# Reactive Environment for Network Music Performance

Dalia El-Shimy and Jeremy R. Cooperstock
Centre for Interdisciplinary Research in Music, Media and Technology
McGill University
Montréal, Québec
{dalia, jer}@cim.mcgill.ca

## ABSTRACT

For a number of years, musicians in different locations have been able to perform with one another over a network as though present on the same stage. However, rather than attempt to re-create an environment for Network Music Performance (NMP) that mimics co-present performance as closely as possible, we propose focusing on providing musicians with novel controls that can help increase the level of interaction between them. To this end, we have developed a reactive environment for distributed performance that provides participants with dynamic, real-time control over several aspects of their performance, enabling them to change volume levels and experience exaggerated stereo panning. In addition, our reactive environment reinforces a feeling of a "shared space" between musicians. Our system—intended for use in more relaxed, informal settings, such as loose rehearsals and jam sessions, rather than live performances before an audience—differs most notably from standard ventures into the design of novel musical interfaces and installations in its reliance on user-centric methodologies borrowed from the field of Human-Computer Interaction (HCI). Not only does this research enable us to closely examine the communicative aspects of performance, it also allows us to explore new interpretations of the network as a performance space. This paper describes the motivation and background behind our project, the work that has been undertaken towards its realization and the future steps that have yet to be explored.

## 1. INTRODUCTION

The relatively recent development of high-speed networks has led to a virtual collapse of geographical distances. As the field of Computer Supported Collaborative Work (CSCW) emerged to address many of the ensuing social and technological effects, the notion of people being apart, yet feeling together, has become quite commonplace. Remote collaboration over a network, however, is not a task without its share of challenges, the most glaring of which is arguably that of latency. Put simply, there are restrictions on the transmission speed of real-time data, which, under idealized conditions, can only reach up to 2/3 of the speed of light when fibre optics are used [6]. This places the theoretical Round Trip Time (RTT) between New York and San Francisco at approximately 40 ms, a figure that does not take into account compression, encoding, and decoding, existing traffic on the network or transmission-error-checking. While this level of latency seems to be relatively quite low, it is considered unacceptable within the context of musical performance, where unidirectional latency must fall below the Ensemble Performance Threshold (EPT) of 25 ms to ensure synchronous play. To cope with this drawback, a number of artists began to approach distributed performance differently, choosing instead to examine the merits of treating the network as a unique collaborative space. Schroeder and Rebelo, for instance, claim that the network is no longer merely a channel for communication and exchange, but rather a "place in its own right, a space for being, a locus for dwelling" [27]. In addition, Shapiro encourages us to shift the emphasis of "being there" towards a greater exploration of "being apart". This notion is echoed by Renaud, who states that the most exciting prospects for NMP lie not in emulating the traditional stage, but in using the network to explore new types of performance and purpose-created music [24]. Therefore, we argue that the network should be embraced as a valid performance space, with all its idiosyncrasies and their implications. However, we also propose taking NMP even further: why not capitalize on the fact that, by definition, the network brings computing technology to the performance setting, and use this to the musicians' advantage by giving them control over certain dynamics of the performance? By providing musicians with interesting functionalities that they do not experience in a traditional performance, we can entice them to explore the network not only as a valid medium, but as one that is full of creative potential worth exploring. In addition, we began to wonder whether these added functionalities could somehow increase the level of interaction between the remote musicians, and compensate for the effect of decreased sociability that remoteness can have on musical performance. However, determining how such functionalities should be chosen and designed is not a trivial task. According to Jordá, the fact that it is not easy to define the role of a computer in live performance the way one can with traditional acoustic instruments is an indication that we are still at the "Stone Age" of technology-aided music creation [19]. To examine this issue, we chose to adopt the User-Centred Design (UCD) methodologies popular within the field of Human-Computer Interaction (HCI). To this end, target users have thus far involved our target users throughout the life-cycle of the project by means of contextual inquiries, observations and interviews. Subsequently, we developed and tested prototypes for the the co-present case, all of which we discuss at length in reference [13]. In this paper, we describe our current efforts at extending our reactive environment from a co-present to a distributed context, and discuss the impressions and feedback of a musical trio who performed with our system as part of the Net-Music 2013 Symposium.


*NIME'13,* May 27 – 30, 2013, KAIST, Daejeon, Korea.


