AMBISONICS SYMPOSIUM 2009
June 25-27, Graz

SPATIAL AUDIO AUTHORING FOR AMBISONICS REPRODUCTION

Frank Melchior[1], Andreas Gräfe[1], Andreas Partzsch[1]

[1] IOSONO GmbH, Germany
Correspondence should be addressed to: frank.melchior@iosono-sound.com

Abstract: *The growing variety of spatial audio reproduction systems makes productions targeting specific loudspeaker layouts and the corresponding channel signals no longer feasible. One solution to this problem is the use of an approach that treats every audio source as a separate entity. As a result, the mix will be transformed into a spatial audio scene representation, which does not depend on a specific playback system. During production or distribution, the desired formats can be generated with appropriate reproduction modules. This contribution focuses on the spatial authoring of the mixing process, explaining the basic concept of the source oriented approach. An implementation of the concept as a software plug-in for an existing digital audio workstation utilizing first order Ambisonics reproduction is presented. The extension to higher order systems will be discussed.*

Key words: spatial authoring, data formats

1 INTRODUCTION

To describe an acoustic field in spherical coordinates, it is convenient to use spherical harmonics [1]. Sound field descriptions using spherical harmonics, also termed as Ambisonics, is well established especially in the context of electronic music. The basic idea is to approximate the acoustic field at a given point (the listener position) with a finite number of spherical harmonics. Spherical harmonics are elementary solutions of the wave equation in spherical coordinates. Such a representation can be used to describe divergent and/or convergent fields. Depending on the degree of approximation (order) low-order Ambisonics and high-order Ambisonics can be distinct. First or second order Ambisonics is often used as a reproduction system independent storage format. Several tools are available for real time audio processing environments like Max/MSP [2] and pure data as well as standalone applications using such environments [3]. Such systems are scene oriented and consist of multiple objects but lack intuitive editing capabilities and high-end processing. In professional audio production Ambisonics is well known, especially for the microphone technique [4][5]. A few mixing tools are available for the integration in digital audio workstations (DAW). An overview can be found in [6] or [7]. The concept of plug-ins allow to calculate low-order Ambisonics signals from mono or multichannel audio files. The low order field representation is fed through a multichannel bus to a decoder generating the desired speaker signals. This DAW approach suffers from difficult routing and channel based mixing by adjusting the parameters track-wise without having an overview of the complete scene and no logical representation of audio events in an auditory scene. Only scenes with first or second order can be handled. As a result, the integration is often not intuitive to handle by the user. For systems having the aim to reconstruct an acoustic field in a larger area like high-order Ambisonics (HOA), object-oriented formats are used in the authoring and reproduction process because of their flexibility and independency from the reproduction set-up. This formats can also be beneficial for spatial audio authoring and storage for low-order Ambisonics. This paper proposes a spatial audio authoring tool which integrates the object-oriented approach for Ambisonics reproduction and mixing. The integration of different reproduction formats into a concept of different abstraction layer is discussed. A concept for authoring low-order as well as higher-order representation has been developed and implemented based on a state-of-the-art digital audio workstation. The paper concentrates on the process of mixing, signal flow and work flow issues of spatial sound design. Without loss of generality all calculations and processing examples will be given using first order spherical harmonics. This kind of processing is implemented in a first version of a spatial authoring tool as a proof of concept.

2 AMBISONICS

Within this paper, the following definitions will be used:

1. Field representation (FR): Representation of an acoustic field in spherical harmonics up to a given order.
2. Source representation (SR): Representation of an acoustic source in spherical harmonics up to a given order.
3. Multichannel representation (MR): Signals used to play back via loudspeakers. This includes standard loudspeaker set-ups like stereo or 5.1 as well as rep-

resentations using a high number of channels.

4. Scene Representation: Representation of a acoustic scene consisting of the raw audio signals for each sound source and corresponding meta data describing the spatial layout and source properties.

2.1. Sound field representations

The derivation in this paper is limited to the two-dimensional case following [8],[9] and [10]. More recently, formulation of the theory and sound field description can be found in [11],[12].

