Wave Field Synthesis and Object-Based Mixing for Motion Picture Sound

By Frank Melchior

Digital Cinema and 3D pictures are the latest innovations for cinemas. Audio reproduction on the other hand, is nearly the same as it was in the late 1980s, when discrete multichannel audio was introduced. Even uncompressed storage and distribution, as well as the addition of a few more channels, did not change things significantly. Wave Field Synthesis (WFS) can change the situation because it offers spatial audio as a complement to 3D pictures. Using WFS, it is possible to focus sound sources inside the audience area. The sweet spot limitation of traditional multichannel audio no longer applies. On the production side, a complete new tool set is available for sound designers and mixers, based on object-based audio post-production. Although the basic principles of Wave Field Synthesis have been well-known for a long time, recent developments have made it possible to bring this technology into today's theaters. This article describes the current state of WFS in motion picture sound, as well as its unique properties and the benefit of object-based reproduction in the audio post-production process.

INTRODUCTION

Since the early days of silent movies, sound in cinemas was considered the poor cousin of the picture. For a long time, mono sound reproduction was predominant. When superseded eventually by stereo there was, for the first time, a spatial sound experience albeit limited to a very small area of the audience. On the way from matrix formats to discrete 5.1 multichannel the situation improved by providing discrete digital storage of formats, which had existed in principle since the late 1970s.[1] Extra channels behind the screen and in the surround improved the situation, but did not solve the underlying key challenges/problems:

- The best spatial sound quality is limited to a small sweet spot in the center of the audience area.
- To achieve an exact reproduction, the speakers must be placed in exactly the same positions as during the mixing process. Only limited adaptations to the actual reproduction environment are possible because of the channel-based sound recording.
- Different quality exists between screen channels and surround channels.
- No true 360° panning is available.
- There is no stable source position, beside three loudspeaker locations behind the screen. Virtual sound sources cannot be placed in the auditorium.

In particular, the first two restrictions leave sound mixers with only basic spatial effects in order to ensure a comparable sound experience in nearly all venues and seats of the auditorium. Using a large number of speakers which surround the auditorium, Wave Field Synthesis technology recreates a plausible perceptual sound field without the earlier described limitations. Furthermore, the representation of audio content is no longer based on the concept of audio channels that correspond to specific loudspeaker locations. Instead, a scene description is used, consisting of audio objects with various properties such as location, volume, size, and the audio signal itself. The scene will then be rendered taking into account the actual sound reproduction system and the distribution of the speakers in real-time at the cinema. First installations in cinemas are now available in Germany and the U.S.

WAVE FIELD SYNTHESIS

At its core, Wave Field Synthesis is based on the Huygens-Fresnel principle. [This states that every unobstructed point on a wavefront acts, at a given instant, as a source of outgoing secondary spherical waves. The resulting net wave amplitude at any position in the scattered wave field is the vector sum of the amplitudes of all the individual waves.]. After early publications in the 1950s by Snow[2], the Delft University of Technology applied the principle of Huygens to sound reproduction by using arrays of closely spaced speakers in order to physically recreate an acoustic field consisting of so called virtual sound sources pure analytical. The scientific work in the early 1990s concentrated in an international consortium[3] further developed the basic technology close to an application stage[4]. This work led to a good understanding of the underlying principles.[5] In addition to the basic idea of sound field reconstruction, further research has brought WFS to the point where reproduction of an auditory scene for a large audience is now possible. This was achieved by integrating psychoacoustic knowledge into a pure physical approach of sound field reconstruction.

There were still challenges with the practical implementation of such a system in terms of perceptual optimization and sound quality. The focus of current developments is on state-of-the-art in

sound quality and bringing spatial quality and sound design possibilities to a new level.

From the point of view of the sound designer, most of today's WFS systems provide three distinct types of sound sources, as illustrated in **Fig. 1**:

- Point Sources which are placed outside the array of loudspeakers, and are perceived by all listeners to originate from the same location.
- Plane Waves which can be interpreted as point sources having an infinite distance from the listener. The result is a constant direction for all listeners.
- Focused Sources which are point sources within the loudspeaker array.