## 2. BACKGROUND AND RELATED WORK

### 2.1 Network Music Performance

While distributed musicians had managed to collaborate successfully with one another via satellite as far back as 1975 [17], it was not until 2001 that the first successful two-way musical collaboration over the Internet took place: Chafe and the SoundWIRE Group at Standford's Center for Computer Research in Music and Acoustics (CCRMA) were able to stream high quality audio bidirectionally between a pianist at Stanford and a cellist at the Internet2 headquarters in Armonk, NY. Although there was very little signal loss, the acoustic latency of 125 ms was "on the hairy edge for an unencumbered performance". Nonetheless, despite the noticeable delay, Chafe reports that musicians were able to "catch-up" during the pauses [6][7]. A much lower delay was experienced when musicians from Stanford University and McGill were joined in a cross-continental jazz session in 2002 using the

Ultra-Videoconferencing system developed at McGill. Surround sound and full-screen video were streamed bidirectionally over the research Internet (CA*net2 and Internet2), with a one-way delay around 45 ms. Although problems with an intervening router sometimes doubled the delay, the musicians involved reported feeling as though they were present on the same stage during the "good moments" [8]. A number of standalone software applications for audio streaming were later created with accessibility to the "average" musician in mind. For instance, the readily-available Soundjack software, developed by Carôt in 2005, can directly access the sound card buffer and send audio data via UDP [4]. Soundjack was successfully used in a number of NMPs attempting to replicate co-present condition all across Europe. As another example, eJamming is a commercial software product available online that promises to "enable musicians to play together in real-time as if in the same room even if they are far from one another". It differs from Soundjack in using MIDI data rather than audio, thereby greatly reducing the bandwidth requirements. In addition, data is transmitted only when an event is triggered by the user, which minimizes the amount of jitter [3][5]. Finally, JackTrip, the application developed by the SoundWIRE group to support their on-line collaborations, is also available for download. Jacktrip uses Jack as its default host audio server, and "has the ability to make audio connections between many different audio clients on the same host", making it a common solution used by those interested in "multi-machine network performance over best-effort Internet" [2].

### *2.1.1 Visual Representations*

While most NMP systems continue to take their lead from traditional videoconferencing, offering full-frontal video as the only solution, we find that this approach does not take into account the network's unique characteristics, but simply attempts instead to replicate traditional performance. In an effort to resolve visual communication during NMP, many researchers believe it is first imperative to examine the implications of being *in the network*. For instance, Schroeder notes that "in the same way that you cannot stare at the network straight in the eye, that you can never directly confront the network, for it is always somewhere else from wherever you may be looking, performers never stare at other players" [26]. In traditional performance, although musicians communicate with each other through various cues and body language, they do not require a full-frontal view of one another. Instead, each musician experiences only "fragments" of the whole performance environment, through glances and peripheral vision [27]. Putting this philosophy to the test, Schroeder et al. started the "Apart Project" in 2007 as a study on various novel network scenarios. Three musicians were asked to perform several songs over a network under five types of scenarios that included "avatars" and standard video conferencing technologies. The avatars were designed as close-up and detailed, yet abstract renditions of performance gestures. Post-test questionnaires revealed that musicians enjoyed looking at the avatars "as a means for visual interaction and potentially for enhancing social interaction." [27]. Furthermore, when iChat was used to stream full-body video capture amongst the musicians, there were remarkably few glances towards the screen in both pieces, supporting Schroeder's theory that performers do not need to stare at one other directly and constantly. Similarly, Kapur et al. experimented with the use of specialized graphics, in addition to full-frontal video, as part of their work on distributed performance. More specifically, the authors developed the *veldt* software, described as a "real-time networked visual feedback software" that can trigger arbitrary text, images, videos or geometric models in response to MIDI events [20]. Mappings are flexible and can be set by the musicians prior to a performance. For instance, when *veldt* was used as part of the Gigapop ritual, players of the Electronic Dholak (EDholak), a multi-player networked percussion controller based on the Indian Dholak, were allowed to interact with one another through a sculptural metaphor. The events they generated by striking the EDholak were "dynamically mapped to a series of geometric operations that generated, deleted, deformed or detached" elements of a visual artefact. Not only did the metaphor render the drummers' actions "visible and distinguishable" to one another, it also encouraged them to interact on a new level through their collaboration to shape the artefact itself.

## 2.2 Other Related Works

Given that our work extends the notion of distributed performance further than the mere exchange of audio, it shares a number common traits with a number of specialized research areas beyond NMP.

### *2.2.1 Interconnected Music Networks*

As our proposed system is largely driven by player-to-player interaction, we are particularly interested in music technology applications that focus on increasing the level of interplay between musicians. One such research area is that of Interconnected Music Networks. IMNs are live performances where players can influence, share and shape each other's music in real-time. Clearly, traditional performance can to a certain extent be considered a form of IMN, as music-playing is a highly interdependent art form. Nonetheless, while co-present musicians can influence each other a great deal, the level of control over this influence is rather limited. For example, a soloist can steer her collaborators towards a musical idea in which she is interested, but this type of influence is more of a suggestion, lacking in any direct control over the other musicians' instruments. As the introduction of New Musical Interfaces (NMIs) facilitated the construct of electronic communication channels between instruments, musicians became able to take a fully active role in determining not only their own musical output, but their peers'. IMNs differ most notably from NMP systems in the fact that they do not necessitate participants to be apart, and can in fact be used in a shared space.

