Developing a Novel Scene-Oriented Framework for Multichannel Program Mixing

Paulo GONÇALVES[1] and Regis R. A. FARIA[2]

[1]Organia Engenharia Musical Tecnologias Audiovisuais
São Paulo, SP

[2]Departamento de Música, FFCLRP, Universidade de São Paulo
São Paulo, SP

pgonsilva@organia.com.br, regis@usp.br

Abstract. *Achieving a stable and correct multichannel rendering is a challenging multistage process that can be performed under a scene oriented perspective, with many advantages over the conventional multi-track orientation. In this paper we present a flexible and scalable framework in development to generate multichannel programs. In addition to novel operational interfaces, it integrates a group of spatial rendering techniques and considers strategies to tackle the management of loudness and the smoothing of the sound objects images, leading to a balanced render of the sound scene regardless the output listening mode. An implementation in the Pd platform has been recently prototyped and is presented.*

1. Introduction

Most available methods for multichannel production rely on multi-bus and track-oriented reasoning, requiring the manipulation and mental control of dozens sliders, buttons, tracks, insert and auxiliary path setups, making difficult the operation for the technician and not-intuitive for the artist or producer. The present work investigates another thinking on the organization of a mixing project, a more intuitive control interface for manipulating parameters, and a task-oriented processing architecture.

A more intuitive representation of a sound scene is approached as a rationale for a novel multichannel mixing system. We propose that a sound scene offers an intuitive metaphor for representing and governing the multi-source audio processing, by allowing the user to interact with a graphical view of the scene where sound objects can be manipulated. The mixing process is controlled in terms of setting up the properties of every sound object and the environment and manipulating the sound scene in space (by annotating objects positions) and time (by annotating objects sounding times).

Although the concept of listener-based spatial scene authoring and rendering is employed in existing systems (e.g. Yamaha surround editor, IRCAM spat, Sony Vegas, etc.) most approaches are limited to control panning functions not integrating the project and scene properties down to the mixing engine.

In the following sections we present a brief overview of an ongoing project addressing the research and development of a novel sound-scene-oriented mixing framework. The underlying architecture and a simplified block diagram of a main rendering pipe are presented, as well as key features that are considered in the design.

2. System Architecture

The main underlying processing architecture adopted in this system design is the AUDIENCE, a modular architecture for auralization proposed by Faria (2005).

Conceived for the spatial audio processing chain, it encompasses the production stage and the final presentation of the spatial sound field rendered from a sound scene description. Briefly, four main groups of processes are executed in this chain. They are represented in table 1.

Table 1. Functional Layers

Scene production – L1 (Layer 1)
The first layer provides the adequate representation of the sound scene entities, such as the sound objects, the receiver (audition point) and environment attributes, including production scripts to control the performance evolution.
Acoustics and Effects Rendering – L2 (Layer 2)
The second layer implements acoustic rendering and simulation processes, such as reverberation, time delay, Doppler effects and sound attenuation due to distance.
Audio Program Generation – L3 (Layer 3)
The third layer gathers all spatial/temporal encoding modules and generates audio streams and final multichannel programs for distribution. Metadata data plus the the audio payload of sound objects are processed in this layer.
Program Monitoring – L4 (Layer 4)
The fourth layer contains modules for decoding formatted streams, and monitoring the multichannel audio.

In the present approach, the main underlying concept of *layer independence* of the reference architecture was straightforward taken into consideration to implement these functional layers. Functions not pertaining to these main processing groups are assigned to an auxiliary layer. The final system is not a particular spatialization effect but instead a framework to permit the integration of any spatialization and effect plugins that can be piped as a processing unit in a specific layer (e.g. the acoustic layer 2).

3. Implementation

The core of the new mixing strategy is a pipe of processes flowing from layer 1 down to layer 4, and has been prototyped in Pure Data [Pure Data 2010]. Several processes are considered in this pipe, and a typical flowchart of the most relevant ones is depicted in the figure 1, exposing the main modules distributed over the functional layers.