The early derivations of Ambisonics theory founding the efficiently of such a system uses the assumption of superposition of plane waves. Sound sources far from the listener can be considered as plane near the listening point. In the two-dimensional case, a plane wave is characterized by its incidence angle according to the origin φ and its signal in the frequency domain $S_\varphi(\omega)$. The pressure of an acoustic field at the origin generated by a plane wave with the wave vector $\boldsymbol{k_\varphi}$ is denoted as $P_\varphi(\omega)$. The wave vector is given by

$$\boldsymbol{k_\varphi} = k \begin{pmatrix} \cos\varphi \\ \sin\varphi \end{pmatrix}, \tag{1}$$

where $k = \frac{\omega}{c}$. The temporal frequency f is represented by $\omega = 2\pi f$, c is the sound velocity. This leads to a pressure of the acoustic field $P_\varphi(\omega)$ in the origin given by

$$P_\varphi(\omega) = S_\varphi(\omega) e^{jk\cos(\phi - \varphi)}. \tag{2}$$

The plane wave can be expanded in terms of spherical harmonics [1],[13]

$$\begin{aligned} P_\varphi(\omega) &= S_\varphi(\omega) J_0(kr) \\ &+ S_\varphi(\omega) \sum_{m=1}^{\infty} 2j^m J_m(kr) \left[\cos(m\varphi)\cos(m\phi)\right. \\ &\left. + \sin(m\varphi)\sin(m\phi)\right], \end{aligned} \tag{3}$$

where J_0 and J_m are Bessel functions of the first kind. If the plane wave assumption holds for the loudspeaker signals at the listening position, each speaker will reproduce a signal according to (2) at the listening position in the origin of the coordinate system. The difference is that the speakers are not in angle φ but in angle ϑ_n, where N is the number of speaker [8]. The pressure $P_n(\omega)$ of the n-th speaker becomes

$$\begin{aligned} P_n(\omega) &= S_n(\omega) J_0(kr) \\ &+ S_n(\omega) \sum_{m=1}^{\infty} 2j^m J_m(kr) \left[\cos(m\vartheta_n)\cos(m\phi)\right. \\ &\left. + \sin(m\vartheta_n)\sin(m\phi)\right]. \end{aligned} \tag{4}$$

The complete field $P(\omega)$ is achieved by a superposition of all speaker signals in the origin:

$$\begin{aligned} P(\omega) &= \sum_{m=1}^{N} S_n(\omega) J_0(kr) \\ &+ \sum_{m=1}^{\infty} 2j^m J_m(kr) \left(\sum_{n=1}^{N} S_n(\omega)\cos(m\vartheta_n)\cos(m\phi) \right. \\ &\left. + \sum_{n=1}^{N} S_n(\omega)\sin(m\vartheta_n)\sin(m\phi) \right). \end{aligned} \tag{5}$$

To reproduce the original plane wave with the set of speakers one only needs to match the terms in the two equations (2) and (4). Since m goes to infinity for an exact reproduction, it is clear that the summation has to be truncated. The maximum value of m used in the calculation is termed as order. In the following, the order will be limited to first order in two dimensions. The required signals are represented by the following equations:

$$\begin{pmatrix} W(\omega) \\ X(\omega) \\ Y(\omega) \end{pmatrix} = \begin{pmatrix} S_\varphi(\omega) \\ S_\varphi(\omega)\cos\varphi \\ S_\varphi(\omega)\sin(\varphi) \end{pmatrix}, \tag{6}$$

where W represents the zero order component, X and Y represent the first order components. This representation is referred to as B-Format in case of a field representation. The n-th speaker signal can be given as

$$\begin{aligned} S_n(\omega) &= \frac{\beta}{N} W(\omega) + \\ &\frac{(1-\beta)}{N} \left[X(\omega)\cos(\vartheta_n) + Y(\omega)\sin(\vartheta_n)\right], \end{aligned} \tag{7}$$

using the concept of virtual microphones. The angle ϑ_n angle of loudspeaker n with respect to the origin and $\beta \in [0, 1]$ can be used to vary the characteristic of the virtual microphone between onmi ($\beta = 1$) and figure-of-eight ($\beta = 0$) Several schemes exist to optimize the result of the speaker signals in terms of perceptual and physical accuracy [10]. The process of generating the speaker signal is termed as decoding. Several publications discuss the optimal decoding for different speaker layouts. The reader is referred to [14], [15] and [10] for a detailed discussion of decoding.