When the commercialization of personal computers began in the 1970s, the League of Automatic Music Composers became the first group to write interdependent computer compositions. Dubbing the new genre "Network Computer Music", the group set up a three-node network, mapping frequencies from one computer to generate notes in another, or mapping intervals from one composition to control rests and rhythmic patterns in another. The League of Automatic Music Composers evolved into The Hub in 1986, improving their communication schemes through the use of MIDI data more accurately exchanged through a central computer. More recent examples of IMNs include William Duckworth's "Cathedral", which in fact was the first piece composed specifically for the Web in 1997. Through a Java Applet, users could trigger sounds by clicking on nodes hidden in the screen. Although the original sounds were composed by Duckworth, players could contribute their own sounds to the mix [11]. Another example is "Variations for WWW", an application introduced by Yamagishi in 1998. The goal of the project was promoting "interactivity as opposed to unilaterality" and "sharing as opposed to monopolizing" [30]. Remote users could access a MAX patch connected to the Internet and manipulate parameters that were then sent to a MIDI synthesizer. The resulting output was then transmitted back to the participant. Users could play the combined manipulations of all the other users, and modify their own contribution in response. Similar to this is the Palette, an online system that not only allows participants to share music in the form of MIDI events, but also to control the "style and "energy" of content uploaded by others [31]. We believe that the philosophy behind IMNs idea bears some resemblance to Interactive Installations, as both encourage interplay between multiple users, all while remaining completely accessible.

### *2.2.2 Interactive Installations*

Given our interest in exploring specialized forms of interaction, we turned to the existing body of work on Interactive Installations for further guidance. We use the term Interactive Installations (IIs) to denote works that are commonly referred to in the literature as "Interactive Sound Installations" or "Interactive Art Installations". Like Interconnected Music Networks, Interactive Installations are an example of highly collaborative interfaces, inviting users to communicate with and influence each other through the system at hand. However, they differ from IMNs in a number of ways. First, while the bar for entry in terms of musical expertise can vary widely for IMNs, IIs are designed with public accessibility in mind. Ideally, participants should be able to walk up to an installation and fully explore it with no prior training or experience. In addition, an II is typically a vehicle for communicating its creator's message or intent by means of audience interaction with the work [18]. IMNs, on the other hand, serve as musical instruments for performance or composition that encourage higher levels of interaction between participating users. An example of Interactive Installations is Iamascope, where a camera captures viewer images and movement that are in turn used by a controlling computer to project corresponding kaleidoscope-like images and create accompanying music [16]. Absolute 4.5 is another example, where participant presence is determined through floor sensors and used to generate a complex soundtrack and a large grid of colours projected on a screen [12][1]. The Intelligent Street System further illustrates the accessible nature of IIs: as an alternative to the often undesirable "Muzak" heard in public spaces, it allows users to request changes via mobile text messages. The overall result is to turn visitors of a space from passive consumers to active participants creating their own aural landscape [22]. Similarly, the Control Augmented Adaptive System for Audience Participation (CAASAP) was a project designed to examine a variety of ways in which audience members could make use of mobile phones to become part of the music-making process [28]. Finally, Feldmier et al. created low-cost wireless motion sensors that enabled them to estimate the level of activity of a large scale crowd. The data could subsequently be used to generate music and lighting effects, thereby essentially allowing members of the crowd to drive the music to which they danced [15].

