J Multimodal User Interfaces (2011) 4:147–156
DOI 10.1007/s12193-011-0069-1

ORIGINAL PAPER

Multi-modal musical environments for mixed-reality performance

Robert Hamilton · Juan-Pablo Caceres · Chryssie Nanou · Chris Platz

Received: 18 January 2011 / Accepted: 11 October 2011 / Published online: 5 November 2011

Abstract This article describes a series of multi-modal networked musical performance environments designed and implemented for concert presentation at the Torino-Milano (MiTo) Festival (Settembre musica, 2009, http://www.mitosettembremusica.it/en/home.html) between 2009 and 2010. Musical works, controlled by motion and gestures generated by in-engine performer avatars will be discussed with specific consideration given to the multi-modal presentation of mixed-reality works, combining both software-based and real-world traditional musical instruments.

Keywords Music · Virtual environments · Mixed-reality · Multi-modal

1 Introduction

Musical performance and the ways in which musicians interact across physical space have traditionally been constrained by physicality: both the innate physicality of sound's motion through air as well as the ability to attend to and communicate with motion and gesture by performer and audience alike. Time and physical space are naturally intertwined, as latencies tied to the motion of sound and light comprise a very real component of musical experience. As such, musical performance practices have evolved throughout history taking advantage of our natural understanding of the physical world, with the evolution of specific meta-languages of musical gesture and subtle communication as the end result.

Flashing forward to a Twenty-First century culture awash in connectivity and instant global communication, musical performance can be liberated from the constraints of static physical location, moving beyond uni-directional musical broadcasting and streaming with fully-interactive poly-directional networked streams of audio, video and control data. Networked musical performance has become a reality thanks to high-quality and low-latency multi-channel audio streaming solutions across research grade and commodity networks [5]. The physicality of distance, still constrained by the laws of physics and the unwavering speeds of sound and light, has been tamed to an extent that acceptable latencies for the transmission of sound and music are routinely achieved. As such, distanced networked performance of traditional analog instruments is not only possible but becoming increasingly common [1, 31].

The use of computer-based three-dimensional rendered multi-user environments and the implementation of enactive or gesture-based musical control systems within such environments offers a multi-modal approach for the connection of and interplay between distanced spaces. Streaming media technologies allow networked or telematic musical performances to span multiple disparate physical locations, conceptually creating an additive musical space existing as a sum of each location. Multi-user rendered environments allow users to interact and communicate in the context of an additional dimensionality, combining the additive musical space of telematic performance with a shared and perceivable space simultaneously belonging to all participants and audience members alike.

R. Hamilton (✉) · J.-P. Caceres · C. Nanou · C. Platz
Center for Computer Research in Music and Acoustics (CCRMA), Stanford University, Stanford, CA 94305, USA
e-mail: rob@ccrma.stanford.edu

J.-P. Caceres
e-mail: jcaceres@ccrma.stanford.edu

C. Nanou
e-mail: chryssie@ccrma.stanford.edu

C. Platz
e-mail: chrisplatz@ccrma.stanford.edu

By combining virtual performers with live musicians, a hybrid or mixed-reality musical environment is created, allowing performers and audience members to combine, perceive and attend to musical stimuli and visual events occurring across both physical and rendered space. Through the use of immersive audio presentation and spatialization techniques [14], audience members can aurally attend to sonic events in the rendered space, transposing their attentional location into the shared space. We believe it is this border—bridging the multi-modal intersection of visual and auditory presentation systems and physical space—that represents the true nature of mixed-reality performance, allowing performers and audiences the opportunity to consciously and intentionally shift their attention to performance sound and action in one "reality" or the other.

2 A history of musical performance networks

While the transmission of music across communication systems has only become truly ubiquitous within the last few decades, the history of network-based musical transmission dates back to the end of the previous century with the introduction of Gray's "Musical Telegraph" and Cahill's "Telharmonium" [7]. While both instruments were designed to generate electronic sound and to transmit that sound over telephone wires, the Telharmonium was established as perhaps the world's first networked musical subscription service, selling live musical performances transmitted in real-time across telephone wires to paying establishments. The ability to create sound in one location and broadcast it for performance in another location is a uni-directional method of communication; there is no musical interplay or bi-directionality involved, constraining any audience member to the role of passive listener.

Musical networks of particular interest can be defined as necessarily bi or poly-directional connected musical streams each representing one or more voices of live musical performance data; networks can exist within a single physical location on a local network or in discrete and disparate physical locations and spaces, making use of both commercial and research-grade internet access. There additionally exists an increasingly well-populated community of network musicians and an evolving performance practice for networked ensemble performance [1].

Early networked performances by The Hub [10] stand out as rich examples of the complex musical constructs formed through communal composition, improvisation and performance. Stanford's SoundWIRE group [24] utilizes multiple channels of uncompressed streaming audio over its JackTrip software to superimpose performance ensembles and spaces alike, with performances of note including networked concert performances with the ensembles Tintinnabulate at RPI (NY) and VistaMuse at UCSD, as well as with performers at Beijing University in 2008s "Pacific Rim of Wire" concert [6]. Both the Princeton Soundlab's Princeton Laptop Orchestra (PLOrk) [26] as well as the Stanford Laptop Orchestra (SLOrk) [29] have displayed the powerful possibilities of collaborative networked compositional form using local area networks for synchronization and communication and distributed point-source spatialization.

