PLaY+SPaCE: An Ultrasonic Gestural MIDI Controller

Steven Campbell
James Cook University
Vincent Campus
Townsville, 4811
Australia

steven.campbell@jcu.edu.au

Abstract

PLaY+SPaCE is an ultrasonic gestural MIDI control system used to detect positions of people moving within a 100sqm sensing space. Eight ultrasonic sensors form the basis of the system, data from the sensors relayed to a proprietary hardware interface, then on to a range of software modules designed in MAX/MSP. Within the software incoming MIDI data is mapped to trigger sound events (samples, sequences and synthesis parameters) as well as mapped to control audio spatialisation, video and lighting events. This paper provides a broad overview of the system, problems encountered in its development and broad descriptions of trialled applications of the system.

PLaY+SPaCE

PlaY+SPaCE has developed since 1999 as a collaborative project between music and electrical engineering researchers at James Cook University (Campbell 2003). The system forms a Digital Musical Instrument (DMI) and is primarily classified as a *gestural controller*, a system in which sensors form "the part of the DMI where physical interaction takes place" (Wanderley 2001). The sensors in the PlaY+SPaCE DMI are ultrasonic proximity detectors, the system utilising up to eight sensors to form a non-tactile and invisible grid in a sensing space of up to 100sqm in size.

Based on the SOUND=SPACE system developed in the mid 1980s by Rolf Gehlhaar (Gehlhaar 1991), and as in the majority of *Human Computer Interaction* (HCI) devices, the PlaY+SPaCE system is comprised of four parts; the gestural controller (the sensors), an interface (to convert sensor voltage outputs to MIDI data), a software mapping device (to map MIDI data to digital representations) and output devices (software and hardware instruments, lighting systems and video).

This paper provides a broad overview of each of the four parts of the PlaY+SPaCE DMI, highlighting relevant backgrounds, developmental procedures and problematic areas. The paper further relates trialled applications of the system and current research into its development.

Gestural Control

Ultrasonic sensing

As a basis for *Human Computer Interaction* (HCI) devices, non-tactile ultrasonic sensor systems have been utilized in numerous ways over the past two decades; (e.g. Gehlhaar 1991; Camurri 1995; Carter 2001;) and further utilised in combination with other sensors (e.g. Waisvisz 1985; Reynolds et al 2001). Ultrasonic systems have certain limitations, such as "no sensitivity past obstructions, narrow beamwidth, and limited resolution" (Paradiso 1996), however the results from experimentation and utilisation of ultrasonic systems, particularly with their non-tactile nature, have provided positive user feedback in the areas of dance, real-time music performance and in work with people with disabilities (e.g. Gehlhaar 1998).

Such wide applications of ultrasonic sensing in various environments have resulted in, and inspired, the current research. Trials of the PlaY+SPaCE DMI in a variety of applications have led to positive results, confirming the benefits of non-tactile ultrasonic sensing.

Background

A direct experience of Gehlhaar's SOUND=SPACE in 1998 provided the catalyst for this research, clear benefits of that system evident from positive responses in workshops with disabilities and special needs groups (Gehlhaar 1998). Further, applications and possibilities of the system in new music composition and performance held considerable appeal: the non-tactile nature of the system allowing access to music creation and performance by people with limited musical capabilities and experience, through to professional performers. The community experiencing the system at the time was enthusiastic to acquire a permanent installation of the system, however as a proprietary system with a cost of c. $35,000, acquisition was beyond financial scope.

Subsequently, research into the availability of components to develop a similar system was undertaken, firstly, bearing in mind the enthusiasm of the community for such a resource, and secondly as a means to expand personal compositional goals in interactive and algorithmic composition systems, a focus of research in preceding years.

A desire to replicate and develop the Gehlhaar system led to investigation of two existing and commercially available ultrasonic systems. The Soundbeam (Carter 2001) from UK company EMS was found to be limited by its maximum sensing range of six metres and a maximum of four ultrasonic sensors running simultaneously.

The I-Cube system (Mulder 2000) from Canadian company Infusion Systems was more feasible, its Far Reach sensors covering an appropriate range, and up to 32 sensors able to work simultaneously. The Far Reach sensors however generate a level of physical noise that is detrimentally compounded in a multi-sensor system, and as each sensor emits continuously, simultaneous use of two or more sensors results in ultrasonic reflections being detrimentally detected by sensors other than the originating sensor. Regulation of the Far Reach sensor emissions was trialled, however these sensors have a long (c. 200-400ms) power-up time that negates any on/off regulation in a multi-sensor arrangement, any data received in this power-up time being erratic and unreliable.