#### *2.2.3 Reactive Environments*

Interactive Installations bring to mind another area offering rich examples of hands-free, highly specialized interactions: Reactive Environments. In fact, we consider the Reactive Environments (also sometimes referred to as Responsive Environments) as the more utilitarian counterparts of Interactive Installations. However, while users partaking in interactive art are typically aware of the process they become one with, the most defining tenet of a reactive environment was perhaps best described by Elrod, who said such a system "should do its job well enough that the occupants are usually not aware of its presence" [14]. Considered by many to be an extension of Ubiquitous Computing, reactive environments gained momentum in the 1990s as a solution to "reduce the cognitive load of the user by allowing the system to make context-sensitive reactions in response to the user's conscious actions" [10]. The concept can in fact be traced back to Elrod, who sought to interconnect Xerox PARC's rich computational infrastructure with a computerized building-management system that could save energy based on office occupancy. Dubbed "Responsive Office Environments", the system made use of small, low-cost sensors to determine whether a worker was present in her office, and made changes to heating, air conditioning, lighting and desktop appliances accordingly. As another example, Wellner's DigitalDesk was designed to merge the advantages of the physical workstation with those of the electronic one. The system could project electronic images onto paper documents, respond to interaction with pens or bare fingers and read paper documents placed on the desk. Thereby, it allowed office workers to transition seamlessly back and forth between the physical and digital desktop. In 1997, Cooperstock et al. created the Reactive Room in response to the frustrations experienced by users interacting with traditional videoconferencing systems. The Reactive Room reacted to a user's high-level actions instead, letting the technology itself manage the low-level operations between the various pieces of equipment [9]. The developers of the Reactive Room believed that the questions they tackled "are not endemic to videoconferencing but apply equally well to other physical environments such as power-plant control rooms, flight decks, and so-called 'smart homes', as well as to software environments such as integrated office suites". We believe they are also applicable to performance environments, where technology can seamlessly augment music-making with functionalities never before available. In fact, such a notion was first explored by Livingstone and Miranda in 2004. The authors developed a novel sonic controller that "regenerates a soundscape dynamically by mapping 'known gestures' to influence diffusion and spatialization of sound objects created from evolving data", and dubbed their system a "responsive sonic environment" [21]. Shortly after, Salter began to explore the use of responsive environments for traditional live performance. The result was Schwelle, a large-scale interactive theatre performance where the rhythm and exerted force of the performers' movements were used to dynamically change a musical composition to "give the impression of a living, breathing room for the spectator" [25].

## 3. SYSTEM OVERVIEW

Our research focuses on the user-centric design and development of a reactive environment for Network Music Performance. As discussed earlier, we regard the network as a unique performance space, and argue that NMP should not have the replication of co-present, traditional performance as its goal. Furthermore, we want to help restore the social aspects of performance, which are too often lost in a distributed setting, by increasing the level of interactions among participants. In order to encourage musicians to delve into new sonic territories, and to make the overall concept of distributed performance more alluring, we believe that our system should enable them to experiment with paradigms that traditional performance does not support. Determining which functionalities our system should afford, however, was not a trivial task. We had a number of criteria in mind. First, we wanted to offer musicians unprecedented control over aspects of their instrumental mix at any given time, a job typically relegated to the soundman prior to the start of a performance. Second, we wanted all controls to be transparent and, therefore, our mappings had to adhere to a clear metaphor. Finally, as a means of extending the social aspects of traditional ensemble music into NMP, we wanted the system to be driven by the interpersonal interactions between the distributed musicians. In order to follow these guidelines, we began by placing the musician, our target user, at the center of our efforts. Subsequently, through extensive user observation [13], we identified two functions for our reactive environment:

1. **Dynamic Volume Mixing (DVM):** Within each local musician's space, the remote participants are given position coordinates, or a *virtual location*. As one musician moves towards another's virtual location, both can experience each other's instruments as gradually increasing in volume. To formalize this feature, assume that $M$ musicians are interacting, each located at virtual location $\vec{x}_i \in \mathbb{R} i = 1, \ldots, M$. Furthermore, assume that the musicians produce the source audio signals $s_i(t)$. Using DVM, we describe the mix $m_i(t)$ that the musician $i$ receives by

$$m_i(t) = \sum_{j \neq i}^{M} a_{ij} s_j(t) \tag{1}$$

where

$$a_{ij} = f(\|\vec{x}_i - \vec{x}_j\|) \tag{2}$$

is given by a function that increases monotonically until the argument falls below a threshold $\theta$. Practically, an exponential function models a linear increase on the decibel scale and matches users' expectations.

2. **Enhanced Stereo Panning (ESP):** As a musician turns his head about, he can experience a stereo panning effect based on his orientation relative to the remote participants' virtual locations within his local space. Our implementation takes the formalism for DVM, and extends the mix for musician $i$ to a 2D-vector $\vec{m}_i = (m_{Li}, m_{Ri})$, representing the left and right audio channels. This allows us to create interactive, spatially structured sound mixes. As we naturally orient towards sound sources in which we are particularly interested, ESP likewise enables an intuitive navigation of the mix. Other musicians deemed to be of less interest are in turn routed to one's spatial periphery, left or right, according to their position. The formalism to create this effect is to compute

$$m_{ki}(t) = \sum_{j \neq i}^{M} a_{kij} s_j(t) \tag{3}$$

where $k \in \{L\text{=left}, R\text{=right}\}$, and $a_{kij}$ are now channel-wise mixing coefficients that depend both on the distances to other musicians $j$ and the orientation of musician $i$. We introduce the unit length vector $\vec{e_i}$, which points from the right to the left ear. An intuitive approach is to set

$$a_{kij} = \frac{a_{ij}}{2}\left(1 + b_k \frac{(\vec{x}_i - \vec{x}_j)}{\|\vec{x}_i - \vec{x}_j\|} \cdot \vec{e}_i\right)^2, \tag{4}$$

with $b_L = -1, \quad b_R = 1$. The scalar product between the difference vector and the ear-connection vector is within the range $[-1, 1]$ and the $(\cdot)^2$ ensures that the overall energy of the source signal remains constant when orienting the head towards a musician.

