A STRATIFIED APPROACH FOR SOUND SPATIALIZATION

Nils Peters[a], Trond Lossius[b], Jan Schacher[c], Pascal Baltazar[d], Charles Bascou[e], Timothy Place[f]
[a] CIRMMT, McGill University, Montréal, nils.peters@mcgill.ca
[b] BEK - Bergen Center for Electronic Arts, trond.lossius@bek.no
[c] ICST, Zurich University of the Arts, jan.schacher@zhdk.ch
[d] GMEA - National Centre for Musical Creation of Albi, pb@gmea.net
[e] GMEM - National Centre for Musical Creation of Marseille, charles.bascou@gmem.org
[f] Cycling '74, tim@cycling74.com

ABSTRACT

We propose a multi-layer structure to mediate essential components in sound spatialization. This approach will facilitate artistic work with spatialization systems, a process which currently lacks structure, flexibility, and interoperability.

1 INTRODUCTION

The improvements in computer and audio equipment in recent years make it possible to experiment more freely with resource-demanding sound synthesis techniques such as spatial sound synthesis, also known as spatialization. For seeking new means of expression, different spatialization applications should be readily combined and accessible for both programmatic and user interfaces. Furthermore, quantitative studies on spatial music (e.g. [12]) remind us that there are great individual and context-related differences in the compositional use of spatialization and that there is no one spatialization system that could satisfy every artist. In an interactive art installation, the real-time quality of a spatial rendering system in combination with the possibility to control spatial processes through a multi-touch screen can be of great importance. In contrast, the paramount features in a performance of a fixed-media composition may be multichannel playback and the compensation of non-equidistant loudspeakers (in terms of sound pressure and time delays). Additional scenarios may require binaural rendering for headphone listening, multichannel recording, up and down mixing, or a visual representation of a sound scene. Moreover, even during the creation of one spatial art work, the importance of these requirements may change throughout different stages of the creative processes.

Guaranteeing efficient workflow for sound spatialization requires structure, flexibility, and interoperability across all involved components. As reviewed in the following section, common spatialization systems too often give no consideration to these requirements.

SMC 2009, July 23-25, Porto, Portugal

2 REVIEW OF CURRENT PARADIGMS

2.1 Digital Audio Workstations - DAW

Many composers and sound designers use DAWs for designing their sound spatialization primarily in the context of fixed media, tape-music, and consumer media production. A number of DAWs are mature and offer a systematic user interface, good project and sound file management, and extendability through plug-ins to fulfill different needs.

DAWs mainly work with common consumer channel configurations; mono, stereo and 5.1. However, through focusing on consumer media products, multichannel capabilities are limited. ITU 5.1 [6], a surround sound format with equidistant loudspeakers around an ideal located listener, is the most common multichannel format. Its artistic use may be limited because 5.1 favors the frontal direction and has reduced capabilities for localizing virtual sources from the sides and back. Recent extensions up to 10.2 are available [1], but are insufficient for emerging reproduction techniques such as Wave Field Synthesis or Higher Order Ambisonics. Also, in art installations or concert hall environments, non-standard loudspeaker setups are common due to artistic or practical reasons, varying in number and arrangements of loudspeakers. These configurations are typically unaccounted in DAWs and therefore often difficult to use.

DAW surround panners often comprise a parameter named *blur*, *divergence*, or *spread* that controls the apparent source width through modifying the distributed sound energy among loudspeakers. Although this parameter enriches the creative possibilities, it is often either missing or only indirectly accessible, e.g. through changing the distance of the sound source.

[1] A comparison of DAWs concerning their multichannel audio capabilities can be seen on `http://acousmodules.free.fr/hosts.htm`.