3 Game-based musical systems

The use of networked/multi-user video game engines for music and sound generation has become increasingly common as generations of musicians who have grown up with readily accessible home video game systems, internet access and personal computers seek to bring together visually immersive graphical game-worlds, wide-spanning networks and interactive control systems with musical systems. Though its graphical display is rendered in 2-dimensions, Small_Fish by Kiyoshi Furukawa, Masaki Fujihata and Wolfgang Muench [9] is a game-like musical interface which allows performers/players to create rich musical tapestries using a variety of control methods. Auracle [8], allows networked users to collaborate and improvise using vocal gesture. Fijuu2 [21], a fully rendered three-dimensional audio/visual installation controlled with a game-pad, tightly marries the videogame and musical worlds through the use of immersive graphics and familiar game control systems.

4 Previous work in virtual worlds

q3apd [19] was an early modification of the open-source 1.32 Quake 3 gaming engine [13] which outputs user coordinate, view and player data to Pure Data (PD) [22] using PD's internal FUDI protocol, designed as the networking protocol used by PD to link GUI and DSP systems. Describing the work, Oliver states "In the installation the movement, position, health, view angle and item status of 4 software agents in combat was sent to the synthesis environment Pure Data and used to make an auralisation of activity in the arena: a musical re-presentation of the flows and gestures of artificial life in combat." [18].

For the work *maps and legends* [12], q3apd was used to track user motion around specially-designed compositional maps, triggering modular components of a multi-channel composition and spatializing each one based on individual user positions. Eight speaker representations were arranged around the periphery of the compositional map for maps and legends, giving visual cues as to the relative panning location of the currently playing sounds in the eight-channel surround sound field. By correlating avatar location with

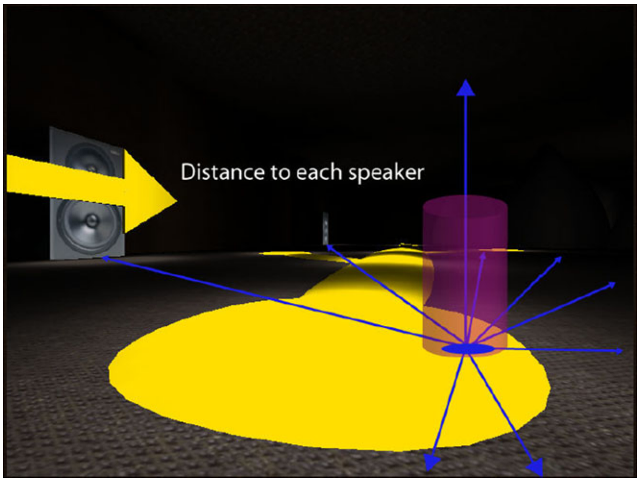

Fig. 1 The distance between an avatar's coordinate position and eight virtual speaker positions is constantly measured in *maps and legends*

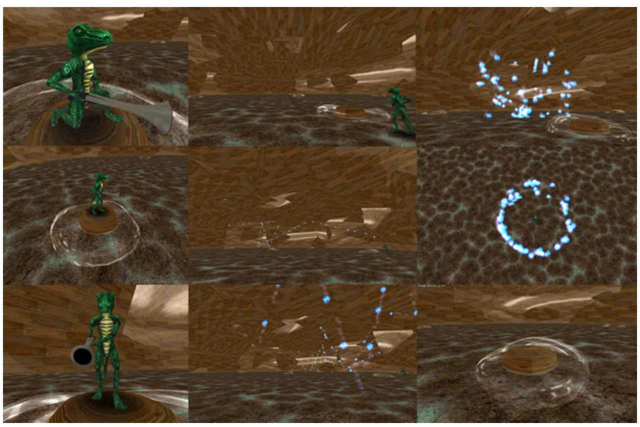

Fig. 2 Virtual performer and projectiles in *nous sommes tous fernando*

respect to each of eight speaker representations to the amplitude of triggered musical events in an eight-channel surround soundfield, an overlay of the compositional map was created in physical space, bringing the experience of action in two environments together (Fig. 1).

q3osc [11] is a heavily modified version of the open-sourced ioquake3 gaming engine featuring an integrated Oscpack [2] implementation of Open Sound Control (OSC) [30] for bi-directional communication between a game server and one or more external audio servers. By combining ioquake3's internal physics engine and robust multiplayer network code with a full-featured OSC packet manipulation library, the virtual actions and motions of game clients and previously one-dimensional in-game weapon projectiles are repurposed as independent and behavior-driven OSC emitting sound-objects for real-time networked performance and spatialization within a multi-channel audio environment.

q3osc was used to create *nous sommes tous fernando…* (Fig. 2) and *jord og himmel*, multi-user works designed for and premiered by the SLOrk. Both compositions explore local and wide-area network performance paradigms to spatialize sound using sets of hemispherical speakers designed for the SLOrk arranged on stage in patterns matching the location of virtual speaker models in q3osc. The in-engine proximity of sound events to the speaker representations was used to spatialize pitched sounds to locations in the sound field while mapping absolute x-coordinate location to frequency. In this manner, the environment itself became a reactive instrument, with which users could engage by firing bouncing projectiles against any surface of the environment.

