

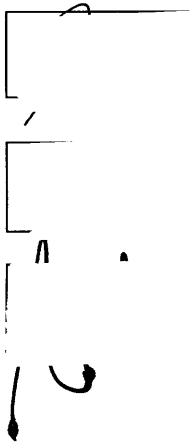
# Cinema Fabriqué : A Gestural Environment for Realtime Video Performance

Justin Manor

S.B. Physics  
Massachusetts Institute of Technology  
June 2000

Submitted to the Program in Media Arts and Sciences,  
School of Architecture and Planning,  
in partial fulfillment of the requirements for the degree of  
Master of Science in Media Arts and Sciences at the  
Massachusetts Institute of Technology  
June 2003

© Massachusetts Institute of Technology  
All Rights Reserved

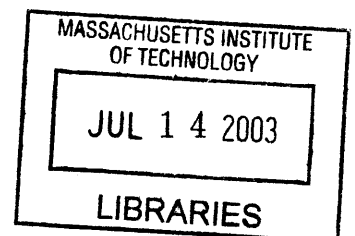


Author : Justin Manor  
Program in Media Arts and Sciences  
May 19, 2003

Certified by : John Maeda  
Associate Professor of Design and Computation  
Thesis Supervisor

Accepted by : Andrew B Lippman  
~~Chairperson~~  
Departmental Committee on Graduate Students

ROTCH





Room 14-0551  
77 Massachusetts Avenue  
Cambridge, MA 02139  
Ph: 617.253.5668 Fax: 617.253.1690  
Email: docs@mit.edu  
<http://libraries.mit.edu/docs>

## **DISCLAIMER OF QUALITY**

Due to the condition of the original material, there are unavoidable flaws in this reproduction. We have made every effort possible to provide you with the best copy available. If you are dissatisfied with this product and find it unusable, please contact Document Services as soon as possible.

Thank you.

Pages 39-42 appear to be missing from the original or there is a miss numbering error by the author.



# **Cinema Fabriqué : A Gestural Environment for Realtime Video Performance**

Justin Manor

S.B. Physics  
Massachusetts Institute of Technology  
June 2000

Submitted to the Program in Media Arts and Sciences  
School of Architecture and Planning  
in partial fulfillment of the requirements for the degree of  
Master of Science in Media Arts and Sciences at the  
Massachusetts Institute of Technology  
June 2003

© Massachusetts Institute of Technology  
All Rights Reserved

## **Abstract**

This thesis presents an environment that enables a single person to improvise video and audio programming in real time through gesture control. The goal of this system is to provide the means to compose and edit video stories for a live audience with an interface that is exposed and engaging to watch. Many of the software packages used today for realtime audio-visual performance were not built with this use in mind, and have been repurposed or modified with plug-ins to meet the performer's needs. Also, these applications are typically controlled by standard keyboard, mouse, or MIDI inputs, which were not designed for precise video control or live spectacle. As an alternative I built a system called Cinema Fabriqué which integrates video editing and effects software and hand gesture tracking methods into a single system for audio-visual performance.

Thesis Supervisor: John Maeda  
Title: Associate Professor of Design and Computation  
This work was supported by the Things That Think Consortium.





# Cinema Fabriqué : A Gestural Environment for Realtime Video Performance

Justin Manor

*Thesis Reader*

Chris Csikszentmihályi  
Assistant Professor of Media Arts and Sciences  
MIT Media Arts and Sciences



# **Cinema Fabriqué : A Gestural Environment for Realtime Video Performance**

Justin Manor

*Thesis Reader*

---

David Small, Ph. D  
Founder, Principal  
Small Design Firm



## **Acknowledgements**

I would like to thank Bernie and Louise, Ralph boy, Ben Fry, Megan Galbraith, Simon Greenwold, Tom White, James Seo, Nikita Pashenkoy, Afsheen Rais-Rohani, and John Maeda, all my unborn children, and every person who I have ever come into contact with.



## Table of Contents

1	Introduction	13
2	Background	19
3	Early Experiments	43
4	Live Performances	57
5	Futuristic Interfaces in Cinema	83
6	Cinema Fabriqué	89
7	Evaluation	100
8	References	105





# 1 Introduction

There are numerous textbooks and university courses available for those wishing to learn the art of movie making or about the history of film. Established conventions and techniques for enhancing a desired mood or message are well documented and the discourse is well-evolved. Figures taken from a popular text on film production illustrating effective methods for cinematic delivery are reproduced below [Zettl90]. Unfortunately, there are no such textbooks or acknowledged standards for improvisational audiovideo performance. Perhaps the field is too young or the context too broad to expect a set of universal principles to have emerged. Indeed those using digital tools to produce realtime visual and sonic entertainment have only recently enjoyed mainstream exposure in Europe and Asia, and hardly at all in the Americas.

Yet the proliferation of audiovisual performers is imminent if not already obvious. Computers and other multi-purpose devices that can capture, store, manipulate, and transmit audiovisual data are easily portable and affordable by consumer standards. Concertgoers are beginning to expect high-quality video to accompany live music, and large projection screens are becoming standard in a great number of musical venues.

In the hopes of adding structure to the ill-defined realm of live



1.1 *The wide-angle lens distortion of the officer's hand adds intensity to his warning*



1.2 *vertical lines are more powerful, exciting than horizontal one.*

audiovisual performance, this thesis introduces a process of realtime content creation and performance tools collectively referred to as *Cinema Fabriqué*. This term translates from French as literally “film manufactured” and encompasses my creations and performances over the past two years.

Relevant historical precedents and contemporary performers and technologies will be surveyed and their relation to my work discussed. Hardware and software devices that I have created will be presented and analyzed, along with the performances they were used in. Lastly, a summary of contributions that *Cinema Fabriqué* has introduced into the field of audiovisual performance will be discussed.

### **1.1 Embedded Entertainers**

In mid March of 2003 many home television viewers tuned into CNN and saw choppy live footage of endless brush plains and destroyed vehicles sailing by at 60 miles per hour. The broadcast was transmitted from the roof of an American tank hauling through Southern Iraq via a satellite video phone in the hands of a lone reporter escorting armed troops into the heat of battle and beyond.

The fact that many media outlets presented Operation Iraqi Freedom as an entertainment spectacle is highly disturbing, but it has prepared the world audience for finely distributed and practically undetectable journalism. People on the move would likely not be startled by a person talking authoritatively into a videophone in a public place.

In this thesis I aim to examine less invasive and more artistic uses of technology to entertain live audiences. I have tried to develop systems and contexts for rich audiovisual expression that require little setup time, are easy to transport, and are adaptable to different scenarios and stylistic requirements.

## **1.2 Recorded Versus Live Entertainment**

If someone is engaged by a movie or a pre-recorded song, they are most likely not concerned with the details of its post-production process, or what kind of equipment it was edited on. By contrast, at a live musical or video performance, the artists and their instruments are on display, and witnessing the act of creation ideally gives the audience an added level of entertainment. Seeing a guitarist fluidly pluck every note of a solo or a DJ simultaneously scratch two records will reinforce that performer's status as master of the trade and confirm that their sound does not depend on post-production enhancements.

## **1.3 Importance of Audience Understanding**

Electronic musicians and video artists commonly use personal computers for controlling live music and visuals because of the systems' great processing power, compact size, and versatility. There are many different software packages used for performance, but their interfaces tend to be complex and difficult for an untrained audience to understand. For this reason, the performer's displays are often turned away from the general view and the computer becomes a visual shield between him and the audience. If the artist's actions are hidden or ambiguous, his level of control is unknown to those who are watching. The audience cannot tell how much the performer is doing, or how much of the output is prerecorded or decided by the computer. An audience that comprehends the artist's live interaction with an instrument ideally has a more complete experience and can be drawn into a performance more readily than listening to a static recording.

## 1.4 Gesture In Performance

Gesture communication is natural to humans: children communicate with gesticulations before they are able to speak, many people always move their hands while talking, and people speaking different languages are often able to communicate simple concepts via gestures. The hand motions of others are also capable of indirectly delivering information to those watching. By seeing how someone operates a musical instrument with their hands, the nature of the device and its control mechanisms can be communicated. Much research has been done in the area of hand based interfaces to introduce gesture control to computer applications. Tasks such as map navigation [Cohen99], 3D painting [Corradini02], and video game control [Freeman98] have been adapted to gestural interfaces with positive results in performance and user satisfaction. My belief is that the naturalness, flexibility, and many degrees of freedom available with gestural inputs will offer advantages over popular interfaces for audiovisual content creation.

## 1.5 Thesis Structure

The goal of the presented research is to produce new systems and contexts for realtime cinematic delivery that communicate the performer's intention and level of control. Relevant historical and contemporary technologies and performers will be analyzed to establish a context with which to evaluate the systems that I have created as well as the performances I have executed. The contribution of this thesis as it relates to this goal are separated into the following six chapters:

*Background* begins with a coarse historical summary of early cinema technology, practices, and cultural relevance. After that, I discuss modern audiovisual performance software packages and the people who use them are discussed. Lastly, interesting gestural interfaces for musical and video entertainment are presented and surveyed.

*Early Experiments* presents the first three devices that I created to explore new visual dialogues between man, machine, and media. Their design and relative merits are discussed.

*Live Performance* is a presentation and discussion of several performances where my systems were used to entertain an audience. I also highlight important details of how the tools evolved to facilitate expressive manipulation of media.

*Futuristic Interfaces in Cinema* examines three gestural interfaces for computer control found in motion pictures, and the lessons learned from depicting the future.

*Cinema Fabriqué* presents my final research project, which employs gestural navigation of media spaces.

*Evaluation* is an overhead view of all presented systems, highlighting both the successes and shortcomings of each. An evaluation of the overall contribution to the area of realtime audiovisual entertainment is also given.



## 2 Background

This chapter divides the relevant background into three sections. The first, *The Moving Image and the Dawn of Cinema*, documents innovations and inventors that made the creation and delivery of movies possible. *Software Tools for Audiovisual Performance and Their Users* is a short survey of commercially available packages for improvisational music and video creation and the established artists who use them. *Gestural Interfaces* examines several hardware devices that were created to harness the expressive quality of hand gestures in order to control sound or visual output.