2.2 Media programming environments

Various media programming environments exist that are capable of spatial sound synthesis, e.g. SuperCollider, Pure Data, OpenMusic, and Max/MSP. In order to support individual approaches and to meet the specific needs of computer music and mixed media art, these environments enable the user to combine music making with computer programming. While aspiring to complete flexibility, they end up lacking structured solutions for the specific requirements of spatial music as outlined in section 1. Consequently, numerous self-contained spatialization libraries and toolboxes have been created by artists and researchers to generate virtual sound sources and artificial spaces, such as Space Unit Generator [26], Spatialisateur [7], or ViMiC [3]. Also toolboxes dedicated to sound diffusion practice has been developed, e.g. BEASTmulch System [2], ICAST [1]. Each tool, however, may only provide solutions for a subset of compositional viewpoints. The development of new aesthetics through combining these tools is difficult or limited due to their specific designs.

2.3 Stand-alone Applications

A variety of powerful stand-alone spatialization systems are in development, ranging from directional based spatialization frameworks, e.g. SSR [4], Zirkonium [19], and Auditory Virtual Environments (AVE), e.g. tinyAVE [2] to sound diffusion and particle oriented approaches, e.g. Scatter [9]. Although these applications usually promote their graphical user interfaces as the primary method to access their embedded DSP-algorithms, a few strategies to allow communication from outside through self-contained XML, MIDI or OSC [25] protocols can be found.

3 A STRATIFIED APPROACH TO THE SPATIALIZATION WORKFLOW

When dealing with spatialization in electroacoustic composition or linear sound editing, the workflow comprises a number of steps in order to construct, shape and realize the spatial qualities of the work. The creative workflow might appear to be different when working on audio installations or interactive/multimedia work. Still, we identified underlying common elements when spatialization is used. For this reason a stratified approach, where the required processes are organized according to levels of abstraction is proposed.

This model is inspired by the Open Systems Interconnection network model (OSI) [3], which is an abstract description for layered communications and computer network protocol design. OSI divides network architecture into seven layers that range from top to bottom between the Application and Physical Layers. Each OSI-layer contains a collection of conceptually similar functionalities that provide services to the layer above it and receives services from the layer below it.

As depicted in Figure 1, six layers have been defined in our model.

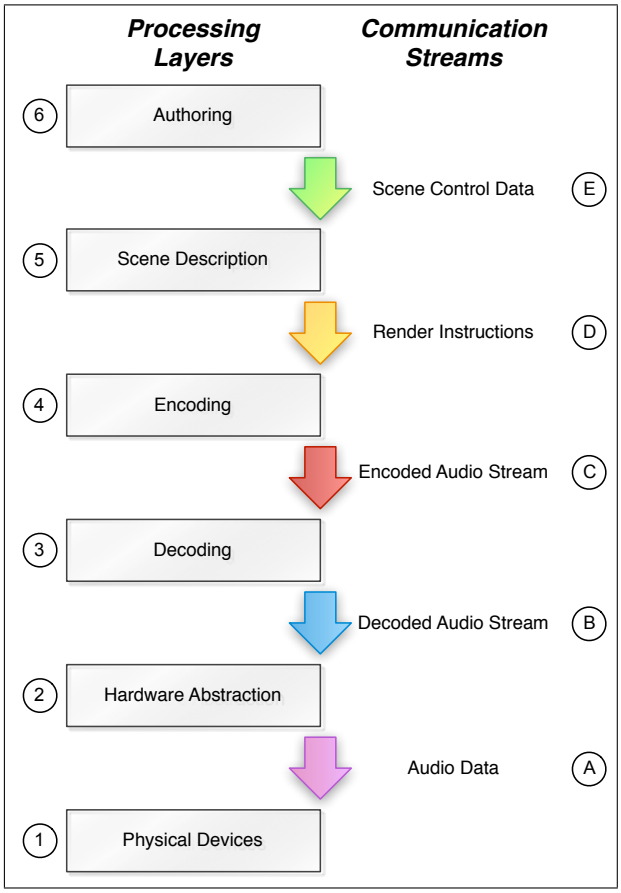

Figure 1. Layers and streams in sound spatialization

3.1 Physical Device Layer

The major functionality of this layer is to establish the acoustical connection between computer and listener. It defines the electrical and physical specifications of devices that create the acoustical signals, such as soundcards, amplifiers, loudspeakers, and headphones.