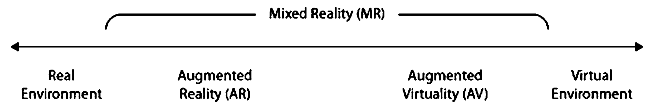

Fig. 3 Milgram-Kishino Virtuality Continuum

5 An attentional definition of mixed-reality

Along the Milgram-Kishino continuum [16] spanning "real environments" and "virtual environments", "mixed-reality" is defined to be inclusive of any combination of real and virtual objects in presentation within a single display (Fig. 3). Similar descriptions [3] extend the presentation from incorporating a single visual display to a more generalized notion of 'overlaying of virtual objects on the real world'. We propose a further extension or generalization, moving away from the restrictive notion of a 'display-based' definition (one inherently biased to visual stimuli and display) to an 'attentional' or 'consciousness-based' definition, where one's attention or conscious attending to events and stimuli are the true attentional focal point. Equal footing can as such be given to non-visual stimuli, most importantly in the context of presentations combining auditory and musical stimuli.

Following an attentional definition of Mixed-Reality, events happening in physical space alongside auditory and visual depictions of virtual events are combined by listeners and viewers. Accurate auditory models of sound events within virtual space can be presented either with or without a corroborating visual presentation, allowing for a hybrid or split attentional model. When combined with presentations of networked 'telematic' [20] musical performance—wherein listeners experience an additive musical space comprised of local and remote musical events—as well as virtual performance environments, a polyvalent set of presentation possibilities and attentional directions is made possible and available to performers as well as audience members. It is this last presentation, a networked musical event combining real-world and virtual actors and musical events spanning disparate geographical and dimensional environments that best describes the style of 'Mixed-Reality' performances.

6 Space and presence for mixed-reality ensembles

One way of addressing the problem of presence or telepresence within and across distributed musical ensembles is through the integration of groups of musicians with avatars that perform in computer-generated interactive distributed worlds. Live acoustic musicians imbue a musical locale with an intrinsic sense of place. Virtual environments, through their networked nature, give a displaced sense of place, i.e., they act as a "meeting place" of the global village. Performances combining distributed musicians with physical instruments encounter the presence problematic in each separate locale—"Are the other musicians that I hear really there?" Even video technology is not able to fully convey this presence. On the other hand, virtual environments create a compelling sense of shared space, but at the cost of losing a connection to local presence. While it's hard to convey the sense of presence in distributed physical ensembles, it's also hard to convey the sense of "locality" in distributed virtual worlds.

7 User and spatio-centric presentation

In video gaming, the "First-Person" viewpoint is a common style of game play, allowing users to tele-operate or control an avatar in rendered space while viewing the environment itself through the eyes of the avatar. In such systems, where a user's self is represented virtually in the form of an avatar capable of interacting "physically" with virtual constructs and actions, the focus of both visual and auditory representations presented to each user has commonly been wholly user-centric: a user in his or her real-world seat is presented with an illusory visual and sonic representation of the virtual environment from the viewpoint of that user's avatar. The illusion is heightened with the use of audio signal processing designed to enhance the related first-person illusions of space, distance and motion. As users listen to audio output through headphones, stereo speakers, or an industry standardized multi-channel configuration such as 5.1 or 8.1, all audio processing done in game-engines tends to attempt to create realistic illusions of motion for one user sitting in the sound-system's centralized "sweet-spot". Such individualistic and user-centric presentation by its very nature restricts the communal sensory experience fostered in the virtual environment from existing anywhere except within the game-world itself.

In a context where live musical presentations are attended by a viewing and listening audience, the use of user-centric video and audio presentation to the audience itself proves problematic. User-centric systems are optimized for a single point of attention and don't translate to multiple points of attention when suddenly co-opted into delivering visual and

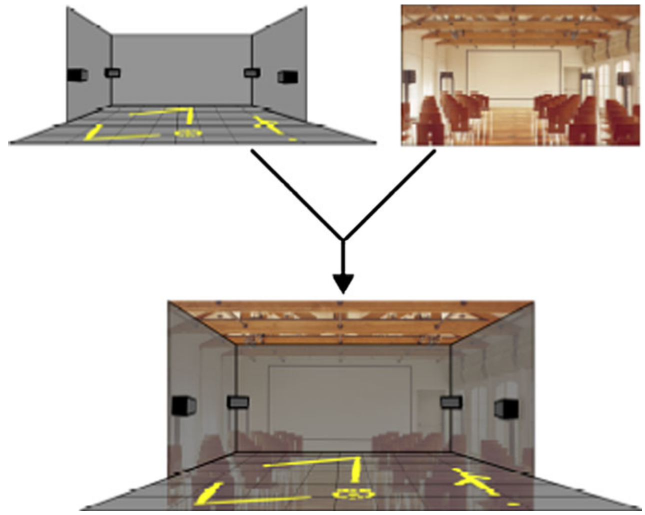

Fig. 4 A spatio-centric presentation can create immersive auditory experiences for multiple listeners

audio cues to a group spread within a viewing and listening space. By inverting these traditional models of sound-presentation and by focusing on a spatio-centric model of sound projection for game-environments, a communal listening experience can be fostered inclusive of all listeners within a shared physical space, including game-users and audience members alike (Fig. 4).