### 2.1 The Moving Image and the Dawn of Cinema

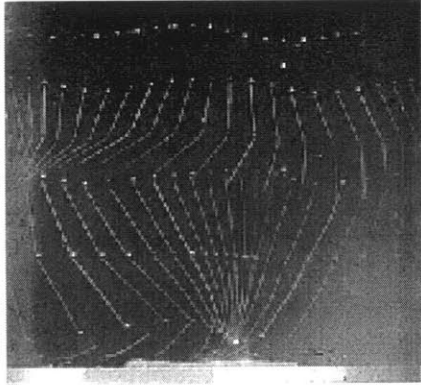
*Plunged into the dark, eyes strained by the flickering light, the jumping image, and the jerky rapidity of any movement, the members of the audience did not feel they were present at the beginnings of a new form of spectacle. The Cinématograph seemed to them cut out for a profound but different calling: the reproduction of life—and maybe its resurrection. [Toulet95]*

#### 2.1.1 Étienne-Jules Marey

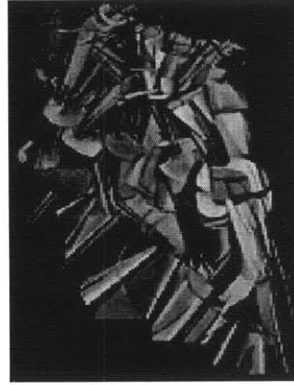
When the first cameras were invented scientists realized their potential as a powerful measuring device. As still cameras were used by thousands of photographers to record historical events and produce aesthetic works, a few curious minds sought to capture the essence of motion and passage of time. Etienne-Jules Marey was among the first and most impactful, producing all of his work in the end of the nineteenth century.

Marey was the inventor of *chronophotography*, a name derived from the Greek, meaning “inscription of time by light”, which in turn gave birth of cinematography, “inscription of movement”. He did many studies of humans in motion wearing a black suit with reflective metallic threadings. The subjects walked or moved around in front





2.1.1 multiple exposure of walking figure



2.1.2 Marcel Duchamp's *Nude Descending a Staircase*

of black panels and multiple exposures of their movements were recorded by one camera of his own design, on a single photographic plate. The results were scientifically important in that they clearly recorded details of human gait and inspired other scientists to develop systems for motion capture. Marey's prints also had an impact on the work of artists such Naum Gabo and Marcel Duchamp, who's *Nude Descending a Staircase* clearly references multiple exposure photography.

Marey's work on his custom built camera laid the foundation for motion pictures, which were introduced by the Lumière brothers in 1895. Motion capture suits almost exactly like the one Marey constructed can be found in hundreds of movie studios and animation



2.1.3 Marey in his motion capture suit



2.1.4 Anthony Serkis as Gollum

firms today. Photoreflctive dots placed on an actor's body can be tracked by several cameras and stored on computer. The paths of the dots through space can be remapped onto the bodies of synthetic characters to be inserted into movie scenes to produce realistic and natural motions. Perhaps the most impressive use of this technology was in the creation of the character Gollum for *Two Towers : Lord of the Rings*.

### 2.1.2 Lumière Brothers

On December 28, 1895 brothers Auguste and Louis Lumière made the first public demonstration of their Cienématographe, in the basement of the Grand Café in Paris. The device was a motion picture projector, which was also the same machine they used to record the films they showed. Although their show opened to a lackluster crowd, within days word of the astonishing spectacle had circulated around Paris and police guards were needed to maintain order as thousands of repeat customers and curious newcomers rushed to the salon doors.

As demand and publicity grew, the brothers hired engineers to take the show on the road and bring the showings to other cities in Europe. Cameramen were trained and sent abroad to film events, locations, and situations of all kinds. Although they were treated like practitioners of witchcraft in some rural destinations, they enjoyed great renown and financial success.

The Lumière shows would consist of a dozen or so "views", films only a few minutes in length capturing a single event or location. Between 1895 and 1907 they collected 1424 "views" divided into 337 war scenes, 247 foreign trips, 175 trips inside France, 181 official celebrations, 125 French military views, 97 comic films, 63 "panoramas," 61 maritime scenes, 55 foreign military scenes, 46 dances, and 37 popular festivals. The "city portraits" made by Lumière operators were intended for a dual audience- spectators eager to discover foreign countries and customs as well as local viewers who took pleasure in recognizing familiar places. [Toulet95]

Shows consisted of several views presented serially. No connection or explanation of the ordering or choice of the shorts was attempted

or implied in the showings. Besides the comedy films, very few of the shorts were scripted in any way.

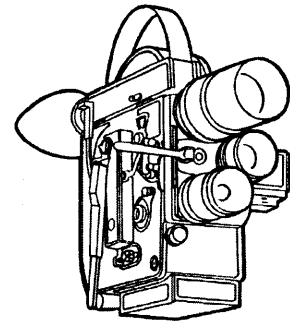
When the brothers brought their Cienématographe to the United States, Americans embraced the brothers for a few years, until patriotism changes their minds. The Lumière brothers were repeatedly harassed and even arrested, and then finally driven out of the market by competitive domestic systems. Even so, the mammoth machines needed to film and project movies were manufactured in low volume and thus available to very few and at great expense.

### 2.1.3 Cinema Verité

The phrase Cinema Fabriqué translates to 'film manufactured' in French and is an allusion to the cinema verité style of movie making, which means 'film truth'. An excellent description of cinema verité can be found in Ken Dancyger's *The Technique of Film and Video Editing* :

Cinema verité is the term used for a particular style of documentary film-making. The post-war developments in magnetic sound recording and in lighter, portable cameras, particularly for 16mm, allowed a less intrusive film-making style. Faster filmstocks and more portable lights made film lighting less obtrusive and in many filmmaking situations unnecessary. The cliché of cinema verité is poor sound, poor light, and poor image. In actuality, these films have a sense of intimacy rarely found in the film experience. Cinema verité was rooted in the desire to make real stories about real people. [Dancyger85];

Thousands of simple and reliable handheld cameras were manufactured during World War II for use by military tacticians and journalists. When the war was over, people all over the world had cameras and the skill to use them. Because cameras were small enough to bring anywhere, film could be taken of anything, any time, and movies could be made by small numbers of people. The films made were not in direct competition with blockbusters from major studios; they belonged to a new genre.



2.1.5 Bolex Paillard 16mm, from [Souto82]

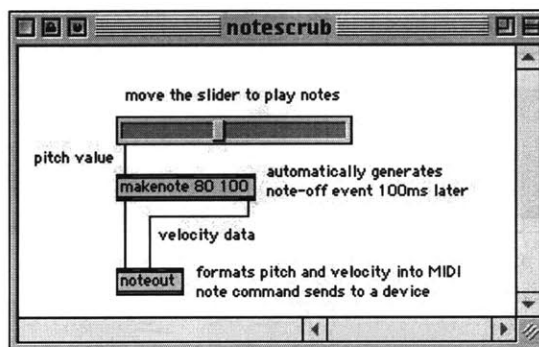
With this thesis I will present tools for providing cinematic entertainment that are easy to operate and very portable. By harnessing the computational power of a computer and the expressive quality of hand gestures I have created a system that allows a great deal of improvisational freedom and control. The intent is not to displace current forms of movie or concert entertainment, but explore a newer form of media presentation, which I have dubbed Cinema Fabriqué.

## 2.2 Software Tools for Audiovisual Performance and Their Users

At this point the discussion will fast forward to the present time, to analyze the publicly available software packages for realtime creation of visuals and sound.

### 2.2.1 Max/MSP

Max is a programming language developed as a toolkit for electronic music production at the IRCAM musical research institute in Paris. It was primarily used by musicians to communicate with and control MIDI devices when it was first released a decade ago. After several releases and the proliferation of free external plug-ins created by users Max became known as the environment that allows 'anything to control anything.' In Max, programs are built by connecting graphical objects together with patch cords. Some of these objects perform calculations and others make up the user interface of your program.



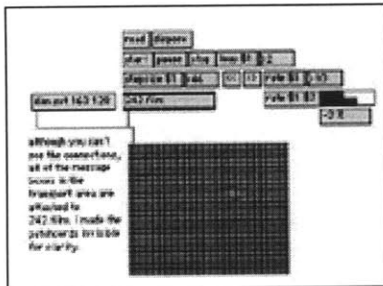
2.2.1 Sample Max Patch for playing MIDI notes with a mouse

In 1999, the company Cycling '74 became the sole distributor of Max and introduced MSP, which adds general audio capabilities to the MIDI environment. With MSP, Max became a fully configurable software synthesizer, sampler and sound laboratory. As with Max, functions are provided to get audio in and out, move audio data around and alter it with filters and envelopes. The process of creating programs is analogous to wiring together different parts of a modular synthesizer.

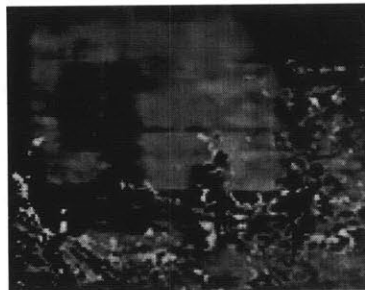
Max is extensible in two ways. Unique configurations of objects that performs a particular task can be saved as 'patches', and loaded at a later time as if it were a built-in function of Max. Drivers, or 'externals' can also be written in C to allow the integration or control of hardware peripherals or software from other vendors.

### 2.2.2 NATO

NATO.0+55+3d modular (NATO from here forward) allows control of QuickTime media (films, images, sound, QuickTime VR, QuickDraw 3D, Flash movies, etc.) from within Max via the familiar object and patch cord environment. MIDI signals or data from any other source can be used to control any of the NATO functions, which include spatial distortions, collage effects, and image generation to name a few.



2.2.2 Sample MSP Patch for playing Audio on MIDI input signal



2.2.3 Example video output

NATO was one of the first software only environments used by VJ's. Using Max and NATO, it is fairly simple to bring in external data

(from a graphics tablet, MIDI keyboard, or mouse and keyboard) and process the input as instructions to play videos, trigger effects, or adjust the composition. Visuals can also be controlled by the audio in Max/MSP or from external microphones. Audio data in the form of frequencies, amplitudes, or other parameters, can be converted into numbers useful to NATO and then used to trigger, alter, or generate images on the fly in response to the sound.

NATO has been around almost as long as Max and used by many people to craft live performances. It ships with very poor documentation, but by being one of the first and cheapest packages allowing live video manipulation, it became quite popular.

### **2.2.3 David Rokeby and Very Nervous System**

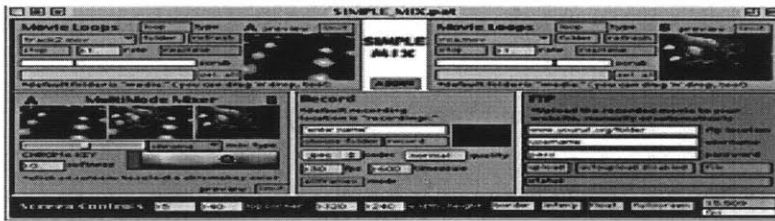
Very Nervous System has a rather complex history. It began as an experiment in visceral communication with computers. When David Rokeby began working with digitized video his Apple II was still not fast enough to analyze an image from an ordinary video camera, so he built his own low-res device: a little box with 64 light sensors behind a plastic Fresnel lens. The box measured the amount of change in light each sensor received from moment to moment and output the data to his computer. He named the device Very Nervous System, or VNS for short because it was able to register the slightest movements and roughly determine where they occurring. VNSIII arrived in the early nineties as a SCSI device, enabling it to communicate to a Macintosh computer in the same way that a hard-drive does.