The lack of feasibility of commercially available sensors led to an investigation of Polaroid sensors, which on power-up remain powered, and their ultrasonic emissions can be regulated. The PLaY+SPaCE system was developed around these capabilities and utilises a programmed microcontroller to regulate emissions of the sensors.

In addition to ultrasonic systems, further sensing possibilities were considered, primarily concentrating on optical systems, though possibilities of microwave and radar sensors were considered and negated on account of high cost and possible human exposure to radiation. Optical systems (e.g. video cameras) have been limited by "obstructions blocking the line of sight, limited angular range, varying reflectance, effects from background light…clutter and changes in the environment" (Paradiso 1996). At the time of initial work on the PlaY+SPaCE system, latency in the use of video was a further factor that negated video use, however as processing speeds and more powerful algorithms have developed, latency has become less problematic, and video sensing is currently considered as a powerful augmentation possibility for the existing ultrasonic system.

Utilisation

The current version of the PlaY+SPaCE DMI is based on eight ultrasonic sensors within a maximum space of 100sqm. The sensors are modified Polaroid sensors (Polaroid OEM Components Group 1999) in which the transmit and receive components are within a single unit. The sensors transmit a 50Khz 'send' signal but only receive a 'return' if the signal is received within a period of 64 milliseconds, i.e. the time taken at the speed of sound (at c. 23°Celsius) for the signal to reach a maximum distance of ten meters and return. Where a signal is received in less than 64ms (i.e. an obstruction of the ultrasonic beam has occurred) an alteration in voltage occurs, this sent as output from the sensor. The eight sensors of the system are wired to the interface, this reliably receiving data from the sensors at a maximum distance of 12 metres.

Current resolution of the sensor beams is equated to milliseconds, allowing 64 subdivisions of each sensor beam, with each subdivision occurring at c. 15cm apart along the beam. With all eight sensors, this equates to 512 mapping points within a 100sqm space, however in practice this is generally reduced by software scaling and limiting of the sensor data.

The sensor configuration of the system is based on a two-dimensional (x y) grid, as shown in Figure 1, allowing considerable flexibility in the mapping of movement within the sensing area. Whilst the system is designed around this, any number of sensors can be placed in any configuration, and within any proximity to one another. One application has, for example, used four sensors placed on the floor in a semicircle around a single performer, the sensors aimed toward the ceiling.

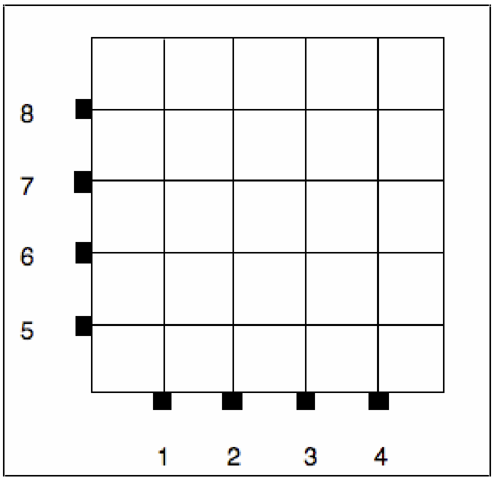

Figure 1 - PLaY+SPaCE Sensor Configuration

Interface

In the current version of the interface the eight sensors are controlled to emit sequentially, resulting in the sensing space being scanned once every 512ms. This overall scan time is problematic and represents a limitation of the system: depending on when, and for how long, a user is in the path of a sensor beam, physical gestures can be missed, or responses to physical gestures can appear to have an unacceptable latency of up to 512ms. Current research is being carried out to reduce latency through investigation of a spread spectrum technique wherein sensors transmit and receive individual signals, excluding recognition of any other sensor signals. Theoretically, such an approach may reduce the scanning time of the sensing space from 512ms to 64ms.

In addition to controlling the sequential sensor emissions, the interface microcontroller is programmed to convert incoming sensor data to MIDI data. MIDI data is formatted as Note-On messages in which the normal pitch/velocity pair is used to output sensor number (1 to 8) and proximity value (1 – 64) respectively. A USB MIDI interface is incorporated into the PLaY+SPaCE interface for output.

