Towards a Spatial Sound Description Interchange Format (SpatDIF)

Nils Peters[1], Sean Ferguson[2], Stephen McAdams[1]
[1]Music Technology Area - Music Perception and Cognition Lab
[2]Music Composition Area - Digital Composition Studios
Centre for Interdisciplinary Research in Music Media and Technology (CIRMMT)
Schulich School of Music, McGill University, Montréal*

1 Introduction

This paper presents the need for a Spatial Sound Description Interchange Format (SpatDIF) with the aim of creating, storing and sharing information for 3D audio applications.

2 Initial Problem

Our research includes a review of several spatialization algorithms and their implementations:

- Vector Based Amplitude Panning (VBAP), [7]
- (Higher Order) Ambisonics, e.g. [8]
- Spatialisateur (SPAT), [5]
- Space Unit Generator (SUG), e.g. [11]
- Virtual Microphone Control (ViMiC), [2]
- Wavefield Synthesis (WFS), e.g. [1]

We found that each spatialization algorithm uses a self-contained syntax and data format, wherein sequences of control messages (e.g. trajectories to move a sound in space) programmed for one application are incompatible with another implementation. Furthermore, like the recently opened *Allosphere*[1] in Santa Barbara, CA more and more concert halls and research facilities are equipped with a large number of loudspeakers. Although these venues might provide a good environment for 3D audio applications, they often differ in terms of room size, their technical specifications and the applied audio rendering concept. This lack of a standardized format for controlling spatialization across different rendering platforms and venues complicates the portability of compositions and requires manual synchronization and conversion of control data - a time consuming affair. Incompatible data formats also prevent collaboration between researchers and institutions. Therefore we call for developing of SpatDIF - a format to describe spatial audio information in a structured way to support real-time and non-real-time applications.

*Correspondence should be addressed to nils.peters@mcgill.ca
[1]http://www.mat.ucsb.edu/allosphere/

3 Implementation

According to the Sound Description Interchange Format (SDIF) [9], SpatDIF should have the following properties:

Platform independence Any 3D audio rendering algorithm on any computer platform should technically be able to understand SpatDIF;

Easily understandable syntax to prevent misunderstandings when stored data are shared;

Extendability Easy adding of descriptors to extend the specification especially as long as SpatDIF is in development;

Free and open source to increase the acceptance and widespread usage of this new format;

Easy to connect with interfaces, controllers and sensors for real-time control of spatialization, e.g. through GDIF [4];

Use of existing standards to focus on conceptual rather than technical development.

SpatDIF builds on the well-known Open Sound Control (OSC) address space standard [10] for real-time purposes. A SpatDIF OSC stream could also be stored as streams in audio sequencer programs, or in SDIF files (see Figure 1). For now, a SpatDIF interpreter receives and converts a SpatDIF OSC stream into the low-level data format according to the applied spatialization algorithm, but we hope that developers will integrate SpatDIF directly into their software.

3.1 Example for SpatDIF

Changing the position of virtual sound sources is the most common controllable parameter in spatialization applications. We briefly demonstrate SpatDIF by using this parameter as an example. There are several ways a position in space is defined:

1. Perspective: a position can be defined either in allocentric (relative to an external framework) or in egocentric (relative to the listener) perspectives;

2. Coordinate system: two or three dimensional, cartesian or spherical coordinate systems;
3. Absolute or relative (normalized) parameter values.

We propose an egocentric and normalized coordinate system as used in VBAP and several Ambisonic implementations. A control message for describing the position of sound source Nr.3 could be defined in cartesian (1) and spherical (2) coordinates. It should also be possible to change just a single coordinate dimension, e.g. the azimuth component in absolute values (3), or in incremental/decremental steps (4):

```
(1) /SpatDIF/source/3/xyz -0.5 0.5 0.0
(2) /SpatDIF/source/3/aed -45.0 0.0 0.0
(3) /SpatDIF/source/3/a -46.0
(4) /SpatDIF/source/3/a:+
```

If absolute positions are defined (as in [11]), a conversion factor needs to be initialized to rescale normalized coordinates. Initial parameters that are necessary for setting up the spatialization algorithm as well as other descriptive meta-information, should also be storable.