Our system is currently intended for use in more relaxed, informal settings, such as loose rehearsals and jam sessions, rather than live performances before an audience. Naturally, to experience the changes described above, musicians must receive their audio mix through headphones, which is typically the case in studio situations. In addition, the instruments used must be all electric or electronic rather than acoustic. This ensures that the modified audio mix played back to each musician is not overshadowed by the actual sound of the instruments themselves. Together, these functionalities enable each musician to create his own individualized mix, simply by moving about his space. While Dynamic Volume Mixing allows performers to notably affect each other's volumes, Enhanced Stereo Panning enables each participant to have a more individual and private focus on other members of the ensemble. Furthermore, they are based on exaggerated properties of sound that we experience everyday: sound sources closer to us are louder in volume, and we already experience a mild form of stereo panning when we move our heads due to the human ear's sound localization abilities. In other words, they satisfy the benchmarks established above. To situate our work within the context of HCI, our system is a reactive environment that responds to a user's actions without necessitating that she detaches herself from the task of music-making. Within the framework of NMP, our reactive environment is a distributed performance space that aims to increase the level of interaction among remote musicians. In this regard, it also shares its ideology with that behind Interconnected Music Networks. It also supplements shared video though the use of a simple Graphical User Interface (GUI) that displays dynamic visual representations of the environment. Finally, the simplicity of our system, as it invites users to shape their environment through clear and direct mappings between interaction and output, exhibits the accessibility typically seen in Interactive Installations.

# 4. DESIGN AND DEVELOPMENT

Having validated some of our functionalities in a co-present setting, as detailed in reference [13], we set about "distributing" the system across three separate locations, as described below.

## 4.1 System Configuration

Our reactive environment is currently deployed across three locations and, as such, can naturally accommodate only three musicians. The hardware configuration of each space can be seen in Figure 1.

### 4.1.1 Video

We opted to create a simple yet stable setup using analog cameras connected directly to Panasonic BT-LH1700W production monitors. Each location includes two monitors, with a camera mounted behind each to maintain a reasonable line of sight across the distributed musicians. Although we are interested in exploring alternative forms of visualizations (further described below), we opted to first supplement them with video in order to better explore the effects of each on distributed performance.

### 4.1.2 Audio

In each location, the musician's instrument is plugged into a Roland Edirol FA-101, which in turn feeds the audio into a computer. The signal is then processed through SuperCollider, where volume and panning are adjusted in accordance with our DVM and ESP system functionalities. Musicians are also able to communicate with one another verbally via Sennheiser MKE-2 lavalier microphones.

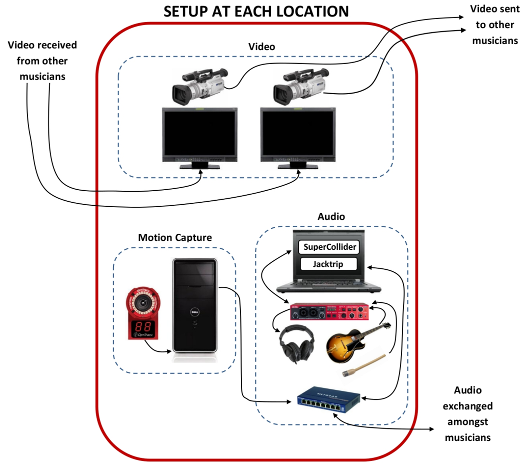


**Figure 1: Overview of system setup at each location.**

**Figure 2: A musician performing within our reactive environment.**

Audio streams (from Supercollider and the microphones) are then shared among all three locations through Jacktrip. We chose Jacktrip as it interfaces easily with SuperCollider through the Jack audio server. To further reduce delay and guarantee sound stability, a real-time kernel is used on all machines executing Jacktrip, and a Local Area Network (LAN) was created to connect them through a Netgear ProSafe 8 Port Gigabit Switch. This setup allowed us to bring down our one-way latency to 11.6 ms, far below the established EPT of 25 ms. Finally, each musician is able to hear his own individual mix through a pair of Sennheiser HD Pro 280 closed headphones.

### 4.1.3 Motion Capture

One of the locations is equipped with a Vicon motion capture system, while the remaining two have been fitted with the more portable Optitrack system. Position and head orientation is determined via markers fitted on a hat worn by each musician, while body orientation is determined through markers attached to an adjustable elastic band strapped across the chest. Such a configuration provides us with all the information needed to implement both DVM and ESP. All motion capture data is then sent via Open Sound Control (OSC) through the gigabit switch to the machines running SuperCollider. A musician interacting with our system in one of the locations can be seen in Figure 2.