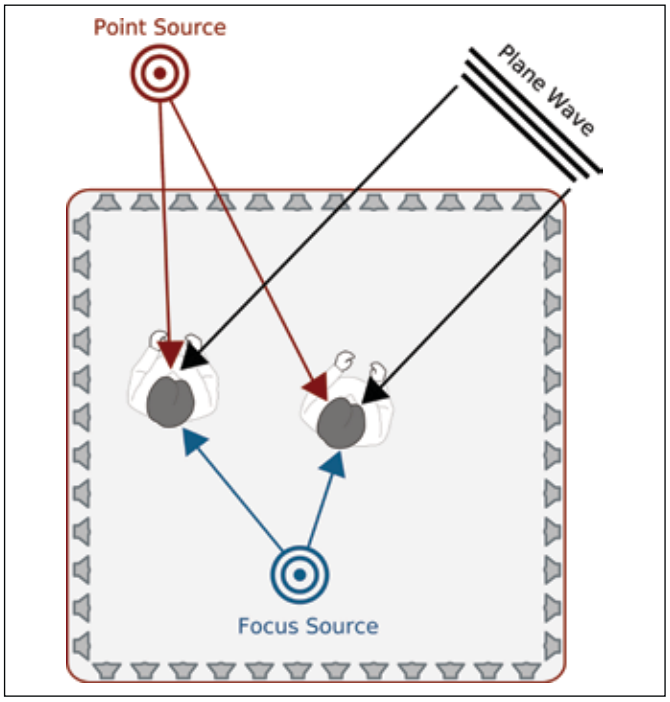

Figure 1. Illustration of different source types in current WFS systems. The perceived direction is shown for 2 listeners. A point source is perceived with a constant position in the complete audience area. A plane wave is perceived with a constant direction over the listening area. A focus source, located inside the listening area, is perceived as a point source for listeners in the upper half of the audience area.

More source types are currently under development and will enhance the creative possibilities of such systems in the future. With these three different types of sound sources, complex audio scenes can be formed. Creatives can use the new potential of WFS to bring the sound experience for the listener to a new dimension. In contrast to current surround 5.1 multichannel sound systems, a WFS system has the following properties:

- Good (spatial) sound experience for a large audience area.
- True 360° panning.
- Stable source position if desired.
- Stable source direction if desired.
- Focus sources inside the auditorium.
- Optimal adaptation to reproduction venues through object-based spatial high resolution masters.

Based on these properties, a system can be designed to take care of today's motion picture production and exhibition needs.

OBJECT-BASED AUDIO POST-PRODUCTION

Since the number of speaker channels depends on the size of the individual WFS installation and can be very high, mixing for a WFS system has to be independent from the number of loudspeaker channels used to reproduce the mix. Therefore, object-based mixing is used for WFS content generation.[6,7] In an object-based production, each audio source is treated as a separate audio object with associated properties like position, and source type, to mention only a few. The basic architecture of such an object-based mixing tool is described in the following section. The architecture and the basic building blocks of an object oriented mixing process are depicted in **Fig. 2**. The general system design is based on a stratified approach for sound spatialization.[8]

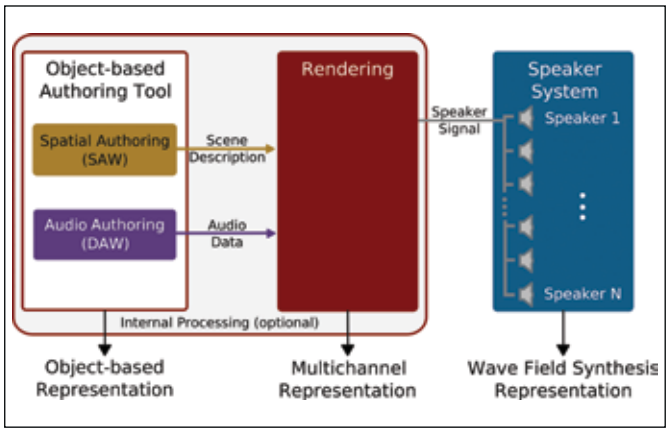

Figure 2. Block diagram of the object-based production.

