SIGGRAPH 2000 Emerging Technologies Proposal

Augmented Groove: Collaborative Jamming in Augmented Reality

I. Poupyrev, R. Berry, J. Kurumisawa
Media Integrations and Communication Research Laboratories, ATR International, Japan

M. Billinghurst, C. Airola
Human Interface Technology Laboratory University of Washington, USA

H. Kato
Hiroshima City University, Japan

Executive Overview

Title: *Augmented Groove: Collaborative Jamming in Augmented Reality*

Abstract:

The Augmented Groove is a musical interface that explores use of augmented reality, three-dimensional (3D) interfaces and physical, tangible interaction for conducting multimedia musical performance. In the Augmented Groove visitors can play music together, with or without use of traditional musical instruments, simply by picking and manipulating physical cards on a table. The physical motions of the cards are mapped to changes in musical elements such as timbre, pitch, rhythm, reverb and others. At the same time, users wearing a lightweight head-mounted display (HMD) can see 3D virtual images attached to the cards whose shapes, color and dynamics reflect aspects of the music controlled by the visitors. Because in augmented reality environments visitors can see physical world, virtual objects and each other, several visitors can easily join around the mixing table and "jam" together, passing musical sequences to each other in the same manner in which we would pass everyday objects. The music in a sense becomes a physical, tangible object, something we can touch and see as part of our physical environment.

Musical control in the Augmented Groove is deliberately imprecise: the performer manipulates short musical sequences, or phrases, rather then individual notes. The resulting musical performance is a "computer-supported improvisation" in which the user arranges, mixes and modifies pre-composed musical phrases on a structural level.

Team Information:

Ivan Poupyrev – Interface Design, Software Development

Dr. Poupyrev is a Researcher in Media Integration and Communication (MIC) Research Laboratories at ATR where he innovates and investigates novel human-computer interfaces. His basic research interests lie at the intersection of interactive computer graphics and interaction, including 3D and multimodal user inter-

faces, virtual and augmented reality, and ubiquitous computing environments. He stayed at the Human Interface Technology Laboratory, University of Washington from 1995 to 1998 as a Visiting Scientist and at the ATR MIC Labs from 1998 to 1999 as a Research Intern. Results of his work have been reported in conferences such as ACM SIGGRAPH, CHI, UIST and EUROGRAPHICS. He co-organized and taught a tutorial on 3D user interface design at IEEE VR99 and VR 2000, also to be presented as a course at SIGGRAPH 2000. His other past projects have included the Smithsonisaur VR dinosaur exhibit shown at the 1997 IEEE Multimedia Systems Conference, in Hiroshima, Japan, and Shared Space Emerging Technologies demo shown at SIGGRAPPH 1999 and Imagina 2000.

Rodney Berry, Music Composition, Sound Engineering

Rodney Berry is an artist and researcher at ATR MIC Labs. Best known as a composer and audio artist, he received his Master of Fine Art degree at University of New South Wales College of Fine Arts in Sydney Australia. Most recently, he has explored ways of using artificial computer generated life-forms to make music in virtual worlds. His artificial life music environment, 'Feeping Creatures' (produced in collaboration with T. Mander, B. Murray and B. Ross of Proximity Ltd. Sydney) was shown in Australian Perspecta 1997, Virtual Worlds 1998 in Paris, ALIFE6 in 1998 in Los Angeles and ECAL99 in Switzerland. His past work has included interactive musical and sculptural installations in a variety of media and was featured in the Third International Symposium on Electronic Art in Sydney 1992, Sound Culture in Sydney 1991 and Tokyo 1993 and Experimenta in Melbourne 1996. He taught Music Composition at the University of Western, Sydney and Digital Audio at KvB Institute of Technology, Sydney in 1997-1999.

Jun Kurumisawa, Computer Graphics Art, 3D Modeling

Jun Kurumisawa is an artist working both in electronic and traditional media in ATR MIC Research Labs. After receiving a B.A. in Western Paintings and Print Art from Tokyo National University of Fine Arts and Music he has been involved in computer graphics design and production in major Japanese television networks and printed magazines. He joined ATR MIC Labs in 1996 where he continued his work in computer generated paintings and interactive electronic art: his prints and interactive art installations were displayed at SIGGRAPH 1998 and 1999 in Art Gallery Exhibitions. He also has been actively involved in research projects on interactive computer graphics and human-computer interaction including Virtual KABUKI presented at SIGGRAPH 1996, 3D Scene Creation by Estimating Hand Postures from Multiple Cameras exhibited at SIGGRAPH '97 and Shared Space presented at SIGGRAPH 1999. Mr. Kurumisawa has been an invited speaker at several conferences and exhibitions on computer graphics art, such as Digital Image Exhibition, Tokyo in 1998 and 1999.

