ASDF: AUDIO SCENE DESCRIPTION FORMAT

Matthias Geier and Sascha Spors
Deutsche Telekom Laboratories, Technische Universität Berlin
Ernst-Reuter-Platz 7, 10587 Berlin
Matthias.Geier@telekom.de

ABSTRACT

The Audio Scene Description Format (ASDF) is an collaboratively evolving format for the storage and interchange of static, dynamic and interactive spatial audio content. This position paper briefly describes the current status and raises a list of open questions which shall be addressed in the panel discussion.

1. INTRODUCTION

Nowadays, several high resolution spatial audio reproduction methods are available, for example Wave Field Synthesis and Higher Order Ambisonics. There are many practical implementations of these and other methods, and there are various venues dedicated to spatial audio reproduction. However, because of the lack of a common file format it is not easy to create content which shall be reproduced on several different systems.

Our contribution to the panel discussion *"Towards an Interchange Format for Spatial Audio Scenes"* is the presentation of a draft proposal for such a format. It is called *Audio Scene Description Format* (ASDF). A preliminary version is already in use within the *SoundScape Renderer* [1], a software for spatial audio reproduction developed at Deutsche Telekom Laboratories. The format is designed to be completely independent of both the audio reproduction algorithm and the implementation platform.

The ASDF is developed within the community of potential users. Everyone is invited to give suggestions and to discuss about any decisions to be made. The format will never be able to accommodate every single feature of every imaginable virtual scene, it should rather try to find its own path between simplicity and universality.

2. FEATURES

There are already several file formats for representing three-dimensional scenes including spatial audio information. Among them are VRML/X3D [3] and MPEG-4 BIFS [2]. However, those formats are mainly aimed at computer graphics and include much information which is not used for audio scenes, like lighting and textures. The ASDF is for audio only and should therefore be much easier to implement in a spatial audio system.

It supports three-dimensional scenes but because many current sound reproduction systems only work in the horizontal plane, it offers a simplified input mode for two-dimensional scenes. The format can be used for simple static scenes but it can also describe dynamic scenes with moving sources and continuously changing scene parameters. It will also support some form of user interaction (see section 3.5).

The ASDF is based on the *eXtensible Markup Language* (XML) [5], which means that scenes are described in a widely-used syntax and stored in plain text files. The files can be edited using a normal text editor or a dedicated XML editor. To include XML support in a software project is easy, because there are many software libraries, written in various programming languages, which facilitate the handling of XML files. Using XML also allows to add custom features to scene files for certain reproduction systems which are simply ignored by others. Audio data is not included in ASDF files but rather saved in traditional audio formats and linked to the scene file.

The syntax of the ASDF is inspired by the Synchronized Multimedia Integration Language (SMIL) [4]. It offers very powerful yet simple ways for synchronizing media, in our case input signals for virtual sources. Because SMIL can neither describe three-dimensional scenes nor trajectories therein, the syntax has to be extended. In the ASDF, source movements can be transparently synchronized with the input signals of the virtual sources and vice versa.

3. OPEN QUESTIONS

There are plenty. In the following paragraphs just a few are introduced briefly. These and many more shall be addressed in the panel discussion.

3.1. Scene Graph

Most 3D-modeling formats use a scene graph to store all objects of the 3D world in a hierarchical structure. This approach is very useful if a scene consists of many small objects which together form more complex objects which can again be combined to even more complex structures and so forth. Virtual audio sources, however, mostly consist of only one or very few objects, therefore it is questionable if a scene graph should be used for the ASDF. Furthermore, it should be easier for the author/composer to place and manipulate virtual sources without having to think about the hierarchical structure of the scene.

The very convenient grouping feature of scene graphs can be made up for by the implementation of a non-hierarchical grouping mechanism.

3.2. Virtual Room(s)

Room acoustics is very important if naturalistic scenes are to be reproduced plausibly. On the other hand, room information can be used in an exaggerated and un-natural manner to achieve certain artistic effects.