For a better technical understanding in this section the post production process is divided into spatial authoring and audio authoring. From a user point of view this separation does not hold because of strong interaction between the spatial layout of an auditory scene and the properties of the underlying audio signals.

AUDIO AUTHORING

The audio authoring part provides all commonly known audio editing and arranging functionalities, which are part of modern digital audio workstations. The audio scenes consist of a multitude of single audio clips or audio tracks, which involve specific automation data. Within audio post-production for motion picture sound,

audio authoring is split into the sound editing and mixing process. These processes are divided into different sections as well, e.g., dialog, effects, music, etc.

SPATIAL AUTHORING

Spatial authoring enhances the audio data with additional spatial properties defined as sound sources. A sound source consists of the audio data from the audio authoring and the spatial properties from the spatial authoring tool. The spatial authoring module also provides functionalities to ease the spatial mixing process:

- Visualization of the sound source position.
- Automation and editing of spatial motion paths.
- Spatial audio effects based on source properties.

The visualization of the sound source positions allows the user to get an overview of the entire scene. The spatial authoring provides functionality to individually record/playback and edit spatial motion paths for each sound source. The timing of motion lines is easily editable to synchronize it with the corresponding audio material. To create complex audio scenes, spatial authoring enables the building of complex hierarchies of sound sources by grouping of sound sources.

INTEGRATION INTO THE CURRENT WORKFLOW

The functionality described in the previous sections is integrated into the IOSONO Spatial Audio Workstation (SAW). The SAW is currently available as a core plug-in in for the Steinberg NUENDO Digital Audio Workstation (DAW). **Figure 3** shows a screenshot of the SAW. The three main areas of the user interface are shown. **Figure 4** shows the stage view in detail. Several audio and group objects can be seen. Beside this the audience area and the screen are added. In this way a spatial orientation is possible without specific loudspeaker locations. The rendering algorithm takes care of the scene scaling on the reproduction side. This takes into account the real speaker position, as well as the audience area and the screen, thus guaranteeing the coherence of image and sound on the screen and the distribution of sources inside and outside the audience area regardless of venue. During the mixing process, the mixer can prepare the sound clips with the DAW or use it to record the final object-based master, utilizing a common mixing environment for all audio processing. During the post-production process, several outputs can be generated as indicated in **Fig. 2**. The most flexible output of the SAW is an audio scene file, which can be used for a variety of applications. It can be reproduced directly with a WFS system, or it can be distributed through a DCP Master and rendered at a WFS equipped theater.[9] The object-oriented mixing approach can also be used for the generation of many varieties of speaker/channel-based surround formats such as 5.1, or even stereo mixes. In the case of multichannel formats, the rendering can be performed directly in the SAW. Even in case of 5.1 production, this opens a future-proof mastering capability because any future format can be rendered directly out of the object-based scene description without up-mixing. An object-based scene description produced by a Spatial Audio Workstation is a high-definition (HD) spatial master. In this context it is important to notice that the rendering is always performed in real-time. As a result, all formats can be monitored in real-time during production.

Figure 3. Screenshot of the spatial audio workstation. The three main areas of the user interfaces are shown.

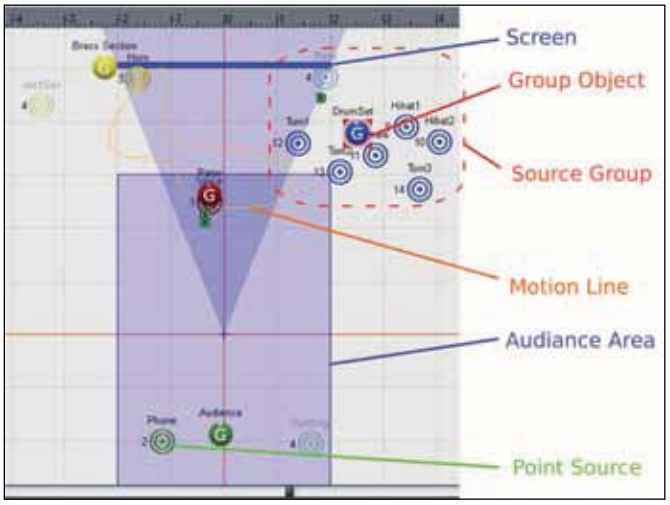

Figure 4. Detailed view of a visualization of an object-based scene description in the Spatial Audio Workstation.

