there is still some work left on basic theory, a lot of the current research topics are application driven.

6.1 Acoustic echo cancellation for WFS

If WFS is used in a communications setup (e.g. high quality video conferencing), AEC (Acoustic Echo Cancellation) it is a necessary part of the system. There is still work going on regarding new methods for AEC specially adapted to WFS.

6.2 Room simulation

For WFS rendering there is always the choice to use data from the actual recording (e.g. surround sound microphones) or to completely synthesize the virtual listening room from an acoustic simulation. In the first case, the real surround sound information is placed into the listening room via plane wave audio objects. In the latter case, the room is simulated via either geometry information and ray tracing or mirror source methods or via perceptually motivated room simulation as defined within the MPEG-4 BIFS framework. For both methods, a lot of work to refine the systems and give best results within the constraints of real time rendering is going on.

6.3 Array equalization

According to the theoretical WFS driving function for the loudspeakers a correction filter must be implemented to get a flat frequency response of the system. Most theoretical papers only focus on virtual sound sources far behind the loudspeaker array. For this a 3dB per octave suppression of low frequencies has to be applied. Taking into account that for virtual source positions exactly on a speaker no frequency correction is necessary, practical implementations require an adaptation of the filter to other virtual source positions. For simple configurations, like rectangular arrays, closed solutions are possible. For real world applications (non-rectangular arrays, irregular gaps between loudspeakers) the problem is more complicated. Because of this investigations are focusing on measurement procedures and automatic calibration procedures. For cost-efficient realization it is essential that such equalization methods do not have to be “perfect” in acoustic sense but “only” have to be “perfect” in psychoacoustic sense. The performance of algorithms will therefore be evaluated with the help of listening tests.

6.4 Listening room compensation

If the acoustics of the listening room overlay the virtual acoustics of the simulated listening space, the sensation of immersion is greatly reduced. One way to get around this problem is of course to use a dry, non-reverberant listening space. In cases where this is not possible or economically feasible, room compensation methods need to be applied. The search for the best compromise to do this is still on. From a psychoacoustic perspective it might be sufficient to cancel the first few reflections of the reproduction room which occur before the reflection of the room to be reproduced. While solutions to reduce reflections coming from the walls, which are equipped with loudspeakers, have been proposed, the reflections from roof and floor are still an unaddressed problem.

6.5 Complex scenes

WFS is ideal for the creation of sound for motion picture or virtual reality applications. In both cases the creation of highly immersive atmospheres is important to give the auditorium the illusion of being a part of the auditory scene. A new approach in designing immersive atmospheres (e.g. rain) using Wave Field Synthesis reproduction is being investigated. New tools and techniques to control and generate these atmospheres have been developed and investigated in listening tests.

6.6 Real 3D

A fundamental shortcoming of the system is that it only generates spatially correct wave fields in the horizontal plane: it is basically 2D[1]. An extension of WFS from 2D to 3D has to be investigated based on a multidisciplinary research program in acoustics, psychoacoustics and signal processing. The challenge is gaining new knowledge by researching and developing methods and tools for immersive, realistic 3D audio production and reproduction.

7. CONCLUSIONS

Wave Field Synthesis can be considered as the next revolution to audio reproduction after going to stereo and multichannel systems. It will find its way into 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.

8. ACKNOWLEDGMENTS

The authors would like to thank all the many contributors to this work within Fraunhofer, at Delft Technical University and many others. 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.

[1] In that sense two and 5.1 channel stereo can be regarded as 1½ D reproduction.

REFERENCES

[1] Brix S., Sporer T., Plogsties J.: CARROUSO – An European Approach to 3D Audio presented at the AES 110th convention, 2001 May, preprint 5314.

[2] Berkhout, A.J.; de Vries, D.; "Acoustic Holography for Sound Control" presented at the AES 86th convention, 1989 March, preprint 2801.

[3] De Vries, D.; Vogel, P., "Experience with a Sound Enhancement System Based on Wavefront Synthesis" presented at the AES 50th convention, 1993 October, preprint 3748.

[4] De Vries, D.: "Sound Reinforcement by Wave-field Synthesis: Adaptation of the Synthesis Operator to the Loudspeaker Directivity Characteristics" J. Audio Eng. Soc., Vol.44, No. 12, 1996.

[5] Theile G., Wittek H., Reisinger M., "Potential Wavefield Synthesis Applications in the multichannel stereophonic world", presented at the AES 24th conference, Banff Canada, 2003 June 26-28.

[6] N. Strobel, S. Spors, R. Rabenstein: "Joint Audio-Video Signal Processing for Object Localization and Tracking", In M. Brandstein, D. Ward (Eds.) 'Microphone Arrays: Techniques and Applications', Springer, Berlin, 2001, pp. 197-219.

[7] Plogsties J., Baum O., Grill B., "Conveying spatial sound using MPEG-4," presented at the AES 24th conference, Banff Canada, 2003 June 26-28.

[8] Melchior, F. et al.: "Wave-Field-synthesis in combination with 2D video projection", AES 24th Conference, Banff, June 2003

[9] Ulrich Horbach, Diemer de Vries, Etienne Corteel: Spatial Audio Reproduction Using Distributed Mode Loudspeaker Arrays presented at the AES 21st conference, St. Peterburg, 2002 June.