3.2 Hardware Abstraction Layer

This layer contains the audio services that run in the background of a computer OS and manages multichannel audio data between the physical devices and higher layers. Examples are Core Audio, ALSA, or PortAudio. Extensions such as JACK, Soundflower, Rewire and networked audio

[2] http://www.beast.bham.ac.uk/research
[3] http://en.wikipedia.org/wiki/OSI_model

streaming can be used for more complex distributions of audio signals among different audio clients.

3.3 Encoding and Decoding Layers

In the proposed model the spatial rendering is considered to consist of two layers. The Encoding Layer produces encoded signals containing spatial information while remaining independent of and unaware of the speaker layout. The Decoding Layer interprets the encoded signal and decodes it for the speaker layout at hand. According to [24, p. 99] this makes the creative process and the created piece more portable and future-proof because different speaker layouts can be used as long as a decoder is available. Examples of such hierarchical rendering methods are Ambisonics B-Format, Higher Order Ambisonics, DIRAC [18], MPEG Surround, AC-3, or DTS.

Not every rendering technique generates intermediate encoded signals, but instead can be considered to encapsulate the Encoding and Decoding Layers in one process. Some examples of such renderers are VBAP [16], DBAP [8], ViMiC [3] and Ambisonics equivalent panning [11]. Processing of sources to create an impression of distance, such as Doppler effect, gain attenuation and air absorption filters, are considered to belong to the Encoding Layer, as does the synthesis of early reflections and reverberation, i.e. as demonstrated by surround effects that employ B-format impulse responses convolution.

3.4 Scene Description Layer

This layer mediates between the Authoring Layer above and the Decoding Layer below through an abstract and independent description about the spatial scene. This description can range from a simple static scene with one virtual sound source up to complex dynamic audio scenes including multiple virtual spaces. This data could also be stored to recreate spatial scenes in a different context. Specific (lower-level) render instructions are communicated to the Encoding Layer beneath. Examples are ASDF [4], OpenAL [5] or SpatDIF [13].

3.5 Authoring Layer

This layer contains all software tools for the end-user to create spatial audio content without the need to directly control underlying processes. Although these tools may remarkably differ from each other through functionality and interface design to serve the requirements for varicolored approaches to spatialization, the communication to the Scene Description Layer must be standardized. Examples are symbolic authoring tools, generative algorithms, and simulations of emergent behaviors (swarms or flock-of-birds); or, more specifically as discussed below, Holo-Edit, and ambimonitor/ambicontrol.

3.6 Concluding remarks

OSI provided the idea that each layer has a particular role to play. The stratified model does not enforce one particular method for each layer; rather, a layer offers a collection of conceptually similar functions. This is analogue to how the TCP and UDP are alternative protocols working at the Transport Layer of the OSI model.

Spatialization processes should be modularized according to the layered model when feasible. With standardized communication to and from the layers, one method for a layer can easily be substituted for another, enhancing a flexible workflow that can rapidly adapt to varying practical situations and needs.

4 STRATIFIED TOOLS

Following, the authors discus several of their developments which strive to establish and evaluate the proposed stratified concept.

4.1 SpatDIF

The goal of the Spatial Sound Description Interchange Format (SpatDIF) is to develop a system-independent language for describing spatial audio [13] that can be applied around the Scene Description Layer to communicate between authoring tools down to the Encoding/Decoding Layers.

Formats that integrate spatial audio descriptors, such as MPEG-4 [23] or OpenAL, did not fully succeed in the music or fine arts community because they are primarily tailored to multimedia or gaming applications and don't necessarily consider the special requirements of spatial music, performances in concert venues, and site-specific media installations. To account for these specific requirements, the SpatDIF development is consequently a collaborative effort that jointly involves researchers and artists.

