A STRATIFIED APPROACH FOR SOUND SPATIALIZATION

First author
School
Email

Second author
Company
Email

ABSTRACT

We propose a multi-layer approach to mediate across essential components involved 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 ([11], [13]) 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. For instance, the requirements of a computer aided spatialization system may vary between a fixed-media composition, an art-installation, and a live diffusion performance. One example is an interactive art installation where the real-time quality of a spatial rendering system is of great importance, in combination with the possibility to control spatial processes through a multi-touch screen. Juxtaposed is a second example: a performance of a fixed-media composition where the paramount features 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 outlined in Section 2, common spatialization systems are too often self-contained, giving no consideration to these requirements.

SMC 2009, July 23-25, Porto, Portugal

2 REVIEW OF CURRENT PARADIGMS

2.1 Digital Audio Workstations - DAW

Currently, 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. Users of DAWs often have former experience with live diffusion systems, or with panning in a hardware mixing console. Their migration to DAWs seems reasonable because the linear time representation of sound material and the audio bus architecture in DAWs originates from the use of multitrack tape recorder and mixing consoles.

A number of (mainly commercial) DAWs are mature and offer a systematic user interface, good project and sound file management, and extendability through plug-ins so that they can fulfill the needs of many users in the described context. 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 the sound energy among loudspeakers. Although this parameter enriches the creative possibilities, it is often either missing or

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

only indirectly accessible, e.g. through changing the distance of the sound source.

2.2 Media programming environments

Beside the paradigm of the DAW, various media programming environments exist that are capable of spatial sound synthesis. These include SuperCollider, Pure Data, OpenMusic, and Max/MSP [16, 27]. 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.

However, in the name of complete flexibility, these environments lack in providing structured solutions for the specific challenges 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, such as the BEASTmulch System [2], or ICAST [1] has been developed. 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 through 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 [20]) 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 embedded 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 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 that are always in play when spatialization is used. For this reason a layered approach is proposed, where the required processes are organized according to levels of abstraction.

[2] http://www.beast.bham.ac.uk/research

This model is inspired by the OSI network model [3], which is an abstract description for layered communications and computer network protocol design. OSI (Open Systems Interconnection) 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 service from the layer below it.

In our proposed model, depicted in Figure 1, six layers have been identified. The adaptation of concepts originally designed for network protocols to computer music systems was legitimized, for instance, in creating the popular Open Sound Control (OSC) protocol [25].

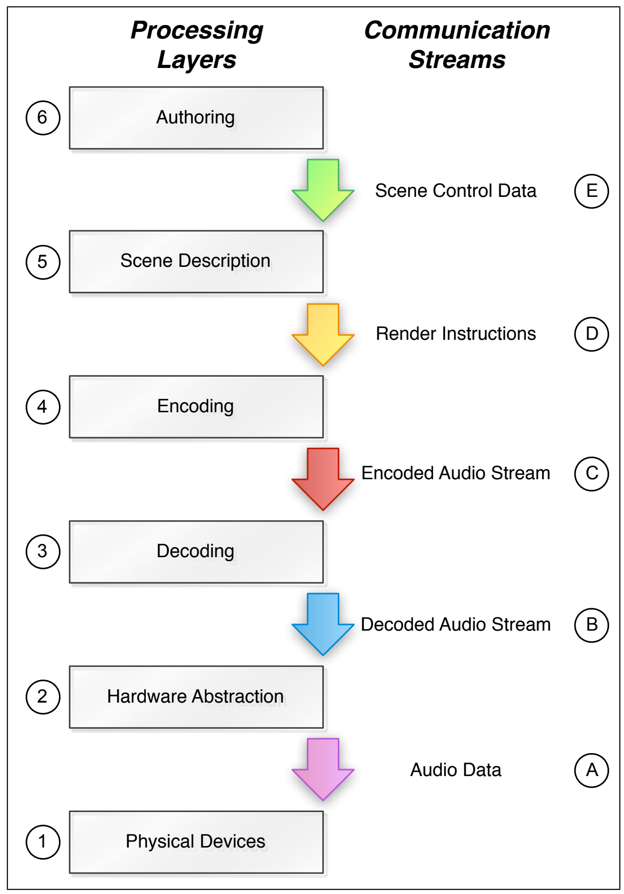

Figure 1. Layers and streams in sound spatialization

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 creates the acoustical signals, such as soundcards, amplifiers, loudspeakers, and headphones.

[3] http://en.wikipedia.org/wiki/OSI_model

3.2 Hardware Abstraction Layer

This layer contains the audio services that runs in the background of a computer OS and manages multichannel audio data between the physical devices and higher layers.