Mark Billinghurst – Interface Design, Software Development

Mr. Billinghurst has a strong background in the development of innovative human-computer interfaces. He is a 5th Year Electrical Engineering PhD Student, University of Washington with over five years experience in virtual reality, augmented reality, multimodal interfaces, collaborative virtual environments and intelligent agents. He has spent time as a visiting scientist at the MIT Media Lab

and the Advanced Perception Unit at British Telecom Labs. Past projects have included the Smithsonisaur virtual dinosaur exhibit shown at the 1997 IEEE Multimedia Systems Conference in Hiroshima, Japan, and Shared Space Emerging Technologies demo shown at SIGGRAPPH 1999, and Imagina 2000.

Chris Airola – Sound and Hardware Engineering

Chris Airola is a Research Engineer at the HIT Lab, University of Washington working within the areas of audio interaction and manipulation, hardware design and simulation. His current research interests are in developing new modes of displaying spatial information via audio. He received his BS in Biomedical Engineering from Marquette University and has completed the computer music series classes at the University of Washington. He is has been working at the HIT Lab, University of Washington on hardware design, designing and developing audio interaction techniques and real time audio rendering software. He is also collaborating with a video artists from Cornish College of the Arts designing a real time video tracking and audio rendering interactive art installations.

Hirokazu Kato – Computer Vision Development, Image Processing

Dr. Kato is working in the area of computer vision, pattern recognition and human computer interaction. He received his D.E. from Osaka University in 1996. He is an assistant professor at the Graduate School of Engineering Science, Osaka University. In 1998, he was a visiting scholar at the Human Interface Technology Lab, University of Washington and worked on wearable interface and augmented reality research projects. His interest is in interface design of computer mediated communication systems.

Level of Innovation

With every leap of technology musicians and engineers have created new ways to produce and play music. The work on electronic music started with the invention of the Musical Telegraph in 1876, which can be considered the first true electronic instrument [9] This started a long engineering quest culminating in today's sophisticated samplers and synthesizers. The interface for making music, however, has not changed much since those days – most of them are still using keyboards, often with familiar layout from the acoustic pianos.

Elisha Gray's Musical Telegraph of 1876

Innovating Musical interfaces

Propelled by the development of MIDI standards in early eighties, as well as emergence of new and increasingly inexpensive interface technologies there has recently been significant growth of innovations in musical interfaces. While some of the new musical interfaces borrow from traditional ways to control music, such as controlling music rhythm by tracking conductor's baton waved by the user, (e.g. [10]), others completely depart from traditional tried and true approaches searching for new methods of artistic expression. In the Dance Space, for example, a video camera captures the dancer's movements,

mapping them so that dancer's head, hands and feet control various musical instruments [12]. MIT Media Lab's Brain Opera and Sensor Chair [9] use the interface based on the electric field sensing techniques, where the musicians as well as novice user can produce and control music simply by moving their hands in space. In Brush de Samba [8] a performer controls the music by drawing a picture on the drawing pad, the music is generated basing on the pen data. Virtual Reality as musical interface has been investigated, for example by Gerard Kim and others [7]: in Musical Motion system, the immersed user controls the individual notes by tapping on an "air piano" with the hand instrumented by the data glove while the other hand controls the general tempo through squeezing gestures. There were also many attempts to combine computer-generated visualization of music with interactive musical performance, e.g. MusiKalscope and Iamascope by Sidney Fels and others [5, 6].

Certainly, with most of these interfaces it is impossible to play music with the level of precision allowed by traditional instruments. They, however, are not designed to be as precise. The focus here is on "computer-supported improvisation" [5], providing freedom of creative artistic expression regardless of the user's musical skills and experiences. By interpreting user motions captured using variety of interface technologies and applying carefully designed sonic mappings, these novel musical interfaces allow for complicated and very enjoyable musical performances [9]. Emergence of new interface technologies invariably opens new, fascinating ways to make and play music.