A database [4] has been created to gather information about syntax and functionalities of common spatialization tools and to identify the lowest common denominator, the "Auditory Spatial Gist", for describing spatialized sound. Beside these essential Core Descriptors, a number of extensions have been proposed to systematically account for enhanced features, e.g. the Directivity Extension, which deals with directivity information of a virtual sound source; the Acoustic Spaces Extension that contains acoustical properties of virtual rooms, or the Ambisonics Extension that handles ambisonics-only parameters. The latter is an example where SpatDIF mediates between the processing layers, starting from Layer 3 to Layer 6.

Although SpatDIF does not imply a specific communication protocol or storing format, at present, OSC for streaming and SDIF [22] as a storing solution are used for piloting.

[4] http://redmine.spatdif.org/wiki/spatdif/SpatBASE

4.2 ICST Ambisonics

The ICST Ambisonics Tools is a set of externals for Max/MSP [21]. The DSP externals ambiencode∼ and ambidecode∼ generate and decode Higher Order Ambisonics and are part of the Encoding and Decoding Layer.

Ambimonitor and ambicontrol complete the set as control tools for the Authoring Layer. Ambimonitor generates coordinate information for the DSP objects, presents the user with a GUI displaying point sources in an abstract 2D or 3D space and is equipped with various key commands, snapshot and file I/O capabilities. Ambicontrol provides a number of methods that control motion of points in the ambimonitor's dataset. Automated motions, such as rotation or random motion, optionally constrained in bounding volumes and user defined trajectories can be applied to single or grouped points. Trajectories and state snapshots can be imported/exported as an XML file, which will be replaced by a SpatDIF compliant formatting in a next release.

A novel panning algorithm [11] was derived from inphase ambisonics decoding and implemented as a Max/MSP external entitled ambipanning∼. It encapsulates the Encoding and Decoding Layer by transcoding a set of mono sources in one process onto an ideally circular speaker setup with an arbitrary number of speakers. The algorithm works with a continuous order factor, permitting the use of individually varying directivity responses.

4.3 Jamoma

Jamoma [5] is a framework [14] for structuring and controlling modules in Max/MSP. Work on spatialization has been of strong interest to several of the developers, and solutions for spatialization in Jamoma have a stratified approach in accordance with the proposed model.

The Max/MSP signal processing chain only passes mono signals, and for multichannel spatial processing the patch has to be tailored to the number of sources and speakers. If Max/MSP is considered a programming environment and the patch is the program, a change in the number of sources or speakers requires a rewrite of the program, not just a change to one or more configuration parameters. Jamoma addresses this by introducing multichannel audio signals between modules with all channels wrapped onto a single patch cord. Jamoma Multicore [6] is being developed as a more robust solution than the current approach for handling multichannel signals which are also used between the Encoding, Decoding and Hardware Abstraction Layers.

Jamoma modules have been developed to convert multichannel signals, play and record multichannel sound files, perform level metering and pass multichannel signals on to the sound card or virtual auxiliary bus. These are supplemented by modules compensating for sound-pressure and time-delay differences in non-equidistant loudspeaker arrangements.

Ambisonics is the only spatialization method implemented in Jamoma that separates spatial encoding and decoding. 1st to 3rd order B-format encoding of mono sources is implemented using the ICST externals[21]. Other modules are available to encode recordings made with the Zoom H2 and to encode UHJ signals. Encoded signals can be manipulated, i.e. the balance between the encoded channels can be adjusted, or the encoded signal can be rotated, tilted and tumbled. The decoding module for up to 3rd order B-format signals uses the ICST externals while a module for binaural decoding uses Spatialisateur [7]. B-format signals can also be decoded to UHJ.

Several other popular spatialization algorithms are available as Jamoma modules: VBAP [17], ViMiC [3] and DBAP [8]. Consequently, one rendering technique can easily be substituted for another, or several rendering techniques might be used in tandem for a variety of spatial expressions, analogues to how an artist will use many different brushes in one artwork.