8 Mixed-reality musical performances

Between 2009 and 2010 members of the "Music in Virtual Worlds" (MvW), Music, Computing and Design (M:C:D) and SoundWIRE research groups from Stanford's Center for Computer Research in Music and Acoustics presented a series of concerts at the Torino-Milano (MiTo) Festival designed to showcase new modes of music creation and performance with the latest media technology and with an effort to re-think musical interaction in multi-modal and distributed environments. Each concert was centered around a series of musical works built within heavily customized virtual world engines: Sirikata [23] in 2009 and the Unreal Engine/UDK [27] in 2010. For each piece virtual performers controlled aspects of sound generation and spatialization though movement and virtual gesture. In a subset of musical works, live performers (both local and remote) played traditional instruments alongside virtual performers.

8.1 Due Serata in Sirikata

In August 2009, Due Serata in Sirikata, two evenings of Mixed-Reality networked musical performances combining live instrumental performers with immersive and interactive multi-user rendered performance environments were presented at the Polytechnico di Milano Bovisa (Milan, Italy) in

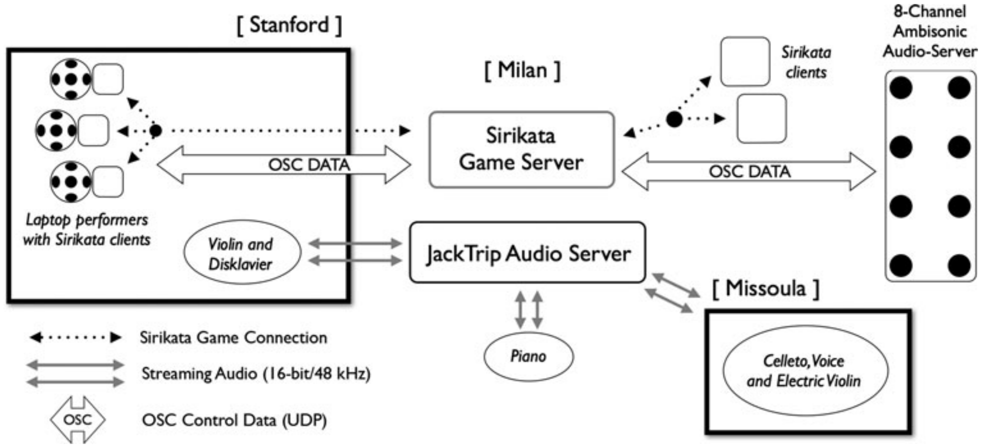

Fig. 5 Networking schematic for *Due Serata in Sirikata*

the Carlo de Carli hall. Surrounding the audience were an array of eight speakers and four projection screens, arranged to provide an immersive audio and video experience. Four musical works were premiered at the event, each one built using a customized implementation of the Sirikata engine connected to multi-channel audio servers using the OSC protocol.

Sirikata is an open source platform designed for the creation of multi-user games and virtual worlds. To integrate Sirikata environments and object motions and events with custom ChucK [28] and SuperCollider [25] sound servers, an OSCPack implementation of Open Sound Control was used. Custom object interactions as well as early-stage networking and physics protocols were implemented as well, allowing client avatars to interact with each other and the environment itself. In this manner, users within the Sirikata engine could move through a given environment, controlling parameters of sound generation and spatialization through their avatar's motion in space or by simply interacting with specific objects in the environment itself.

As seen in Fig. 5, live instrumental performers located in Stanford, California and Missoula, Montana made bi-directional audio connections to the concert hall using JackTrip [4] across research-grade Internet2 network connections, affording high speed and bandwidth for each connection. 16 bit/48 kHz uncompressed audio was sent over JackTrip and selectively spatialized throughout the concert hall, giving great clarity to each instrumental performance. Three laptop performers in California connected to a Sirikata server running in Milan, joining three local Sirikata performers. OSC data representing coordinate and control data from each individual performer was then streamed from the Sirikata server to an array of sound-generating audio servers (Fig. 6).

The musical goals and the technology designed to achieve them are described below for each piece.

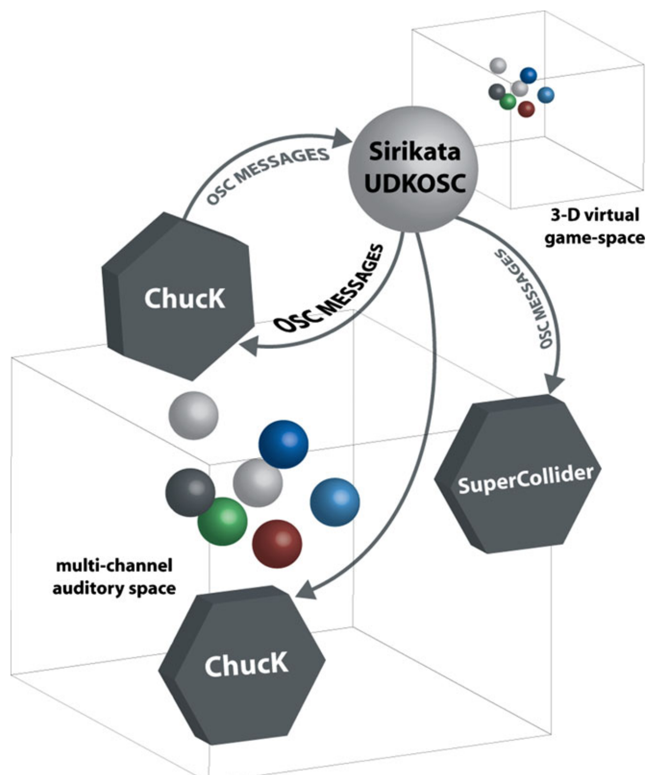

Fig. 6 Flow of OSC data between game space and auditory space

8.2 Musical form and virtual architecture: *In C*

Terry Riley's *In C*, a metronomic piece composed for an indeterminate and unspecified number of musicians, presents a unique set of challenges in the context of distributed performance. We have described the musical implementation in detail in previous work; the pulse of the piece is calculated and distributed to match network delay between connected sites, ensuring that the rhythmic patterns that comprise the piece are tightly synchronized, albeit shifted a number of metric beats, ensuring the musical result is different at each end of the performance (see [6] for details).

