VR-RoBoser: Real-Time Adaptive Sonification of Virtual Environments Based on Avatar Behavior

Sylvain Le Groux
Laboratory for Synthetic, Perceptive, Emotive and Cognitive Systems
Technology Department
Universitat Pompeu Fabra
Ocata, 1
08003 Barcelona, Spain
+34 935 421 366
slegroux@iua.upf.edu

Jônatas Manzolli
Interdisciplinary Nucleus for Sound Studies (NICS)
University of Campinas (UNICAMP)
Rua da Reitoria, 165. C.P. 6166
Campinas, 13091-970, BRAZIL
+55(19) 3521 4966
jonatas@nics.unicamp.br

Paul F.M. J. Verschure
Laboratory for Synthetic, Perceptive, Emotive and Cognitive Systems
ICREA & Technology Department
Universitat Pompeu Fabra
Ocata, 1
08003 Barcelona, Spain
+34 935 421 366
paul.verschure@iua.upf.edu

ABSTRACT
Until recently, the sonification of Virtual Environments had often been reduced to its simplest expression. Too often soundscapes and background music are predetermined, repetitive and somewhat predictable. Yet, there is room for more complex and interesting sonification schemes that can improve the sensation of presence in a Virtual Environment. In this paper we propose a system that automatically generates original background music in real-time called VR-RoBoser. As a test case we present the application of VR-RoBoser to a dynamic avatar that explores its environment. We show that the musical events are directly and continuously generated and influenced by the behavior of the avatar in three-dimensional virtual space, generating a context dependent sonification.

Keywords
Real-time Composition, Interactive Sonification, Real-time Neural Processing, Multimedia, Virtual Environment, Avatar.

1. INTRODUCTION AND MOTIVATION
It is widely acknowledged that music is a powerful carrier of emotions [1, 2], and that audition can play an important role in enhancing the sensation of presence in Virtual Environments [3]. Yet, music generation and sonification for environments such as games have received relatively little treatment in comparison to the sophistication of real-time 3D graphics rendering. Most often musical events are simple repetitive sound samples triggered at specific stages of the game combined with a looping background musical score. There is no linear continuity in the way the musical discourse follows the game evolution. Being aware of this insufficiency, the game industry has started to show a growing interest for adaptive and interactive audio [4]. It is calling for a model of sonification where background music reacts appropriately to gameplay thereby improving the overall gaming experience [5].

The automatic generation of music has a long history intimately related to the invention of automated instruments ranging from the music box, the Yamaha Disklavier, to MIDI-controllable robots such as the Lemurbots [6]. Those automated instruments can be used as playback machines but also as part of interactive music systems. For instance EyeWeb [7] integrates movement, music and visual languages in a multimodal perspective. In these approaches the transformation of sensory data to sound is often mostly driven by a reactive approach. Basically, the space of possible interactions is previously labeled with particular sonic reaction where, for instance, one specific gesture will trigger one specific sonic reaction.

In contrast to these approaches our project focuses on interactive composition. Our aim is to integrate sensory data from the environment in real time and interface this interpreted sensor data combined with the internal states of the control system to a composition engine. In this way unique emergent musical structures can be generated. In our previous work on Roboser, a Real-World composition system, we have shown how the dynamics of a real-world system induces novelty in the micro-fluctuations of sound control parameters [8]. Here our goal is to use the Roboser framework to transform the interplay between and avatar and the environment into new musical structures. In other words, sonic structures emerge as result of the continuous interaction with a dynamic environment.

In the past, we have explored several aspects of such systems for interactive composition, interactive dance, and communication between humans and large-scale interactive installations [9]. We used the Roboser system as the main controller of interactive soundscape that expressed the behavioral mode and emotional state of Ada, an immersive interactive space that has been exposed to the public as part of the Swiss national exhibition Expo.02 [9].

Here we describe an application of RoBoser, called VR-RoBoser, where the complex and dynamic sensor data “perceived” by an avatar in a 3D virtual environment is processed by a neural control system, that transforms this dynamic sensor data into meaningful cues on the behavior of the avatar that is used to control a real-time interactive composition engine. In this way a dynamic and context dependent sonification of the interaction is autonomously generated.

2. SYSTEM OVERVIEW

Real-time operations performed by real-world music systems can generally be described in terms of a sensor interface, a central-processing stage, and a transformation into sound [10], also referred to as sensing, processing, and response [11].