Augmented Groove: An Augmented Reality Musical Interface

The Augmented Groove is a musical interface that employs augmented reality and a three-dimensional tangible interface metaphor to allow several people to jam, conducting multimedia musical performances together. It differs from the existing musical interfaces in several important aspects:

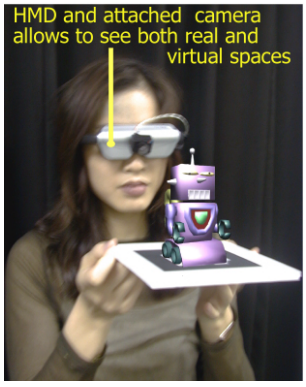
- *Augmented Groove is an augmented reality musical interface*. Augmented reality (AR), i.e. overlaying of 3D virtual objects on the real world, allows us to integrate computer-generated imagery into surrounding physical reality [4]. Unlike VR environments that replace the physical world, in AR environments virtual objects are mixed with physical objects to become in a sense, an equal part of the world. In the Augmented Groove, performers interact and control music by manipulating musical sequences that are represented visually as 3D virtual objects. Changes in music result in visual changes of the appearance of virtual objects, i.e. their color, shape and animation. Furthermore, in the AR environment, the performer, while interacting with virtual musical objects, has also full access to the conventional musical equipment, such as keyboards and synthesizers, and at the same time can easily collaborate with other musicians. Technological details are described in the next section.
- *Augmented Groove is a physical, tangible musical interface*. In the Augmented Groove the performer plays music by physically manipulating simple physical cards with fiducial markers attached to them. Individual musical sequences and corresponding visual representations are assigned to each card. As the user manipulates cards, the computer vision tracking software maps card movements into changes in corresponding musical parameters, such as pitch, distortion, amplitude, filter cut-off frequency, filter resonance, delay mix and others. Since the system can track and

recognize several cards at the same time, performers can manipulate several musical elements simultaneously. The physical, tangible nature of the Augmented Groove makes it easy and intuitive to use: all the user needs to do is to pick up and move the cards.

- *Augmented Groove is an interface for mixing electronic music*. Most of the novel musical interfaces usually allow either global control of global musical parameters, such as tempo of the whole composition, which makes them unexpressive, or control of individual note progression, which makes them difficult to use. The Augmented Groove takes quite different approach to music composition and performance, which more closely resembles the performance of a DJ rather than a conventional musician. In Augmented Groove, the musical performance is composed from a collection of short looped musical sequences, or phrases. Several performers can interactively add, mix and deform these musical sequences by manipulating physical cards. Each musical phrase is carefully composed to fit others, furthermore cards movements would affect different musical phrases differently: a composer writing music for the Augmented Groove chooses sonic deformations and distortions individually for each musical phrase to ensure a high quality of musical performance. The Augmented Groove, therefore, is an interface that allows the performance of quite complex modern electronic music, such as techno, and yet retains a significant level of user control.
- *Augmented Groove is a collaborative musical environment*. Making and playing music has been an ultimate collaborative activity for thousands of years. Yet, the collaborative aspect is often missed in musical interface work. The Augmented Groove is designed to allow several people to join and jam together, combining their efforts in producing enjoyable musical performance.

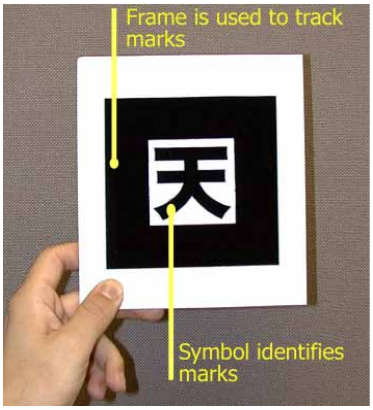
Augmented Reality Technology

This section describes AR technology used in the Augmented Groove. The performer wears a pair of Olympus Eye-Trek FMD-100 head mounted display (HMD) with a small color CCD camera attached. The Olympus HMD is lightweight, high-resolution, full color display that is worn just like a pair of normal glasses. The camera output is connected to an SGI O2 workstation, where it is processed using computer vision tracking and registration techniques [2], the 3D virtual images are rendered in each video frame and, finally, sent back to the HMD. Hence, when the user wears the glasses he or she can see both the real world and 3D virtual objects at the interactive frame rate (15-20 frames/second).