The modules in the first layer implement a simple view of a scene, where the user can freely move the receiver and sound objects, and also provide a simple timeline where transport messages of the sound source players of each object can be sequenced. Each change in the scene setup is transmitted to the L2 modules that render the acoustical simulation using the objects' spatial data, and then produce a render matrix (RM) with gains derived from an optimized implementation of the VBAP algorithm [Pulkki 1997]. We are particularly interested in 2D modes, as the 5.1 surround [ITU 2002] adopted as default multichannel mode in the Brazilian digital TV system [ABNT 2007].

The choice of the VBAP algorithm in this implementation is due to its low computational cost and flexibility to deal with non-regular speaker layouts, as the ITU 5.1. However other spatial rendering schemes can be applied (e.g. Ambisonics) requiring to

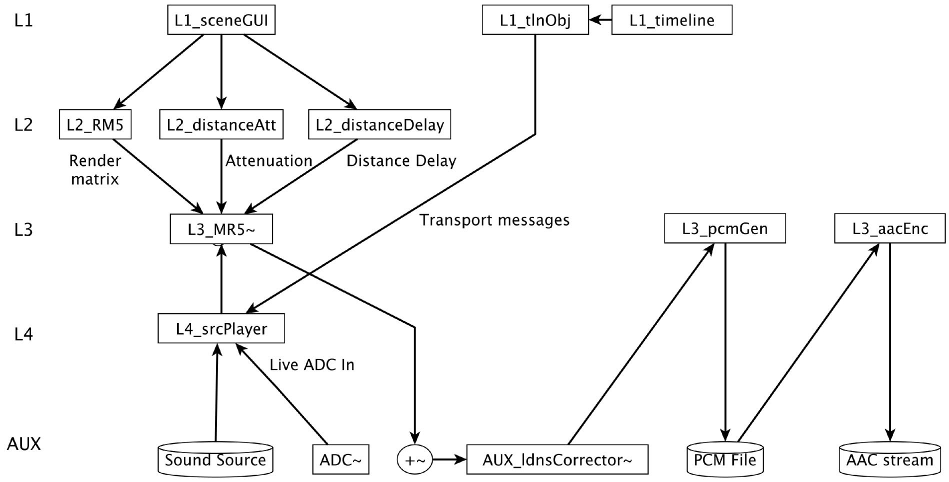

Figure 1. Example flowchart of one sound object mixing pipe

change only components in layers 2 and 3. To increase the smoothness of virtual sound sources, the RM generator was further modified to allow the control of sound sources spreading as in the MDAP approach [Pulkki 1999].

Figure 2 shows a segment of a Pd patch implementing this mixing pipe based on rendering matrices. The RM generated in the layer 2 is applied by the module L3_MR~ to the audio streams fed by the object players in L4. The spatially-encoded multichannel audio generated by all objects at L3 is mixed and applied to a loudness managing module at L4. Loudness management is a two-step process to measure and control the relative loudness of each sound object in scene and the unified loudness of the output multichannel program. The loudness measurement is compliant with the algorithm recommended by ITU BS.1770 [ITU 2006]. Additionally, a gating system was adopted in order to achieve consistent loudness measurement over audio streams with a variety of dynamic profiles [Grimm et al 2010].

4. Preliminary Results and Future Work

We currently achieved a beta-prototype of this sound-scene oriented mixing system, which has successfully passed through tests to validate acoustical rendering coherence, sound image stability and loudness management. Preliminary operational tests with professional sound engineers and artists have shown the advantages and challenges of adopting a scene oriented control paradigm. The realistic gain on spatial impression by using acoustic rendering techniques integrated with a intuitive mixing steering mechanism has been pointed out systematically, specially for the 5.1 layout.