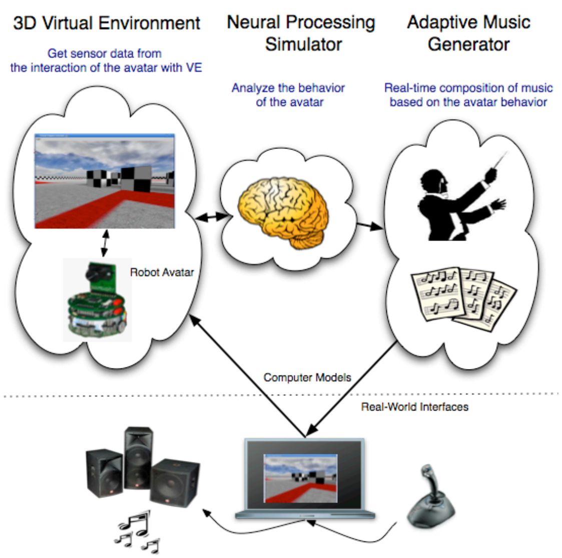

Figure 1. The system follows the Sensing/Processing/Response architecture

The system we implemented follows this decomposition and is made up of three main interconnected modules (see Figure 1). The first module consists of a virtual environment and an avatar "behaving" in this environment. The second module consists of a neural network simulator, which goal is to analyze the data "sensed" by the avatar in its environment. Finally, the third module deals with the interactive composition of musical structure and the generation of sounds. The musical output "illustrates" the behavior of the virtual robot/avatar in the environment.

2.1 The Virtual Environment

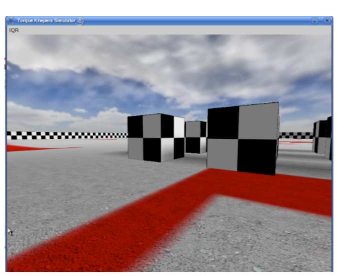

Figure 2. 3D Virtual Environment simulated with Torque. The camera view is that of an avatar that is exploring this environment.

We used the Torque Game Engine SDK 1.4 [12] as our main tool for the creation of the virtual environment. The virtual space we built is a three-dimensional bounded space where we have placed some obstacles (walls) and have painted color marks on the ground (see Figure 2)

2.2 The Robot Avatar

Here, we define an avatar as the virtual entity moving in the 3D virtual space while being controlled by a user in the real world. In comparison a robot avatar is behaving autonomously in the space and no user input is given. The robot avatar living in the virtual space is equipped with a set of sensors that detect properties of the virtual environment, and effectors that act upon this environment. The disposition of the sensor system is loosely modeled after the Khepera robot [13]. The avatar is endowed with eight distance sensors disposed as shown in Figure 3 and a video camera. Additionally, by sending the appropriate control parameters to the virtual Khepera Engine, we can control the avatar's motion upon the two-dimensional surface of the environment's ground. We then analyze the sensor data to extract meaningful information.

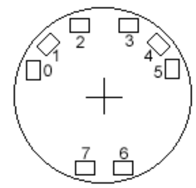

Figure 3. Distance Sensor Topology for the Robot Avatar

2.3 A Real-Time Neural Network Simulator

A simulation environment named IQR [14] handled the real-time processing of sensor inputs. IQR is a flexible tool for creating and running simulations of large-scale neural models and provides a user-friendly graphical interface to design multilevel neural models (see Figure 4.). In IQR, the top level is the system, which contains an arbitrary number of processes, and a process can comprise several cell groups, each group being defined as a number of neurons of identical type. Processes, or rather the groups within a process, can be connected to a module and a module can exchange data with an external entity. This can be a sensor (camera, microphone, infrared, ultrasounds, among others), an actor (robot, motor) or an algorithm.

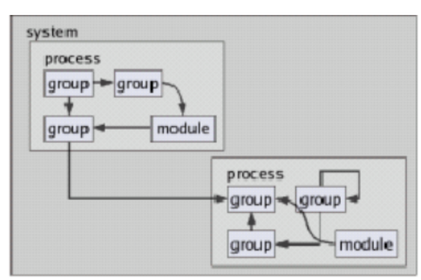

Figure 4. Structure of the neural model in IQR: In this example we distinguish two processes that each contain a neuronal circuit made up of three neuronal groups. One of these groups receives input from a module that is a user defined shared object that is dynamically linked to the simulation.

Connections are used to feed information from group to group and connections can be made between two groups within a process as well as between two groups from different processes. Due to the prime importance of connectivity for neuronal computation, IQR offers various possibilities to define and influence synaptic connections.

For our project, we used IQR to transform the avatar's sensor data into higher-level relevant control parameters for an interactive music generator modeled after the Roboser [8]. IQR allowed us to model a biologically plausible neural network for sensor data analysis. Musical parameters are assigned to specific neuronal groups, which activity controls the musical output. (See Section 4. for a more detailed description of the mapping between sensor information and musical control parameters).

