RENDERING OF HETEROGENEOUS SPATIAL AUDIO SCENES

Nicolas Bouillot, Michał Seta, Émile Ouellet-Delorme, Zack Settel, Emmanuel Durand

Société des arts technologiques [SAT]
Montréal, Canada
[nbouillot,mseta,eodelorme,zack,edurand]@sat.qc.ca

ABSTRACT

We present the ability of our Spatial Audio Toolkit for Immersive Environment (SATIE) to render simultaneously real-time audio scenes composed of various spatialization methods. While object oriented audio and Ambisonics are already included in SATIE, we present a prototype of a directional reverberation method based on Impulse Response computation and describe how this method will be included in SATIE.

1. INTRODUCTION

A growing number of computer music performance venues are now equipped with large loudspeaker configurations [1], and therefore provide new opportunities for artists using 3D audio scene environments for composition and sound design. This, along with the recent rise of affordable spatial audio recording devices and increased interest in virtual reality experiences, gives rise to a growing need of combining multiple spatialization methods: captures (live or not) made in different ambisonic formats, mono object-based audio sources as well as flexible & adaptable speaker configurations. We anticipate the evolution of spatial audio composition — targeting performing arts, installations or any other immersive experiences — involving different types of audio sources such as live audio capture, field recordings and synthetic audio, and where visual[2] and haptic[3] correlates with the audio part.

Moreover, innovation from the game industry is pushing forward virtual and augmented realities, approaching spatial audio with an object oriented manner: sources are *sound objects*, located in space and controlled with low level parameters such as gain, equalizer and spread. This approach, although effective for speaker array systems, is missing architectural acoustical responses and adapts poorly to non clearly located sound sources such as the sound of a river. The 3D graphic world is now entering audio and provides methods for the simulation of sound based on physics of soft body vibration and sound propagation [4]. Although such simulations are probably hard to achieve in real-time, simulations of acoustic responses of 3D environment may improve significantly the coherence of the integration of audio sources with the virtual space, while still allowing a real-time & 6-DoF navigation [5]. The use of ray tracing algorithms for real-time rendering is appropriate [6] and has the advantage of including the direction of the sound during auralization [7], allowing real-time calculation of directional sound reflections.

One of the main challenges today for spatial audio render is to support the multiplicity of the i) audio display methods, ii) spatial audio algorithms and iii) spatial audio authoring and 6-DoF navigation in spatial audio [8]. To date however, many existing real-time 3D audio scene rendering systems, such as COSM [9], Blender-CAVE [10], Spatium [11], Zirkonium [12], CLAM [13], 3Dj [14], Panoramix [15] and the spatDiff library [16] mostly focus on trajectory based composition with object oriented audio and sound fields with ambisonics. The challenge of navigating in heterogeneous spatial audio content is illustrated with Figure 1, where the spatial audio scene is constituted from 360° audio/video footage where the sound field captured using an ambisonic microphone[1] is mixed with synthetic audio is spatialized through an object oriented approach and correlated with 3D objects on screen (the white bubbles coming out from the white vase).

Figure 1: Example of an augmented reality application where a combination of several spatialization algorithms (ambisonics and object oriented audio): a 360° audiovisual capture is rendered simultaneously with synthetic objects, the bubbles coming out from the white vase.

In this paper, we present how our Spatial audio Toolkit for Immersive Environments (SATIE[2]) addresses the challenge of several approaches to audio scene rendering, possibly combining simultaneously object based audio, ambisonic formats and architectural based acoustical spatialization.

2. SATIE

The development of SATIE (with the SuperCollider language [17]) was first motivated by the need to render dense and rich audio scenes the Satosphere, a large dome-shaped audiovisual projection space at the Society for Art and Technology [SAT] in Montreal, and to compose real-time audio/music scenes consisting of hundreds of simultaneous sources targeting loudspeaker configurations of 32 channels or more, and sometimes with two or more different audio display systems [18]. In fact, SATIE easily adapts to different audio display configurations and supports plugins architecture which makes it easily extensible to new situations. As such, it fills the role of a rapid

[1] The Zylia ZM-1 microphone.