SoftVNS is Rokeby's latest implementation and is a collection of patches for Max/MSP. SoftVNS was developed with a focus on realtime video tracking, and includes a broad range of tools for tracking, including presence and motion tracking, color tracking, head-tracking, and object following.

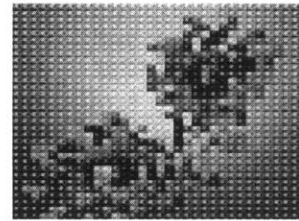
Rokeby's has used his various systems for installations and performances that produce realtime audio and visual in response to bodily movements. He has also sold his hardware devices and software packages to many other artists and musicians.

## 2.2.4 Jitter

Jitter is Cycling '74's response to NATO and SoftVNS. The package is set of 133 new video, matrix, and 3D graphics objects for the Max environment. The Jitter objects extend the functionality of Max4/MSP2 with means to generate and manipulate matrix data; any data that can be expressed in rows and columns, such as video and still images, 3D geometry, as well as text, spreadsheet data, particle systems, voxels, or audio. Jitter is aimed at anyone interested in real-time video processing, custom effects, 2D/3D graphics, audiovisual interaction, data visualization, and analysis.



2.2.5 Example video mixer interface built with Jitter



2.2.6 Example output from Jitter

Although the Jitter architecture is general, it is supposedly optimized for use with video data. A diverse set of mathematical operators, keying/compositing, analysis, colorspace conversion and color correction, alpha channel processing, spatial warping, convolution-based filters, and special effects are the building blocks for custom video treatments.

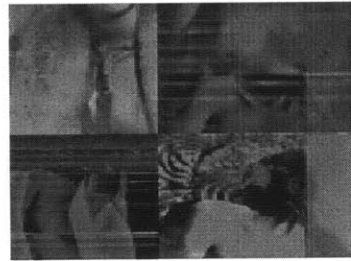
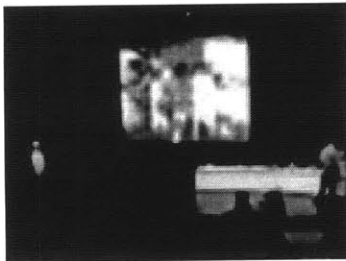
Because Jitter is fairly new in comparison with NATO and SoftVNS, it has not been used in as many performances yet, but because it was released by the company that ships MAX/MSP and comes with professional documentation and technical support, it is expected to become quite popular.

### 2.2.5 242.pilots

242.pilots is HC Gilje, Lukasz Lysakowski, and Kurt Ralske. They have created their own performance software in NATO, which they use to generate music and imagery for their concerts. Improvising as a group, the three artists respond and interact with each other's images and sounds or individually control different aspects of the composition. Images and audio tracks are layered, contrasted, merged, and transformed by the three. The degree of interplay and unspoken communication between the artists is akin to a free jazz ensemble.

The end product is a complex 'visual conversation': A quasi-narrative that explores degrees of abstraction in an immersive journey through diverse landscapes. The experience provides a range of styles and sensory excitation; from natural unaffected footage to media so digitally mutated it defies classification.

The trio have received numerous awards and accolades in the press, including the Transmediale Image prize. Unfortunately, the only action to witness at a 242.pilots concert is on the projection screen. By limiting their controls to mice and laptop keyboards, there is no physical dynamic presented to the audience to communicate who is doing what, and when.



*2.2.7 Snapshots from a 242.pilots performance*

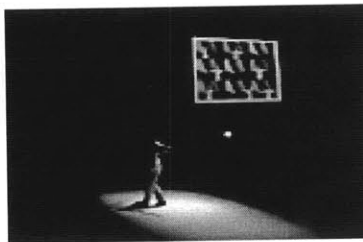
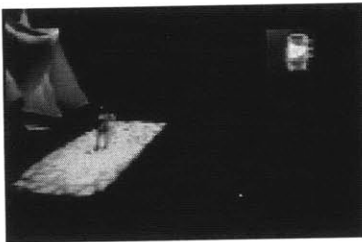


### 2.2.6 Todd Winkler

Todd Winkler is a composer and multimedia artist on the faculty at Brown University, where he is Director of the MacColl Studio for Electronic Music and Chair of the Music Department. His work explores ways in which human actions can affect sound and images produced by computers in dance productions, interactive video installations, and concert pieces for computers and instruments.

He employs Rokeby's VNS system extensively in his collaborations with choreographed dancers. This requires many hours of experimentation and testing for each new work in order to discover music and response mechanisms that feel "right" to the dancers for specific types of movement. He states that his greatest artistic challenge is to go beyond the novelty of producing music "out of thin air," by finding links between the body and sound that are both convincing to an audience and which serve the expressive purpose of the dance [Winkler03].

On March 29, 2003 I witnessed Winkler's performance titled "Falling Up", which dealt with concepts of flight, gravity, and spacetime. Narrative and historical movie clips of planes, science fiction films and physics documentaries punctuated with dance movements that were remapped in the audio and visual domains in realtime. While each movement of the performance was quite evocative, the audiovisual transitions between them were jarring or non-existent. This could have been an artifact introduced by Max as new patches were loaded and old ones removed. It was also difficult to determine exactly what aspects of the music and projected video Winkler was controlling or were pre-programmed.

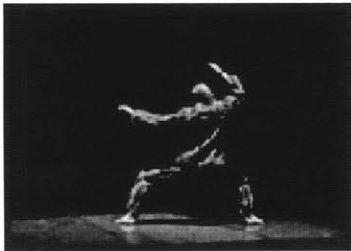


2.2.8 Photographs from "Falling Up"

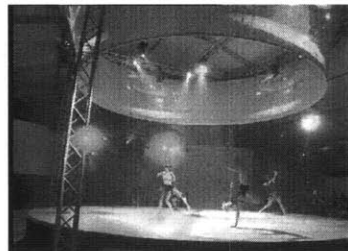
### 2.2.6 Palindrome and EyeCon

Palindrome is a performance collective in Germany that develops and uses interactive computer systems for controlling musical and visual events through body motion. They have released a product called EyeCon which can be controlled by video inputs, body monitor electrodes, or both.

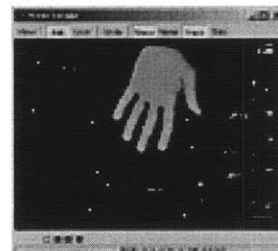
By frame-grabbing a dancer's movements, and processing them with software, it is possible to convert their movements into music (or other media). For their performances a number of small video cameras were placed around the stage and connected to the computer. There they are analyzed by the *EyeCon* system.



2.2.9 Palindrom performance photo



2.2.10 another performance photo



2.2.11 EyeCon screengrab

EyeCon runs on a PC platform and outputs MIDI information by analyzing, separately or together, the following movement parameters:

- changes in the presence or absence of a body part at a give position in space

- position of the center of the body on stage or in vertical space and direction and pathway of travel

- degree of expansion or contraction in the body

- relative positions of multiple dancers (using color recognition and tracking)

- degree of right-left symmetry in the body, how similar shape the two sides of body are

The electrodes are small electrically-conductive pads or strips which, when pressed or glued onto the skin, allow electrical signals from within the body to be received. These signals are quite weak, but can be used to generate a visual or auditory representation of what is happening inside the different parts of body.

The heartbeat system uses specially modified electrodes and software to isolate the electricity created by the heart muscle each time it beats. The dancers' heart beats can be heard, each as a different musical note.

### **2.3 Gestural Interfaces**

During the past twenty years there have been tremendous innovations in the development of interfaces and methods for composing live music with computers. However, most of these systems have not been widely adopted by performing musicians. One explanation for this lack of use is that these technologies do not yet convey the most deeply meaningful aspects of human expression. That is, they do not capture and communicate the significant and emotional aspects of the gestures that are used to control them. Perhaps this is why so few musicians have used computer technology to replace, enhance, or transform the capabilities of their guitars, violins, and conducting batons. [Nakra00]

In many real-time devices (a hammer, a bicycle, a clarinet, or a drum-kit) the human operator has to inject energy or 'excite' the system before it will operate, and must continue to supply energy to keep it going. Then, the energy is steered through the system or damped (dissipated) in order to achieve the task such as playing a note or climbing up a hill. These two operations (inject/excite & steering/damping) are often carried out by different conscious body controls.

Should the task of 'creating a new sound or image' be entirely a case for navigating menu options and entering parameter values? Would it not be better to allow the user to have access to the same type of explorative direct manipulation techniques as outlined above? This section explores several electronic instruments that create deeper connections with the performers' bodies and actions., and the people who use them.

### 2.3.1 The Theremin

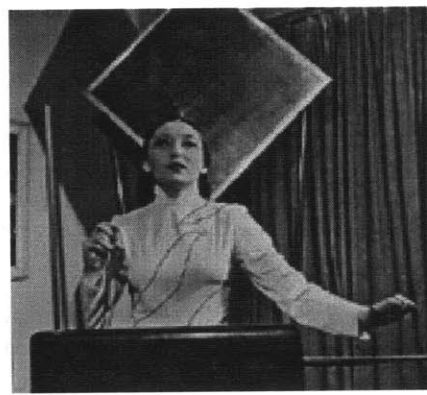
The theremin is a unique instrument in that it is played without being touched. Two antennas protrude from the theremin; one controlling pitch, and the other controlling volume. As a hand approaches the vertical antenna, the pitch gets higher. Approaching the horizontal antenna makes the volume softer. Because there is no physical contact with the instrument, playing the theremin requires precise skill and perfect pitch.

The device was originally dubbed the “aetherphone” by Leon Theremin, who invented it in Russia around 1920. He licensed the design to RCA, which mass produced and marketed the units. As a result, the Theremin became the first commercially successful electronic musical instrument. Many pieces were commissioned for it as solo performances or as part of an ensemble. Clara Rockmore, a violinist by training, quickly became the most renowned Theremin player, and helped to drive both the popularity of the instrument and the refinement of its design by Theremin himself.

Originally, the theremin was intended to replace entire orchestras with its “music from the ether.” While that never happened, it has been used in many recordings over the years. During the 60’s and 70’s, bands like Led Zeppelin brought the theremin back into the public eye for a short time.



2.3.1 *Leon Theremin and his theremin*



2.3.2 *Clara Rockmore*

### 2.3.2 Michel Waisvisz and Edwin van der Heide

Michel Waisvisz and Edwin van der Heide have developed very similar handheld musical devices at the STEIM foundation in Amsterdam. They often perform together with Atao Tanaka in a group known as SensorBand. Waisvisz was one of the first to develop and perform with what is now called gestural controllers. van der Heide now teaches sound design at the Interfaculty Sound and Image at the Royal Conservatory and the Royal Academy of Art in The Hague.

