Real-Time Sound Source Spatialization as used in *Challenging Bodies*: Implementation and Performance

JJ Nixdorf
Department of Computer Science
University of Regina
Regina, Saskatchewan, Canada
nixdjosh@cs.uregina.ca

David Gerhard
Departments of Computer Science and Music
University of Regina
Regina, Saskatchewan, Canada
gerhard@cs.uregina.ca

ABSTRACT

In this paper we will report on the use of real-time sound spatialization in *Challenging Bodies*, a trans-disciplinary performance project at the University of Regina. Using well-understood spatialization techniques mapped to a custom interface, a computer system was built that allowed live spatial control of ten sound signals from on-stage performers. This spatial control added a unique dynamic element to an already ultramodern performance. The system is described in detail, including the main advantages over existing spatialization systems: simplicity, usability, customization and scalability

Keywords

surround sound, sound spatialization, sound localization, sound architecture, real-time systems, performance systems, live systems, `pd`, GEM

1. INTRODUCTION

In June 2005, researchers from various faculties of the University of Regina, including Music, Theatre, Dance, Kinesiology and Computer Science, came together to collaborate on *Challenging Bodies*[1]. Billed as "A Multi- disciplinary Performance for Variously-Abled Artists", the purpose of the project was to challenge the notion of who is abled and who is disabled. The main performer of this project, Craig Fisher, has cerebral palsy and is confined to a wheel chair. A custom computer system called Cool Moves was developed locally by Music Therapist Doug Ramsay based on David Rokebys Very Nervous System (VNS)[2]. Using this, Craigs limited range of controllable movement was mapped to a virtual instrument. Other facets of the performance included an interpretive dance group featuring a performer confined to a wheel chair (a "sit-down dancer"); a visualization system that projected graphics based on Craigs musical output; A video-mixing artist using previously recorded motion-capture footage of the dancers, and a full-theatre sound spatialization system.

The spatialization system, built in Pure Data (`pd`) [4], took all of the sound input from each of the ten instruments (played by six performers) and mapped those input channels to a position in a virtual representation of the performance theatre. Using fundamental sound localization techniques, these virtual locations were spatialized to a corresponding position in the actual theatre. Most importantly, this interface operated in real-time, allowing the user to animate a sound throughout the theatre, thus giving the impression of a moving musician who could also respond to the motion of the virtual source through which he was playing. The spatialization system user and the instrumentalists therefore needed to be very attentive to each other, and the spatialization system user became like a member of the ensemble. The interaction brought to the fore some very interesting reactions to live spatialization in performance, which will be described in Section 4 after the system itself is described in the following sections.

[1] `http://uregina.ca/~challbod/`
[2] `http://www.davidrokeby.com/`

2. BACKGROUND

Sound spatialization (localization, sound architecture) focuses on the position and motion of sounds through an actual or virtual performance space. Sound localization for musical expression is a familiar concept in current new interface studies. Most presentations of new music and interface technology use some form of spatialization to enhance the performance. Often, this spatialization is recorded beforehand and the movements of the individual sound sources are fixed during performance. *Diffusion* refers to (among other things) the live spatialization of mono or stereo sources by fading or panning them between speakers in a multi-channel environment, causing the sounds to appear to be coming from different angles. The interfaces to diffusion systems are often based on the physical mixers and connections to the speakers themselves, and diffusing more than one source at a time is a difficult task, let alone diffusing ten or more sources independently [7, 3].

The use of spatial music in performance is not new. Classical music examples include spatial antiphony in choir performance (Willaert and Gabrieli); Mozart's pieces for multiple orchestras (K. 239 and K. 286); and Verdi's Requiem (Tuba Mirum), wherein trumpets are positioned offstage. Notwithstanding these, localization is a relatively uncommon technique in popular classical music, and is often considered a novelty rather than an expressive technique. This is not surprising given the cost and effort required to produce any complementary result. Historically, implementing localization on any grand scale would have been a cumbersome endeavor.