2.2. High-order field representations

For high-order field representations, a huge amount of computational power is required. These can not be handled directly on the authoring workstation. A dedicated rendering PC cluster is required, which communicates with the authoring tool via network connection. The required processing chain is dicussed in Section 3.

2.3. Source representations

The outgoing field of a radiating source can also be represented in spherical harmonics. This was termed O-Format for low-order Ambisonics in [16]. The representation is equivalent to the incoming field represented using the B-Format and was extended more recently in [17]. The source representation can be used in an acoustic scene representation by decoding the radiated field for a given direction to the origin of the desired listening position. In case of a source representation a single audio stream is extracted which corresponds to the angle of incidence for the listening position. A detailed description can be found in [18]. By applying manipulation techniques, the source representation can be adapted in the spatial authoring process. Furthermore, room simulation applications all required directions for a room model can be extracted from the source representation.

2.4. Manipulating representations

If a field representation or a source representation should be included in an auditory scene the modification of the signals is desirable. For the spherical harmonic representation, the following manipulations are well known and implemented in several devices [19].

- Rotation in azimuth and elevation
- Mirroring the representation
- Dominance/Width

These modifications can easily be implemented in a matrix transforming the original signals $W(\omega)$,$X(\omega)$,$Y(\omega)$ to the transformed versions $W'(\omega)$,$X'(\omega)$,$Y'(\omega)$ using the dedicated matrix $\boldsymbol{M}_{(\cdot)}$ [19].

$$\begin{pmatrix} W(\omega) \\ X(\omega) \\ Y(\omega) \end{pmatrix} = \boldsymbol{M}_{(\cdot)} \times \begin{pmatrix} W'(\omega) \\ X'(\omega) \\ Y'(\omega) \end{pmatrix} \tag{8}$$

In case of a rotation in azimuth direction, the coefficients of the manipulation matrix $\boldsymbol{M}_{\alpha}$ become

$$\boldsymbol{M}_{\alpha} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(\alpha) & \sin(\alpha) \\ 0 & -\sin(\alpha) & \cos(\alpha) \end{pmatrix}, \tag{9}$$

where α denotes the rotation angle in azimuth direction. For the modification of the dominance, which is equivalent of using a different directivity of the W signal, the matrix $\boldsymbol{M}_{\Delta}$ is used. This process reduces the amount of energy from a certain direction. If the dominance is defined as the factor Δ the matrix becomes

$$\boldsymbol{M}_{\Delta} = \begin{pmatrix} \sqrt{2}\Delta & 1 & 0 \\ 1 & \frac{1}{\sqrt{2}}\Delta & 0 \\ 0 & 0 & 1 \end{pmatrix}, \tag{10}$$

to modify the front-back dominance. One can also think about extracting a single audio stream out of the field representation, which can be positioned as a single monophonic source in the further processing. This is equivalent to decode an Ambisonics signal but can also be defined as a virtual microphone at the origin of the field representation [20]. This is more adequate in terms of spatial sound design.

3 OBJECT-ORIENTED SPATIAL AUTHORING

In case of object-oriented sound design, the auditory scene is not produced for a specific multichannel representation (speaker layout). The mix will be converted into an abstract scene representation and is not bound to a fixed loudspeaker set-up or reproduction-technique. During playback, the scene representation will be adapted to the actual reproduction system using specialized rendering modules for each format. Figure 1 illustrates the basic components of an object-oriented production as well as the output of every processing stage.