A virtual object registered on the card

The AR environment includes a number of marked physical cards with square tracking patterns and identifying symbols in the middle of the patterns. The position and orientation of 3D virtual objects are defined using these markers: When the user looks at the card, computer vision techniques identify specific markers, calculate head position and orientation relative to the marks, and display 3D virtual im-



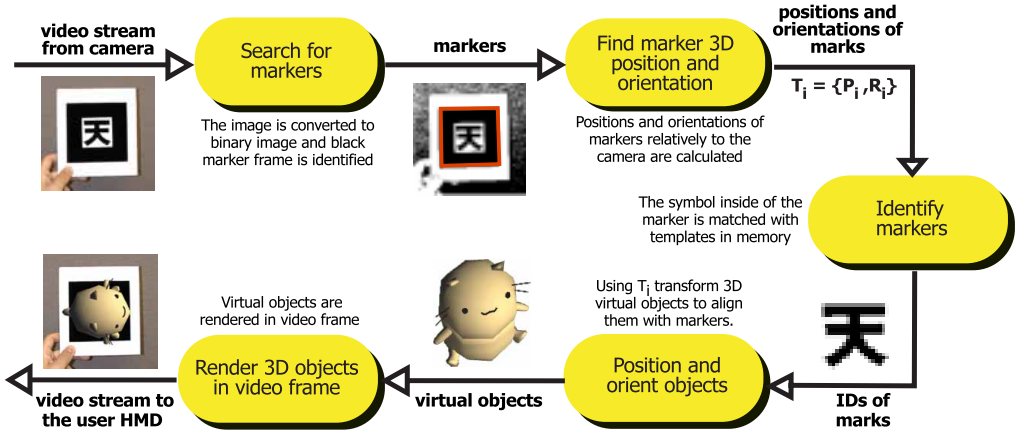
A card with marker used for tracking

ages so that they appear precisely registered with the physical objects. The camera worn by the user defines, however, only local tracking information, relative to the user's viewpoint: if, for example, the user does not look at the card the system could not know the position of the card in space. Hence, in Augmented Groove we define an absolute tracking space by using a fixed overhead camera to resolve ambiguities and improve quality of tracking.

In Augmented Groove each card has a musical sequence assigned to it, changes in the card's position and orientation are mapped into MIDI commands that are sent to a professional sampler and synthesizer through cascaded MIDI merging boxes. The MIDI commands control which musical sequences are played by the samplers and synthesizer at any given time as well as how they are played, e.g. which filters are applied.

Since all users share the same database of cards, sounds and virtual objects, they all see the same virtual objects attached to the markers and therefore can easily collaborate in producing musical performance. The user can pick up a card, show it to the other participants, pass or request musical sequences in the same manner that we request physical objects in normal face-to-face communication. The collaborating users can use a full range of verbal and non-verbal cues during the interaction, which greatly enhances communication between participants. The details of the tracking techniques are briefly outlined in the figure below, for more details see [2].

The core of AR technology used in the Augmented Groove has been developed in the Shared Space project that we demonstrated at SIGGRAPH 1999 [3]. In the Augmented Groove we expand and improve the AR technology developed for Shared Space and investigate it in a new different application: from interacting with visual information to interacting with music.



The AR tracking and registration technique

Interactive Experience

The Augmented Groove environment includes a large round table with a rack of LP records, i.e. round black paper disks, in the middle. Each disk has a unique pattern in the place of a record label that is used for tracking. When visitors approach the table they will be given one of the two video see-through head-mounted displays. When the user picks a record from the table the corresponding sound piece will come up from the four speakers located above the table.

An artist's sketch of the
Augmented Groove interface

The visitors can control music by manipulating physical cards and for each musical phrase a unique highly stylized animations will be displayed on top of the record, which animation reflect changes of the music. The virtual objects and their animations will be highly stylized and entertaining: for example a large stylized CG lips will represent vocal, drums will represent percussion etc. A number of musical compositions, mostly modern electronic music, will be composed for Augmented Groove by project members and collaborating experimental musicians in Japan. Each composition will consist of a number musical sequences attached to the different records, so if the user opens all cards a complete composition will be played.