## 4.2 Feature Implementation

Despite the musicians' lack of shared presence, we had to maintain the concept of "physical distance" between them in order to implement DVM. As a result, all three motion capture systems have been calibrated such that the spaces they cover form the Cartesian space depicted in Figure 3. This also allows the musicians to feel that they are in essence sharing a space, an idea that is further reinforced through our GUI's visual representations of the environment, as further described below.

# 5. GRAPHICAL USER INTERFACE

In addition to invisibility, Cooperstock et al. list two other factors deemed critical to the usability of reactive environments: feedback

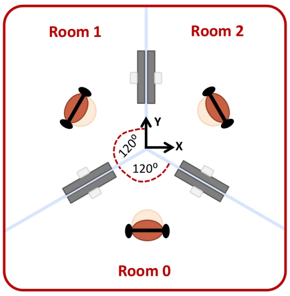


**Figure 3: An overview of the "shared space" created by superimposing all three locations.**

and manual override [9]. Therefore, we wanted even our most non-technical users to be able to operate our system entirely on their own, without relying on the help of experts to carry out what should, in theory, be simple operations such as setting connections, starting and stopping the system, or changing preferences. We created the GUI seen in Figure 4 to allow musicians to connect with remote participants and to adjust their own settings. Once a musician clicks the 'Connect' button, we consider him to be "online". He is told whether any other musicians are also online, and once all three participants are connected, they can start their session. The knob on the left-hand side allows each musician to adjust his own base volume as heard by all participants. The slider on the bottom allows him to personally determine the "sensitivity" of DVM, with a larger value leading to more dramatically noticeable results. As described in Section 2.1.1, Schroeder noted that musicians barely made use of the full-frontal video made available to them through iChat during distributed performance. Therefore, we were highly inclined to consider alternative forms of visualization to supplement shared video. Most importantly, we wanted to evoke the sense of a "shared space" by showing musicians where they stood in relation to one another, in spite of their remoteness. In addition, while the effects of DVM and ESP can be heard, we also wanted to provide graphical feedback to further reinforce each participant's state at any given point. Seeing as we wanted to avoid having our musicians focus on a computer screen for any significant amount of time mid-performance, we knew that all visual information had be easily understandable in a matter of seconds. Therefore, we conducted a brief user experiment where we polled subjects on their preferred simplest visual representations of position, orientation, volume and panning, and used the results to drive our design. As illustrated in Figure 4, the musicians are represented as "avatars", whose heads and shoulders can be seen from a bird's eye view. Such avatars implicitly provide body orientation and position information. Head orientation, and the resulting level of stereo panning, are illustrated through the head marker sliding towards either side of the shoulders, almost representing a crossfader. Finally, volume levels are illustrated through the number of crests in the "waves" emanating from the avatars.

## 6. USER FEEDBACK: THE NET-MUSIC 2013 SYMPOSIUM

A trio of rock musicians, consisting of a lead singer and rhythm guitarist, a lead guitarist, and a bassist, performed using our system as part of a demo for "Net-Music 2013: The Internet as Creative Resource in Music", an international, multi-site telematic symposium. Overall, the musicians required very little training to become familiar with the system and its functionalities. When asked if the latency was perceptible, they reported not finding it at all problematic. They were easily able to find suitable volume levels, and dynamically adjusted them throughout the performance through the DVM functionality. In comparison, however, ESP was used far less often. The bassist explained that he did not quite understand the use for ESP, as DVM seemed to provide him with enough volume control. He added, however, that he could perhaps grasp its usefulness after using the system over more extended periods of time. Interestingly, the bassist reported not really making use of the video displays, but this is consistent with Schroeder's findings, as described earlier. He attributed this lack of attention to video, in part, to the fact that the musicians were already quite familiar with the songs and one another. Finally, the musicians commented on the GUI several times, especially with regards to the visual representations, often joking when their avatars were getting increasingly closer to one another as they moved about, even though the musicians were obviously in different rooms. This helped indicate their perceived sense of shared space. However, when asked about the usefulness of the avatar representation, the bassist explained that he was unsure of its purpose. He described the variations in volume triggered by the DVM function as being a "good feedback for distance" and, therefore, he did not find it necessary to gauge that information from the avatars themselves. He added, however, that with time, he might find the visual representations more useful. The lead guitarist reported finding the visual layout of the GUI, especially the volume knob and DVM sensitivity slider, to be "very simple to use and very responsive". Overall, the musicians found our performance environment enjoyable and easy to use, with the lead guitarist adding that he "saw great potential in the arrangement".