3.1. Basic Concept

In object oriented mixing, each auditory event is represented by a sound source object. A sound source consists of the actual audio data and additional meta data describing the source properties. The sound engineer is able to manipulate the audio as well as the source properties of each sound source separately. Automations with respect to time can be applied, e.g. to create a moving sound source [21]. Due to the object-oriented representation, the sound engineer is able to create higher abstraction levels by grouping sound sources and edit these groupings as if they were single sound sources. For example, the spatial arrangement of the sources of a drum set (hi-hat, snare, cymbals, etc.) could be put together into one group object [22]. After finishing the mix, audio and descriptive data are exported into a object-oriented scene representation. This format needs to save the source properties aligned in time to the audio data to ensure consistency of audio playback and sound source position. The scene representation can be read by an arbitrary playback/rendering module. This module generates the actual speaker-wise audio data for the specific reproduction system or the desired field representation.

The main advantage of object oriented mixing is that an artist does not have to care about a specific reproduction situation. Instead, he can create and edit the auditory scene with all contained sound sources directly. The specifics of the intended playback system can be largely neglected and the artist can concentrate on the spatial composition. In comparison of using low order field representation an optimal reproduction quality can be ensured for the given rendering system. In summary, the mixing work flow for different reproduction formats is unified and specific playback system characteristics are hidden to some extent to the user.

3.2. System structure

Referring to Figure 1 the integral parts of a tool for object oriented spatial mixing are, Audio authoring module, Spatial authoring module, Rendering module. These parts will be discussed in detail in the following sections.

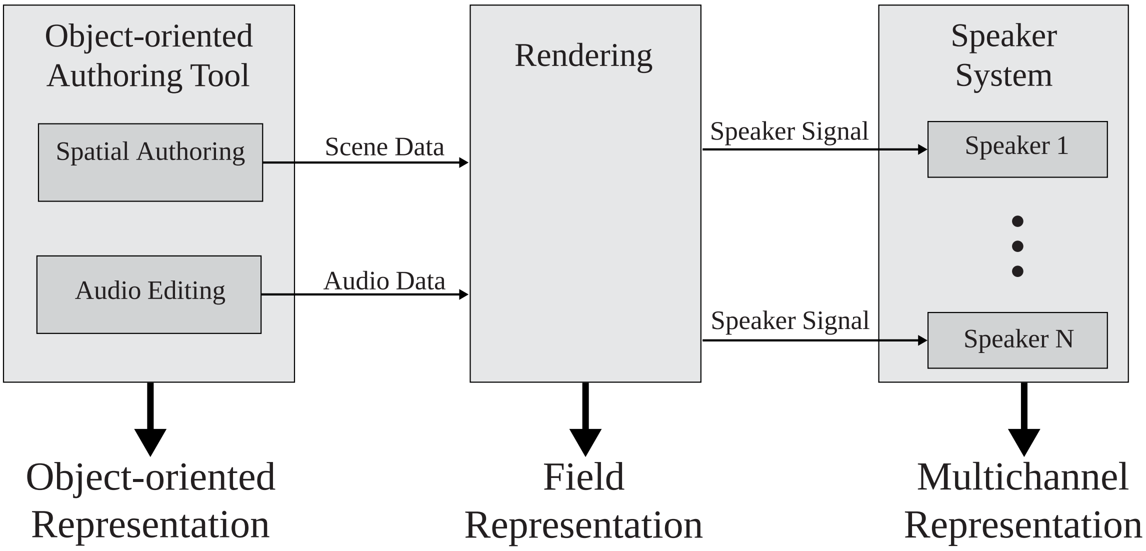

Figure 1: General signal and control flow of an object-oriented spatial authoring system.

3.2.1 Audio authoring

The audio authoring contains the audio editing and arranging functionalities, which are supplied by a digital audio workstation. Thereby, the audio scenes consist of a multitude of single audio clips which involve clip-specific meta data (automation).

3.2.2 Spatial authoring

The spatial authoring module communicates with the audio module. In this way, it connects the audio-clip related data with additional spatial properties to so called sound sources. Such a sound source consists of the audio-clip from the audio authoring module and the spatial properties from the spatial authoring tool. The spatial authoring module enriches the common mixing process with the following functionalities:

- Visualization of the sound source positions
- Editing/recording of spatial motion paths and timing adjustment.
- Creation of hierarchies to provide complex connected motion lines .