Prior to rendering, additional modules offer Doppler, air absorption and distance attenuation source pre-prosessing. All modules operating at the Encoding Layer are SpatDIF-compliant and hence provide the same interface to controlling modules operating at higher layers.

At the Scene Description Layer, a module provides a simple interface for defining the position of sources. The same module can be used to set loudspeaker positions for the Decoding Layer.

At present, two modules operate at the Authoring Layer; Boids simulation of co-ordinated animal motion and a scene manipulator allows geometric transformations (e.g. scaling, skewing, rotation) and stochastically driven manipulations of the whole scene in three dimensions. In addition, Jamoma can be bridged to Holo-Edit as discussed in the next section.

4.4 GMEM Holo-Edit

Initiated by L. Pottier [15], Holo-Edit is part of the GMEM Holophon project and conceptualized as an authoring tool for spatialization.

This standalone application uses the timeline paradigm found in traditional DAWs to record, edit, and play back control data. The data is manipulated in the form of trajectories or sequences of time-tagged points in a 3D space, and the trajectories can be generated or modified by a set of tools allowing specific spatial and temporal behaviors including symmetry, proportion, translation, acceleration, and local exaggeration. Different scene representation windows allow the user to modify data from different (compositional) viewpoints: *Room* shows a top view of the virtual space, the *Time Editor* shows the traditional DAW automation curve view

[5] http://www.jamoma.org

[6] http://code.google.com/p/jamulticore/

and, finally, the *Score Window* represents the whole composition in a multi-track block-based view. Holo-Edit's space and time representations are generic and can be adapted to any render at the Encoding Layer. To allow precise alignment of sound cues to desired spatial movements, waveform representations of sounds and associated trajectories are displayed together.

Holo-Edit uses OSC for communicating with the desired spatial sound renderer. Here, the main challenge is to adapt and format the data stream that fits the specific rendering algorithm syntax (e.g. coordinate system, dimensions, units). To overcome this challenge, a Holo-Edit communication interface that handles sound file playback and position data of loudspeakers and sound sources through its standardized OSC-namespace was developed for the Jamoma environment. Therefore, Holo-Edit can be used as the main authoring tool for spatialization, while all DSP audio processes are executed in Jamoma (Figure 2). The communication between Holo-Edit and Jamoma is full-duplex, thus also enables the recording of trajectories in Holo-Edit from any real-time control interface addressable through Jamoma.

5 DISCUSSION & CONCLUSION

The examples from the previous section illustrate that a stratified model can be fruitful for development within media programming environments. The modular framework TANGA [20] for interactive audio applications reveals a related separation of tasks.

A few stand-alone applications are designed with a similar layered approach that allows control of different spatial rendering algorithms from one common interface, e.g. [4]. Artists and researchers would benefit greatly if all these "local solutions" could be accessed by any desired authoring tool and integrated into existing environments.

After an ICMC 2008 panel discussion on interchange formats for spatial audio scenes [7] and informal discussion showed that adequate spatialization tools for working in DAWs are missing, but strongly desired. The proposed stratified approach would be more flexible than the current DAW architecture where tools for spatialization are tied to a number of consumer channel configurations. The object oriented mixer approach proposed in [10] suggests that stratification can be employed in DAWs. A potential limitation might be imposed by the fact that automation in DAWs generally is represented as time-tagged streams of one-dimensional values while spatial information is generally multi-dimensional.

One keystone may be to define and agree on a meaningful communication format for spatialization. Therefore SpatDIF needs to be further developed which will culminates in an API that easily integrates in any spatialization software.

[7] http://redmine.spatdif.org/wiki/spatdif/Belfast_2008

6 ACKNOWLEDGMENT

The concepts proposed in this paper were to a large degree developed during a workshop at GMEA Centre National de Création Musicale as part of the Virage research platform funded by the French National Agency for Research. This work is also partly funded by the Canadian Natural Sciences and Engineering Research Council (NSERC), the Canada Council for the Arts, The COST IC0601 Action on Sonic Interaction Design (SID) and the Municipality of Bergen.