2.4 Music Generator

Roboser [8] is a real-time music composition and performance system that accepts input from a variety of sources to guide a composition process. Input activity ranges from raw sensory data (for example, video images, audio events, and floor load) to high-order control parameters like biased neural oscillators, models of circadian rhythms, or models of behavior control. For this project, Roboser received real-time input from the IQR controller system. Technically, the Roboser composition engine synthesizes a stream of MIDI data upon simulated neuronal input. Roboser composes music on up to 12 performance tracks. This is similar to tracks of a multi-track tape recorder, the difference being that the performance of every track is synthesized as well as performed in real time.

Musical parameters that are interactively and independently controlled in each track include the MIDI parameters for instrument, velocity, volume, pitch bend, tempo, and articulation. In addition, predefined fragmented note sequences, rhythm lines, and note onset dynamic sequences are interactively selected for each single performance track. Each track's output is delivered on a single MIDI channel. These sound control parameters are described by a Style File in which the musical parameters are stored and used for real time interaction.

2.5 Sound Synthesis

The composition engine Roboser sends MIDI data to the sound generation modules. The modules can be constituted of a standard sampler (in our case an AKAI S6000 midi stereo professional digital sampler) and/or of synthesizers. When the use of standard samples was not desired, we designed our own sound synthesis patches within the Pure Data environment [15].

2.6 Communication Protocol

The different modules of the projects are interconnected and communicate information via the TCP/IP protocol thanks to two external IQR modules called NetSend and NetReceive. This design allows us to distribute the computational load and run different processes on different computers on a local network. In our setting, the virtual environment and data analysis were carried out on a first computer, while the generation of musical structures and sound synthesis were accomplished on a second one.

3. MAPPINGS

The mappings from the sensing, neural processing and composition modules are presets. The fine adjustment of the different mappings and parameters is a heuristic process. It is one of the main places where the designer of the system has the liberty to express her creativity. For instance in Ada we defined compositional heuristics that were psychoacoustically mapped to emotional experience [9].

3.1 From Sensor Data to Behavioral Information

We carried out the processing of the sensor data within the IQR neural network simulator. By injecting the visual sensor input (in the form of live RGB data) to specific neural network topologies, we were able to detect edges, motion and track different colors as well as luminance in real-time (see Figure 5.). Similarly, feeding the distance sensor data to an appropriate network structure, we were able to infer information about the direction, position of the avatar, as well as to detect collisions with obstacles in the 3D virtual environment.

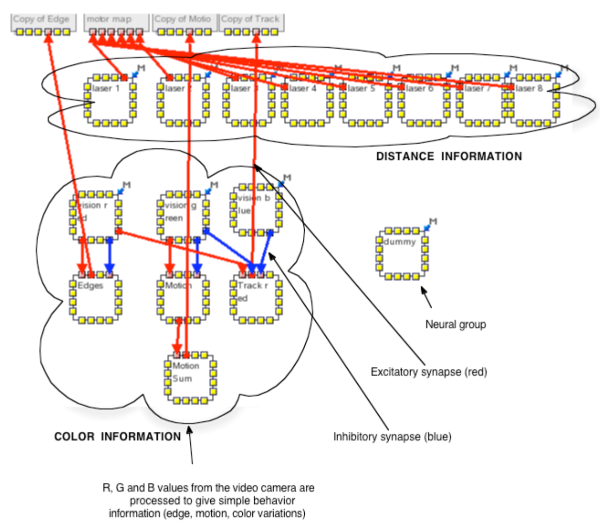

Figure 5. In the IQR simulator, we send sensor data from the robot avatar to specific neuronal groups designed to detect edges, colors and motion.

3.2 From Behavioral Information to Compositional Parameters

To transform the neural data into musical structures, we used a simple set of MIDI parameters that serve as controls to a sampler or a real-time synthesizer in Pure Data [15]. Each real-time MIDI event is defined by a set of predefined note number, velocity change, duration patterns, and additional global parameters such as BPM, pitch shift and volume. The changes in parameters are driven by the state of the avatar within its environment. This produces an interactive music composition reacting in real-time to the avatar's behavior. For this project, the detection of a specific color was mapped to a change of predefined musical style file. We used the luminance to vary the brightness of the synthesized sounds. Motion was mapped to a global increase of tempo. Collisions of the robot with a wall triggered a sudden increase of tempo and volume as well as high pitch shifting. The proximity of an obstacle was mapped to an increase of the global volume (see Table 1.)