ARCHITECTURE OF REPRODUCTION SYSTEMS

Figure 5 shows the basic architecture of a WFS system for sound reproduction in the digital cinema environment. The audio scene is stored on the digital cinema server and will be played out to the desired renderer synchronized to the picture using the digital audio interfaces. These signals are fed into the control PC. The WFS renderer analyzes the audio scene and generates the signal for each individual speaker of the reproduction system. For the calculation of the individual signals the parameters of the exact position of each individual speaker and the acoustics of the room are included. Afterward the signals are distributed to the individually driven speaker array. A comparison to the common channel-based

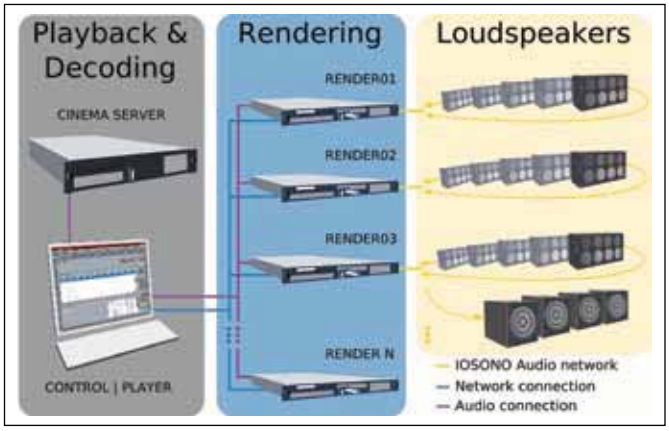

Figure 5. Block diagram of a WFS reproduction system.

structure can be found in **Fig. 6**. The important difference is the calculation of the speaker signals through the rendering process on the reproduction venue. In a discrete multichannel system, the rendering is simply performed by panning the audio signal to the desired speaker.

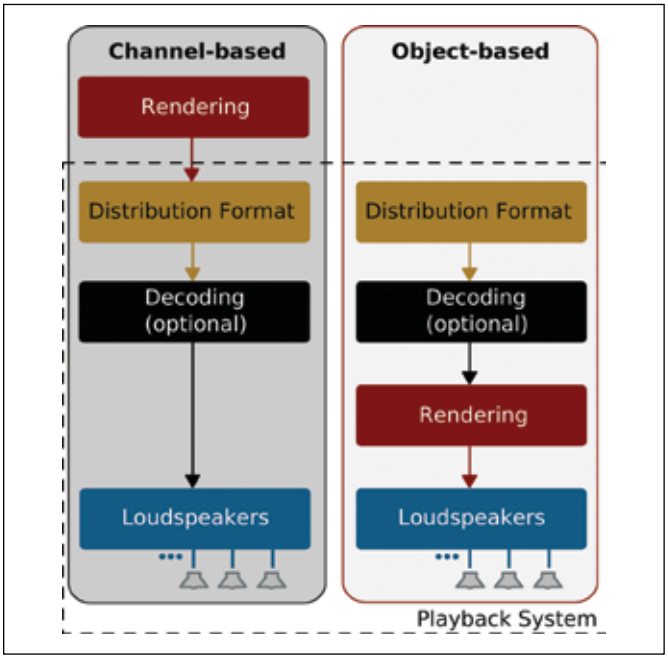

Figure 6. Comparison between channel-based reproduction and object-based reproduction.

Figure 7. WFS installation at Mann's Theater, Hollywood, CA.

