

SPATIAL AUDIO AUTHORIZING AND RENDERING: FORWARD RESEARCH THROUGH EXCHANGE

Gabriel Gatzsche

Fraunhofer IDMT
Ehrenbergstraße 29
D-98693 Ilmenau, Germany
gze@idmt.fraunhofer.de

Frank Melchior

Fraunhofer IDMT
Ehrenbergstraße 29
D-98693 Ilmenau, Germany
mor@idmt.fraunhofer.de

ABSTRACT

Spatial audio authoring and rendering becomes more and more important for sound designers, musicians, artists but also for audio engineers, industry and entertainment. To answer and discuss the many interesting research questions in that field a common technical basis for spatial audio is required. Such a basis can only be provided by defining a standardized and well accepted *authoring format for spatial audio scenes* on the one hand and a *general audio rendering control protocol* on the other. From experience an effective introduction of such a format requires not only the discussion of the format's and protocol's semantic and syntactic elements, it is also required to think about a feasible way to standardize the outcome of this process.

1. INTRODUCTION

In [1] an ICMC panel discussion is proposed “with the intention to set on the development of a file format to create, store and share spatial audio scenes across 2D/3D audio applications and concert venues. This discussion shall include composers, sonic artists, researchers and developers in order to make such a format widely acceptable.” This paper explains background, attitude and goals of the authors' position.

2. POSITIONS

Fraunhofer IDMT strongly supports the idea of developing an interchange format for spatial audio scenes. The goal of such a process must be to bring object oriented spatial audio to a wide acceptance and to accelerate the answer of research questions like scene scaling, layering, rendering, etc.

It is not enough to discuss an authoring exchange format (incl. room simulation, dedicated spatial audio processing and user interfaces/ usability) only. It is also required to incorporate the rendering of spatial audio scenes with best possible reproduction quality using approaches like wave field synthesis (WFS) or higher order ambisonics (HOA). *Additionally* to the development of a authoring exchange format a *general rendering control protocol (GRCP)* has to be discussed. The goal of this protocol format is to connect different spatial audio authoring systems to different spatial audio rendering systems.

Based on MPEG-4 XMT a format for storing object oriented WFS-Scenes – called XMT-SAW – had been developed already in 2003 [2]. Furthermore Fraunhofer

can provide wide experience in the field of development, implementation and standardization of object oriented spatial audio formats [3], [4] as well as coding and compression of spatial audio (e.g. [5]). Furthermore Fraunhofer did many contributions in the development of spatial audio reproduction systems (e.g. [6]) but also in the development of appropriate authoring systems (e.g. [6], [7]).

Fraunhofer IDMT recommends to use one of the existing authoring standards (MPEG-4 BIFS [10], MPEG-4 XMT [11], AAF [8] / OMF [9], ...) as starting point for the discussion. These formats provide many concepts for object oriented spatial audio scenes, animation, interaction, linkage of audio streams etc. These standards are well documented and software development tools to process these formats already exist. The possibilities and drawbacks of these formats have to be discussed. Beside technical aspects the current state of the integration of these formats into existing audio products has to play a role.

The MPEG-4 object oriented audio tools (Audio BIFS and Advanced Audio BIFS) were not able to gain a high degree of acceptance. A reason for this can be found in the difficulty to implement all of the proposed audio nodes, missing or low quality software bindings, and the fact that there were some parts standardized at an early stage which where partly future subject. A trade-off between standardization and viable implementation has to be found. It should be paid attention that software tools are provided to the community so that everybody can easily use the data format.

Fraunhofer IDMT recommends the standardization of an interchange format for spatial audio scenes. But the standardization should not be done in advance or during the implementation and development of the appropriate reproduction systems. Experience shows that this can lead to a lack of acceptance and adoption of the standard. The proposed way would be to establish a basis standard. New description elements are provided to the community as user extensions. If a user extension has been discussed and widely accepted and adopted it is officially integrated into the standard.

3. FORMATS

The discussion of a common spatial audio authoring exchange format should start with a detailed review of existing work within that area. This chapter provides a short overview on existing formats that are related to spatial audio:

MPEG-4 XMT-O [11]: The Extensible MPEG-4 Textual Format XMT-O is an XML based high-level multimedia description language. It is closely related to SMIL. Fraunhofer used XMT-O as a starting point for the storing format of the Spatial Audio Workstation (SAW).

MPEG-4 XMT-A [11]: Like XMT-A is like XMT-O an XML based multimedia description language. XMT-A is more low-level oriented and incorporates descriptive elements that regard aspects of transmission and object identification.

MPEG-4 BIFS: The MPEG-4 Binary Format for Scenes BIFS [10] is a 1:1 binary representation of the XML based XMT-A. MPEG-4 BIFS make it possible to convey object oriented multimedia scenes using well organized small bitstreams.