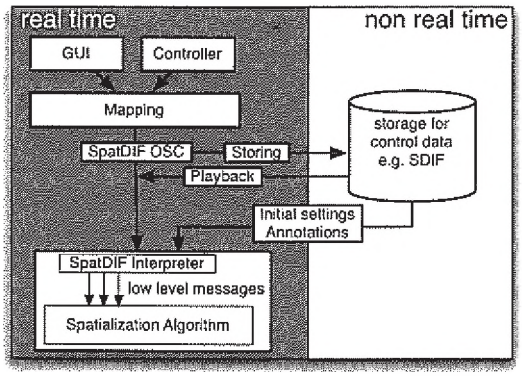

Figure 1: SpatDIF in real-time and non-real-time applications.

4 Conclusion and Acknowledgment

An interchangeable namespace for describing, storing and sharing data for 3D audio applications was presented. For now SpatDIF handles elementary parameters, but we also want to describe higher level data, such as reverberation or complex sound scenes. SpatDIF is being tested in three projects related to live electronics: Integra [3], Jamoma [6], and an NSERC/CCA New Media Initiative project at CIRMMT. Researchers from McGill University, the University of Oslo and the Bergen Center for Electronic Arts are involved in the development - others are welcome to join.

This work is funded by the Canadian Natural Sciences and Engineering Research Council (NSERC) and the Canada Council for the Arts. Special thanks to Alexander Jensenius, Trond Lossius and Joseph Malloch.

References

[1] M. A. Baalman, T. Hohn, S. Schampijer, and T. Koch. Renewed architecture of the WONDER software for Wave Field Synthesis on large scale systems. In *proceedings of the Linux Audio Conference*, Berlin, Germany, 2007.

[2] J. Braasch. A loudspeaker-based 3D sound projection using virtual microphone control (ViMiC). In *Convention of the AudioEng. Soc. 118, Barcelona, Spain, Preprint 6430*, 2005.

[3] J. Bullock and H. Frisk. libIntegra: a system for software-independent multimedia module description and storage. In *Proceedings of the International Computer Music Conference*, Copenhagen, Denmark, 2007.

[4] A. Jensenius, T. Kvifte, and R. Godøy. Towards a gesture description interchange format. In *Proceedings of the 2006 conference on New interfaces for musical expression*, pages 176–179. Paris, France, 2006.

[5] J.-M. Jot. The Spat Processor and its Library of Max Objects - Introduction. Technical report, IRCAM, Manual, Ver. Oct. 2006.

[6] T. Place and T. Lossius. Jamoma: A modular standard for structuring patches in Max. In *Proceedings of the International Computer Music Conference 2006*, New Orleans, US, 2006.

[7] V. Pulkki. Virtual sound source positioning using vector base amplitude panning. *AES Journal*, 45(6):456 – 466, 1997.

[8] J. C. Schacher and P. Kocher. Ambisonics spatialization tools for max/msp. In *Proceedings of the 2006 International Computer Music Conference*, pages 274 – 277, New Orleans, US, 2006.

[9] M. Wright, A. Chaudhary, A. Freed, D. Wessel, and X. Rodet. New Applications of the Sound Description Interchange Format. In *Proceedings of the International Computer Music Conference*, pages 276–279, Ann Arbor, US, 1998.

[10] M. Wright, A. Freed, and A. Momeni. OpenSound Control: state of the art 2003. In *Proceedings of the 2003 conference on New Interfaces for Musical Expression*, pages 153–160, Singapore, 2003.

[11] S. Yadegari, F. R. Moore, H. Castle, A. Burr, and T. Apel. Real-time implementation of a general model for spatial processing of sounds. In *Proceedings of the 2002 International Computer Music Conference*, pages 244 – 247, San Francisco, CA, USA, 2002.