The visualization of the sound source positions provides the sound engineer with the possibility to easily check the spatial position of a single sound source and to get an overview of the entire scene. The spatial authoring module should furthermore provide the functionality to individually record/playback and edit motion lines (spatial motion paths) for each sound source. The timing of motion lines should also be easily editable to get in sync with the corresponding audio material. To create complex auditory scenes, the spatial authoring should also provide the functionality to build up complex hierarchies of sound sources. Therefore, single sound sources should be groupable.

3.2.3 Rendering

The rendering part is responsible for mapping the sound sources to specific playback-system dependent audio data. Therefore, it has to use both the pure audio from the audio authoring module and the corresponding scene description from the spatial authoring module. One possibility is to render the acoustic scene into a field representation. Another solution is to render the acoustic scene directly to a specific speaker layout for example using higher order Ambisonics. In the latter case, the rendering is normally done on a dedicate rendering system consisting of a plurality of audio processing devices.

4 AMBISONICS IN OBJECT-ORIENTED SCENE REPRESENTATIONS

In case of low-order Ambisonics, the rendering can be performed inside the object-oriented mixing tool. The output is a field representation, which can be exported or decoded directly into a specific speaker layout as part of the audio processing of a digital audio workstation. This scenario will be discussed in the following. Figure 2 shows a detailed view on the signal flow. The rendering block is implemented as part of the digital audio workstation using the audio processing based on plug-ins. These plug-ins are controlled by the spatial authoring tool. Depending on the input format coming from the audio editing stage, one can think of different source representations which are rendered to the field representation. Mono signals are represented by a single source and encoded into a field representation. In case of multichannel representations like stereo or surround, the signal are represented as a group of mono sources and can be treated as a separate mono sources from the signal processing point of view. For the user this sources are groups. If the multichannel representation is already coded into a field representation, the spatial authoring offers controls to modify this representation to fit into the current auditory

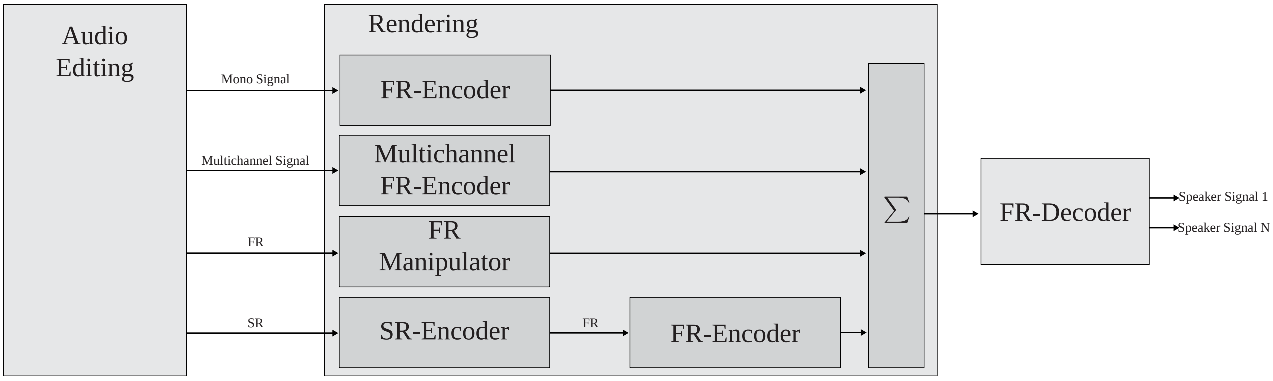

Figure 2: General signal flow for low-order rendering and processing using field representations (FR) and source representation (SR) data.

