A VR Interface for Collaborative 3D Audio Performance

Martin Naef, Daniel Collicott
Digital Design Studio,
The Glasgow School of Art
Glasgow, UK

[m.naef | d.collicott]@gsa.ac.uk

ABSTRACT

This paper presents a novel interaction paradigm to support musical performance using spatial audio. This method reduces the interface bottleneck between artistic intent and spatial sound rendering and allows dynamic positioning of sounds in space.

The system supports collaborative performance, allowing multiple artists to simultaneously control the audio spatialization. The interface prototype is built upon standard virtual reality software and user interface technology. Tracked data gloves are used to manipulate audio objects and stereoscopic projection to display the virtual 3D sound stage.

Keywords

Spatialized audio, virtual reality interface, 3D audio performance.

1. INTRODUCTION AND MOTIVATION

Contemporary artists have explored the power of electroacoustic spatial audio over the last few decades. Composers such as Edgar Varese or Karlheinz Stockhausen have used multi-speaker systems to immerse the audience in artificial sounds, extending upon earlier ideas of positioning the musicians freely around the audience. Today, an increasing number of artists are regularly using multi-channel speaker arrays to immerse the audience even further. Still, spatial audio is far from becoming mainstream.

While today's digital systems support matrix mixing to enable many sound sources and complex speaker arrangements, the user interface severely limits the performer and composers. There is little support to position sound sources in space beyond the traditional 2D joysticks on high-end mixing desks with surround capabilities; 3D capabilities are almost unheard of. Instead, a performer has to manually distribute sounds to speakers, typically using traditional channel-fader interfaces. Dynamic sound sources require complex automation, rendering live performance almost impossible. Compositions and recordings are limited to playback systems with closely matched speaker arrangements.

This situation arises because of the high dimensional requirement placed upon the interface: the intrinsic physicality of spatialized sound sources is best represented by four dimensions: three for position and one for source volume. If, for the sake of the discussion, we supposed the artist used eight speakers in a performance, they would be attempting to control an intrinsically four dimensional representation with eight channel faders.

Analyzing this interaction in more detail, the input device space has eight dimensions, with faders being operated for the most part sequentially; representation feedback is poor (the slider positions themselves) and related through a counter-intuitive mapping to the underlying physical representation, which is itself four dimensional. In effect we have created an 'interaction bottleneck': dimensionally impoverished control devices act upon an inefficient and counter-intuitive mapping, the only feedback and representation being the input device itself. Such an interface could not easily support multiple dynamic sound sources.

The project presented here tackles the bottleneck in two ways. Firstly, by separating data and control representation; the concept of a sound source in space is abstracted from the actual mixing process onto the output audio channels. Secondly, via a richer input device and simple mapping: the abstract sound sources are moved in space using a 3D input device; the mapping is now intuitive, directly relating the visualization to the sound spatialization, with source volume being represented by the orientation of the geometric sound object.

A first performance prototype was implemented to test the concept, provide a user interface and support collaborative performance. The underlying technology is derived from existing virtual reality technology and adapted to meet the requirements of music performance. Future work will aim to extend the capabilities of the initial prototype beyond limitations imposed by a VR system that is optimized for realism in simulations instead of artistic expression.

The remainder of the paper is organized as follows. Related work is acknowledged in Section 2. Section 3 then presents the system overview including all the components. Section 4 presents the interaction paradigm. Section 5 presents the extension of the interface for collaborative performance. Conclusions are drawn in Section 6 and future research directions are identified.

2. RELATED WORK

This work combines the fields of spatial music performance with virtual reality technology. It builds upon technology and interaction paradigms developed independently in the two fields.

There is a wealth of research literature concerning novel musical interface devices; [14] includes a comprehensive overview. Many artists have successfully employed VR-gloves and non-contact sensing in live performance; examples of early innovation being the work carried out at STEIM, and by Jaron Lanier and Tod Machover.

In the areas of mapping [3],[5] and visualization of sound, experimental research is relatively sparse, with mapping research also tending to focus on instrumental richness, rather than the intelligibility of the mapping.