Michel Waisvisz's instrument, *The Hands*, consists of a number of sensors and keys mounted on two small keyboards that are attached to the players hands. The device slightly resembles the middle of a trumpet. Over a score of buttons and different sensors are used to capture the movements of the hands, fingers and arms of the user in a design that has seen many revisions over its lifetime.

Edwin plays the *MIDI-Conductor*, a pair of machines worn on his hands that look like the flight sticks in fighter jets. The *MIDI-Conductor* uses ultrasound signals to measure his hands' relative distance, along with mercury tilt sensors to measure their rotational orientation. There is also a movement sensor, and a number of switches. All of these sensors are connected to the STEIM Sensor-Lab, an embedded, programmable microcontroller that translates the signals into MIDI data to be turned into musical notes by a hidden computer.



2.3.3 *Waisvisz and The Hands*



2.3.4 *van der Heide and the MIDI-Conductor*

### 2.3.3 Atau Tanaka and the BioMuse

On stage, Atau Tanaka will appear to be dancing or practicing Tai Chi while the audience hears sounds and sees moving images projected behind him. His body is in fact controlling what the audience sees via his “instrument”, the BioMuse.



2.3.5 *Tanaka and the BioMuse*

The BioMuse is a biosignal musical interface developed at Stanford University’s Medical School in conjunction with the Electrical Engineering Department. The intent was to create a musical instrument driven by bioelectrical signals from the brain, skeletal muscles, and eyeballs. The device translates these weak analog signals into MIDI and serial digital data.

Atau’s musical work with the BioMuse uses its four EMG channels, which measure tension. Each input channel receives a signal from a differential gel electrode. Armbands with these electrodes are worn on the inner and outer forearm, tricep, and bicep. These signals are filtered and analyzed by the digital signal processor on-board the BioMuse, and converted to MIDI controller data as output.

The resulting data is routed to a laptop computer running Max. Max patches created by Atau transform the control data so that it can enter a compositional framework, and then be dispatched to MIDI synthesizers and real-time computer-graphics performance software.

Even trained, concentrating humans have only a limited amount of control over their internal biosignals. For this reason, the output of the BioMuse is far from deterministic. While Atau can add synthetic rhythms and melodic shapes through his MAX patches, it is fairly obvious to the audience that he is only directing the performance, and not controlling it in every way. Atau of course understands this and is more interested in alternately letting his body speak for itself, then physically reacting to the audiovisual output in an improvisational manner.

Waisvisz, van der Heide, and Tanaka have all been using the same systems for over five years. They aim to master their respective instruments in the same way a classical virtuoso would after decades of intense practice.

#### **2.3.4 Gestural Music Systems from MIT**

MIT, and especially the Media Lab, has a rich history of producing works that employ cutting edge technologies to augment or inspire creative expression. Three recent projects are presented and their merits discussed in the following sections.

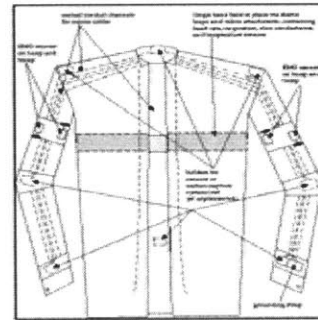
#### **2.3.5 Conductor's Jacket**

The Conductor's Jacket is a wearable physiological monitoring device similar to the BioMuse, which has been specifically designed to provide data on the bodily states of a conductor during rehearsals and performances. This jacket was designed by Teresa Marrin to provide a testbed for the study of musical and emotional expression and was used in a series of data-acquisition experiments by several professional conductors and musicians in Boston in 1999. The project is intended to answer certain fundamental questions about the nature of musical expression and how it is conveyed through gestures.

The sensors used in the Jacket measure respiration, heart rate, temperature, skin conductance, and electromyography (for each bicep and tricep). The Conductor's Jacket has been worn most extensively by Benjamin Zander, a noted Boston-area conductor. He wore the jacket for several rehearsals of the Youth Philharmonic Orchestra at the New England Conservatory, and the data from those sessions has been analyzed visually for features. Correlations between conductor heart rate, muscle tension and dramatic musical events have been found by Marrin. So far, this device has only been used to record and characterize body information and thus far has not been employed as an input device or instrument in its own right.



2.3.6 *Marrin in the Conductor's Jacket*



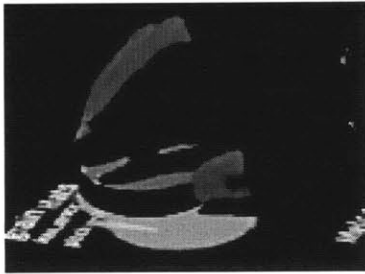
2.3.7 *Conductor's Jacket schematic*



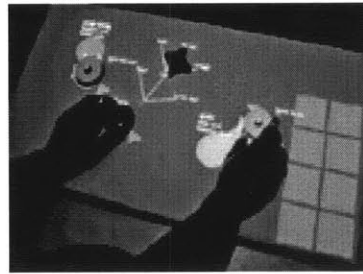
### 2.3.6 AudioPad

James Patten and Ben Recht have developed Audiopad, an interface for musical performance that aims to combine the modularity of knob based controllers with the expressive character of multidimensional tracking interfaces. Audiopad uses a series of electromagnetically tracked objects, much like the objects in Musical Navagatrics, called “pucks,” as input devices. The performer assigns each puck to a set of samples that he wishes to control. Audiopad determines the position and orientation of these objects on a tabletop surface and maps this data into musical cues such as volume and effects parameters. Graphical information is projected onto the tabletop surface from above, so that information corresponding to a particular physical object on the table appears directly on and around the object. Their exploration suggests that this seamless coupling of physical input and graphical output can yield a musical interface that has the flexibility of a large rack mount synthesizer with the expressiveness of a simple display.

Anemone like structures containing the names of musical tracks emerge from the control pucks as the user builds the score. Pushing the pucks in circles can speed up or slow down the tempo. The audience sees the live computational embodiment of song elements.



2.3.8 *Selecting a track on AudioPad*

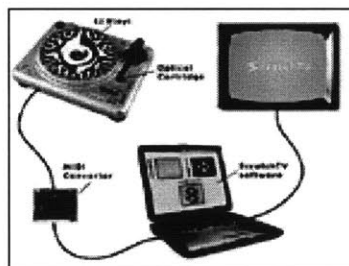


2.3.9 *overhead view of entire AudioPad interface*

### 2.2.7 EJ / ScratchTV

The EJ system, developed by Justin Kent while an undergrad at MIT, allows a user to manipulate and mix video clips with DJ turntables. Instead of standard vinyl records, a special optical record is placed in the player. To read the patterns on the disc, the needle stylus is replaced with a custom built optical pickup. The printed disc is read in much the same way as a barcode scanner reads UPC symbols. The turntable then outputs a digital signal, which is converted to MIDI and fed into a computer.

The computer then deduces the rotation speed of the record, and maps that to the playback speed of video clips stored on the computer. Video can be scratched in realtime with the same motions an analog DJ uses to scratch music. Kent sees the familiarity of the interface as the main advantage of his system. He argues that since a large community of users exists that already have a high degree of mastery of the turntable, users can expect to hear and feel with ScratchTV what they have come to expect from scratching a record [Kent03].



2.3.10 *ScratchTV setup*

In practice the user must do a fair amount of manipulation on the computer to change and fade between tracks. While the video can be scratched analogous to vinyl records, most of the user's time goes into manipulating the software and its controls. If more of the functions could be accessed through manipulating or changing the record, the audience might be drawn into the performance more.



# PAGES (S) MISSING FROM ORIGINAL

Pages 39-42 appear to be missing from the original  
or there is a miss numbering error by the author.

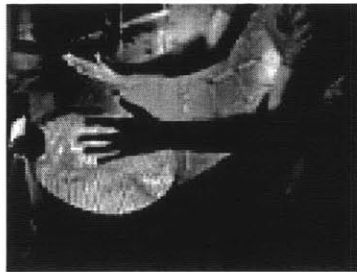
## 3 Early Experiments

This section outlines the first three projects I created while at the Media Lab and represent my first attempts to create visual dialogues between man, machine, and media at the whole-body level.

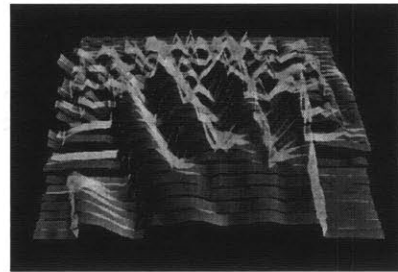
### 3.1 Xtrusions

Naturalist and painter Abbot Thayer coined the term countershading, which refers to the fact that animals tend to be “painted by nature darkest on those parts which tend to be most lighted by the sky’s light, and vice versa” [Meeryman99]. He argued that on average, animal pelts were naturally pigmented in such a way to create uniform distributions of color regardless of illumination, thus camouflaging them from predators or potential prey.

In my first completed project at the Media Lab, Xtrusions, I aimed to develop a display that would have the opposite effect of countershading; exaggerating human forms and motions, separating them from their environment. A projector was housed in a barrel shaped enclosure capped with a projection surface and a camera. The camera signal was fed to a hidden computer, modified, and projected onto the top of the enclosure. The installation acted as a digital mirror or reflecting pool that like all mirrors, is quite uninteresting until someone peers into it.



3.1.1 *photo of Xtrusions in use*



3.1.2 *screenshot of imaged hand*

The image seen on the device was not a direct reflection of the surroundings, but a reinterpretation. Color was stripped of the image, but still used in deciding the form of the resultant output. Instead of a flat representation of the visible scene, the video feed was transformed into a three dimensional landscape of geometric shapes. The brightness of the objects would decide their displayed size and distance from the user. The mapping was simple and linear; the brighter the body part, the closer it would appear.

Due to the projected light emanating from the translucent acrylic screen, objects close to it would become illuminated. The result was an oddly convincing illusion of depth detection by the system. Faces were shaped like faces because the hollows of the eyes, nostrils, and crease of the lips became sunken into the reflection due to their darker hues. Fingertips closest to the screen grew in size and obscured other parts of the scene because they picked up the most light.

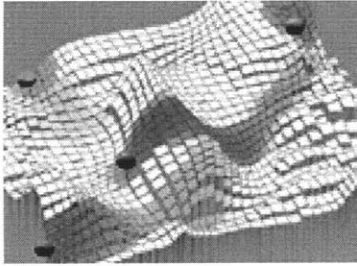
### **3.2 Mad Marble**

Created for exhibition at the London Institute of Contemporary Art, Mad Marble is a painting system done in reverse. Users do not directly control a digital paint brush, but instead change the shape of the canvas by moving their bodies in front of the display. The three dimensional form of the canvas is a direct reflection of the environment in front of the installation. Colored marbles that inhabit the space roll around and leave trails of paint according to the shape and movement of the terrain.