The 2009 performance of *In C* at the MiTO Festival was distributed between Milan, Stanford California and Missoula Montana. While previous performances have included traditional instruments and laptop performers, this new rendition added virtual world avatar performers with custom visual art and a novel method of moving one's instrument through Riley's composed musical pathways. The orchestra consisted of Piano and Avatar/Laptop performers in Milan, Avatar/Laptop performers and violin in Stanford and Electric Cello and Electric Violin in Missoula. In keeping with earlier networked performances of *In C*, the ensemble was synchronized through a metronome that generates audio and OSC messages.

In C is designed in a manner that lets musical performers move linearly through composed musical cells, choosing repetition and cell-onset timings while remaining strictly locked to the underlying metronomic pulse. The virtual environment artwork was created as an allegory of the work, sequentially following a set of patterns or cells though motion down a rendered pathway. A clear linear pathing model was used, so that each cell of the modular music was likewise represented visually as the cell of a pathway (Fig. 7). The scale of the path and the visual affordance was clearly laid out in the lattice railways that lead from one area of the environment to the next.

As multiple performers were connected from disparate global locations, separate paths were created so that the physics bounding box would hold everyone on the paths, and prevent anyone from literally "falling off" the performance path. The journey took the audience through a surreal floating city landscape with an end tower clearly in site for all 3 paths. This conforms with the instructions of the piece where at the end the whole ensemble "lands" on the last pattern.

This rendition of *In C* adds different layers of musical and visual meaning that conform with the original intention of the piece, and it's an example of multi-modal expansion of existing repertoire. The idea of local synchronization is expanded to distributed musical synchronization. The idea of advancing to a series of musical cells is not only illustrated in the virtual world, but literally this is performed, i.e., physically advancing means musically advancing in the cells.

8.3 Positional spatialization: *Dei Due Mondi*

In the work *Dei Due Mondi*, eight Sirikata performers act as spatializing agents for an eight-channel pre-composed "tape" piece, with their location in virtual space mapped to a panning location in listening space around the audience. Rendered environments representing the Milano concert hall and the connected concert space in California were displayed, with avatars freely able to move between the two spaces, bringing their sound with them. Virtual rendered speaker images within both the Milan and the Stanford environments were designed to match the physical speaker layout in each concert hall. Performers created patterns of motion throughout each concert space, effectively traveling between the two spaces through their virtual representation. As user avatars move through three-dimensional coordinate space, each avatar's X, Y and Z location data is streamed over OSC to a sound-server. Performers following a loosely-arranged choreography of motion control amplitude levels and sounding direction of their assigned part. An additional set of speaker locations, spread "outside" the representations of the concert halls in a fantastical world was used as well.

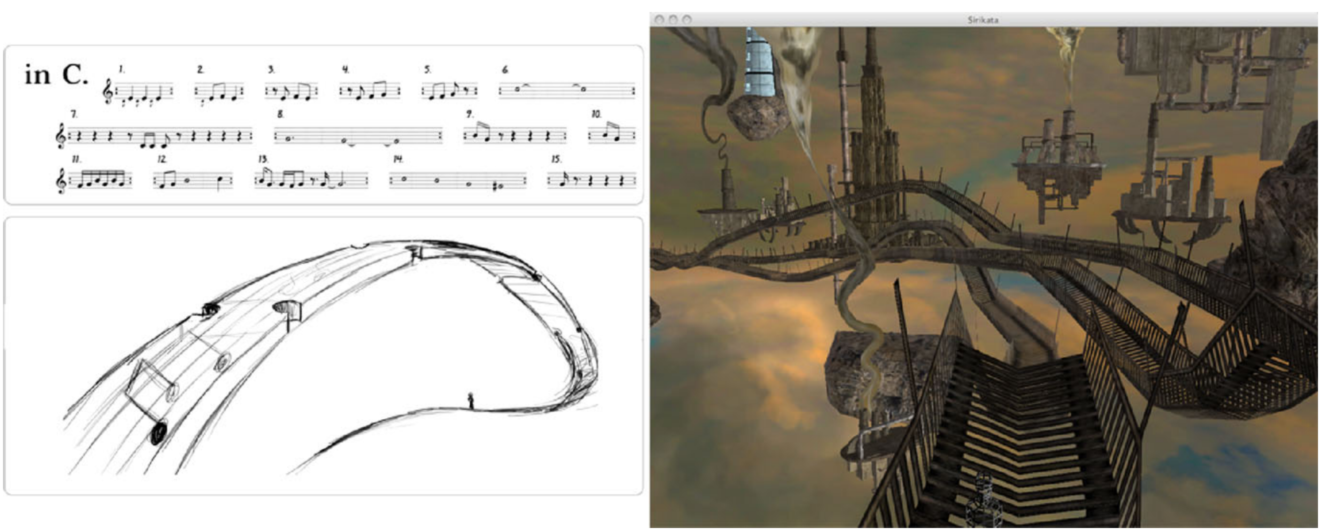

Fig. 7 Visual pathways as analogues to the musical paths of *In C*

Fig. 8 Outdoor and indoor rendered environments from *Dei Due Mondi*

In this manner, performers could control sound with a direct mapping (within their own relative concert space), with a remote mapping (within the other concert hall) or with an abstract mapping.