NIME 06, June 4-8, 2006, Paris, France.

Historically, the use of VR as an expressive medium has concentrated on visual art, e.g. [8]. To date, there has been little experimental research on the use of VR interfaces for sound control [1],[10]. The DIVA system [6] was among the first to use VR for musical performance. A VR evaluation framework has been built at the Helsinki University of Technology, introducing concepts such as a virtual air guitar [9].

Spatial perception and rendering of sound is well understood, ranging from amplitude panning approaches [16], Ambisonics [4] to large speaker arrays for rendering wave fields or head-related transfer function methods predominantly used with headphones. Virtual reality toolkits usually include some form of spatial sound rendering, often based upon standalone audio servers or DSP hardware (e.g. Lake Huron). These VR systems employ sound as a tool to achieve a realistic simulation and increase the feeling of immersion. Little emphasis, however, has been given to tackling the difficult interaction issues surrounding the visualization and control of sound spaces for artistic purposes.

This paper addresses this area of interaction design, offering intuitive post-processing of the sound stream for spatialization.

3. SYSTEM OVERVIEW

The system consists of three major components: the scene-graph that stores the positions and parameters of the audio sources including their visualization; the user interface to modify the scene; and the audio rendering system (see Fig. 1). The audio rendering system as well as the "glue"-code required to combine the elements is provided by the blue-c API [12], a virtual reality toolkit originally designed for collaborative and tele-presence applications.

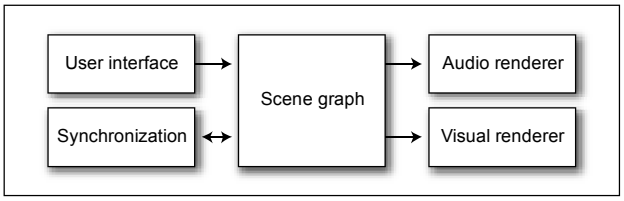

Fig. 1. System overview.

3.1 User Interface

The user interface module handles all user interaction with the scene graph and its embedded audio objects. The interface module enables the user to pick and move objects using tracked virtual reality data gloves. Rotating objects changes the volume of the point source; this user interface is described further in section 4. The interface also handles object locking to avoid concurrency problems when the system is used collaboratively.

3.2 Audio Rendering System

The audio rendering system spatializes the audio source objects using a volume panning approach, deriving the data from the scene graph. All audio sources are rendered using the blue-c API sound rendering system [13] that supports spatialization of a large number of sound sources with arbitrary speaker configurations. The audio system supports audio file playback either from memory (e.g. for short loops), streaming from disk (e.g. longer audio tracks) or from live sources (e.g. microphones or synthesizers). The audio rendering system was chosen mostly for convenience reasons as it performs well and is directly integrated into the virtual reality software development toolkit. Plugging in a different audio renderer or transmitting audio source positions to a different spatialization server (e.g. a Lake Huron system) would be straight forward.

3.3 Hardware Environment

The prototype implementation uses a standard virtual reality environment with a single wall-type stereoscopic projection surface, head- and hand-tracking, gloves with bend sensors for all fingers (see Fig. 2b) and a 14 speaker audio rendering system. Additional tests were conducted on a stereoscopic workbench environment with a slightly smaller 8 channel audio system (see Fig. 2a). Both systems are driven by a standard PC.

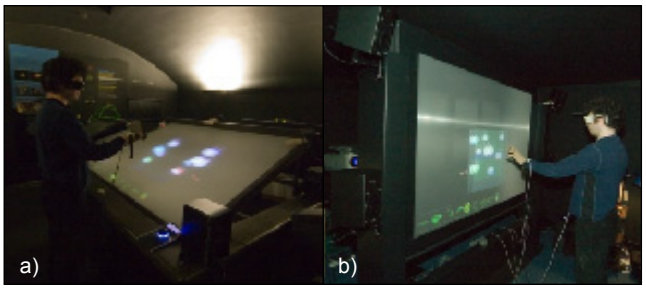

Fig. 2. Pictures of the hardware setup used for testing.
a) Workbench display system. b) Wall display system.