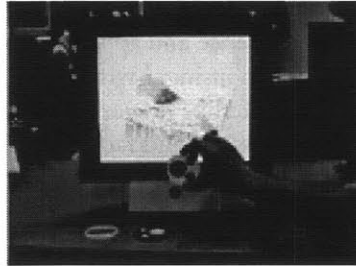
The input device for this piece is also a camera. A landscape similar to the one in Xtrusions is created, where bright parts of the video capture will create higher regions, and darker areas produce sunken terrain. The colored marbles that populate the landscape obey the laws of physics. They roll down hills and into valleys and bounce off each other. The marbles are also constantly leaking paint onto the canvas terrain.

Users interact with the system by simply moving around in front of the canvas. Peaks and troughs are created on the canvas out of the light and dark areas of the user's clothing and skin. Marbles and paint can be pushed or pulled with natural body motion.

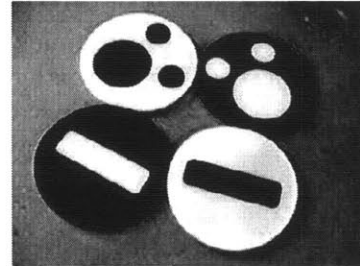
To increase the "grabbing power" that the users had over the marbles, small disks of black and white acrylic with interior shapers were constructed. Their high contrast design produced very sharp peaks and wells in the terrain and made it easier to push and pull the marbles around the canvas.



*3.2.1 screenshot showing marbles painting the video landscape*



*3.2.2 photo depicting usage of disks to control marbles*



*3.2.3 high-contrast acrylic disks used to manipulate the marbles*

### 3.3 Key Grip

As experimentations in video manipulation continued, I noticed that viewers were most intrigued by the previous projects' ability to reconfigure any given live scene into a three dimensional space in realtime. I set out to create a system that would allow the operator to control many parameters of this new expansive video environment. The project is called Key Grip, which is the name of the head lighting engineer on a movie set.

#### 3.3.1 Physical Design

In the previous projects I minimized the presence and appearance of the camera by integrating it into the display housing. I wished the Xtrusions and Mar Marble screens to act more like mirrors, and the illusion was fairly effective. People put their hands and faces very close to the displays in both cases even though it would not reflect them. Users expected the screens to act as direct reflectors, and not just affected output driven by an offset camera.

The form factor of an old camera was chosen for reasons of practicality, durability, and irony. Retrofitting an existing camera housing eliminated the need to design a working encasement and ensured that the final product would feel like a graspable, pointable tool. A camera also communicates to an audience what the user is doing without explanation; capturing image information. And by choosing a camera that is over thirty years old, the audience must do a double take when they see the operator controlling three dimensional graphics with an antiquated device.



3.3.1 camera body



3.3.2 added controls

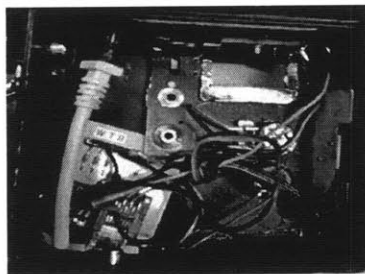


Specifically, the shell of a Canon Motor Zoom 8mm camera was used to house a digital camera and control panel. The body is made of stainless steel and aluminum, which has protected its contents during several hard falls and extensive travel. Everything on the interior of the camera was removed by hand, or with a Dremel cutting tool. The film rolls, lenses, winding mechanisms as well as all external controls were discarded. The only electrical hardware inside was the zoom mechanism, which was controlled by rocker toggle switch. This item was spared because it could be rigged as a pair of digital switches in the new setup.

### 3.3.2 Hardware

A USB webcam with its plastic housing removed was glued in place where the original lens was located. Potentiometers with ridged knobs were fastened to the holes where the film rewind and record speed dials were. A slide potentiometer replaced the f-stop selector. A large toggle button was placed in an extra hole. When the makeover was complete, three analog potentiometers and three digital buttons had been added.

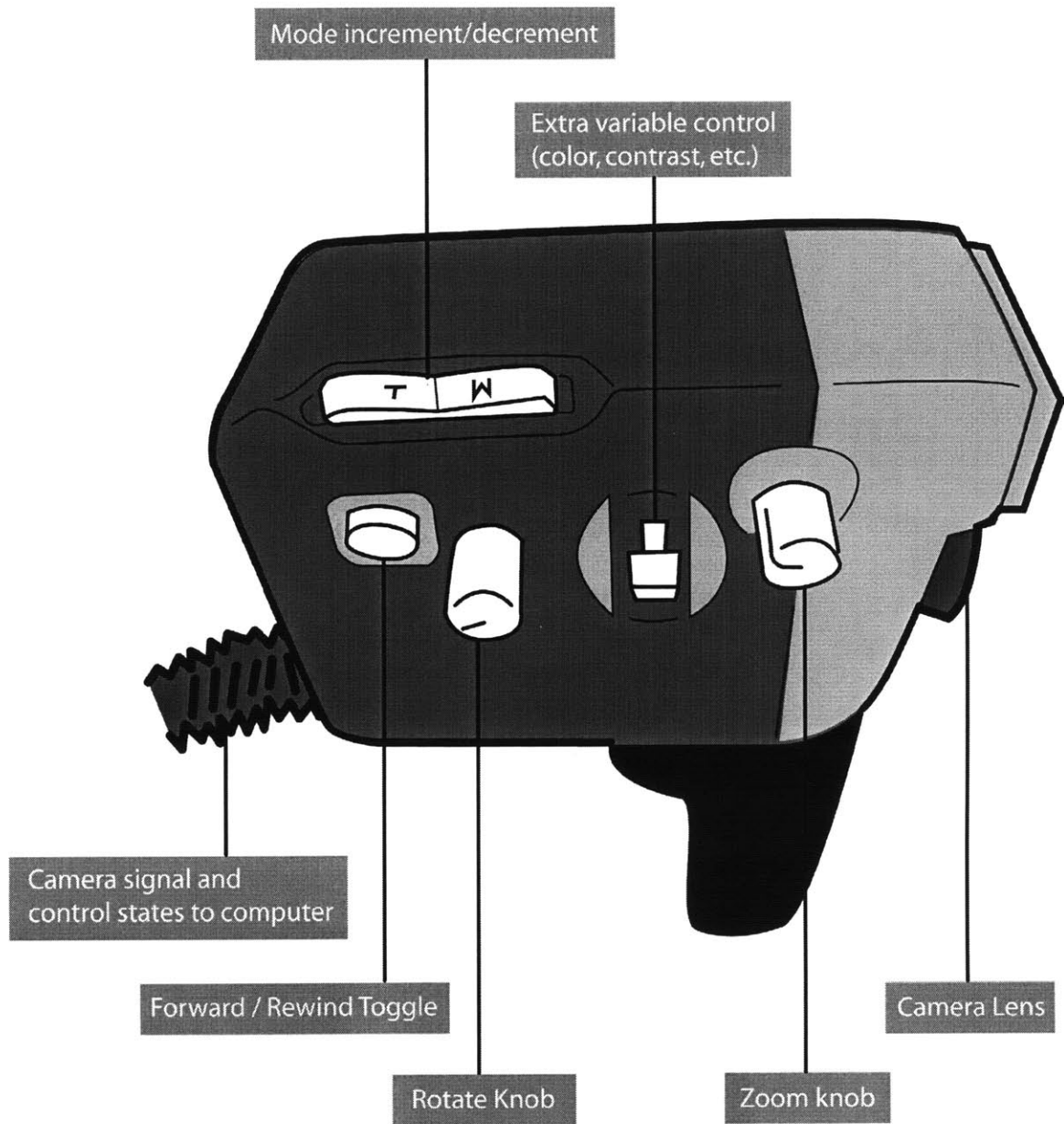
A WildCat analog to digital conversion board manufactured by Z-World was used to read the voltage levels of the camera controls and pass the data to the computer via a serial cable. An eight wire CAT-5 cable was used to carry power and ground to voltage dividers on the controls and the six remaining wires carried the voltage information back to the Wildcat.



3.3.3 Interior of camera



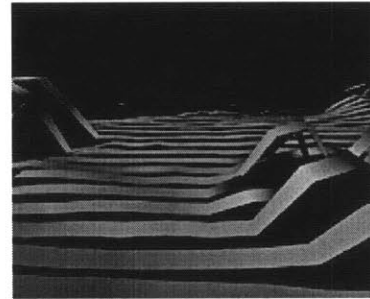
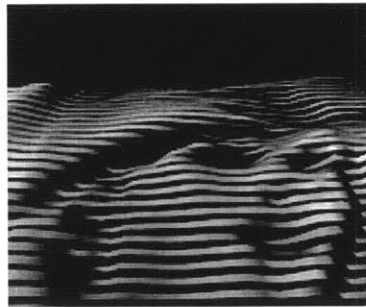
3.3.4 Box protecting A/D board



3.3.5 Diagram of camera controls

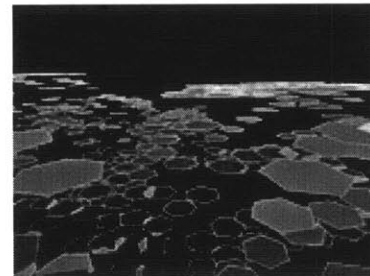
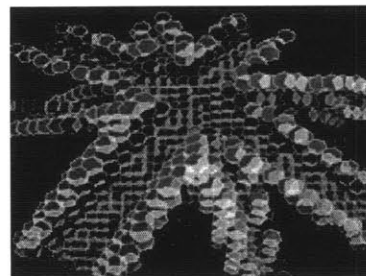
### 3.3.3 Software

#### 3.3.3.1 Spatial Manipulations



3.3.6 *Keyslice mode*

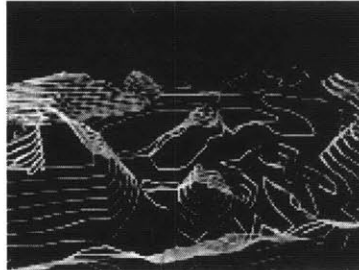
Adapted from the Xtrusions project, these modes filleted the video into pasta like strips or diced it into rotating disks that varied in height and width according to the brightness of the image portion they were representing. Color from the video feed was added to help sensibly communicate the nature of objects and their features. Objects in motion are easy to decipher when rendered in a single color at low resolutions, but static features can appear quite ambiguous. The two knobs on the camera rotate the scene around the  $X$  and  $Z$  axis to allow scenes to be displayed as an undulating landscape .



3.3.7 *BubbleGrip mode*

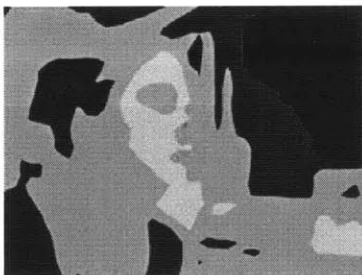
The rotation speed of the disks is a function of how quickly the image intensity was changing at that location.