Core Audio, ALSA, or PortAudio are examples of such services. Extensions such as JACK, Soundflower, Rewire and networked audio 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* produce encoded signals containing spatial information while remaining independent of and unaware of the speaker layout. The decoding process interprets the encoded signal and decodes it for the speaker layout at hand. According to [24, p. 99] this makes the creational 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 [19], MPEG Surround, AC-3, or DTS.

Not every rendering technique generate 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 [17], DBAP [8], ViMiC [3] and Ambisonics equivalent panning [10].

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. Likewise early reflections and reverberation belongs in the encoding layer. The Waves IR360 surround convolution reverb internally use B-format reverb impulse responses.

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 sources up to complex dynamic audio scenes including multiple virtual spaces. These data could also be storable 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 [12].

3.5 Authoring Layer

This layer contains all software tools for the end-user to create spatial audio content without the need to control underlying processes directly. Although these tools may differ remarkably through functionality and interface design from each other 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, simulations of emergent behaviors (swarms or flock-of-birds); or more specifically, ICST ambimonitor/ambicontrol, and Holo-Edit.

3.6 Concluding remarks

The idea we take from OSI is that each layer has a particular role to play. The layered model do 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

In the following, several developments by the authors are discussed 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 [12] that can be applied around the *Scene Description Layer* of Figure 1, to communicate between authoring tools down to the *Encoding/Decoding Layers*.

Formats that integrate spatial audio descriptors such as MPEG-4 Advanced Audio BIFS [23] or OpenAL did not fully succeed in the music or fine arts community because they are primary 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 researcher 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.

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

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 parameter. The *Ambisonics Extension* is an example where SpatDIF mediates between the processing layers, starting from Layer 3 to Layer 6 (see Figure 1).

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.

4.2 ICST Ambisonics

The ICST Ambisonics Tools comprise a set of four externals for Max/MSP [21].

The two DSP externals ambiencode∼ and ambidecode∼ generate and decode Higher Order Ambisonics and are being part of the Encoding and Decoding Layer in Figure 1.

The two externals ambimonitor and ambicontrol complete the set as control tools for the Authoring Layer. Ambimonitor presents the user with a GUI displaying point sources in an abstract 2D or 3D space, various key commands, snapshot and file I/O capabilities and it generates coordinate information for the DSP objects. Ambicontrol provides a number of methods that control motion of points in the ambimonitor's dataset. Automated motions such as rotation, 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 a XML file, which will be replaced by a SpatDIF compliant formatting in a next release.

A novel panning algorithm [10] was derived from in-phase 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, which permits to apply individually varying directivity responses. This external understands the same Render Instructions and is therefore interchangeable with the two ICST Ambisonics Tools DSP objects.

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 layered approach in accordance with the proposed model.

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

The Max/MSP signal processing chain only pass 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 require a rewrite of the program, instead of just a change to one or more configuration parameters. Jamoma address this by introducing multichannel audio signals between modules with all channels wrapped onto a single patch chord. 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.

A number of Jamoma modules have been developed for converting multichannel signals, play and record multichannel sound files, doing level metering and for passing multichannel signals on to the sound card or virtual auxiliary buses. 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 encoding of mono sources is implemented using the ICST externals[21]. Other modules are available for super stereo encoding, encoding of tracks from the portable Zoom H2 recorder and for encoding of UHJ signals. Encoded signals can be manipulated, e.g. the balance between the encoded channels can be adjusted, or the encoded signal can be rotated, tilted and tumbled. The decoding module for 1st to 3rd order signals use the ICST externals while a module for binaural decoding use Spatialisateur [7]. B-format signals can also be decoded to UHJ.

Jamoma supports several other other popular spatialization algorithms: VBAP [18], ViMiC [3] and DBAP [8]. Prior to rendering, additional modules offers Doppler, air absorption and distance attenuation source pre-prosessing.

All modules operating at the Encoding Layer are SpatDIF-compliant and hence provides 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 are implemented that operate at the Authoring Layer; Boids simulation of co-ordinated animal motion and a 3D scene manipulator offering rotation, repositioning and scaling of the whole scene. In addition Jamoma can be bridged to Holo-Edit as discussed in the next section.

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

4.4 GMEM Holo-Edit

Initiated by L. Pottier [15], Holo-Edit is part of the GMEM Holophon project and conceptualized as a 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 and finally, the *Score Window* represents the whole composition in a multi-track block-based view. Holo-Edit's space and time representations are intentionally generic and able to adapt to any render at the Encoding Layer.

An important feature of Holo-Edit is the ability to display the waveform representation of a sound next to its associated trajectory. This allows precisely alignment of sound cues to desired movements. Holo-Edit uses OSC to communicate with various media programming environments presented in section 2.2 and control various spatialization methods and AVEs. A major challenge in these scenarios 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 was developed for the Jamoma environment that handles sound file playback and position data of loudspeakers and sound sources through its standardized OSC-namespace. Therefore, Holo-edit can be used as the main authoring tool for spatialization, while all DSP audio processes are executed in Jamoma. The communication between Holo-Edit and Jamoma is full-duplex, thus allows also to record trajectories in Holo-Edit from any real-time control interface addressable through Jamoma.