[2] `https://gitlab.com/sat-metalab/satie`, accessed Dec. 2018

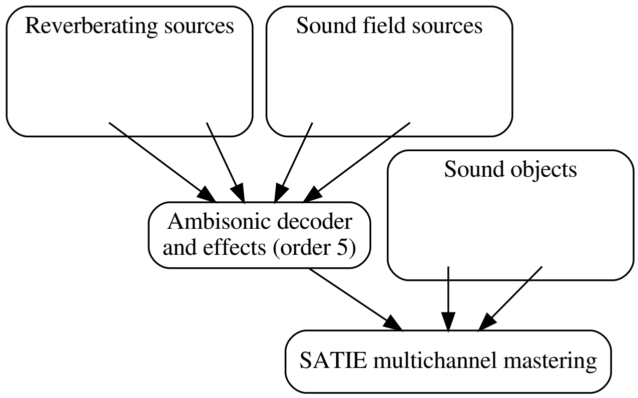


Figure 2: Example pipeline of a spatial rendering involving heterogeneous audio sources: reverberating sources, sound field sources and sound object.

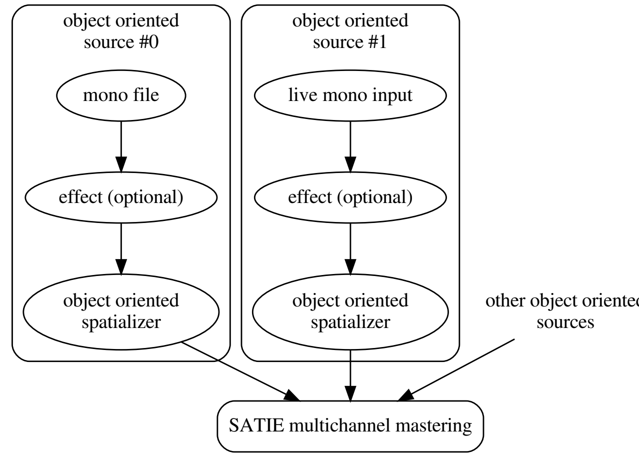


Figure 3: Internal pipeline for object oriented sound spatialization

prototyping tool for spatial audio composition.

Control of sound sources in SATIE is done through unified OSC [19] messages allowing for life management of each sound sources, along with (possibly custom) parameters control.

3. RENDERING METHODS

Facing a variety of approaches to composition with dense audio structures and a variety of audio displays, SATIE implements a flexible rendering pipeline allowing mixing of different audio input formats and multichannel mastering and is easily adaptable to various audio displays. We rely mainly on SuperCollider's *supernova* rendering engine for multi-threading operation. Consequently, we have access to parallel groups[20] which solve some real-time related issues with synth instantiating and bus allocation. SATIE structures different types of audio processors in layers, represented by a hierarchy of parallel groups (ParGroups):

- audio sources
- effects
- post-processors.

Audio sources are different types of mono or multichannel audio generators and players. On the second level are effects which usually do not generate sound but modify the signal of audio sources. Finally, post-processors are meant as mastering stage, where the final stages of DSP are done. In the actual implementation, the post-processors are divided in two groups: one for b-format signals and one for traditional mono/multichannel signals.

The signals between audio sources and effects pass through busses, i.e. the user allocates *auxiliary* busses and manages the bus access on both, the generator and effect side. If any post-processors are present, all signals are collected there, otherwise, they bypass directly to the spatializer. Multiple spatializers can be used, in which case SATIE will create appropriate number of output channels.

Figure 2 shows a rendering pipeline that combines object based audio sources, sound field sources and reverberating sources into heterogeneous mix.

3.1. Object Based Audio

Object audio (Figure 3) is what is most commonly used in various entertainment industries where a sound source has a clearly defined position within the coordinate system [21]. SATIE supports different types of object based audio sources, such as mono audio, mono live input sources and synthesized sounds [22]. The spatializers handling object audio expect azimuth, elevation and gain for panning each audio object.

SATIE was initially designed to render large numbers of mono audio sources, optionally with effects, to large multi-channel loudspeaker systems. Audio sources and effects can be placed in groups and controlled either per group or on individual basis. Similarly, spatializers take mono signals and place them on different channels according to azimuth, elevation and gain parameters. The post-processing audio object is comparable to mastering effects in a studio or live pipeline, typically limiting, compressing or normalizing signals.