An artist's sketch of a 3D model
and animation for Augmented Groove

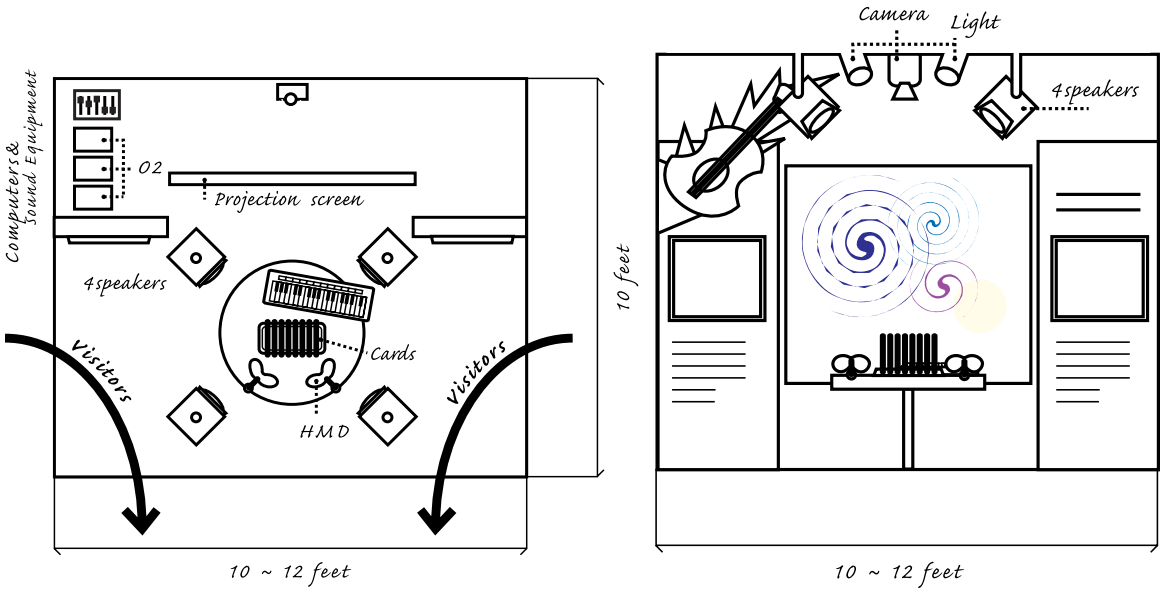
A member of the project will be facilitating the demo at all times guiding visitors through the experience and encouraging them to play music together. The table will also be equipped with additional MIDI keyboard that can be used by facilitator to join participants in musical performance, which be used by one of the visitors.

The experience of the users will be projected on the large screen behind the users as well as two monitors on both sides of the booth to involve spectators into the demonstration. The projector on the back wall will project video mixed with computer graphics to create interesting and entertaining video show. It is anticipated that two users at a time will be able to experience the Augmented Groove and they will spend no more than 5 minutes in the head mounted displays. This should enable a throughput of 30-40 people per hour through the demonstration, not including spectators.

Appearance

Figures below show a proposed booth layout for the Augmented Groove demonstration.

The table with a record rack and HMDs in the middle will dominate the space in the center of the booth. To either side and behind the table will be decorative props, which recreate an atmosphere of musical studio with wires, billboards, photographs and musical instruments, hanging from the wall. Stage lights above the exhibition will add to the atmosphere of the performance as well as control adequate lighting for exhibition. Above the table 4 speakers will provide sound and a projector will hang above the table and display on the back wall. Two large flat screen panels will be hanged on two sides of the booth with view from both participants, allowing the spectators to be part of the exhibition.



Top and front views sketches of the proposed booth layout (on the left and right respectively)

History

The Augmented Groove is a musical interface and interactive installation developed in collaboration between MIC Labs International and HIT Lab, University of Washington. It has been developed under the auspices of the Shared Space project, designed to explore various aspects of collaborative augmented reality and was initiated by Mark Billinghurst at the HITL, University of Washington in 1996. The arrival of Dr. Kato in 1998 lead the way for the development of the very accurate vision based registration techniques demonstrated in the current Augmented Groove application. In 1999 the Shared Space was joined by Ivan Poupyrev, from ATR MIC Labs, who also briefly worked in the Shared Space project in 1996 before starting his Ph.D. on 3D user interfaces. Since then a number successful demonstration systems and interactive installations have been developed and demonstrated to the public. This includes most recently a collaborative AR game shown at SIGGRAPH 99 (http://www.hitl.washington.edu/share) and Imagina 2000

(http://www.imagina.fr). Rodney Berry, a music artist from ATR, joined the project in late 1999 in exploring yet another, musical application of the Shared Space augmented reality technology.

Impact

Network enabled wearable computers will be used in the future as a communications device to enable people to collaborate anywhere and at any time, [11]. The development of augmented reality devices will also result in convenient, lightweight and inexpensive information displays that provide information anywhere and anytime. Furthermore, the environment in which we are living will become increasingly "smart"; equipped with sensors and wireless networks allowing a wearable computing system to understand and interpret our actions and our relations to our surrounding environments.