## TopoGrip and KeyMesa



3.3.8 *TopoGrip effect*

These two modes work by transforming the video signal into a luminosity contour map. A marching squares algorithm was used to extract closed paths of uniform brightness in the camera feed. In this way, visual detail in the display was concentrated in the areas of highest contrast, unlike the previous modes, where a constant number of shapes was displayed regardless of the scene. TopoGrip is so named because of its visual similarity to topographic maps. KeyMesa is a contour map with each of the contours filled in as solid forms and extruded from the contour below.

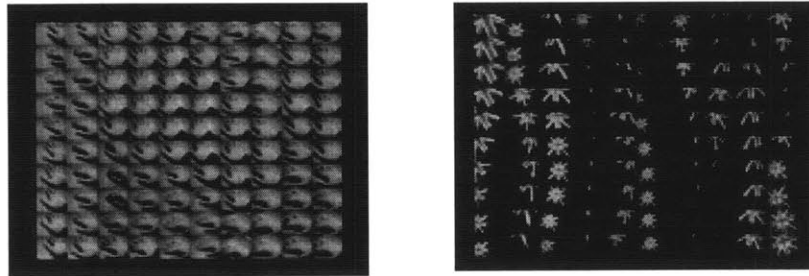


3.3.9 *KeyMesa effect*

One knob moves the viewing position and angle of the rendered scene through an elliptical path. The second knob controls the amount of extrusion or three dimensionality of the scene. The linear fader adjusts the number of contours that are drawn, up to a maximum of sixteen.

### 3.3.3.2 Time Based Effects

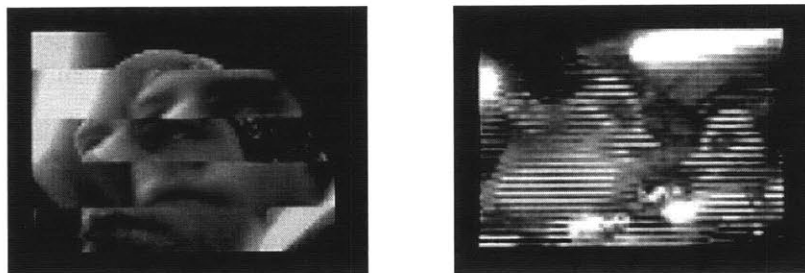
#### TimeGrip



3.3.10 *TimeGrip effect*

The idea behind the TimeGrip effect is quite simple, the last 100 frames of video are always buffered and displayed in a ten by ten grid, with the most current frame at the upper left. Frames flow down, then across in a waterfall of snapshots that undulate with the motions of the past. All recorded frames in the Key Grip system are saved and displayed in black and white to assist the user in separating present events from those in the past. This was a partial homage to old films and the work of Muybridge. The knobs allow the cameraman to tweak the hue and saturation of the current frame, which is preserved as it travels across the screen. It is possible to create multifaceted chromatic portraits of a scene reminiscent of Warhol prints.

#### TimeSlice



3.3.11 *TimeSlice effect*

This TimeSlice mode could display the present, the past, or both at the same time. The present view was in color and the past in black and white. One knob controlled how old the footage displayed in past was, and the other controlled the ratio of past to present shown. When both are being displayed they are drawn in alternating stripes.

### 3.4 Nice filters, so what?

One of the many conveniences that computer visualization affords is the ability to view virtual objects from any angle and scale. Software for CAD design, 3D animation, and medical imaging allow the user to translate, rotate, and duplicate any of the items being displayed with relative ease. The dynamic nature of systems can be studied by constraining object behaviors to the laws of physics and probability and defining appropriate initial conditions. Simulations of physically realistic scenarios are used to help develop new consumer goods, produce special effects in motion pictures, and research the origins of the universe.

The spatial and temporal command over a single object or an entire scene afforded by computer visualization provides a level of control and freedom not allowed in the physical world. Decades ago cars and planes had to be built and crashed to learn how well they crashed. New nuclear bombs had to be assembled and detonated to discover the merit of recent scientific advances and design innovations. While real world testing of ideas and products will always be necessary to fully evaluate their success, engineers have harnessed the computational power of computers to save time, money, and resources to rule out flawed designs and maximize desired qualities such as speed and durability.

In the entertainment industry, comprehensive software packaged to model human bodies, explosions, and space travel are used to produce footage that can be convincingly realistic or fantastically absurd. There are many complaints to be made about the overuse of digital effects, but they have been used effectively to enhance plotlines or present unique views of events.

### 3.4.1 Bullet Time in *The Matrix*

In the mega popular geek-hero movie *The Matrix*, a new special effect technique known as “bullet time” was introduced. This convention has already spread to dozens of other films and is characterized by seamlessly introduced slow motion playback combined with the illusion of total camera freedom. The main characters in *The Matrix* moved very quickly and new techniques were invented to portray their motions effectively. When combat begins in the movie, the movie playback speed would slow down as much as 99% to allow the audience to witness every gunshot, evasive action, and counterpunch as discreet events.

The vantage point given to the audience would routinely circle or travel through dramatic events over the course of tens of seconds, when only a small fraction of a second had transpired in ‘bullet time’. The viewer is allowed to experience action scenes in super slow motion as the heroes and villains would.

The technique was used to communicate the extent of the characters’ superpowers, and stylistically reinforced the plotline, which asserted that life was merely a flawless simulation. Humans in the world of *The Matrix* led normal lives and moved about in a natural environment to the best of their knowledge. In fact, all of humanity was kept in embryonic isolation by a race of intelligent computers. All experience from birth to death was only a simulation of life fed directly into the brains

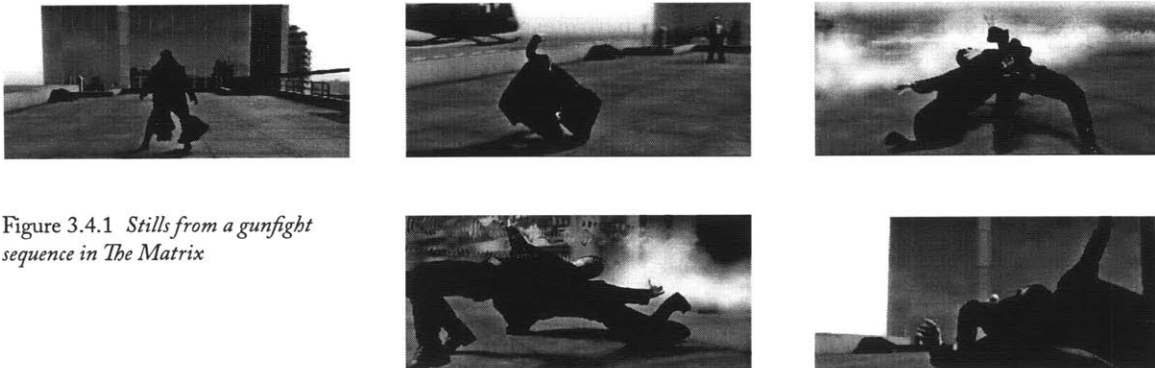


Figure 3.4.1 *Stills from a gunfight sequence in The Matrix*

of the captive population.

Events recorded with impossibly complex camera paths and across several time scales reinforced the notion of simulation, where the camera is but another easily manipulated object in the plastic world of digital effects. Any hint of a cameraperson is removed as the viewpoint follows bullets with uncanny precision or flies through exploding buildings.

It only takes a few camera corkscrews and time stretches before a viewer can visualize total control and freedom of observation. Even though it is the directors decisions and actors' actions that dictate the details of each shot, the use of 'bullet time' creates a world that is 'immersive'. The fakeness of the artificial Earth is delivered in such hyper-realistic clarity that the audience is absorbed into the environment with the on screen actors. Events seem less passively observed than directly experienced.

By programming command over spatial and temporal presentation of live video into the Key Grip system, I aimed to combine the control convenience of computer modelling software with the expressive possibilities of direct scene manipulations. Key Grip users can loop interesting footage or zoom into a tiny portion of a scene from any angle or velocity. With the ability to introduce time changes and three dimensional distortions to live video subtly or abruptly, viewers can be drawn into a world that is simultaneously real and unreal. The shared surroundings of performer and audience are easily reinterpreted at will to exaggerate prominent features or introduce new meaning.







## 4 Live Performances

Any researcher or product developer can confirm that the laboratory is an idealized setting for demonstrating new technologies to others. Power outlets provide high amperage current with high-rated fuses, spare parts and knowledgeable colleagues are within reach when systems fail, and detailed explanations from the inventor can instruct the viewers and defend design choices.

Since the stated goal of my projects is to provide tools that facilitate human expression for a live audience, the most obvious and useful testing ground is an entertainment venue. Having no prior performance credentials or fan base, the first deployments the Key Grip system were used to accompany other acts. As the hardware, software, and my confidence evolved, I controlled larger portions of events, and finally gave a few solo performances.

### 4.1 Section Format

The format for this section will be to toggle between descriptions of individual performances and the hardware and/or software improvements that resulted. Many features were inspired by problems that arose at specific showings, so intertwining the evolution of the Key Grip system with the acts it was used in is perhaps the clearest way of justifying design decisions. At the end of this section I will summarize what I believe to be the most interesting trends seen after many performances.

#### 4.2.1 Quincy Cage - April 20, 2002

The world premiere of the Key Grip system was in the basement of Harvard University's Quincy Hall. The generous use of chain-link fence to demark separate "rooms" in the subterranean expanse earned the venue its common name, Quincy Cage. Harvard students are allowed to hold concerts there themselves or to invite local acts to perform.



4.2.1 Screenshots from video taken at event

After learning of a hip-hop and dance competition at the 'Cage', I volunteered to provide visual mixing during the event. The turntablists, rappers, and dancers were to perform at the front of the space, and I set up my gear next to the sound equipment. As the show started I quickly assumed the role of a video journalist. I pointed the camera at the act on stage or particularly interesting attendees, and the video screen behind the performers broadcasted the Key Grip output.

I was primarily matching visual rhythms with the beats being played by the turntablists. Using my time dilation functions I could mimic the scratching effects of the DJ's. The potentiometer knobs were well suited for controlling motion through time or speed of playback. Like a mini-record platter, video could be drawn back slowly and evenly, or see-sawed between normal and retrograde playback. The potentiometers on the camera were not continuous turn, so they were only effective for one or two twists through a few seconds of footage.

The three dimensional effects could also be linked to the beat of the music. The graphics can be rotated an integer number of times each measure or switched between clockwise and counterclockwise motion on the beat to emphasize the rhythmic structure of the song.

As catchy as Hip-Hop beats may be, they are typically simple and repetitive, and therefore predictable. Within minutes, I was beat-matching time changes, scene rotations, and color mixing of the visuals to the songs being played. I tried toggling the output modes on the beat, but they are so visually and stylistically dissimilar that the result was too chaotic to be pleasing.