The system is incorporating resources for intuitive loudness and sound space operation and user interfaces for aiding the sound engineer in creating sound scenes projects. Other important future development includes the integration of automatized resources, for tackling the dynamic of moving sources and automatic calibration of sound sources parameters to avoid typical studio problem, such as phase cancellation.

The most important features of this system might be its systematic decoupling of the processing modules and the sound realization of a scene regardless the output rig layout, through an adaptive recalibration of the multichannel program playback when the auditory acoustics and loudspeaker rigs change. There are relevant challenges regarding the complete operation of a mixing project with the new interfaces and its integration with

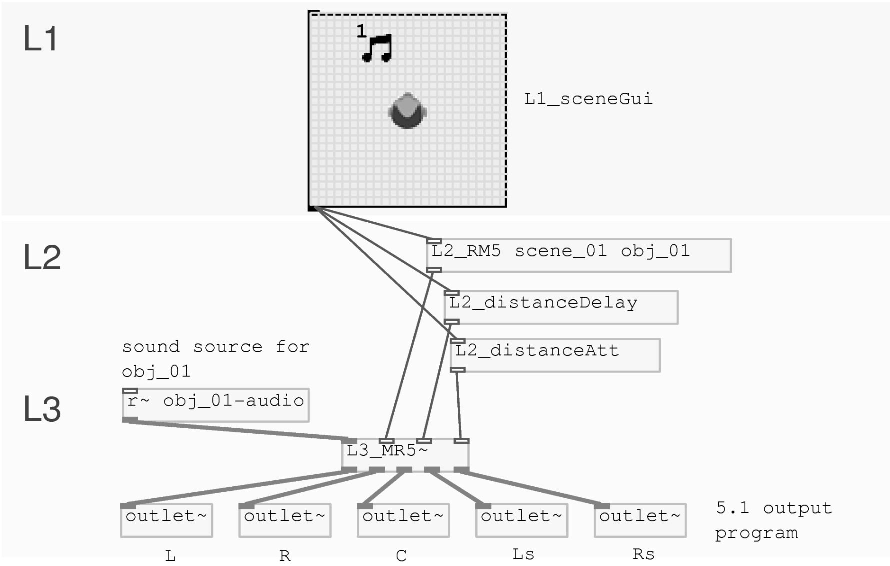

Figure 2. A simple rendering matrix engine implementation in the Pd prototype

track-oriented processes under a scene-oriented paradigm. A comprehensive technical approach and limitations evaluation is expected in a future paper.

5. Acknowledgements

This research has been supported by CNPq through the process 573424/2008-0.

References

Associação Brasileira de Normas Técnicas - ABNT. NBR 15602-2 (2007). Digital terrestrial television – Video coding, audio coding and multiplexing Part 2: Audio coding. Associação Brasileira de Normas Técnicas.

Faria, R.R.A. et al. (2005) “Improving spatial perception through sound field simulation in VR”. In: Proceedings of VECIMS 2005 - IEEE Conference on Virtual Environments, Human-Computer Interfaces and Measurement Systems, Giardini Naxos, Italy.

Grimm, E. , Skovenborg, E. and Spikofski, G. (2010) “Determining an optimal gated loudness measurement for TV sound normalization,” AES 128th Convention, London.

ITU-R. Recommendation BS.1770, Algorithms to measure audio programme loudness and true-peak audio level. International Telecommunications Union, 2006.

ITU-R. Recommendation BS. 775.1, Multi-channel stereophonic sound system with or without accompanying picture. International Telecommunications Union, 2002.

Pulkki, V. (1999) “Uniform spreading of amplitude panned virtual sources,” In: Proc. IEEE Workshop on Apps. of Signal Processing to Audio and Acoustics, New Paltz, NY.

Pulkki, V. (1997) “Virtual source positioning using vector base amplitude panning,” Journal of the Audio Engineering Society, 45(6):456–466.

Pure Data – Pd Community Site, http://puredata.info/, Access on Oct, 22, 2010.