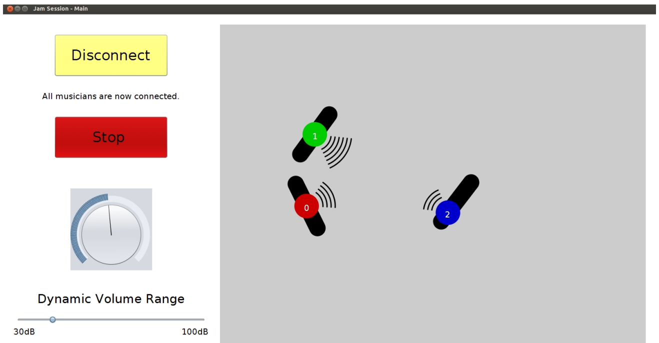


**Figure 4: The Graphical User Interface, seen from the perspective of Musician 0 (red). Musician 1 (green) is closer to him than Musician 2 (blue) and, as a result, the former's louder volume is illustrated through his larger number of waves. In addition, Musician 2 (blue) has activated stereo panning by turning his head to the left, as depicted through his circular head marker leaning in that direction.**

## 7. FUTURE WORK

Given that our system was designed with ease of use in mind, and that we would like to distribute it to a large number of musicians, we are currently implementing a version that tracks user position and orientation via the Microsoft Kinect, a much more affordable and widely available alternative to traditional motion capture systems.

In terms of further refining the functionality of the system, the next step in our user-centric approach is to conduct formal user experiments. The results from such tests will heavily drive the design of our next development iteration, which we hope will meet musicians' expectations even more closely than our current design. In particular, while the usefulness of DVM seems to be immediately apparent, we believe that more extended test sessions are needed to determine whether the musicians might eventually gauge the advantages of ESP. Formal tests with our earlier prototypes have thus far utilized a large number of customized data collection techniques, described in reference [13], namely: logged position and orientation data, detailed questionnaires after each session to assess flow, creative engagement and self-expression, and audio and video footage for contextual analysis of the performances. In addition to those, we would like to explore holding non-leading interviews with musicians at the end of each test day to obtain a more detailed overview of their impressions of the system, along with any suggestions for improvement.

We are also interested in evaluating the benefits of our system after longer periods of regular usage: will DVM and ESP continue to be features that musicians enjoy or find interesting? Or will they simply tire of the system's novelty? As a result, we have invited a composer to interact with our reactive environment and write a musical piece that specifically exploits the features of our reactive environment. Throughout the process, he will also be encouraged to share his thoughts on the system's design, particularly how it could be improved to better sustain long-term engagement. Finally, we look forward to inviting a group of musicians as part of an "artist residency" program, where they will be encouraged to explore new sonic interactions using our reactive environment throughout an extended period of time.

## 8. CONCLUSION

We described our efforts thus far towards the development of a reactive environment for Network Music Performance, where participants can affect each other's volumes and experience enhanced stereo panning. Our aim is to guide musicians towards perceiving the network as an exciting performance milieu that is inherently different from the traditional, co-present case. In addition, we believe that our system can increase the level interaction between remote musicians by conveying a perception of shared space. By maintaining our user-centric approach, which dictates an iterative process of formal user tests and subsequent prototype refinements, we are confident that we can arrive at a final system that helps increase the distributed performers' sense of flow, creativity and self-expression. Ultimately, we hope that our work will encourage more musical interface designers to consider the merits of UCD methodologies.

## 9. ACKNOWLEDGEMENTS

The authors would like to thank Oliver Sullivan, Dan Clahane and Niall Skinner for their participation in the Net-Music 2013 demo, and their input throughout this project. In addition, they would like to thank Marcelo Wanderley and Frank Ferrie of McGill University for their their invaluable help and advice. The research described here was supported by the Natural Sciences and Engineering Research Council of Canada.

## 10. REFERENCES

[1] Z. Bilda, E. Edmonds, and L. Candy. Designing for Creative Engagement. *Design Studies*, 29(6):525–540, Nov. 2008.

[2] J.-P. Cáceres and C. Chafe. Jacktrip: Under the hood of an engine for network audio. *Journal of New Music Research*, 39(3):183–187, 2010.

[3] G. Cantor. eJamming audiio. http://www.ejamming.com. Accessed July 21, 2011.

[4] A. Carôt. Soundjack. http://www.soundjack.eu/. Accessed July 21, 2011.

[5] A. Carôt, P. Rebelo, and A. B. Renaud. Networked Music Performance: State of the Art. In *Audio Engineering Society Conference 30th International Conference: Intelligent Audio Environments*, March 2007.

[6] C. Chafe, S. Wilson, A. Leistikow, D. Chisholm, and G. Scavone. A Simplified Approach to High Quality Music and Sound over IP. In *Proceedings of the COST G-6 Conference on Digital Audio Effects (DAFX-00)*, pages 159–164, 2000.