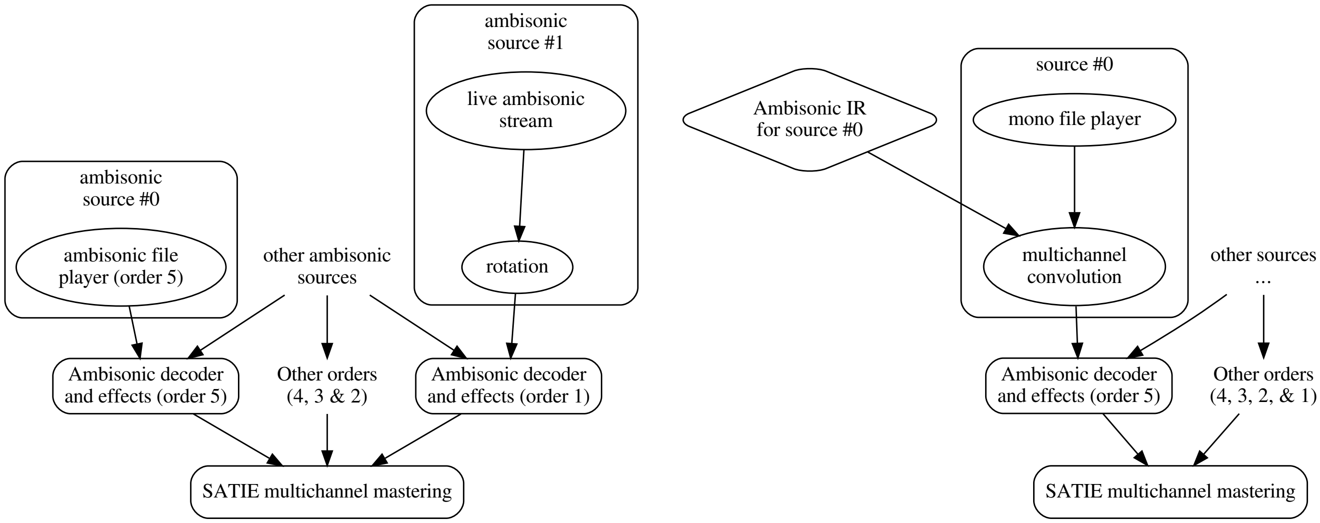
While all parameters (audio object specific as well as spatialization) can be modified either directly from the SuperCollider language, SATIE supports OSC and our preferred method is using a 3D engine for "volumetric" control of the sources as well as actual geometry computation. In line with this object based approach and load balancing physical computation we were able to use particle swarms of hundreds simultaneous sound sources.

3.2. Ambisonics

Ambisonic pipeline, implemented via SC-HOA plugins/quark [3] (Figure 4(a)) provides means to play multichannel files, live audio inputs, encode mono signals into b-format signals and transcode between different ambisonics formats (ACN and FuMa). It supports b-format up to order 5.

SATIE supports ambisonics with the same approach to signal path. The ambisonic audio input can be sent to ambisonic effects and post-processors such as rotation, mirroring, and beamforming filtering. The significant cost of ambisonic decoding is payed only once since not embedded in each ambisonic source pipeline, but rather at

[3] `https://github.com/florian-grond/SC-HOA`



(a) Pipeline for ambisonic sources.

(b) Pipeline for convolution based spatialization.

Figure 4: SATIE pipeline involving ambisonics

the post-processor stage. We can also transcode between different ambisonic orders.

3.3. Reverberating Sources with Convolution Reverb

Having various audio rendering methods driven by 3D engines opens doors to the desire of simulating acoustic spaces. Consequently, we have started developing a tool for real-time generation of impulse responses through ray tracing with the idea of integrating the IR workflow with SATIE. Figure 5(a) shows a screenshot of a real-time rendered frame where the listener is facing a sound source represented by a cube at the end of the hallway. Figure 5(b) shows a wireframe view of a simple model (not related to the picture on the left) showing what is actually going on. The black dots on the inner faces of the model represent the impact points of the rays on the walls of a 3D model. Sound sources and the listener are not shown, it simply shows a point cloud mapped on the model for reference. This implementation uses another custom software, VARAYS [4], which shares the 3D model with the 3D engine (in this case we're using EIS), receives the coordinates of the sound sources and the listener and writes IR files to disk. The IR files are read by SATIE which continuously replaces the buffer read by SuperCollider's PartConv UGen. A crude prototype of this process (using mono convolution) is demonstrated in the following video `https://vimeo.com/306202441`.

Besides mono IR, we can also generate Ambisonic IR (AIR), although at the time of the writing, this process has not yet been integrated into SATIE.