OMF: "The OMF is a common interchange framework developed in response to an industry led standardisation effort." "Like Quicktime the primary concern of the OMF format is concerned with temporal representation of media (such as video and audio) and a track model is used." [9]

AAF: The Advanced Authoring Format AAF has been developed by a industrial consortium [8]. AAF is a file format that allows the exchange of different kind of essence data along with metadata. There are several professional AV workstations that adopt AAF.

X3D is the official successor of VRML [13]. The format targets the creation of 3D virtual worlds, games, visualizations and interactive learning applications in real-time [12].

SMIL [14]: "The Synchronized Multimedia Integration Language (SMIL, pronounced "smile") enables simple authoring of interactive audiovisual presentations. SMIL is typically used for "rich media"/multimedia presentations which integrate streaming audio and video with images, text or any other media type."

iXMF [15]: "The IASIG has developed, and continues to maintain, standards important to the interactive audio community, including the I3DL2 spec for interactive 3D audio, and the DLS family of file formats for musical instrument definitions."

Other references that deal with 3D Audio formats are [16], [17] and [18].

4. REFERENCES

[1] Gary Kendall, Nils Peters, Matthias Geier, *Towards an interchange format for spatial audio scenes*, Proposal for an ICMC panel discussion, International Computer Music Conference 2008, Belfast

[2] Katrin Münnich (Reichelt), *Untersuchung und Implementierung von Speicher- und Übertragungsformaten für Mischwerkzeuge der Klangfeldsynthese*, Diploma thesis, 2003, Technische Universität Ilmenau

[3] Plogsties, Jan; Baum, Oliver; Grill, Bernhard, *Conveying Spatial Sound Using MPEG-4*, 24th

International AES Conference: Multichannel Audio, The New Reality, Banff, Canada - June 26-28, 2003

[4] Dantele, Andreas; Schuldt, Michael; Reiter, Ulrich; Baum, Oliver; Drumm, Helge, *Implementation of MPEG-4 Audio Nodes in an Interactive Virtual 3D Environment*, 114th AES Convention, Amsterdam, The Netherlands - March 22-25, 2003

[5] Herre, Jürgen; Disch, Sascha, *New Concepts in Parametric Coding of Spatial Audio: From SAC to SAOC*, Proceedings, IEEE International Conference on Multimedia & Expo (ICME) 2007, July 2 – 4, 2007, Beijing, China

[6] Dausel, Martin; Deguara, Joachim; Gatzsche, Gabriel; Melchior, Frank; Reichelt, Katrin; Strauss, Michael, *Universal System for Spatial Sound Reinforcement in Theatres and Large Venues - System Design and User Interface*, 120th AES Convention, May 2006, Paris

[7] Brix, Sandra; Melchior, Frank; Roder, Thomas; Wabnik, Stefan; Riegel, Christian, *Authoring Systems for Wave Field Synthesis Content Production*, 115th AES Convention, New York, 2003

[8] AAF Association, *AAF An industry-driven open standard for multimedia authoring*, http://www.aafassociation.org/html/techinfo/aaf_dev_overview.pdf

[9] *Open Media Framework Interchange (OMFI) Format*, <http://www.cs.cf.ac.uk/Dave/Multimedia/node296.html>

[10] ISO/IEC, *Scene description (BIFS) and Application engine (MPEG-J)*, ISO/IEC, MPEG-4, Part11

[11] Michelle Kim, Steve Wood, Lai-Tee Cheok, *Extensible MPEG-4 textual format (XMT)*, Proceedings of the 2000 ACM workshops on Multimedia, Los Angeles, California, United States

[12] Web3D Consortium, "What is X3D?", <http://www.web3d.org/about/overview/>

[13] ISO/IEC 14772-1:1997 and ISO/IEC 14772-2:2004 — *Virtual Reality Modeling Language (VRML)*

[14] W3C, Interaction Domain, "Synchronized Multimedia Integration Language", <http://www.w3.org/AudioVideo/>

[15] Interactive Audio Special Interest Group, "Interactive XMF: File Format Specification", http://www.iasig.org/pubs/releases/pr_ixmf-public.shtml

[16] J. Herder. Sound Spatialization Framework: An Audio Toolkit for Virtual Environments. Journal of the 3D-Forum Society, Japan, 12(3):17–22, 1998.

[17] H. Hoffmann, R. Dachselt, and K. Meissner. *An independent, declarative 3D audio format on the basis of XML*. In ICAD, 2003

[18] G. Potard and I. Burnett. Using XML Schemas to Create and Encode Interactive 3-D Audio Scenes for Multimedia and Virtual Reality Applications. Lecture Notes In Computer Science, pages 193–203, 2002.