Lighting was the main showstopper. Cheap cameras cannot see well in low lighting. Several lamps were scattered around the stage and I would reposition them as needed to illuminate the scene, often setting the camera unit down. The ad-hoc process was quite dynamic and totally exposed to the audience. For a first try the whole system worked very smoothly. Since I had little idea how the event was going to develop from the outset, small lapses of beat matching or real time scene adjusting were not big disappointments.



4.2.2 *Screengrabs from video taken at event*

#### 4.2.2 Lighting and Output Flip Upgrades

A standard filming light was added to the camera to add illumination when needed. While the camera was over three decades old, film lights and tripod standards are apparently older, for they both attached to the Canon with ease.

A small detail I noticed when filming the audience was that the output was naturally reversed in the horizontal when I turned toward them. It was slightly awkward for people to see themselves or others 'backwards' in realtime. A simple toggle was incorporated into the system that would reverse the horizontal output, so I could rotate freely while filming, and not depart from left-right parity with audience expectation.

#### 4.3.1 Bucky Spins Concert - May 24, 2002

MIT grad students Clark Kemp (guitar + vocals), Ben Vigoda (bass), and Dan Paluska (drums) make up the local rock band Bucky Spins. They performed a private concert in a Cambridge warehouse and invited me to provide visual accompaniment to their music. I filmed the band from the front and the computer output was projected behind the band on a white wall.



4.3.1 *Bucky Spins in front of Key Grip produced visuals*

The audience was much more captive than the one at Quincy Cage, with most people seated and paying close attention to the show. For this reason my visual delivery style was much more deliberate than it had ever been. Only six effects were available at the time, and to prolong the novelty of the whole visual treatment, they were revealed to the audience one at a time, roughly one for each new song. The conceptual space of each filter was explored and fleshed out over the course of a song, with controlling variables adjusted in time with the rhythm of the music.

The biggest problem was again lighting. I positioned a spot light

on the stage before the band started, but Dan the drummer had set up in the rear outside the light's throw. The camera mounted light helped slightly, but the restriction of the cord length kept me from illuminating or filming him well.

Due to the downsampling of the video feed in many of the video modes, higher contrast shapes are represented much clearer. Charlie was at the front of the stage and was the easiest subject for my system to portray. His black t-shirt helped his head, arms, and guitar pop out of the scene, especially in the 3D representations. Ben was not lit as well and was wearing an earthy patterned sun-dress and wig. It was difficult to build coherent scenes focused on him due to his optically busy outfit.



4.3.2 *Bucky Spins in front of Key Grip produced visuals*

As a result, the left/right symmetry of the projection was fairly constant, being mainly focused on Charlie. I used crude hacks to create variety, like turning the camera upside-down.

There were several lessons learned at the concert. A cord is very limiting. Like a dog on a leash, I quickly inspected everything within my available radius and became most annoyed at the opportunities just beyond my reach. Restricted by the presence and length of the camera and control cords I routinely snagged on speakers and my own legs. Even if I could have lit the drummer well, I would not have been able to get close enough to film him. I resolved to make the system wireless. If I could reposition myself in space as easily the graphics could be, I would have a far greater control over how the environment was captured and presented.

Even though the band members were restricted to the sounds their one instrument could create, if they wanted to communicate facts, ideas, or emotions directly, all they had to do was open their mouth and sing. Concepts as varied as free donuts, mid-term elections, and bicycle crashes can be presented through song with great ease and clarity, if not surprising wit. After visually reworking essentially the same scene from the same angle for about an hour, I realized that the addition of outside material into the show could have added a lot of variety and meaning. I made the decision at that point to add playback of prerecorded media clips to system.

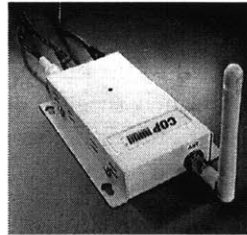
### 4.3.2 Key Grip Unwired

Building a wireless camera was quite simple. I used a small analog video transmitter marketed for spying purposes connected to a common CCD 'camera on a chip' and a rechargeable 12 volt battery. These were all placed inside another Canon Motor Zoom 8MM body. The video signal can be transmitted over 50 feet indoors and is carried in the 2.4 GhZ band. The signal is received by a small box with an antenna and an NTSC output cable. This is then fed into a video capture card on the PC.

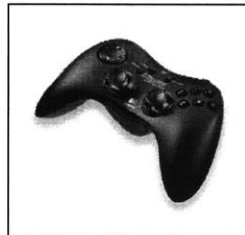
The building of reliable remote controls proved to be much trickier than rigging up the wireless camera. After attempts to hack into the closed caption portion of the video signal and build an independent send and receive system for the knobs and buttons were failures, I resorted to an out of the box solution; a wireless video game joystick.



4.3.3 *wireless camera*



4.3.4 *NTSC receiver*



4.3.5 *wireless video game controller*



The gamepad is manufactured by Logitech and provides two analog joysticks, a linear throttle, ten buttons, and an eight directional thumbpad. As the nature of my visuals became more complex and three dimensional, the use of a videogame input device became very natural and useful. PC and console games now take place in accurate reconstructions of real cities, richly decorated landscapes, and across whole galaxies. And with this explosion of scale and realism, the ability of users to fluidly navigate and view their surroundings has blossomed with the advent of creative control metaphors and high-bandwidth input devices. Trying to manipulate two degrees of freedom smoothly with two separate knobs on the Key Grip camera was as frustrating as trying to draw a circle on an Etch-a-Sketch.

After watching a seasoned gamer who has been locked into an all-niter of networked Doom III, the notion of proprioception surfaces. The player ceases to register individual button presses and joystick twiddles; they simply 'become' the character in the game and proceed to jump, dodge, and shoot their way towards victory. This control transparency coupled with the ease of three dimensional navigation made the gamepad a natural choice for realtime video manipulation.

### **4.3.3 Prerecorded Media**

To import digital movies and sounds into the performances, Microsoft's DirectShow libraries and video code written by Simon Greenwold were utilized. DirectShow allows for integration of MPEG movies, MP3's, AVI files, and older QuickTime formats. As long as movies are initialized at run-time, they can be played back at any resolution or speed, or have those values change dynamically over time with no loss of output performance.

The number of clips that can be loaded for a performance is proportional to the system RAM. Each clip has a certain amount of memory overhead, and this is not related to its length. I have routinely loaded over a hundred clips into the system, but have noted considerably performance degradation, especially after the software for several hours, when loading more than sixty.

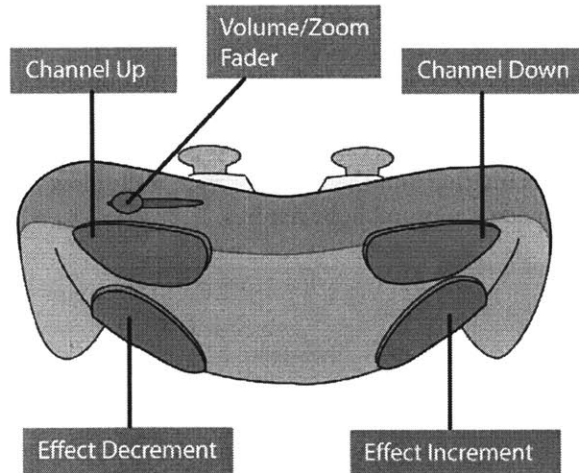
The introduction of prerecorded media and many new degrees of control with the gamepad presented a wealth of new decisions to be made. The final assignment of editing and navigation functions can be seen in the figure on the opposing page.

With the availability of many movie clips, the metaphor of television remote control was quite obvious using the gamepad. It became conceptually convenient to consider each clip as a separate channel, with the live video input being channel '1'. Channels were incremented and decremented with symmetric buttons on the front of the pad. The visual filters are scrolled through in an analogous way with another pair of buttons on the controller's front.

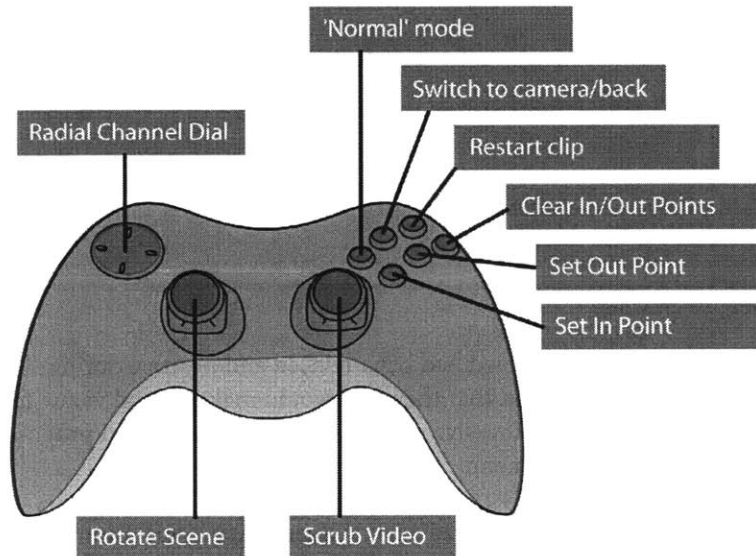
A few functions native to video-editing hardware or software systems were also incorporated. The right joystick was turned into a jog/shuttle device, so that when rotated, playback could be sped up, slowed down, or run in reverse. The motion of the joystick was also buffered, so that 'scratches' of the video could be recorded, played back, or looped. The clips were looped by default, starting over from the beginning when finished, but subsets of each clip or any traversal of their content could also be set to repeat. Pressing the 'in point' button set the start position of the loop and recording of temporal control began. Pressing the 'out point' button ended the recording session and playback of the produced loop began. To return to the unedited clip and delete the recorded video scratch, the 'clear points' button is pressed.

Rotation and panning of the scenery were controlled with the left joystick and throttle lever. In general, parameters that could hold a range of values were altered using the joysticks and throttle and binary mode changes were toggled with the buttons.

4.3.6 front view of joystick:  
clip and effect mode scrolling



4.3.7 top view of joystick:  
spatial and temporal controls



#### 4.4.1 Enormous Room - August 10, 2002

The Enormous Room is an eatery/lounge in Cambridge that opened its doors in April 2002 and quickly became known for its very low lighting and a menu with two choices, carnivore or vegetarian. The elegantly spartan decor is inviting and the exposed brick walls glow in the wash of custom made sculptural light fixtures.



4.4.1 flyer for event

The atmosphere and crowd had been relaxed and positive during several visits. I imagined that the place would make a good venue for experimental but non-invasive audiovisual performance and proposed a show to the management.

To deliver the video to the audience, sixteen vintage televisions were scattered about the room. Some were on the floor, tables and in between couches, and a few were stacked into totems or pyramids. All sets shared the same signal, but the color, tint, and contrast knobs on all were adjusted to add variety to the array.