As early as 1956 [5], sound installations capable of spatializing sound through a loud speaker setup begin to appear. These installations vary wildly in form and function. Due to the hardware involved, the compositions for these arrangements are usually specific to the installation they were composed for, thus the installation is little more than a glorified instrument. In 1990, a Stanford Researcher, Marina Bosi, produced an entirely digital spatialization system [1], which was an early instance of using multiple speakers in a connected way, as opposed to a forest of independent sound sources. This system accepted MIDI signals as input and spatialized these signals into an adjustable quadraphonic setup. This was a step towards allowing a composer to use spatialization in any performance, however MIDI was still a limiting factor on the performance characteristics of the system. The more recent *GMEM* spatialization system, *Holo-Spat*[3], has improved functionality but will spatialize any sound file, not just MIDI. *Holo-Spat* has an extremely limited interface for spatialization control and requires it's sister program, *Holo-Edit*, in order to perform any complex spatialization. *Holo-Edit* records the desired spatialization to a MIDI control file, which *Holo-Spat* then uses to localize the sound sources. The only downside to this arrangement is that since *Holo-Spat* and *Holo-Edit* are uncoupled, it is impossible for any complex spatialization to be played as it is composed, or to be performed in real-time. Although beneficial for composing and recording purposes, this is certainly not advantageous for live performance. Another current system is the IRCAM[4] *SPAT* , which has many of the same features as *Holo-Spat*. *Holo-Spat* and *SPAT* both require the purchase of Max/MSP[5] in order to customize the system, and Max/MPS licenses are expensive for individual users.

The system discussed in this paper has a unique interface which allows for the real-time manipulation of sound sources in the virtual environment. The spatialization features implemented are based on the requirements for the *Challenging Bodies* application, and do not make use of such straightforward spatialization techniques as distance cues or doppler shift. Earlier versions of the system did have all of the standard spatialization cues, and they are still present within the code of the system.

3. IMPLEMENTATION

The spatialization system was implemented from the beginning based on three primary concepts: scalability, usability, and non-proprietarity (no monetary cost). Each of these concepts represents an advantage over existing systems. When we began this work we could not find a complete, usable, open-source spatialization interface for `pd`. All of the systems described earlier have an associated cost (our system is freely available[6]) and are not as usable as they could be. For instance, the *Holo-Spat* + *Holo-Edit* pair relies heavily on modes, considered to be a disadvantage for usability [6]. Modern usability theory has yet to be applied to the majority of computer-music interface applications.

The scalability of the system can be seen in Figure 1, where each sound source signal is applied to a router, which compares the virtual location of that source with the virtual location of all speakers, and routes the source to each speaker, with the appropriate attenuation. The addition of a new source or a new speaker is theoretically trivial, corresponding of the addition of a new router or mixer respectively. As indicated by the dashed line in the diagram, the GUI currently does not alter the position of the speaker, but there is no theoretical reason why the speakers cannot be manipulated in the GUI.

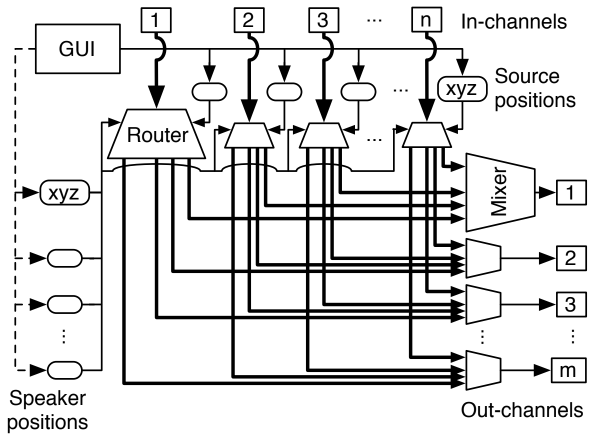

Figure 1: The flow layout of the system. Each router takes the relative position of its source and all speakers and calculates the relative amplitude of that source for each speaker.