7 REFERENCES

[1] S. D. Beck, J. Patrick, B. Willkie, and K. Malveaux. The Immersive Computer-controlled Audio Sound Theater: Experiments in multi-mode sound diffusion systems for electroacoustic music performance. In *Proceedings of International Computer Music Conference 2006*, New Orleans, US, 2006.

[2] C. Borß and R. Martin. An improved parametric model for perception-based design of virtual acoustics. In *AES 35th Int. Conference*, London, UK, 2009.

[3] J. Braasch, N. Peters, and D. L. Valente. A loudspeaker-based projection technique for spatial music applications using virtual microphone control. *Computer Music Journal*, 32(3):55 – 71, 2008.

[4] M. Geier, J. Ahrens, and S. Spors. The SoundScape Renderer: A Unified Spatial Audio Reproduction Framework for Arbitrary Rendering Methods. In *124th AES Convention, Preprint 7330*, Amsterdam, The Netherlands, May 2008.

[5] G. Hiebert. *OpenAL 1.1 Specification and Reference*, 2005.

[6] ITU. Recommendation BS. 775: Multi-channel stereophonic sound system with or without accompanying picture, International Telecommunications Union, 1993.

[7] J.-M. Jot. *Etude et Réalisation d'un Spatialisateur de Sons par Modèles Physiques et Perceptifs*. PhD thesis, France Telecom, Paris 92 E 019, 1992.

[8] T. Lossius, P. Baltazar, and T. de la Hogue. DBAP - Distance-Based Amplitude Panning. In *Proceedings of 2009 International Computer Music Conference*, Montreal, Canada, 2009.

[9] A. McLeran, C. Roads, B. L. Sturm, and J. J. Shynk. Granular sound spatialization using dictionary-based methods. In *Proceedings of the 5th Sound and Music Computing Conference*, Berlin, Germany, 2008.

[10] S. Meltzer, L. Altmann, A. Gräfe, and J.-O. Fischer. An object oriented mixing approach for the design of spatial audio scenes. In *Proc. of the 25th Tonmeistertagung*, Leipzig, Germany, 2008.

[11] M. Neukom and J. Schacher. Ambisonics equivalent panning. In *Proceedings of the 2008 International Computer Music Conference*, Belfast, UK, 2008.

[12] F. Otondo. Contemporary trends in the use of space in electroacoustic music. *Organised Sound*, 13(01):77–81, 2008.

[13] N. Peters. Proposing SpatDIF - the spatial sound description interchange format. In *Proceedings of the 2008 International Computer Music Conference*, Belfast, UK, 2008.