In this setting we will be able to develop unique interfaces that allowing us to manipulate information, music, sounds and images, not through keyboards, dials and handlers, but through virtual, augmented reality interfaces, by manipulating physical objects, cards, and etc instead of mouse and keyboards. In case of musical production, for example, musicians will be able to move sound samples from the synthesizer or sound sampler simply by taking a card and bringing it next to the marker on the sound sampler and making a simple gesture. The card can be then faxed or e-mailed to another musician. The composition and mixture of sounds can be done by moving physical cards on the table and each of the sounds will be seen and visible at the same time. Resulting music can be moved back to the recording equipment in the same way – by moving the physical cards next to the digital recording facilities.

The augmented reality technology used and developed by the Augmented Groove project explores a new dimension of collaborative digital music creation and production, helping the people who make the music come together and work on it. These scenarios show how the prototype Augmented Groove interface could be used to develop compelling applications in the future. Such applications will change the way that people collaborate, work and play in the future.

Technical

Equipment

Hardware:

- 4 SGI O2 computers
- 2 Olympus or Sony head mounted displays
- 3 miniature color cameras on the glasses
- 1 camera on top of the table
- 1 projector
- 1 projection screen
- 1 Yamaha 5000 sampler
- 2 MIDI mixing boxes
- 1 sound mixer
- 1 MIDI keyboards

1 audio amplifier
4 loud speakers
4 stage projectors
5 desktop lamps
Assorted power, networking and video cables
Round table

We currently have 3 SGI O2 computers and will be looking toward commercial donations to provide the remaining systems.

Software
Custom applications

Physical

The exhibit will require a minimum of a 12 foot by 12 foot space, preferably on a corner location so bystanders can watch from two sides. A larger 15 foot by 15 foot booth would allow for a more spacious layout. Figures 4 and 5 show the proposed layout of our hardware and physical props.

Each of the O2s will require power and at least 10 mb/s ethernet connectivity. We will also require additional power outlets for the head worn cameras, desktop lamps, projector, and audio equipment.

There will be as little outside noise generated as possible and in order to have good quality computer vision based tracking, our work requires controlled lighting conditions.

Given the small amount of physical equipment, it is anticipated that our demonstration will be able to be set up and pulled down within 4-5 hours.

Network

The project will require a 10 mb/s network between the four machines used for graphics rendering. Each of our SGI O2 machines will have 128 mb of ram and will be running the IRIX 6.5 operating system. One of the O2 machines will support our server needs. The application and networking software will be custom developed software.

References

1. Billinghurst, M., Bowskill, J., Dyer, N., Morphett, J., An evaluation of wearble information spaces. *Proceedings of VRAIS'98*. 1998. IEEE. pp. 20-27.
2. Billinghurst, M., Kato, H., Collaborative Mixed Reality. *Proceedings of International Symposium on Mixed Reality (ISMR '99)*. 1999. Springer Verlag. pp. 261-284.
3. Billinghurst, M., Kato, H., May, R., Kraus, S., Otsuka, T., *et al.*, Shared Space: Collaborative Augmented Reality. *Proceedings of SIGGRAPH'99, Emerging Technologies*. 1999. ACM.

4. Feiner, S., MacIntyre, B., Seligmann, D., Knowledge-Based Augmented Reality. *Communications of the ACM*, 1993. 36(7): pp. 53-62.
5. Fels, S., Mase, K., Iamascope: a graphical musical instrument. *Computer & Graphics*, 1999. 23: pp. 277-286.
6. Fels, S., Nishimoto, K., Mase, K., MusiKalscope: A graphical Musical Instrument. *IEEE Multimedia*, 1998. 5(3): pp. 26-35.
7. Kim, G., Hwang, J., Musical Motion: A medium for uniting visualization and control of musuc in virtual environment. *Proceedings of VSMM'99*. 1999.
8. Kotani, A., Maes, P., An environment for musical collaboration between agents and users. *Proceedings of Lifelike Computer Characters*. 1995. pp. 54.
9. Paradiso, J., New Ways to Play, Electronic Music Interfaces. *IEEE Spectrum*, 1997(12): pp. 18-30.
10. Segen, J., Kumar, S., Zhang, X., Gluckman, J., Cenek, E., *et al.*, Visual conductor. *Proceedings of SIGGRAPH'99 Emerging Technologies*. 1999. ACM. pp. 180.
11. Siegal, J., Kraut, R., John, B., Carley, K., An Empirical Study of Collaborative Wearable Computer Systems. *Proceedings of CHI95*. 1995. ACM. pp. 312-313.
12. Sparacino, F., Choreographing media for interactive virtual environments, in *Medial Lab*. 1996, MIT.