The main interface of the spatializer consists of two primary windows, both of which exist on screen simultaneously. The first is the primary graphical interface of the spatializer, the GEM[7] [2] window. It paints a simulated speaker environment to the screen and allows for mouse manipulation of the spatial location of the sound sources. The GEM window, shown in Figure 2, is intended to be a representation of the physical speaker arrangement of the performance hall as illustrated in Figure 3. The speakers are arranged in an array behind the performers, with five speakers at audience level from stage right to stage left, and two speakers in the catwalk for elevation simulation. The decision was made not to have speakers behind the audience in an effort to increase the size of the "sweet spot," the audience location with the ideal listening conditions. The speaker layout in the GEM window is therefore different from the standard speaker-ring configurations of current systems. Instead, the system is presented as if from the audience's point of view. This shows that spatial interfaces can be designed from a multitude of perspectives, and that this system is capable of localizing sound sources in any general speaker arrangement. Indeed, whereas the presented layout is limited to 2-D by the single orthogonal view on the screen, multiple views or a single oblique view would provide opportunities for manipulating sources in 3-d in real-time. The current implementation encodes the location of the speakers and the sources in 3-d, but because a single layer of speakers is used, it can be projected onto a 2-d interface.

The second window is a `pd` patch, `[rtsss.pd]`, which has the master output volume control as well as individual sliders controlling the input level of each sound source. It should be noted here that while the speaker arrangement

[3]`http://www.gmem.org/`
[4]`http://www.ircam.fr/`
[5]`http://http://www.cycling74.com/`
[6]`http://armadilo.cs.uregina.ca/rtsss`
[7]`http://gem.iem.at`

was fixed for this performance, in general the speakers can be moved and positioned anywhere in 3-space. This is a significant advantage of this system over current implementations which normally assume a ring of 5 or 8 speakers in a standard configuration. Standard configurations are ideal for recorded spatialization playback since composers can rely on a known arrangement of speakers, however, this requirement is less important than the actual speaker arrangement for live performance of spatial source positioning, and variable speaker placement allows the accurate representation of the playback space, since speaker arrangements in real performance venues are rarely perfectly circular.

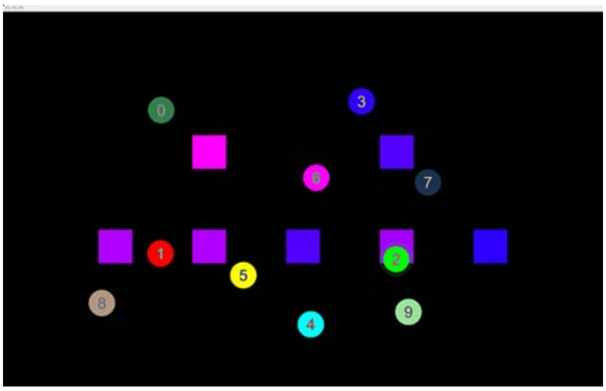

Figure 2: The GEM window. Squares represent speakers. Circles represent sound input signals.

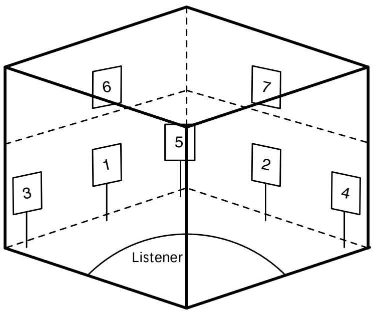

Figure 3: The physical speaker arrangement in the concert hall.