The current version of the interface does not allow for MIDI input to the interface, however work is currently being carried out to allow MIDI input to control the number of sensors in use, sensor output resolution, and calibration of the sensors to suit the size of the sensing space.

Software and Mapping

Generic Patches

The PLaY+SPaCE software component was developed in MAX/MSP, the input area of the software encapsulated within a main interface shown in Figure 2.

Figure 2 - PLaY+SPaCE Main Interface

The interface consists of three generic patches; a Control patch, primarily for selecting the MIDI input device, a Calibration patch allowing scaling and limiting of incoming data from the sensing space, and the 'Noids' patch that provides a virtual sensing space, allowing the simulation of movement and activity of up to eight people within the space. A further component of the main interface is a monitoring window, a representation of the sensing space, as shown in Figure 3. The grid lines represent the beams of the eight sensors, the dots here representing virtual activity as generated by the Noids patch.

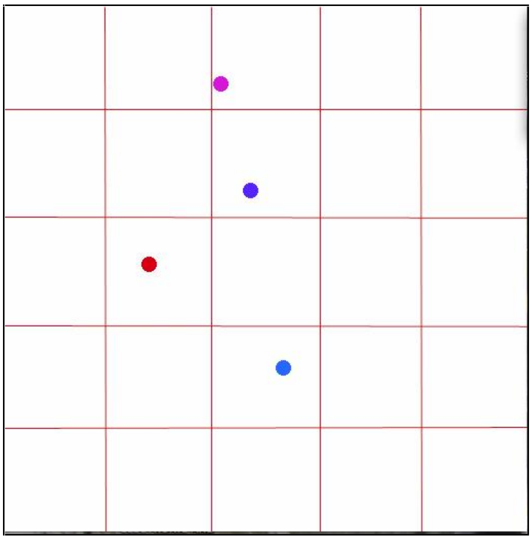

Figure 3 PlaY+SPaCE Monitor Window

The software provides two basic modes of operation labelled as Triggering Mode and Activity Mode. In the former, data from the sensing space is scaled and limited in the Calibration patch before being sent directly from the main interface to trigger musical events in subsequent proprietary MAX patches. In the latter, data from the space is summed every two seconds to measure the level of activity in the space on a scale of 0 to 10. As in the Triggering Mode, this data is sent on to subsequent patches to control musical output. Activity Mode is selected and monitored in a generic patch selected from a menu in the main user interface. Further generic patches available from main user interface menus include instruments (a basic software sampler and loop player), a simple audio mixer and an elementary editor patch that enables the programming of MIDI data to be triggered by input from the sensing space.

Proprietary Patches

The elementary nature of the generic editor patch precludes its use in many situations where complex mappings of sensing space data are required for interactive music performance. Generally, such mappings are developed within proprietary MAX patches suited to the requirements of specific musical environments. Proprietary patches are used in combination with the generic patches of the main user interface, a range of proprietary patches having been developed as exemplary works for the system.

An elementary example is a patch for a work titled *BeLLS+PLuCKS*, this work primarily used in workshop situations as a very simple introduction to the system. The sensing space is divided into an inner and an outer zone, as shown in Figure 4, trigger points (shown for sensors 1 and 5 only) in the outer zone assigned to simple General MIDI bell sounds (eg. tubular bells), and triggers in the inner zone assigned to plucked string sounds (guitar, banjo etc.). Pitch materials are diatonic, with the outer zone assigned the tonic chord and the inner zone assigned the dominant. As users move in a clockwise direction around the space, ascending scales and arpeggios are heard, these descending with anti-clockwise movement.