4. CONCLUSION

This paper outlined some of our approaches to heterogeneous audio scenes consisting of different types of audio input sources and multichannel displays. We described some SATIE functionalities with regard to heterogeneous spatial audio scenes. We have also described our approach to Ambisonic Impulse Response (AIR) in VARAYS in order to enable ambisonic acoustic simulation. VARAYS is still at very early stages of development, it needs proper support for material based diffraction and diffusion. Figure 4(b) shows the general workflow, where AIR is applied to a mono sound source and is spatialized using the usual SATIE pipeline. There is still some work left to do in order to fully integrate vaRays into SATIE pipeline (both IR and AIR). One of the areas to explore is in the interpolation of IR instances in order to compensate for real-time changes in the listener and the sound source location. This process can be mixed with types of rendering which provides sufficient creative liberty to the user. There is also some work left to provide IR and AIR to SATIE as files I/O are not the most optimal. We will be looking into sending OSC blobs. Another path would be sharing buffers between SATIE and vaRays using out shared memory library SHMDATA [5]. Another desired functionality is rendering VBAP spatialization into b-format signals.

5. ACKNOWLEDGEMENTS

This project would not be possible without the support of the Ministère de l'Économie et de l'Innovation du Québec. We also wish to thank Julien Wantz who spent time with us at the [SAT] Metalab, starting implementation of VARAYS, Mylène Pardoen for sharing the Paris 3D model used in Figure 5(b) and Florian Grond for his SuperCollider Higher Order Ambisonics library, SC-HOA.

6. REFERENCES

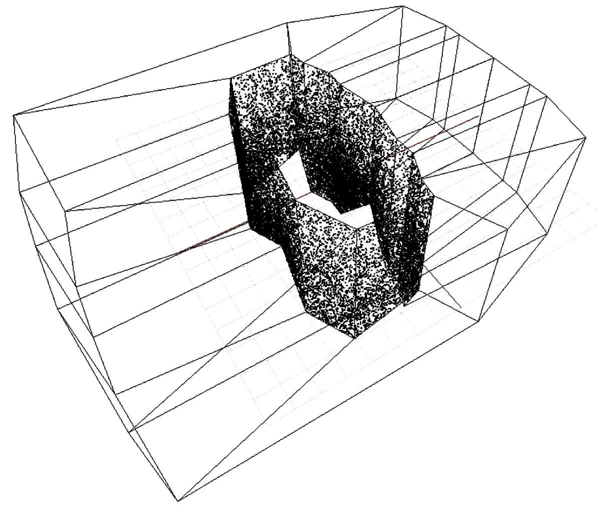
[1] N. Barrett, "A musical journey towards permanent high-density loudspeaker arrays," *Computer Music Journal*, vol. 40, no. 4, pp. 35–46, Dec 2016.

[2] Zack Settel, Nicolas Bouillot, and Michal Seta, "Volumetric approach to sound design and composition using SATIE:

[4] `https://gitlab.com/sat-metalab/varays`

[5] `https://gitlab.com/sat-metalab/shmdata`

(a) The cube floating at the end of the hallway in the 18th century Paris model represents a sound object. The image is from a prototype developed by Metalab using EIS for visual rendering and navigation, VARAYS for real-time impulse response processing and SATIE for audio spatialisation.



(b) Visualisation showing the impacts (black dots) of sound sources (not shown) on the walls of a 3D volume for a listener (not shown) placed inside the same volume. For this example, 2000 rays were thrown with a maximum of 3 reflections. A point cloud representing the impacts was saved by our software VARAYS and rendered in BLENDER.

Figure 5: Example of directional reverberation approach with our prototype based on conjoint use of SATIE and vaRays. We used the Bretez 3D of the 18th century.

a high-density 3D audio scene rendering environment for large multi-channel loudspeaker configurations," in *15th Biennial Symposium on Arts and Technology*, Ammerman Center for Arts and Technology at Connecticut College, New London, feb 2016, 8 pages.

[3] Nicolas Bouillot and Michał Seta, "A scalable haptic floor dedicated to large immersive spaces," in *Proceedings of The Linux Audio Conference*, March 2019.

[4] Jui-Hsien Wang, Ante Qu, Timothy R. Langlois, and Doug L. James, "Toward wave-based sound synthesis for computer animation," *ACM Trans. Graph.*, vol. 37, no. 4, pp. 109:1–109:16, July 2018.