4. PERFORMANCE TESTING

Although the spatializer was completed only a short time before the performance date, it had undergone extensive testing and two complete versions during development. This ensured that the more fundamental issues had been discovered and corrected, and that the system was functionally correct, but this did not ensure performance success. Testing during development typically consisted of spatializing two or three simple sound sources (`[osc~]`) in a lab environment. This ensured that the system performed the spatialization tasks, but gave no guarantees of successful performance use. Final testing came during the dress rehearsal, when ten instruments where used as simultaneous input and spatialized around the performance hall in real-time. Because the *Challenging Bodies* project incorporated performance features from many disparate disciplines, the dress rehearsal was the first time all pieces were assembled into a cohesive whole. It speaks to the professionalism and preparation of all parties that the show came together as expected, but there were a few kinks that were worked out during or immediately after the dress rehearsal.

One technical issue that was resolved during the dress rehearsal was the time delay on the system. The default time delay for `pd` is 50 milliseconds, and the longer the time delay, the larger the audio buffers and the less just-in-time processing is required by the computer system. 50 milliseconds, however, is on the cusp of perceptibility, especially when professional musicians are playing through the system in real time. All of the musicians mentioned that there was a small but perceptible delay in the playback of the sound through the spatializer. The time delay was reduced but at 10 milliseconds there was audible clicking and a reduction in performance. It was discovered that a delay of 25 milliseconds was a reasonable compromise that allowed the performers to interact with the system and still allow the system to perform well.

The performance was a success: the spatializer performed flawlessly from a technical standpoint The show itself, however, shed light on two performance issues that had been overlooked, and could not be fixed between the dress rehearsal and the show proper. These issues were the effectiveness of spatialization of multiple sources and performer interaction.

4.1 Spatialization effectiveness with multiple sources

During the testing phase, only a small number of sounds were spatialized and they proved to be easily discernible by the listeners. In the *Challenging bodies* show, however, there were ten independent sources which were often spatialized to ten different locations. Listeners were able to discern the localization of a single source and follow it around the room in sparse pieces where only a few sources were present, but in pieces where many instruments were localized at the same time, it was more difficult to discern the location of one source in the midst of the soundscape. Spatialization was most effective when only a few musicians were playing, and was especially effective during solos. At the very beginning of the show, during the initial blackout, Craig introduced the show (using his text-to-speech system) and this introduction was spatialized around the room in a very effective use of the technology.

4.2 Monitoring real time live spatialization

A key feature of performance-based sound reproduction systems is monitoring: Performers must be able to hear themselves in the context of the performance in order to be able to interact with the other performers and to fine-tune their own sound. In the *Challenging Bodies* show, the speakers were positioned behind the artists, and it was expected that this would provide sufficient monitoring, given the size of the performance space and the nature of the instruments being played. In a static system (with no spatialization) this would be the case although independent monitor mixing would be impossible. The spatialization of the source sounds added a dynamic element to the output produced by each speaker. The artists, being used to static systems, expected that the speakers would produce a consistent reproduction of their music, but because the sounds were moving around the performance space, the sounds coming from the individual speakers faded in and out. This reduced the monitoring effect of the speakers and reduced the interaction of the musicians with each other and the musical whole, although again because of the smaller performance space, the musicians were also able to interact directly with each other. The overall time delay between the performers actions and the sound from the speakers to their ears varied depending on the virtual location of the sound, but it was the change in amplitude that was most noticeable and most difficult to deal with. Delay between action and sound has an interfering effect when the delay is not correlated with the perceived distance from the sound to the ear. Since the perceived distance was also changing, there was no overt delay effect in the monitoring.

The issue of the monitoring of real-time sound source spatialization is important and difficult, since musicians are more able to interact with the soundscape if they can hear where their source has been moved to, but as evidenced in the experiences of *Challenging Bodies*, direct individual monitoring is also important. More study needs to be done on the balance between these two reinforcement techniques.

5. FUTURE WORK

The system presented herein was developed for a particular show, and as such it had some limitations that will be easily removed in future versions. The main development next in line is to implement the proposed 3-d interface, incorporating multiple views of the speaker environment, and an oblique view. The development of this 3-d interface means exploring means of manipulating sources in 3-d with a standard 2-d screen window or windows. Several implementations of 3-d object manipulation exist for other applications, and these will be studied in order to inform this system.