Table 1. Mappings between avatar behavior and musical parameters

Behavior	Musical Parameters
Motion Variation	Tempo Variation
Proximity of Obstacle	Increase the Global Volume
Detection of Color	Change of Musical Style File
Collision	Sudden Increase of Tempo, Volume and high pitch shifting
Luminance Variation	Sound Brightness Variation

4. SCENARI

This automatic composition system can be used in three different modes. In the main mode, the music "illustrates" the interaction of the user with the environment in order to improve the sensation of presence in a video game. The user controls the avatar's trajectory in the virtual environment via a game controller such as a joystick, and the music is responsive to the action of the user. In the second mode, the user consciously modulates the result of the music sonification by directing the behavior of the avatar. The joystick can be understood as a gestural control and the whole system is a computer-based musical instrument played by the user. Finally, the last mode is that of an autonomous avatar, or robot avatar, that moves around in the space on its own. To this scenario corresponds a never-ending automatic generation of a musical stream based on the autonomous trajectory of the robot. In our current implementation, the robot behavior is simplistic and based on avoidance of collisions, but the model of behavior can be refined and would produce more complex and interesting sonic structures as for instance in Emotobot [8].

5. CONCLUSIONS AND FUTURE WORKS

In this paper, we have presented a complete system that can generate complex sonification of a Virtual Environment based on the behavior of an avatar in that space. We proposed an architecture that allows for seamless communication and interaction in real-time between a virtual world, a neural network simulator for sensory data analysis and a music generator. The results are promising but we can see several research directions that would prove useful for further enhancements to our system.

One of the most interesting possibilities is to operate on a higher level of integration between the neural control and the musical structure. We want to empower the use of the Distributed Adaptive Control, or DAC, methodology [16] as an interactive music composition strategy called behavioral music composition. DAC distinguishes three levels of control in a cognitive behaving system: reactive, adaptive and contextual. In the current system, the mappings are made directly from the sensor data. By relying on the DAC paradigm to extract the avatar's mood and emotional states from its sensory inputs, we can use this higher-level information as a composition parameter as shown in [8].

We would also want to use adaptive sonification in games to induce a specific emotional state in the player in relation to her behavior in the virtual world. For this purpose, a complete and detailed study of the effect of certain musical parameters on the emotional state of a person is necessary.

At the software level, one of our next objectives is to integrate in one single package: composition, sound synthesis and analysis, MIDI control and music visualization. This system will be applied to a large-scale immersive mixed reality space to sonify the interaction between physically present humans, virtual humans and synthetic characters.

6. ACKNOWLEDGMENTS

The authors thank Sergi Bermudez, Ulysses Bernardet and Armin Duff for their support in developing and designing the different parts of the system, as well as anonymous reviewers for helpful comments and suggestions. This project is supported by PRESENCCIA (FP6 IST-FET 27731).

7. REFERENCES

[1] P.N. Juslin, and J.A. Sloboda, Music and emotion: theory and research, Oxford University Press, Oxford; New York, 2001.

[2] L.B. Meyer, Emotion and Meaning in Music, University Of Chicago Press, 1961.

[3] R.H. Gilkey, and J.M. Weisenberger, The Sense of Presence for the Suddenly-Deafened Adult: Implications for Virtual Environments. Presence 4 (1995) 357-363.

[4] The Interactive Audio Special Interest Group. http://www.iasig.org/.

[5] W. Guy, Design With Music In Mind: A Guide to Adaptive Audio for Game Designers. http://www.gamasutra.org.

[6] E. Singer, J. Feddersen, C Redmon, B. Bowen. LEMUR's Musical Robots. Proceedings of the 2004 Conference on New Interfaces for Musical Expression, Hamamatsu, Japan.

[7] A. Camurri, B. Mazzarino and G. Volpe. Analysis of Expressive Gesture: The EyesWeb Expressive Gesture Processing Library. 2003

[8] J. Manzolli, and P.F.M.J. Verschure, Roboser: A Real-World Composition System. Comput.Music J. 29 (2005) 55-74.

[9] K. Wassermann, K. Eng, and P.F.M.J. Verschure, Live Soundscape Composition Based on Synthetic Emotions. IEEE Multimedia 10. (2003) 82-90

[10] M.M. Wanderley, N. Schnell, and J.B. Rovan, ESCHER - Modeling and Performing Composed Instruments in real-time, IEEE Int. Conference on Systems Man and Cybernetics, San Diego - CA, USA.

[11] R. Rowe, Interactive music systems: machine listening and composing, MIT Press, Cambridge, MA, USA, 1992.

[12] Garage Games, Torque Game Engine. http://www.garagegames.com.

[13] K. Team, K-Team Corporation. Lausanne, Switzerland.

[14] U. Bernardet, M. Blanchard, and P.F.M.J. Verschure, IQR: a distributed system for real-time real-world neuronal simulation. Neurocomputing 44-46 (2002) 1043-1048

[15] Pure Data: http://www.puredata.org

[16] Verschure, P.F.M.J and Voegtlin, T and Douglas, R.J. Environmentally mediated synergy between perception and behaviour in mobile robots, Nature, 425: 620-624, 2003