[7] J. R. Cooperstock. History of Spatially Distributed Performance. http://www.cim.mcgill.ca/sre/projects/rtnm/history.html. Accessed July 28, 2011.

[8] J. R. Cooperstock. McGill Ultra-Videoconferencing. http://www.cim.mcgill.ca/sre/projects/rtnm/. Accessed July 28, 2011.

[9] J. R. Cooperstock, S. S. Fels, W. Buxton, and K. C. Smith. Reactive Environments: Throwing Away your Keyboard and Mouse. *Communications of the ACM*, 40:65–73, 1997.

[10] J. R. Cooperstock, K. Tanikoshi, G. Beirne, T. Narine, and W. A. S. Buxton. Evolution of a Reactive Environment. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '95, pages 170–177, New York, NY, USA, 1995. ACM Press/Addison-Wesley Publishing Co.

[11] W. Duckworth. Cathedral. http://www.monroestreet.com/Cathedral/main.html. Accessed July 22, 2011.

[12] E. Edmonds. Abstraction and Interaction: An Art System for White Noise. *International Conference on Computer Graphics, Imaging and Visualization*, 0:423–427, 2006.

[13] D. El-Shimy, T. Hermann, and J. R. Cooperstock. A reactive environment for dynamic volume control. In *New Interfaces for Musical Expression (NIME)*, May 2012.

[14] S. Elrod, G. Hall, R. Costanza, M. Dixon, and J. Des Rivières. Responsive Office Environments. *Commun. ACM*, 36:84–85, July 1993.

[15] M. C. Feldmeier. *Large Group Musical Interaction Using Disposable Wireless Motion Sensors*. PhD thesis, Massachusetts Institute of Technology, 2002.

[16] S. Fels and K. Mase. Iamascope: A Graphical Musical Instrument. *Computers and Graphics*, 2:277–286, 1999.

[17] K. Galloway and S. Rabinowitz. Satellite Arts Project. http://www.ecafe.com/getty/SA/. Accessed August 22, 2011.

[18] K. Höök, P. Sengers, and G. Andersson. Sense and Sensibility: Evaluation and Interactive Art. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '03, pages 241–248, New York, NY, USA, 2003. ACM.

[19] S. Jordà. FMOL: Toward User-friendly, Sophisticated New Musical Instruments. *Computer Music Journal*, 26:23–39, September 2002.

[20] A. Kapur, G. Wang, P. Davidson, and P. R. Cook. Interactive Network Performance: A Dream Worth Dreaming? *Org. Sound*, 10:209–219, December 2005.

[21] D. Livingstone and E. Miranda. Composition for Ubiquitous Responsive Sound Environments. In *Proceedings of the International Computer Music Conference*, pages 321–324, 2004.

[22] H. Lörstad, M. d'Inverno, and J. Eacott. The Intelligent Street: Responsive Sound Environments for Social Interaction. In *Proceedings of the SIGCHI International Conference on Advances in Computer Entertainment Technology*, pages 155–162, New York, NY, USA, 2004.

[23] A. Renaud. Dynamic Cues for Network Music Interaction. *7th Sound Music and Computing Conference*, July 2010.

[24] A. B. Renaud and P. Rebelo. Network Performance: Strategies and Applications. *Proceedings of the 2006 Conference on New Interfaces for Musical Expression (NIME)*, 2006.

[25] C. Salter, M. Baalman, and D. Moody-Grigsby. Between Mapping, Sonification and Composition: Responsive Audio Environments in Live Performance. In *Computer Music Modeling and Retrieval: Sense of Sounds*, pages 246–262. Springer-Verlag, Berlin, Heidelberg, 2008.

[26] F. Schroeder and R. Pedro. Sounding the Network: The Body as Disturbant. *Leonardo Electronic Almanac*, 16, 2009.

[27] F. Schroeder, A. B. Renaud, P. Rebelo, and F. Gualda. Addressing the Network : Performative Strategies for Playing Apart. In *International Computer Music Conference*, 2007.

[28] K. Tahiroglu. Dynamic Social Interaction in a Collective Mobile Music Performance. In *Proceedings of the 2009 International Conference on Computational Science and Engineering*, pages 854–859, Washington, DC, USA, 2009.

[29] G. Weinberg. Interconnected Musical Networks: Toward a Theoretical Framework. *Computer Music Journal*, 29:23–39, June 2005.

[30] S. Yamagishi. "variations for WWW: Network Music by MAX and the WWW. In *Proceedings of the 1998 International Computer Music Conference*, pages 510–513, 1998.

[31] J. Yu. Computer Generated Music Composition. Master's thesis, Massachusetts Institute of Technology, 1996.