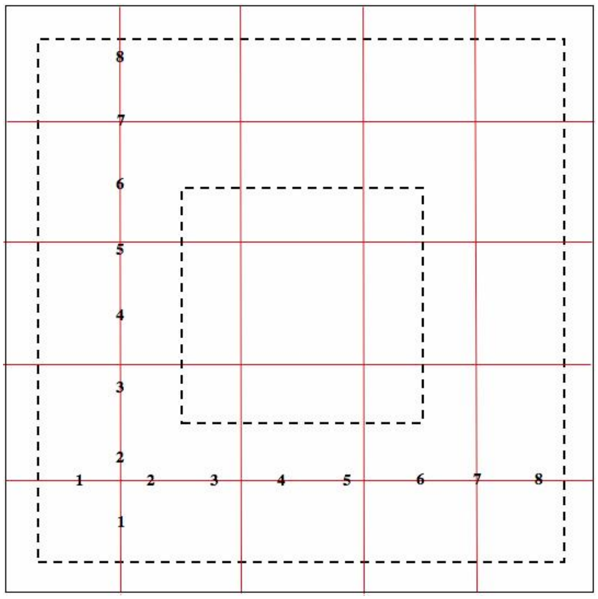

Figure 4 – BeLLS+PLuCKS Sensing Space Zones

The proprietary MAX patch for *BeLLS+PLuCKS* is shown in Figure 5. On opening the patch the 'sdivall' object sends a command to the generic Calibration patch, to divide the incoming sensor input from each sensor into eight subdivisions, as seen in Figure 4. This results in the sensing space having 64 possible triggering points. The 'r S*n*' objects receive incoming scaled sensor data from the main user interface, the eight trigger points mapped to MIDI note numbers within the 'coll' objects, such as in the list below in which a C major arpeggio is mapped and assigned to Sensor 1.

```
1, 48;
2, 52;
3, 55;
4, 60;
5, 64;
6, 67;
7, 72;
8, 76;
```

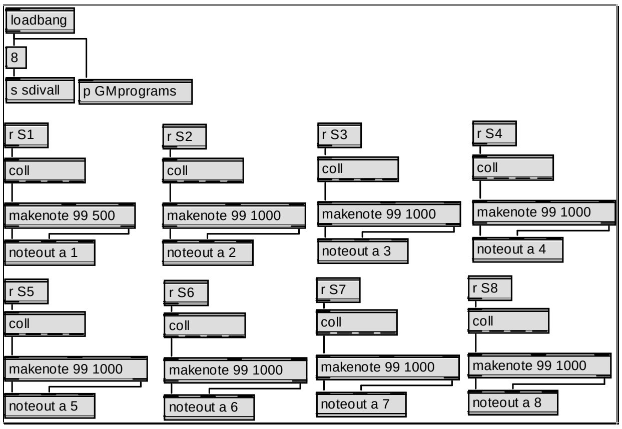

Figure 5 – BeLLS+PLuCKS Proprietary Patch

The *BeLLS+PLuCKS* patch example represents an extremely simple mapping of sensing space data, each trigger point in the space simply assigned a MIDI note number, with no provision for rhythmic control, timbre, tonality, or form, and no utilisation of data regarding the number of users or activity levels within the sensing space. Considerably more complex mappings are used in further exemplary patches and in patches designed for specific applications of the system.

Output

Output from proprietary patches is used to trigger MIDI and audio events, control synthesis parameters, lighting and video. MIDI events may be sent to the system's generic sampler or loop player patches, with subsequent audio output to the generic mixer patch. Alternatively, VST instruments can be incorporated into patches to receive MIDI data, or the data sent to external hardware devices including MIDI lighting desks. Synthesis subpatches can be incorporated into proprietary patches, data from the sensing space controlling the direct triggering of generated audio events and synthesis parameters.

Most commonly in proprietary patches, a VST instrument, for example a soft sampler, is utilised for audio output, and a range of signal processing VST instruments further incorporated for audio processing. In complex works, where a range of proprietary patches are developed for separate sections or scenes of a work, an independent output patch is devised, one that can be used by a succession of proprietary patches. Figure 6 illustrates, a four-channel output used and a VST sampler (Halion2) receiving MIDI data from proprietary patches. Output from the sampler may here be direct to a quadraphonic panning system, or via either a VST granular synthesis plug-in or a simple delay effect. Settings for, and selection of, the audio processors are received from the succession of proprietary patches. Spatialisation of the audio output is controlled in this example by sensor data output from the main user interface, i.e. by the locations of people within the sensing space. Triggering points within the space are relayed to the Output patch to pan the audio to four speakers, the four large square objects representing the quadraphonic space for four channels, and the inner squares representing the location of the audio within the quadraphonic space. In Figure 6, all channels are centred. Output from all channels is also sent to a filtered (Low Frequency) fifth channel, essentially providing a 4.1 surround sound output.