Although technically not required, picking in 3D is significantly easier with stereoscopic projection. Similarly, using a 6DOF tracking system allows for much smoother and more intuitive interaction than using a mouse that is inherently 2D.

The hardware environment was chosen based on availability at the lab. It is obviously not well-suited to live performances due to portability restrictions. The software, however, is flexible enough to run on a variety of platforms, including laptops.

3.4 Scene Graph: Representing 3D Audio

The scene graph is subdivided into a static and a dynamic section. The dynamic section (Audio group in Fig. 3) includes the sound sources and their visual representation. This dynamic section is synchronized and distributed among the different machines in the collaborative setting (see section 5). The static section (Stage and UI groups in Fig. 3) is used to provide the performer with guidance elements, such as a reference coordinate system, the position of the speakers, or an abstract representation of the performance environment and user interface elements.

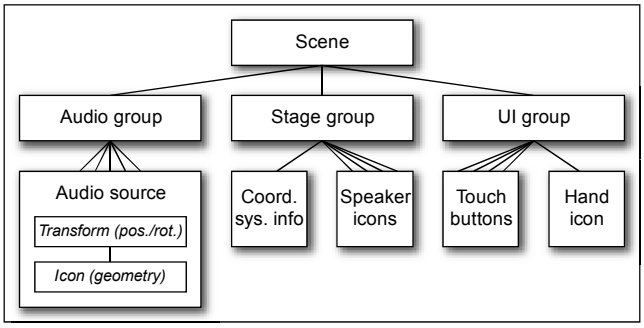

Fig. 3. Scene graph structure.

Each sound source is represented as a simple geometric object. The associated structure in the scene graph consists of the position transformation node at the top, an additional rotation transformation node and attached geometry. The

position given by the transformation node is transferred to the audio renderer to control spatialization of the respective audio object. The rotation transformation affects only the small pointer visually. The roll component is transformed into a gain value. The current prototype only implements gain control; the two additional rotation axes, however, could be used to control source-dependent effects (e.g. reverb send, filter frequency, etc.). Care has to be taken not to overload the user interface, though. The use of a purely geometry-based approach exploits the scene synchronization features of the underlying VR toolkit without requiring additional modification.

4. THE PERFORMANCE INTERFACE

The performance interface concept is derived from a 3D visualization and interaction environment "AutoEval" [1] originally developed for design review in the automotive industry and adapted to the needs of 3D audio manipulation.

4.1 Glove Interface

All editing is conducted using a tracked virtual reality glove. For the prototype implementation, we used a Polhemus Fastrack 6DOF magnetic tracking system and an Immersion CyberTouch glove with vibration devices on the fingers and palm to provide haptic feedback.

Editing operations are initiated by picking an audio object with the tracked glove. Picking virtual objects in 3D is often difficult for the untrained user, especially for those with limited stereoscopic depth perception. The interface system therefore supports the user with additional cues: if the user touches an object, the object is visually highlighted and small vibration motors inside the VR glove (similar to those in mobile phones) provide a haptic sensation. A touched object can be picked by pinching the thumb and index finger.

The audio object is moved by picking and dragging the object to the desired location. The attached audio source is continuously updated during move operations, enabling the performer to "animate" sound in real-time.

Volume is changed by rotating the hand while the object is grabbed, which is essentially the same gesture as turning a physical knob. While the move operation follows a 1:1 mapping, twisting the knob is accelerated by a factor of two. Tests revealed that a scaled mapping reduces fatigue without significantly sacrificing precision. Additional editing modes are available to avoid accidental side effects.

4.2 Latency Issues

In this system, latency is introduced at various stages. The audio rendering engine itself can run with the smallest possible buffer size allowed by the audio interface. Significant latency, however, is introduced through the user interface handling that runs synchronized to the visual rendering system, with frame times typically between 16 to 22 ms. Magnetic tracking systems also introduce a delay due to their limited update rate and required noise filtering. All factors included, the time delay between an actual event and its effect on the audio output may well go beyond 50ms. It is therefore clear that this interface is not suited to control a percussion performance. Preliminary tests, however, suggested that the system is fast enough for all practical uses.