5 DISCUSSION & CONCLUSION

We propose a multi-layer approach to mediate across essential components involved in sound spatialization. The examples of work on spatialization done at ICST and in Jamoma and HoloEdit illustrates that a layered model can be fruitful for development within media programming environments.

An informal discussion after an ICMC 2008 panel discussion on interchange formats for spatial audio scenes [7] revealed that adequate spatialization tools for working in DAWs are strongly desired. The question remains whether this demand can be accommodated by employing the multi-layer approach in DAWs. It remains to be researched how and if a similar layered approach could also be adapted and become fruitful in DAWs. A potential limitation might be imposed by the fact that automation in DAWs generally is represented as timetagged streams of one-dimensional values while spatial information is generally multi-dimensional.

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

One keystone may be to define a meaningful and accepted interchange communication format for spatialization. Therefore further works needs to be done on SpatDIF which culminates in a API to be easily integrated in any spatialization software.

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

6 ACKNOWLEDGMENT

The concepts proposed in this paper was 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*, 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 Réalization d'un Spatialisateur de Sons par Modèles Physiques et Perceptifs*. PhD thesis, France Telecom, Paris 92 E 019, 1992.

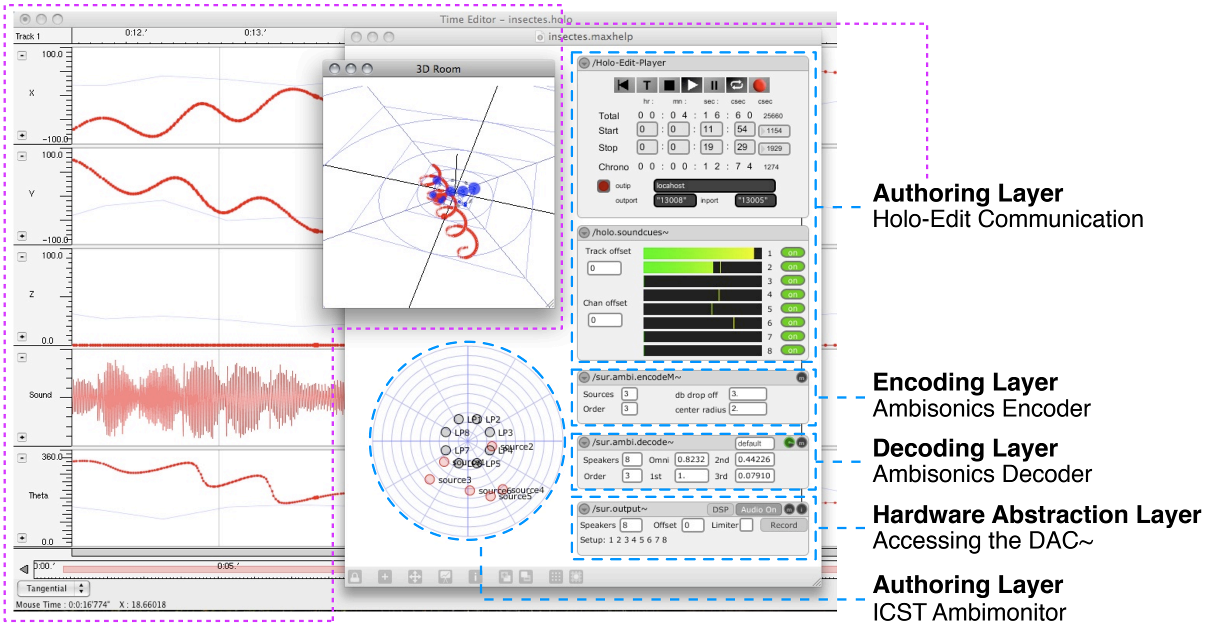

Figure 2. Holo-Edit, Jamoma and ICST Ambisonics Tools unified

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

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

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

[13] N. Peters, G. Marentakis, and S. McAdams. Current technologies and compositional practices for spatialization: A quantitative analysis. *in preparation*, 2009.

[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 *International Conference on Digital Audio Effects, DAFX98*, Barcelona, Spain, 1998.

[16] M. Puckette. Max at seventeen. *Computer Music Journal*, 26(4):31–43, 2002.

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

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

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

[20] 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.

[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. *Proceedings of the International Computer Music Conference, Berlin, Germany*, pages 481–484, 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. *Proceedings of the 1997 International Computer Music Conference*, pages 101–104, 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.

[27] D. Zicarelli. An extensible real-time signal processing environment for Max. In *Proc. of the 1998 International Computer Music Conference*, pages 463–466, San Francisco, US, 1998.