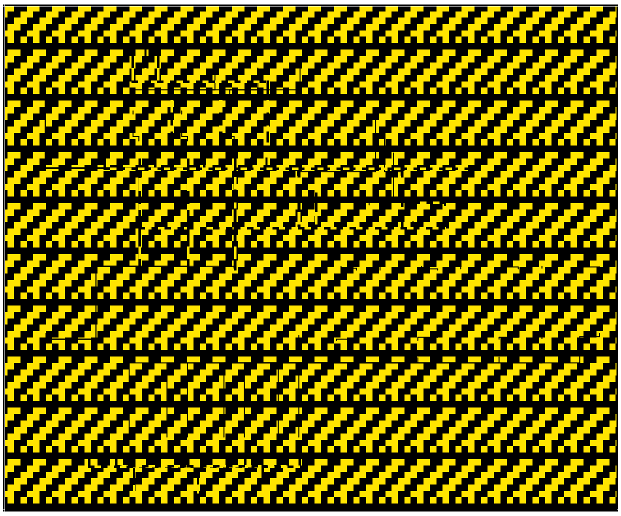

Figure 6 – PlaY+SPaCE Independent Output Patch

Video output patches have also been developed to trigger and control video parameters, primarily within Jitter. As in the above audio output patch, video patches have been devised in which settings are relayed from proprietary patches to an independent video patch. Again, triggering points within the sensing space are used to control image sequences and parameters of video effects.

Applications

To date, PlaY+SPaCE has been used in a range of applications, two works, entitled *FloW* and *RiVeRSCaPE* created for installations. *FloW*, for a 2002 'River Festival', utilised the system's Activity

Mode to trigger water samples, along with sampled vibraphone, marimba and untuned percussion. As user activity increased in the space, density of texture and rhythm increased, moving from 'a trickle through to a torrent'. *RiVeRSCaPE*, for a community arts installation, was designed to heighten visitor experiences of art works, developed as part of a community project, through the triggering of recorded readings of texts and a range of samples derived from the community environment. Problems encountered were primarily related to obstructions to the sensor beams within the space, with numerous artworks incorporated into a limited area. Careful planning of, and alterations to, the layout of the space was required to suit the use of PlaY+SPaCE system.

Applications in workshops for people with disabilities were initially explored in 2003. The workshops were made open to people with a broad range of disabilities, many participants gaining valuable and enjoyable experiences. Exemplary proprietary patches (such as *BeLLS+PLuCKS*) were utilised within the sessions. High functioning clients gained most from the sessions, with an ability to grasp the movement to sound (action/reaction) basis of the system. Those unable to grasp this gained little from the sessions, indicative of the need to limit sessions to higher functioning clients, and a need to further research possibilities of using the system with such a broad range of clients.

Applications in dance were explored in a work for Townsville-based company Dance North in 2004. The production was comprised of six scenes, the dancers having full control over the triggering and density of sounds, the form, and also the spatialisation of the audio output of the work. The choreography of the work ranged from being very strict through to allowing the dancers considerable freedom. Accordingly a range of mapping strategies were utilised in the work's proprietary patches, with considerable success achieved in sections where dancers were most free to explore the space.

In musical performance, various works have been developed, most recently a collaborative work entitled *Neonderthals*, in which two laptop performers control textures, video and text samples, and one performer uses PlaY+SPaCE to control real-time granular synthesis parameters with percussive samples. A further work entitled *SaX4MaX* involves a saxophonist and PlaY+SPaCE, the performer triggering a range of pre-recorded saxophone samples and controlling a range of effects processors and granular synthesis parameters through movement and location within the space.

Aesthetic Considerations

As a proprietary system PlaY+SPaCE is unfamiliar, its non-tactile nature leading to the possibility that physical gestures made in the space (action) will not be directly recognised as resulting in output (reaction). As mentioned, with some clients in disabilities workshops, the inability to grasp the action/reaction concept will occur regardless of the simplicity of the application mappings, and further research is needed to address the needs of such clients. Where an action/reaction concept is not clearly grasped by the user (and by extension an audience), the ability to control and manipulate musical output is lost, and hence the purpose of the system is defeated. Consequently, a basic aesthetic premise in developing works for the system is adopted, an aesthetic in which gesture (action) must result in musically cohesive output (reaction), the level of complexity of reaction dependent on the level of understanding of the system and its output by the intended user.

As an elementary example, a work entitled *ANiMaL FaRM* is addressed to c. 3-5 year old children, a hybrid of the system's Triggering and Activity Modes utilised. Here the sensing space is divided into four zones, each triggering different farmyard animal sounds. Simple background music supports the animal sounds, changing in harmony and volume as activity in the sensing space increases. The musical output here is secondary, the primary aesthetic emphasis being to empower children with the ability to "find" the different animal sounds available within the space.