4.3 Working Volume

Previous research suggests that the optimal actual working volume of the hands is relatively small compared to the volume defined by the fullest reach and is maintained at a fixed distance relative to the body. The standing position was chosen to enhance performance aesthetics; in this position, fine motor control is best achieved with a hand position 50-100mm above elbow height and within the 'normal working area' [15] which equates to approximately one forearm's span from the body. Hence the working volume chosen is a cube of approximately 0.4m centered directly in front of abdomen, and the audio scene is scaled accordingly. In this region the most accurate motor control is achieved and additionally, muscle fatigue is greatly reduced. In contrast, typical wall-type VR display environments are designed for a large interaction volume, and consequently offer less precision for a small interaction volume as used for this prototype. The workbench type display fits the preference for a smaller working volume better and proved to be less tiring to work with.

5. COLLABORATIVE PERFORMANCE

The system supports collaborative performance in two ways. By using the live audio streaming sources, a performance can be split along functional lines: one performer is responsible for the spatialization of the sounds from the other musicians, much in the same way that a front-of-the-house mixing engineer takes care of the band's sound (Functional separation in Fig. 4). This option is not further discussed here.

The system also supports concurrent editing of the 3D sound stage by several users (Parallelization in Fig. 4). For this type of operation, the virtual sound stage is distributed and synchronized among several computers, one per spatial audio performer. 3D audio objects can be edited independently on all connected computers, while a locking system ensures that no two users try to modify the same object at the same time.

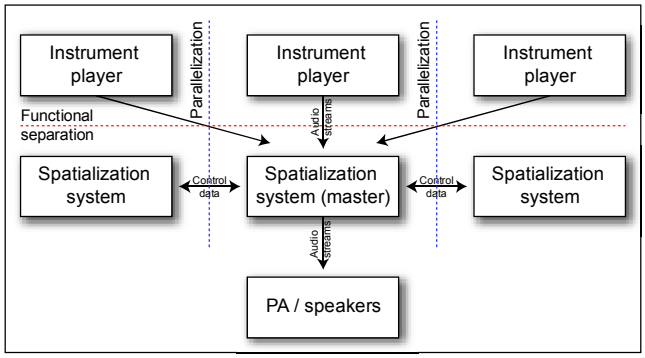

Fig. 4. Collaborative performance. Functional (sequential) parallelization between instrument players and spatialization vs. parallelization within the spatialization post-process.

5.1 Synchronization between Machines

The system enables several users to edit the sound stage concurrently. All performers use their own computer, each displaying the stage from an arbitrary viewpoint and polling the input device. Each machine holds an independent copy of the stage group and user interface, and the software ensures that all machines share the same position and state for all sound sources. This essentially provides multiple instances of the user interface to a single audio rendering system.

The synchronization is based upon the blue-c Distributed Scene Graph (bcDSG) [11] that synchronizes the scene graph data structure across multiple machines and manages concurrency issues including locking to make sure no two users can modify the same object concurrently. Although the bcDSG was designed with graphical applications in mind, the

synchronization system did not require any adaptation since all audio states have a direct scene graph representation. Building a distributed application therefore required only little additional development effort over a single-user solution.

5.2 Distributed Audio Rendering

In a typical live performance situation, only a single computer actually processes audio data; the other machines are used for visualization and interaction only. If desired, every machine in the distribution group could run their own local audio renderer with an arbitrary speaker setup, providing individual monitoring for each performer.

6. CONCLUSIONS AND FUTURE WORK

The work presented here forms the first step towards an intuitive performance system for spatial audio and music performance. We demonstrated the usefulness of virtual reality tools in the context of music performance, and introduced a first concept prototype for visualizing sound sources in a 3D environment.

The interaction paradigm has proven its effectiveness over previous multiple-fader techniques; due to the improved efficiency of the interface, one performer is now able to sequentially alter the spatialization of multiple sources, whereas previously this sequential or 'time-sharing' capability had been consumed by poor interaction. In addition, the collaborative interface allows parallel spatialization of sound sources with multiple performers. Thanks to scene distribution features inherent in the underlying VR toolkit, enabling a collaborative performance only required minimal additional development effort.