REAL-WORLD IMPLEMENTATIONS

A recent IOSONO WFS installation is the Mann Chinese Theater Sixplex screen six in Hollywood, CA. This cinema has roughly 300 seats, with a stadium-like seating area. **Figure 7** shows the theater with the WFS system installed. The speaker panels are located on the left and right side of the room. There are speakers behind the screen and in the back as well. The WFS system consists of 380 discrete audio channels that are mechanically grouped into units of four audio channels, called a panel. The panels are equipped with active crossover and include an eight-channel class D amplifier. For signal distribution, a proprietary audio network is used. This network allows the transport of128 audio channels over a single Cat-5 cable, using a daisy chain configuration. All nodes in the speaker ring can be reached from both sides, which provides redundancy for a single node failure. Several of such rings can be combined to build larger systems.The rendering unit, based on a PC cluster, consists of units generating signals for 64 speaker channels. The object-based audio scene is transmitted from the digital cinema server to the control PC, which distributes the signals to the renderer PCs. Each renderer PC handles the signals for assigned speaker channels. The rendering is done in real-time with latency low enough to allow direct interaction for mixing purposes. During installation of the system, the position of each speaker of the system is exactly measured and fed into the rendering system. Afterward the system is acoustically optimized by defining filters for each speaker audio channel, which are processed in the rendering unit.

CONCLUSION

Wave Field Synthesis and object-based production offer new dimensions of sound for motion pictures. The creative potential allows mixers to create a more intense and immersive sound experience. Object-based production guarantees future-proof mastering of auditory scene data without the limitation of specific reproduction systems. Although an object-based description is required for WFS sound tracks, current and future multichannel setups can benefit from the universal high spatial resolution storage of auditory scenes. The complete production and distribution chain for the use of WFS in the cinema environment is available. First installations have been deployed with positive reactions from visitors. Wave Field Synthesis can be the perfect audio companion for 3D pictures.

ACKNOWLEDGMENTS

The author would like to thank Stefan Meltzer for writing the original manuscript published at IBC 2009 and all his colleagues for their fruitful comments on the early version of this article.

REFERENCES

1. Thomlinson Holman, *5.1 Surround Sound: Up and Running*, Focal Press: Burlington, MA, 2000.
2. William B. Snow, "Basic Principles of Stereophonic Sound," *J. SMPTE*, 61:567-587, Nov. 1953.
3. A. J. Berkhout, "A Holographic Approach to Acoustic Control," *J. Audio Eng. Soc.*, 36(12):977-995, Dec. 1988.
4. Thomas Sporer, Jan Plogsties, and Sandra Brix, "CARROUSO—An European Approach to 3D-Audio," presented at the 110th AES Convention, Amsterdam, The Netherlands, May 2001.
5. Sascha Spors, Jens Ahrens, and Rudolph Rabenstein, "The Theory of Wave Field Synthesis Revisited," paper presented at the 124th AES Convention, Amsterdam, The Netherlands, May 2008.
6. Stefan Meltzer, Lutz Altmann, Andreas Gräfe, and Jens-Oliver Fischer, "An Object Oriented Mixing Approach for the Design of Spatial Audio Scenes, presented at the 25th VDT Convention, Leipzig, Germany, Nov. 2008.
7. Frank Melchior, Thomas Röder, Stefan Wabnik, Sandra Brix, and Christian Riegel, "Authoring Systems for Wave Field Synthesis Content Production, presented at the 115th AES Convention, New York, NY, Oct. 2003.
8. Nils Peters, Trond Lossius, Jan Schacher, Pascal Baltazar, Charles Bascou, and Timothy Place, "A Stratified Approach for Sound Spatialization," presented at the SMC 2009, Porto, Portugal, July 2009.
9. Gabriel Gatzsche, Benoit Michel, Jerome Delvaux, and Lutz Altmann, "Beyond DCI: The Integration of Object Oriented 3D Sound into the Digital Cinema, presented at the NEM Summit, Saint Malo, France, Sept. 2009.

A contribution received by the author, March 2010. *Copyright © 2010 by SMPTE.*

Frank Melchior received a Dipl.-Ing. in media technology from Technical University Ilmenau. After finishing his thesis on "Wave Field Synthesis for Cinema Applications" in 2003, he worked as a researcher and project manager at the Fraunhofer Institute for Digital Media Technology. Melchior is currently finishing his Ph.D. project "Spatial Sound Design" at the University of Technology Delft, The Netherlands. Since 2009 Melchior has been director research and development at the IOSONO GmbH, Erfurt. He has published more than 20 papers in journal and conference proceedings. His research interests include spatial array signal processing, spatial audio reproduction, and user interface technology for audio systems. Melchior is member of the Audio Engineering Society (AES) and the Deutsche Gesellschaft für Akustik (DEGA).