To create perceivable movements of sound through space each channel of *Dei Due Mondi* was spatialized using a first-order Ambisonics endoder and decoder in SuperCollider. With Ambisonics, all 8 speakers could be used to spatialize individual sounds, creating a stable and localized soundfield surrounding the audience. Sounds moved through the virtual space were corrolated to movements through the physical concert halls: as an avatar traversed the virtual Milan concert hall, that performer's associated sound stream were perceived by the audience as moving across the physical hall (Fig. 8). This spatio-centric approach meant that members of the audience were immersed in an audio rendering designed to both fill the space and show the 1:1 relationship between avatar motion in virtual space and sound motion in physical space.

8.4 Spatial improvisation: *Canned Bits Mechanics* and *Dialoghi*

A different approach to the role of spatialization was devised for the other two pieces presented in 2009. *Canned Bits Mechanics* was inspired in part by the 100-year anniversary of the 1909 publication of the Futurist Manifesto [15]. In keeping with the turn-of-the-century futurist aesthetic, the visual and sonic components of the piece are metallic and mechanical (Fig. 9). In addition to a local pianist performing from Milan, two remote Yamaha Disklaviers located in California were controlled using interactive real-time algorithms from the stage in Milan. The use of mechanical acoustic instruments that are not available locally is another potential of network performance. An avatar in the Sirikata environment

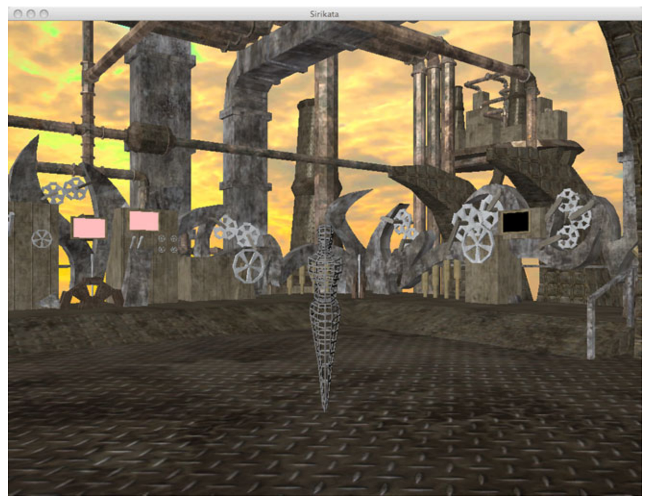

Fig. 9 Mechanical constructs as visualized in *Canned Bits Mechanics*

completes the ensemble, providing both a film-like visual narrative as well as controlling pre-recorded metallic and percussive sound events through speed of motion and coordinate location.

The virtual world is an industrial setting of diamond plated walkways, and large i-beam columns (Fig. 9). As the performer avatar moved through the environment, sonic manipulations consisting of variable-speed and direction sample playback and gain control were triggered. The aural results depended on the speed of motion of the avatar and its position, giving great flexibility to interact with the other performers. Flying gestures created simultaneous visual and aural sweeps, creating an engaging experience for performer and audience alike.

These same sample-based motion gestures were used in *Dialoghi* a networked improvisation combining four remote instrumental performers, the local piano and Sirikata avatar

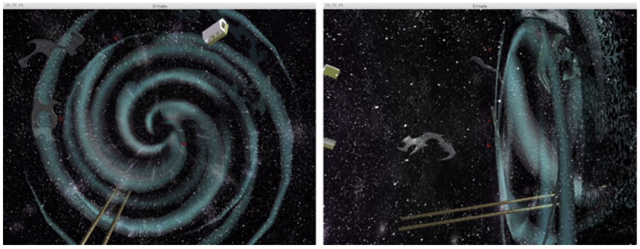

Fig. 10 Space field with "Cosmic Debris" from *Dialoghi*

Fig. 11 Pianist Chryssie Nanou improvises with live networked musicians from Stanford, Montana as well as in virtual space

performer (Fig. 11). Designed as a culminating piece for the concert, the visual setting consisted of "cosmic debris" from each the previous pieces in the concert (*In C*, *Canned Bits Mechanics*, *Dei Due Mondi*) serving as a visual and musical deconstruction of the event itself (Fig. 10). To musically illustrate this deconstruction the avatar controlled altered samples of *In C* and other sounds that formed the basis of the previous three pieces of the concert.

8.5 MiTo 2010

For the 2010 MiTo Festival, musical works were conceived that would require a more complex game-based control system and a richer visual ecology than was currently available in the still-young Sirikata platform. To incorporate features such as dynamically-controllable projectiles and complex visual shader and lighting treatments, the Unreal Engine in the form of the freely available Unreal Development Kit (UDK) was used. Two works using the UDK were presented in 2010, *Tele-harmonium* and *Perkussionista*, both of which made heavy use of Ambisonic spatialization and OSC driven control systems.