[5] David Poirier-Quinot, Brian FG Katz, and Markus Noisternig, "EVERTIMS: open source framework for real-time auralization in architectural acoustics and virtual reality," in *Proceedings of Digital Audio Effects (DAFx)*, New York, NY, USA, 2017.

[6] Samuli Laine, Samuel Siltanen, Tapio Lokki, and Lauri Savioja, "Accelerated beam tracing algorithm," *Applied Acoustics*, vol. vol. 70, no. no. 1, pp. pp. 172–181, 2009.

[7] Mendel Kleiner, Bengt-Inge Dalenbäck, and Peter Svensson, "Auralization-an overview," *J. Audio Eng. Soc*, vol. 41, no. 11, pp. 861–875, 1993.

[8] Axel Plinge, Sebatian J. Schlecht, Oliver Thiergart, Thomas Robotham, Olli Rummukainen, and Emanuël A. P. Habets, "Six-degrees-of-freedom binaural audio reproduction of first-order ambisonics with distance information," in *Audio Engineering Society Conference: 2018 AES International Conference on Audio for Virtual and Augmented Reality*, Aug 2018.

[9] Graham Wakefield and Wesley Smith, "COSM: a toolkit for composing immersive audio-visual worlds of agency and autonomy," in *Proceedings of the International Computer Music Conference 2011*, University of Huddersfield, UK, aug 2011.

[10] David Poirier-Quinot, Damien Touraine, and Brian F.G. Katz, "BlenderCAVE: A multimodal scene graph editor for virtual reality," in *Proceedings of the 19th International Conference on Auditory Display (ICAD2013)*, Lodz, Poland, jul 2013, Georgia Institute of Technology & International Community for Auditory Display.

[11] Rui Penha and Joao Pedro Oliveira, "Spatium, tools for sound spatialization," in *Proceedings of the Sound and Music Computing Conference*, Stockholm, Sweden, 2013.

[12] Chikashi Miyama and Götz Dipper, "Zirkonium 3.1-a toolkit for spatial composition and performance," in *Proceedings of the International Computer Music Conference*, 2016, vol. 313, p. 312.

[13] Natanael Olaiz, Pau Arumi, Toni Mateos, and David Garcia, "3D-audio with CLAM and blender's game engine," in *proceedings of The Linux Audio Conference*, Parma, Italy, 2009.

[14] Andres Perez-Lopez, "3Dj: a supercollider framework for real-time sound spatialization," in *Proceedings of the 21th International Conference on Auditory Display (ICAD–2015)*, Graz, Austria, jul 2015.

[15] Thibaut Carpentier, "Panoramix: 3D mixing and post-production workstation," in *42nd International Computer Music Conference (ICMC)*, Utrecht, Netherlands, Sept. 2016.

[16] Jan C. Schacher, Chikashi Miyama, and Trond Lossius, "The spatdif library - concepts and practical applications in audio software," in *ICMC*, 2014.

[17] James McCartney, "Rethinking the computer music language: SuperCollider," *Computer Music Journal*, , no. 26, pp. 61–68, 2002.

[18] Zack Settel, Peter Otto, Michal Seta, and Nicolas Bouillot, "Dual rendering of virtual audio scenes for far-field surround multi-channel and near-field binaural audio displays," in *16th Biennial Symposium on Arts and Technology*, Ammerman Center for Arts and Technology at Connecticut College, New London, February 2018, 5 pages.

[19] M. Wright, "Open sound control 1.0 specification," Published by the Center For New Music and Audio Technology (CNMAT), UC Berkeley, 2002.

[20] Tim Blechmann, "Supernova, a scalabale parallel audio synthesis server for SuperCollider," in *Proceedings of the International Computer Music Conference 2011*, University of Huddersfield, UK, aug 2011.

[21] Ben Shirley, Rob Oldfield, Frank Melchior, and Johann-Markus Batke, "Platform independent audio," *Media Production, Delivery and Interaction for Platform Independent Systems: Format-Agnostic Media*, pp. 130–165, 2013.

[22] Nicolas Bouillot, Zack Settel, and Michal Seta, "SATIE: a live and scalable 3d audio scene rendering environment for large multi-channel loudspeaker configurations," in *New Interfaces for Musical Expression (NIME'17)*, Copenhagen, Denmark, 2017.