The second improvement that will be added to the current system is the ability to manipulate the location of the speakers. The problem with allowing the speakers to be manipulated is that they may be inadvertently moved during performance. One possibility to alleviate this problem is to use a speaker manipulation mode and a normal operation mode, with obvious visual feedback when in speaker manipulation mode (e.g. a change in background colour). Since usability theory suggests that modal interaction should be avoided unless necessary and obvious, a second option is to have a lock on each speaker what would reduce the likelihood of inadvertent manipulation.

A third improvement would be adding the ability to record source motion paths and re-play them, similar to existing spatialization systems. A related improvement would be the ability to manipulate groups of sounds, for example, by moving stereo pairs or clusters of sources.

The system was developed for musical performances, and it was shown to be a useful and powerful tool for live performances. Another area of future work will be to apply this system to other application domains such as sound effects for movies and video games, and for spatialization of sound effects for live plays or musicals. Having the ability, in a usable interface, to manipulate sound effects or foley in real-time for these applications has considerable potential. For instance, it would be quite useful to be able to produce off-stage voices in a live play. We will investigate opportunities to introduce the system to other live venues and applicaitons.

6. CONCLUSIONS

A real-time sound-source spatialization system was developed with three main aims: scalability, usability and open-source. It was implemented with a specific application target, that of the *Challenging Bodies* transdisciplinary project. The advantages that it has over existing spatialization systems are that it allows real-time spatialization of live sound sources, that it is open-source, is intentionally simple in order to enhance usability, and that it allows general speaker arrangements in vertical as well as horizontal layouts. The use of spatialization in the context of live performance was studied, and a number of issues were identified and flagged for future research. The system as developed will be enhanced to provide 3-D interaction, speaker location modification, cluster movement (movement of multiple sources in relation to each other), and implementation in other performance contexts (e.g. theater, games, virtual reality).

Although spatialization has been used in musical performance in the past, it is still just beginning to catch on as a creative aspect of musicality. By employing spatialization technologies such as SPAT, Holo-Spat, or the system described in this paper, spatialization in music can be further explored. Non-real time systems allow for an individual acting as a 'spatial composer' to later add a fixed sequence of effects to a performance, but real-time sound source spatialization as described in this paper turns this composer into a performer.

7. ACKNOWLEDGMENTS

The project was funded by the University of Regina Transdisciplinary project fund, The Natural Sciences and Engineering Research Council of Canada, the Social Sciences and Humanities Research Council of Canada, and the Canadian Foundation for Innovation. The authors would like to thank all who participated in Challenging Bodies, and in particular Doug Ramsay, Daryl Hepting, Ann Kipling-Brown, Kathleen Irwin, and John Barden.

8. REFERENCES

[1] M. Bosi. An interactive real-time system for the control of sound localization. In S. Arnold and G. Hair, editors, *Proceedings of the ICMC*, Glasgow, September 1990.

[2] M. Danks. Real-time image and video processing in gem. In *Proceedings of the ICMC*, pages 220–223, Thessaloniki, 1997.

[3] J. Harrison. Sound, space, sculpture: Some thoughts on the 'what', 'how', 'why' of sound diffusion. *Organised Sound*, 3:117–127, 1998.

[4] M. Puckette. Pure data: another integrated computer music environment. In *Internation Computer Music Conference*, 1996.

[5] C. Roads. *Computer Music Tutorial.* MIT Press, Cambridge, MA, 1996.

[6] M. B. Rosson and J. M. Carroll. *Usability Engineering: Scenario-based Development of Human-Computer Interaction.* Morgan Kaufmann, 2001.

[7] B. Truax. Composition and diffusion: space in sound in space. *Organised Sound*, 3:141–146, 1998.