8.6 UDKOSC

UDKOSC is a customized implementation of the Unreal Development Kit linking a modified object and event codebase with an Open Sound Control OSCPack Windows.dll library. Perhaps best seen as an extension of q3osc incorporating lessons learned with the customized Sirikata implementation, UDKOSC enhances the basic user coordinate tracking and customized projectile behaviors found in those systems with greater granularity of control as well as new environmental object classes capable of reporting data on their interactions with user avatars and projectiles alike.

Some of UDKOSC's interactive features include:

- OSC tracking of continuous projectile location as well as upon collision with environment
- dynamically controllable projectiles with homing/swarm behaviors
- projectile homing behaviors can be targeted to user-generated locations in space as well as to other moving projectiles
- environment object classes capable of outputting OSC messages upon collision/interaction with projectiles and/or client avatar
- projectile size and speed can be changed dynamically
- external OSC control over projectile direction/homing behaviors with custom iPad interface.

8.7 *Tele-harmonium*

Tele-harmonium for live piano and virtual UDKOSC performer was the first work written using UDKOSC and was designed to incorporate the architecture of the rendered environment itself as an instrument capable of performing a duet with a live pianist. Visually, the Tele-harmonium was conceived as a giant machine-building, similar in concept to an immersive musicbox with rotating gears and mechanisms to mechanically reproduce sound and music. The storyline of the work describes the building as an instrument capable of playing back musical memories and as such, the virtual performer is seen traveling, entering and interacting with structures within the building to create and control sound and music. At the center of the space is a rendered organ-like console, intended as the virtual manifestation of the live pianist, though perhaps more appropriately, the pianist should be seen as the live representation of the virtual instrument.

Musically, *Tele-harmonium* is loosely built upon the harmonic and melodic structure of the *Sonata for keyboard in D minor*, K. 213 (L. 108) "The Lover" by Dominico Scarlatti. Excerpts from the Sonata are heard both in manipulated recording as well as in live performance by the pianist. Additional musical materials that comprise the piano score were composed as derivative and allegorical variations on particular phrases and structures found in the original work. The virtual performer controls the structure of the composition as a whole, moving though specific locations and triggering recorded excerpts of the original Scarlatti in conjunction with the live pianist's playing.

Fig. 12 Homing projectiles revolving around larger target projectiles

To allow the virtual performer the ability to successfully duet with an instrument as melodically and harmonically flexible as a piano, both projectile collision and continuous OSC data streams were connected to sound-generating virtual instruments in SuperCollider and encoded using first-order Ambisonics. As projectiles collide with specific objects in the rendered environment, simple pitched sounds are generated using bursts of noise passed through tuned filter banks. Settings for filter frequency/pitch and amplitude are extracted from the UDKOSC stream: frequency was mapped to a Cartesian distance from the central rendered keyboard while amplitude was mapped to the size of the projectile itself. Individual or groups of projectiles can be dissolved at any given time by the performer, allowing great control over musical texture and additive dynamic level.

Continuous sounds are created using a modified version of the projectile object, mapping spatial coordinates to frequency, amplitude and specific timbral characteristics of a simple synthesized instrument. Swarming and following behaviors are extremely effective in generating rich soundfields when mapped to a simple continuous sound source, with slight fluctuations in pitch and timbre working together to create a timbrally rich musical sound (Fig. 12).

As each individual projectile is tracked through coordinate space over OSC, their relative locations within virtual space can be corrolated to locations in physical space as rendered through an Ambisonic decoder. Musical spatial gestures can easily be created in this manner, allowing sound to smoothly move throughout the audience, creating an immersive auditory experience.

8.8 *Perkussionista*

Perkussionista is a wholly new musical work built on the 2009 infrastructure originally used in *Canned Bits Mechanics*. Reworked using the UDK, the visual environment was updated with user-controllable levers and additional faux-mechanical constructs hooked into the SuperCollider sound engine with OSC. The goal of ther Perkussionista performance was to manually move across platforms, pulling levers and rotating wheels in a pre-choreographed pattern. Each movement and action in virtual space triggered pre-composed recorded snippets which were in turn spatialized with Ambisonics throughout the concert hall.

9 Conclusions and future work

The use of computer-based rendered visual environments as the core component to multi-modal networked musical presentations has proven to be an exciting and rewarding research path. As technology becomes more open and available with the introduction of platforms such as Sirikata and the UDK, artists and musicians are given valuable tools for building immersive and enactive systems for nuanced musical composition and performance. Indeed, the very nature of composing music to be performed through such novel interfaces has necessitated a concerned look at methodologies for composing modular and flexible musical works and systems. That being said, there exists a great deal of room for improvement and refinement of current practices. New interfaces allowing controllers of in-engine avatars to interact more naturally and compellingly with virtual representations of instruments and sound-controlling processes are under consideration, ranging from touch-based mobile device interfaces capable of tracking subtle gesture to force-feedback controllers capable of delivering physical haptic feedback, to fully-engaged body tracking using commodity devices such as the Microsoft Kinect.