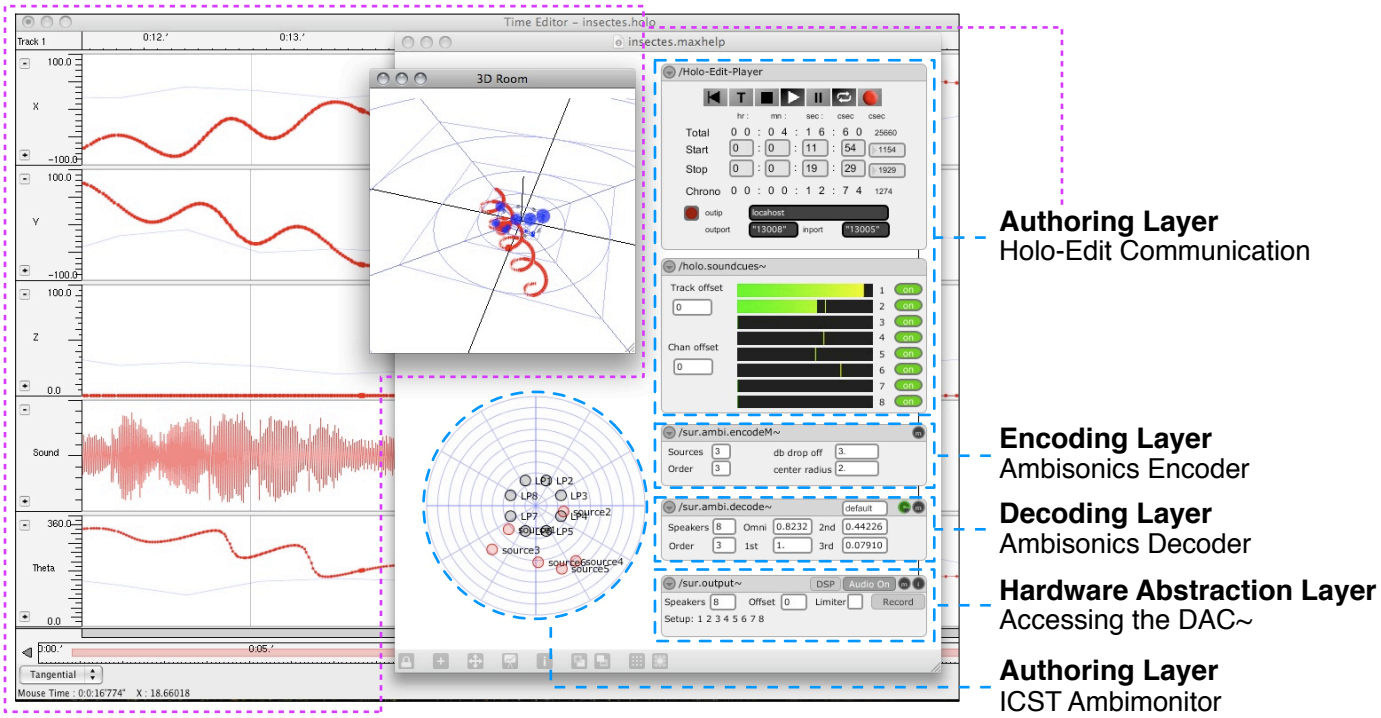

Figure 2. Holo-Edit, Jamoma and ICST Ambisonics Tools unified

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

[15] L. Pottier. Dynamical spatialisation of sound. holophon: a graphical and algorithmical editor for sigma 1. In *Proceedings of the International Conference on Digital Audio Effects, DAFX98*, Barcelona, Spain, 1998.

[16] V. Pulkki. Virtual sound source positioning using vector base amplitude panning. *J. Audio Eng. Soc.*, 45(6):456–466, 1997.

[17] V. Pulkki. Generic panning tools for MAX/MSP. In *Proceedings of 2000 International Computer Music Conference*, pages 304–307, Berlin, Germany, 2000.

[18] V. Pulkki. Spatial sound reproduction with directional audio coding. *J. Audio Eng. Soc.*, 55(6):503–516, June 2007.

[19] C. Ramakrishnan, J. Goßmann, and L. Brümmer. The ZKM Klangdom. In *Proc. of the 2006 conference on New Interfaces for Musical Expression*, pages 140–143, Paris, France, 2006.

[20] U. Reiter. TANGA - an interactive object-based real time audio engine. In *Proceedings of the 2nd Audio Mostly conference*, Ilmenau, Germany, 2007.

[21] J. C. Schacher and P. Kocher. Ambisonics Spatialization Tools for Max/MSP. In *Proc. of the 2006 International Computer Music Conference*, pages 274–277, New Orleans, US, 2006.

[22] D. Schwarz and M. Wright. Extensions and Applications of the SDIF Sound Description Interchange Format. In *Proceedings of the 2000 International Computer Music Conference*, pages 481–484, Berlin, Germany, 2000.

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

[24] B. Wiggins. *An investigation into the real-time manipulation and control of three-dimensional sound fields*. PhD thesis, University of Derby, Derby, UK., 2004.

[25] M. Wright and A. Freed. Open Sound Control: A New Protocol for Communicating with Sound Synthesizers. In *Proceedings of the 1997 International Computer Music Conference*, pages 101–104, Thessaloniki, Greece, 1997.

[26] 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, Goteborg, Sweden, 2002.