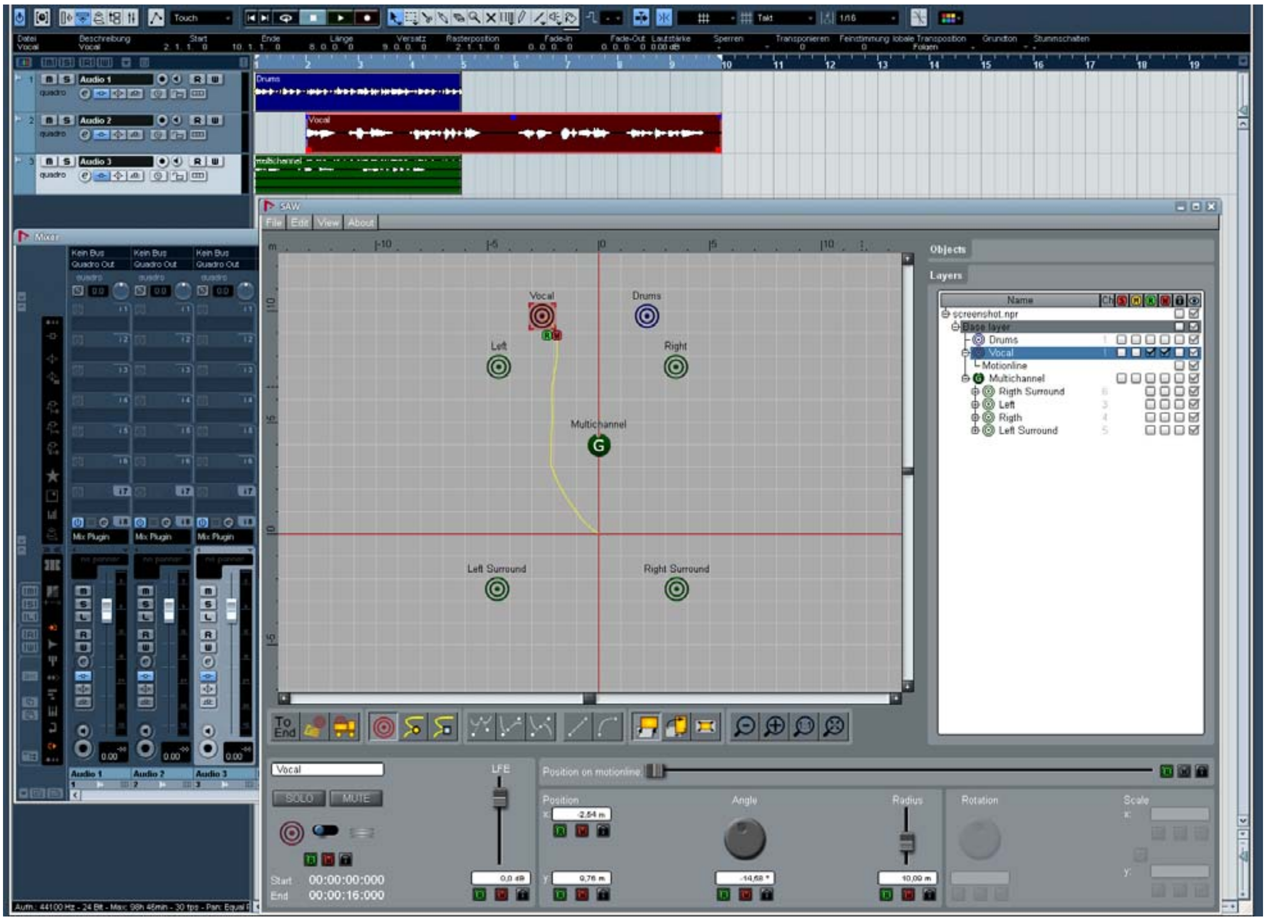

Figure 3: Realization of an object-oriented Ambisonics spatial authoring tool.

scene. The last possibility is a multichannel audio file consisting of a source representation. In this case, a single audio stream is extracted according to the desired position and orientation of the source. The resulting audio signal is encoded to a field representation afterwards. The use of an encoding plug-in as part of the processing of the DAW enables the direct monitoring of the mix in the given loudspeaker layout. This concept is realized in a prototype system. Figure 3 presents a screen shot of the system. By inserting a specific plug-in into an audio track this track can be used by the object-oriented authoring tool shown on the right side. Audio events in the track are automaticly represented as sources in the object-oriented authoring tool. These mono sources can be automated by having a view on the complete auditory scene. The audio processing of the events is performed in the track and a B-Format signal is the output of the track. These signals are summed and fed through a decoder plug-in into the master bus.