With adult users, the complexity of reaction may be increased. The work entitled *FLoW*, discussed previously, takes as a basis the use of the system by the general public. The compositional premise of the work is to provide a simple reaction to activity level within the space, increases in activity relative to increases in music texture, rhythm and timbre density. Here the users are able, through their different activity levels, to slow and accelerate the programmatic flow of an imagined river.

In music performance the aesthetic potential of PlaY+SPaCE is manifold. In addition to the system being used as a standalone DMI, it may, for example, be utilised to enhance an acoustic performance through the provision of an accompaniment (or independent part), controlled by the performer's own movements within the sensing space, or by a second performer; another musician, a dancer or actor for example.

Formation of such an accompaniment is relative to current technological possibilities: simple triggering of pre-recorded instrumental or electronic samples and loops, real-time generation of synthesised sounds with control over synthesis parameters, and control over signal processing parameters as applied to a live audio input. In each case, the system can empower the user/s with control over a wide range of output, the non-tactile nature of the system providing access to such outputs without the need for any extensive knowledge of signal processing or synthesis parameters. In these applications of the system, the aesthetic premise concerning action/reaction is extended, the mapping of signal processing and synthesis parameters needing to be at a level of complexity that is dependent on both the knowledge and aesthetic requirements of the user.

Directions

Research with the PlaY+SPaCE DMI has primarily focused on stand-alone applications of the system; in efforts to gain reliable data from the system's hardware, in order to explore a range of mapping strategies of the sensing space within the software, and in order to fully explore fundamental applications of the system. As a non-tactile system, PlaY+SPaCE is generally utilised in a stand-alone mode, but may also be augmented with video sensing systems, and a range of tactile sensors, e.g. pressure, light, and temperature sensors, all representing avenues of further exploration.

Current research is concerned with the stand-alone system, simultaneously addressing issues regarding the speed and resolution of the hardware, greater software-based user control of the hardware, and further development of the system's generic software patches. The applications of the system briefly discussed above represent a point of departure for ongoing research, work that includes the exploration of more diverse applications, and the development of further mapping strategies to suit those applications.

References

Campbell. S. 2003. "Bats, Max, Boids and Music." *Proceedings of the 2003 Apple University Consortium Conference*. Available from auc.uow.edu.au/conf/conf03/papers/AUC_DV2003_Campbell.pdf [Accessed 18th March, 2005].

Camurri, A. 1995. "Interactive Dance Music Systems." *Proceedings of the 1995 International Computer Music Conference*. San Francisco: International Computer Music Association: 245–252.

Carter, C. 2001 "Music and Movement" *Sound on Sound*, October, 2001: 144-148.

Gehlhaar, R. 1991. "SOUND=SPACE: An Interactive Musical Environment". *Contemporary Music Review 6*, No.1:59-72.

Gehlhaar, R. 1998. "SOUND=SPACE Workshops for Disabled Children". Available from http://easyweb.easynet.co.uk/nourrolf/resfelr.html [Accessed 18th March 2005].

Mulder A. 2000. "Towards a choice of gestural constraints for instrumental performers". *Trends in Gestural Control of Music*, Marcelo Wanderley and Marc Battier (eds.), Paris, France: IRCAM.

Paradiso. J. and Gershenfeld, N. 1996. "Musical Applications of Electrical Field Sensing." *Computer Music Journal,* 21(2):69-89.

Polaroid OEM Components Group 1999. "Technical Specifications for 6500 Series Sonar Ranging Module". Available from: www-inst.eecs.berkeley.edu/~ee40/ calbot/sensor_datasheets/6500series.pdf [Accessed 18th March 2005].

Reynolds, M., Schoner, B., Richards, J., Dobson, K., Gershenfeld, N. 2001. "An Immersive, Multi-User, Musical Stage Environment." *SIGGRAPH Proceedings,* New York: ACM Press.

Waisvisz, M. 1985. "The Hands, a set of Remote MIDI-controllers". *Proceedings of the International Computer Music Conference* (ICMC), Burnabey, BC, Canada.

Wanderley, M. 2000. "Gestural Control of Music". *International Workshop Human Supervision and Control in Engineering and Music.* Kassel, Germany: 101-130.