In the same light, the design of compelling gestural mapping schemata, intended to aid in the correlation of actions between physical and virtual space is necessary to allow musical performers in such systems to reach for virtuosity. The perceptual coherence between gesture and resultant musical action is of great importance, necessitating a closer look at methods for creating compelling immersive experiences capable of nuanced gesture and reaction. Towards these ends we are experimenting with a series of user-based studies aimed at fostering a better understanding of the perceivable relationships between virtual gesture and physiological auditory response.

Acknowledgements The authors would like to thank everyone who helped make the MiTo Festival concerts a huge success including the Sirikata Team: Daniel Miller, Patrick Reiter Horn, Daniel Horn, Ewen Cheslack-Postava, Jeffrey Schnapp, Henrik Bennetsen; the Sirikata and UDKOSC art teams: Jason MacHardy, Ben Nadler, Hadidjah Chamberlain, Justin Herbst, Will Harper; the Musicians: Debra Fong, Chris Chafe, Charles Nichols, Luke Dahl, Michael Berger, Visda Goudzari, Spencer Salazar, Jorge Herrera, Carr Wilkerson and Ge Wang.

References

1. Barbosa A (2003) Displaced soundscapes: a survey of network systems for music and sonic art creation. Leonardo Music J 13:53–59

2. Bencina R (2006) Oscpacka simple c++ osc packet manipulation library. URL http://code.google.com/p/oscpack/
3. Bilinghurst M, Kato H (1999) Merging physical world and virtual world. In: Proceedings of the international symposium on mixed reality, Yokohama, Japan. URL ftp://ftp.hitl.washington.edu/pub/publications/r-91-2/r-98-36/r-98-36.html
4. Cáceres JP, Chafe C (2009) JackTrip: under the hood of an engine for network audio. In: Proceedings of international computer music conference, International Computer Music Association, San Francisco, CA, pp 509–512
5. Cáceres JP, Chafe C (2010) JackTrip: under the hood of an engine for network audio. J New Music Res 39(3):183–187. doi:10.1080/09298215.2010.481361
6. Cáceres JP, Hamilton R, Iyer D, Chafe C, Wang G (2008) To the edge with china: explorations in network performance. In: ARTECH 2008: proceedings of the 4th international conference on digital arts, Porto, Portugal, pp 61–66
7. Chadabe J (1996) Electric sound: the past and promise of electronic music. Prentice Hall, New York
8. Freeman JEA (2005) The architecture of auracle: a voice-controlled, networked sound instrument. In: Proceedings of international computer music conference, Barcelona
9. Furukawa K, Fujihata M, Muench W (2000) Small_fish. URL http://hosting.zkm.de/wmuench/small_fish
10. Gresham-Lancaster S (1998) The aesthetics and history of the Hub: the effect of changing technology on network computer music. Leonardo Music J 8:39–44
11. Hamilton R (2008) q3osc: or how i learned to stop worrying and love the game. In: Proceedings of international computer music conference, Belfast, Northern Ireland
12. Hamilton R (2008) Maps and legends: designing fps-based interfaces for multi-user composition, improvisation and interactive performance. In: Computer music modeling and retrieval. Lecture notes in computer science, vol 13, pp 478–486
13. id software: Quake 3 (2000) URL http://www.idsoftware.com/games/quake/quake3-arena/
14. Malham D (1995) 3-D sound spatialization using ambisonic techniques. Comput Music J 4(19):58–70
15. Marinetti FT (1909) The futurist manifesto
16. Milgram P, Kishino F (1994) A taxonomy of mixed reality visual displays. IEICE Trans Inf Syst E77-D(12):58–70. URL http://etclab.mie.utoronto.ca/people/paul_dir/IEICE94/ieice.html
17. Mito: Settembre musica (2009) URL http://www.mitosettembremusica.it/en/home.html
18. Oliver J (2003) q3apd: making music with bots. URL http://vimeo.com/1380686
19. Oliver J, Pickles S (2003) q3apd. URL http://www.selectparks.net/archive/q3apd.htm
20. Oliveros P, Chafe C, Dresser M (2009) Telematic music: six perspectives. Leonardo Music J 13:95–96
21. Pickles S (2004) Fijuu 2. URL http://www.fijuu.com
22. Puckette M (1996) Pure data. In: Proceedings of international computer music conference, San Francisco, pp 269–272
23. Sirikata (2009) A bsd licensed open source platform for games and virtual worlds. URL http://www.sirikata.com
24. SoundWIRE Group (2010) SoundWIRE research group at CCRMA, Stanford University. URL http://ccrma.stanford.edu/groups/soundwire/
25. Supercollider (2006) URL http://supercollider.sourceforge.net
26. Trueman D, Cook P, Smallwood S, Wang G (2006) Plork: Princeton laptop orchestra, year 1. In: Proceedings of international computer music conference
27. Unreal development kit (2010) URL http://www.udk.com
28. Wang G, Cook PR (2003) ChucK: a concurrent, on-the-fly, audio programming language. In: Proceedings of international computer music conference, Singapore
29. Wang G, Bryan N, Oh J, Hamilton R (2009) Stanford laptop orchestra (slork). In: Proceedings of international computer music conference, Montreal
30. Wright M (2005) Open sound control: an enabling technology for musical networking. Organ Sound 10:193–200
31. Xu A, Cooperstock JR (2000) Real-time streaming of multichannel audio data over Internet. In: Proceedings of the 108th convention of the audio engineering society, Paris, pp 627–641