5 SUMMARY AND CONCLUSIONS

In this paper, a concept for the integration of Ambisonics authoring into a object-oriented spatial authoring tool was presented. Various representations of acoustic sources and fields have been discussed. Due to the integration into a reproduction system independent scene description, the sound designer is able to prescind from the specific reproduction techniques. Furthermore, the spatial arrangement of sources in a complete scene description corresponds well to the mental model of a sound designer during the spatial authoring process. Ambisonics can be used as source and field representation in the object-oriented mixing process to represent complex audio sources as well as pre-produced field representations from external sources or recordings. A well defined exchange format which can be embedded into an object-oriented scene description is highly desirable.

REFERENCES

[1] Philip M. Morse and K. Uno Ingard. *Theoretical Acoustics*. Princeton University Press, 1968.

[2] *Ambisonic externals for Max/MSP*. Internet resource, available at `http://www.grahamwakefield.net/soft/ambi~/index.htm` (accessed: 15th May 2009).

[3] Winfried Ritsch, Thomas Musil, Johannes Zmölnig, and Robert Höldrich. IEM Report 28/05 3D soundmixer. Technical report, IEM, 2005.

[4] *Soundfield microphone*. Internet resource, available at `http://www.soundfield.com/soundfield/soundfield.php` (accessed: 15th May 2009).

[5] *TerraMic*. Internet resource, available at `http://www.core-sound.com/TetraMic/1.php` (accessed: 15th May 2009).

[6] *Ambisonic tools overview*. Internet resource, available at `http://www.optofonica.com/kb/index.htm` (accessed: 15th May 2009).

[7] *Ambisonic Net*. Internet resource, available at `http://www.ambisonic.net/` (accessed: 15th May 2009).

[8] Jeffery S. Bamford and John Vanderkooy. Ambisonic sound for us. Presented at the 99th AES Convention, October 1995.

[9] Michael A. Gerzon. Periphony: With-height sound reproduction. *Journal of the AES*, 21(1):2–10, February 1973.

[10] Jerome Daniel, Jean-Bernard Rault, and Jean-Dominique Polack. Ambisonics encoding of other audio formats for multiple listening conditions. Presented at the 105th AES Convetion, September 1998.

[11] Sascha Spors and Jens Ahrens. A comparison of wave field synthesis and higher-order ambisonics with respect to physical properties and spatial sampling. Presented at the 125th AES Convention, October 2008.

[12] Franz Zotter, Hannes Pomberger, and Matthias Frank. An alternative ambisonics formulation: Modal source strength matching and the effects of spatial aliasing. Presented at the 126th AES Convention, May 2009.

[13] E.G. Williams. *Fourier Acoustics*. Academic Press, 1999.

[14] Michael A. Gerzon. General metatherory of auditory localisation. Presented at the 92nd AES Convention, March 1992.

[15] Martin Neukom. Decoding second order ambisonics to 5.1 surround systems. Presented at the 121st AES Convention, October 2006.

[16] Dave G Malham. Spherical harmonic coding of sound objects - the Ambisonic 'O' Format. Presented at the 19th International Conference: Surround Sound - Techniques, Technology, and Perception, June 2001.

[17] Dylan Menzies and Marwand Al-Akaidi. Ambisonic synthesis of complex sources. *JAES*, 55(10):864–876, 2007.

[18] Dylan Menzies. W-Panning and O-Format, tools for object spatialization. Presented at the 22th International AES Conference:Virtual, Synthetic, and Entertainment Audio, June 2002.

[19] D. G. Malham. Computer control of ambisonic sound fields. Presented at the 82nd AES Convention, March 1987.

[20] *Visual virtual mic*. Internet resource, available at `http://mcgriffy.com/audio/ambisonic/vvmic/` (accessed: 15th May 2009).

[21] 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 25. VDT Convention, 2008.

[22] Frank Melchior, Thomas Röder, Stefan Wabnik, Sandra Brix, and Christian Riegel. Authoring systems for wave field synthesis content production. Convention paper presented at the 115th AES Convention, New York, 2003.