Additional work will be required to increase the dimensionality of the control interface. The current system only supports position and gain parameters, whereas a fully fledged performance system should include effect control. Quantitative and qualitative HCI testing will be used to determine which interaction paradigm(s) represent the most expressive musical interface.

7. ACKNOWLEDGEMENTS

Thanks to Alistair MacDonald for fruitful discussions about 3D audio performance, to ETH Zurich for providing the blue-c API, and the developers of the original AutoEval system at the Digital Design Studio for inspiration of the user interface paradigm.

8. REFERENCES

[1] Anderson, P., Kenny, T., Ibrahim, S. "The role of emerging visualisation technologies in delivering competitive market advantage." *2nd International Conference on Total Vehicle Technology*, Institute of Mechanical Engineers, University of Sussex, Brighton, UK, pp. 87-97, 2002.

[2] Choi, I. "A Manifold Interface for Kinesthetic Notation in High-Dimensional Systems." in *Trends in Gestural Control of Music.* Battier and Wanderley, eds., IRCAM, Paris. 2000.

[3] Garnett, G., Goudeseune, C. "Performance Factors in Control of High-Dimensional Spaces." In *Proc. 1999 Int'l Computer Music Conf.* San Francisco. 1999.

[4] Gerzon, M. "Ambisonics in multichannel broadcasting and video." *Journal of the Audio Engineering Society 33, 11*, 859–871. 1985.

[5] Hunt, A., Wanderley, M., Paradis, M. "The Importance of Parameter Mapping in Electronic Instrument Design." In *Nime 2002 Proceedings*. 2002.

[6] Huopaniemi, J., Savioja, L., Takala, T. "DIVA virtual audio reality system." In *Int. Conf. Auditory Display*, Palo Alto, CA, 111-116, 1996

[7] Jot, J.-M. "Real-time spatial processing of sounds for music, multimedia and interactive human-computer interfaces." *Multimedia Systems 7, 1*, 55–69. 1999.

[8] Keefe, D., Acevedo, D., Moscovich, T., Laidlaw, D.H., LaViola, J. "CavePainting: A Fully Immersive 3D Artistic Medium and Interactive Experience." In *Proceedings of ACM Symposium on Interactive 3D Graphics 2001*, pages 85-93, March 2001.

[9] Mäki-Pataola, T. "Experiments with Virtual Reality Instruments." *Proceedings of the 2005 International Conference on New Interfaces for Musical Expression [NIME05].* Vancouver, BC, Canada, 2005.

[10] Mulder, A. "Design of Virtual Three-Dimensional Instruments for Sound Control." PhD Thesis, Simon Fraser University. 1998.

[11] Naef, M., Lamboray, E., Staadt, O., Gross, M. "The blue-c distributed scene graph." In *Proceedings of the IPT/EGVE Workshop 2003.* J. Deisinger and A. Kunz, Eds. 2003.

[12] Naef, M., Staadt, O., Gross, M. "blue-c API: A Multimedia and 3D Video Enhanced Toolkit for Collaborative VR and Telepresence." *Proceedings of ACM SIGGRAPH International Conference on Virtual-Reality Continuum and its Applications in Industry 2004*. VRCAI 2004, Singapore, June 16-18, 2004.

[13] Naef, M., Staadt, O., Gross, M. "Spatialized audio rendering for immersive virtual environments." In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology 2002*. H. Sun and Q. Peng, Eds. ACM Press, 65–72. 2002.

[14] Paradiso, J. "Electronic Music Interfaces: New Ways to Play." *IEEE Spectrum*, *34(12)*, 18-30, 1997.

[15] Pheasant, S. "Bodyspace: Anthropometry, Ergonomics and the Design of the Work." pp52-57, CRC; 2nd edition. 1996.

[16] Pulkki, V. "Uniform spreading of amplitude panned virtual sources." In *1999 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics.* 1999.