The problem with room acoustics simulation is that every technical implementation has an own set of parameters to manipulate acoustic properties of virtual rooms. These parameters depend on both the applied simulation algorithms and the actual implementation.

In general, there are two approaches for specifying room acoustics in a virtual scene. Firstly, an actual room (or several rooms) can be modeled virtually by specifying walls and other acoustic obstacles and their acoustic parameters like (frequency dependent) absorption, transmission and diffraction. Secondly, high level parameters, both instrumentally measurable and based on subjective perception, can be specified.

MPEG-4 Advanced AudioBIFS already specifies a plethora of such parameters like `reverbTime`, `reverbFreq`, `sourcePresence`, `envelopment`, `modalDensity` and many more [2].

It shall be discussed if one or both approaches should be included into the ASDF and if room acoustics shall be part of the core features of the format of if it shall be an optional extension.

3.3. Scene Scaling

As soon as one virtual scene shall be reproduced on more than one system, the problem of scene scaling arises. There are very different spatial reproduction setups from ordinary headphones to several-hundred-channel Wave Field Synthesis installations. A virtual scene which seems appropriate for headphone-based reproduction could be much too concentrated in space for a large auditorium.

How can we specify a scene that can be automatically scaled based on the system it is reproduced with?

3.4. Trajectories

Many of the now existing systems save dynamic content as a time-stamped series of parameter updates. Although this is fairly easy to implement it lacks flexibility when a virtual scene needs to be modified later.

The ASDF uses trajectories for source movements, rotations and other parameter animations. These trajectories can e.g. be scaled, transformed, repeated and concatenated. This allows for a more modular and more easily editable scene setup.

However, the exact syntax of trajectories still has to be discussed as well as the decision if they shall be piecewise linear or based on splines or similar smooth curves.

3.5. Interaction

As mentioned earlier, the ASDF shall not only be able to describe deterministic scenes but also allow interaction during the runtime of a virtual scene.

External events can be defined in the scene description and they can be used to start and stop soundfiles, trajectories and other scene elements. It shall be discussed how these external events can be triggered by software tools or hardware controllers.

3.6. Hardware/Software Connections

Source signals may not always be available as pre-recorded soundfiles. They might also be generated by real-time software or come directly from hardware interfaces.

A system- and device-independent way of specifying the various possibilities of sound input and output should be found in discussion with potential users.

3.7. Data Based Rendering

Most of the aforementioned features are dealing with model based rendering. However, one might want to include high resolution spatial recordings to a virtual scene, for example as an ambience track. It shall be discussed if it is feasible to include such data, possibly in the form of an Ambisonics B-format recording.

4. OUTLOOK

For an interchange format to be successful, it should serve the needs of as many potential users as possible. Therefore, everyone who is interested is invited to join the ASDF mailing list. Just write to the e-mail address given above to be added to the list.

5. REFERENCES

[1] M. Geier, J. Ahrens and S. Spors. "The SoundScape Renderer: A unified spatial audio reproduction framework for arbitrary rendering methods". In *124th Convention of the Audio Engineering Society (AES)*. Amsterdam, The Netherlands, May 2008.

[2] R. Väänänen and J. Huopaniemi. "Advanced AudioBIFS: Virtual acoustics modeling in MPEG-4 scene description". *IEEE Transactions on Multimedia*, 6(5):661–675, Oct. 2004.

[3] Web3D Consortium. *eXtensible 3D (X3D)*, 2004. `http://www.web3d.org/x3d/`.

[4] World Wide Web Consortium. *Synchronized Multimedia Integration Language (SMIL 2.1)*, Dec. 2005. `http://www.w3.org/TR/SMIL2/`.

[5] World Wide Web Consortium. *eXtensible Markup Language (XML 1.0, Fourth Edition)*, Aug. 2006. `http://www.w3.org/TR/xml/`.