The majority of the visuals were driven by pre-recorded movie clips. This was the first show where I controlled audio as well as visuals. It was not a solo performance, but I was contributing samples and pre-recorded clips into the audiospace as well as creating all the visuals. Electro-vocalist Cephalopod and DJ Chris Parlato provided the main audio entertainment, and I would periodically fade in the soundtracks from the stored movie files.



4.4.2 Pictures taken at the event by Joe Unander

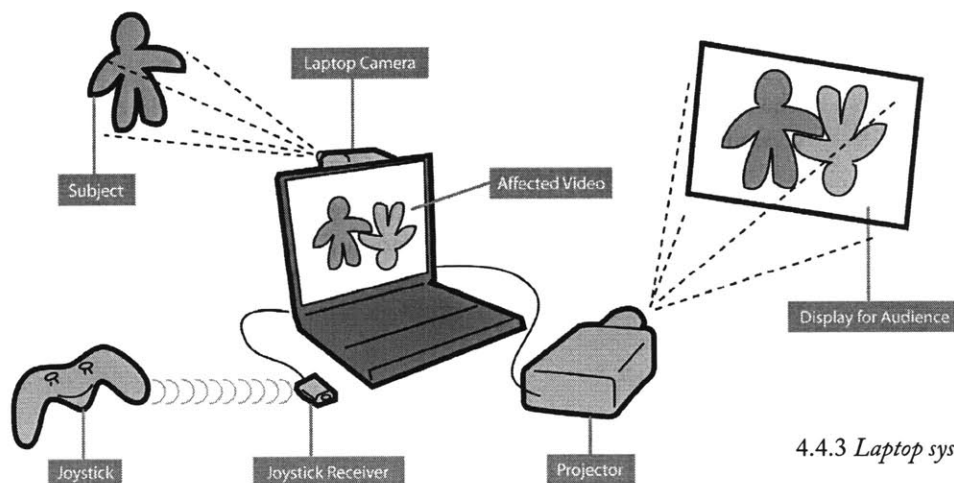
The sixteen televisions and cables necessary to link them to the computer were an ordeal to transport and set up, but the computer itself was the source of most of the problems throughout the show. The IBM workstation I used had a dozen peripherals to connect and troubleshoot. During travel, the video card dislodged itself partially, allowing the system to boot, but not display high color or high resolution graphics. This took about an hour to figure out.

The whole system would turn off periodically until a higher amperage wall outlet was switched to. The whole event was made possible by the temporary donation of a pickup truck and over six hours of setup. If I was going to perform outside the Boston area or by myself, I would need to streamline the hardware requirements in terms of size, weight, and simplicity.

Thus the decision to move the software to a laptop was made. Fitting in any backpack with room to spare for controls, camera, and a small projector, I could potentially become a one-man show with a setup time comparable to that of a guitarist or DJ.

#### 4.4.2 Truly Mobile

The transfer of code to the laptop went smoother than predicted and the whole suite of effects and media controls ran nearly as smooth as on desktop. The resolution and color depth were reduced, but not unbearably so. Video input was limited to a special clip on camera, because the IBM laptop I used only had a single USB connector, which needed to be used for the joystick. USB hubs were tested but they did not work well with cameras, possibly for bandwidth reasons. A schematic of the new system is presented below.



4.4.3 Laptop system schematic

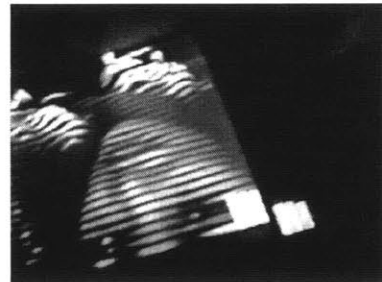
#### 4.5 Museum of Sex Opening Parties - September 17-18, 2002

The portability of the new system allowed me to take the show on the road to New York City for the opening party at the Museum of Sex. The chaotic pace of New York life and business was quite apparent when I arrived three hours before the museum was to open and the walls of the lobby were still wet with fresh paint and live electrical wiring dangled from the ceilings.

The whole building was transformed into a clean, safe, party environment by a dozen hectic workers and I was given two square feet to set up five minutes before the doors were to open. The mobile Key Grip system was up and running when the first visitors arrived.

I had prepared a whole collection of new source material for this particular show. The footage consisted of low to mid grade porn filmed directly off a television screen as well as clichéd Freudian shots of rockets, volcanoes, and trains. I was not responsible for the audio, which was provided by a local DJ hired for the event. I sometimes matched the visual rhythm to his track selection, but quickly lost interest in choreographing long pieces for a transient audience. Visitors were allowed to wander around all three floors of the museum, so traffic was constant and people tended to circulate rather than linger.

I started presenting the system to visitors as *Sex Box*, as if it was another item in the museum's collection. People were much more interested when I allowed them to control the output and pass it around amongst themselves. Fairly soon, the controller was being



4.5 photos taken at MOSEX opening

passed around the room and museum goers were explaining the controls to each other. I was almost not needed, once a critical mass of intrigued partiers collected.

The most frequent complaint by those who tried the system was the lack of smooth transitions between clips and modes. Only jump cuts and drastic geometric changes were possible. On that note I implemented optional fading between clips and gradual changing of modes, accessed by holding down the increment or decrement buttons instead of tapping them.

#### **4.6.1 Metamorphique Fashion show - October 27, 2002**

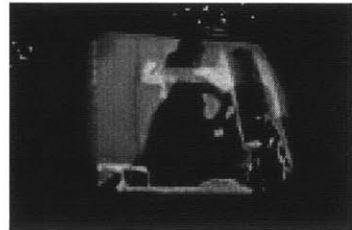
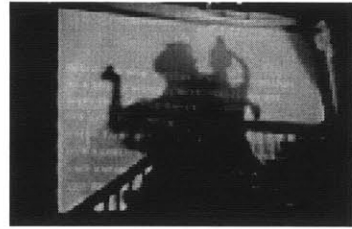
I provided visuals for a hair and fashion show sponsored by Biyoshi Salon and Puma. The show was held at Felt, a lounge, restaurant, and pool hall in downtown Boston. The theme was international dress through the times, with traditional styles juxtaposed against modern designs. Nine regions were represented by two models apiece, one wearing historic garb and the other donning a custom outfit designed for the show.

There was no natural runway at the venue. A large spiral staircase was used as the presentation structure. Models entered at the top of the stairs from the second floor and descended into the audience. A large scrim at the top of the spiral blocked the spectators' view of entering models. The unlit screen blocked the view of both models as they struck poses, and then captured their silhouettes when a backlight was activated by a technician.

By front projecting on the scrim I visually narrated the event. As the models remained motionless, quotes from historic regional figures would cascade down the scrim. The backlight would fade away, leaving only the projections of the Key Grip system and the first model would descend. The imagery changed at this time to pictures from the area represented; prayer flags for Tibet, sprawling marketplaces for China, and the like.

The tone of the projections was subdued and their tempo quite slow. With no announcer, the visuals informed the audience of which





#### 4.6.1 *Models behind the projection screen*

nation and era was being represented, and what to expect next. The visuals then faded to the background as the models came into view. No special effects or geometric distortions were applied to the visuals. Simple fades and pans were used to transition between clips to avoid jarring flashes that might detract attention from the models and the outfits.

The *Metamorphique* show was the most standard, ‘analog’ use the Key Grip system has seen. Many other packages could have delivered identical results, but the wireless aspect made preparation and control painless. The laptop and projector were set up on a tiny shelf in the middle of the crowd, and I was free to operate the visuals from anywhere in the room.

The hardware footprint and setup time were minimal, and there was no need to extensively scout the venue beforehand. The computer did not need to be in a convenient location because it was only receiving and processing data, and not acting as a physical input device. The camera on the laptop proved to be quite useless because the models moved a great distance and I did not want to ruin the seriousness of the affair by chasing after them with a laptop ‘pointed’ in their direction. I resolved to integrate the wireless camera solution into the laptop setup.

## 4.6.2 Modular Camera and Control Design

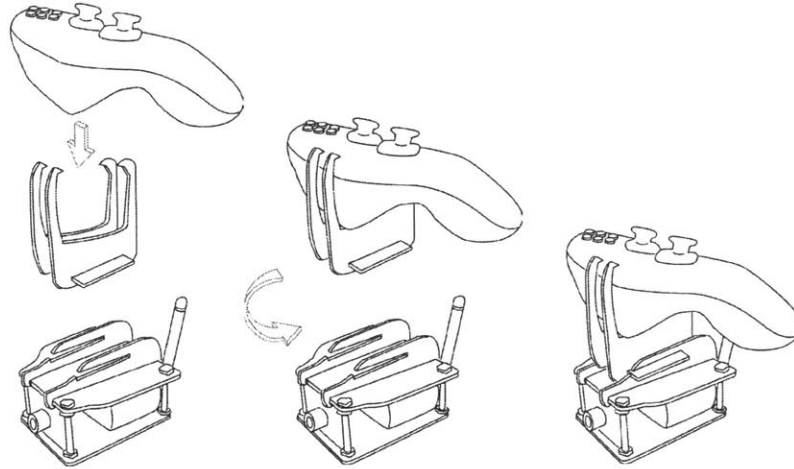
A unique feature of the first Key Grip system was the collocation of the camera and its controls for realtime effects. Available video performance software such as Max or EyeCon have accommodated video-in functionality for years, but situations with long term or great camera movements required a separate camera operator, because the software controller can not stray far from the computer. If only a single person was available, only the software or the camera could be adjusted at any one time.

When several people are responsible for creating a live broadcast quality video output signal, redundant information capture, limited task responsibility, and communication are the rule. The use of several cameras attached to predictable cameramen helps the editor of news program, sports broadcast, or sitcom to produce coherent realtime content.

But in the age of embedded journalism, often only a single reporter is available to cover any given event on location. They might be far from editing equipment or their situation too hectic to boot up the necessary laptop. A camera with onboard editing functions would be quite convenient in a fast paced environment.

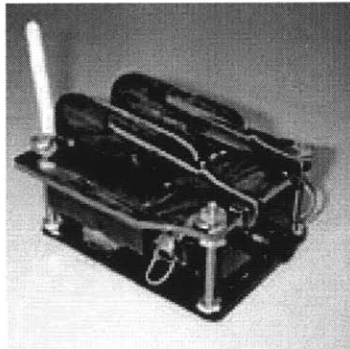
Every venue I had performed in had been a hectic environment. Often there wasn't a space to set up my equipment until minutes before the show was to begin, and it wasn't necessarily ideal for camera placement. By making the camera and controls both wireless, I was able to move about space more freely. But because each required the use of both hands, I couldn't simultaneously move the camera and affect its output.

I built a new housing for the wireless camera that could be attached or removed from the joystick controller easily. The camera could be fastened to the controls for walkabout filming, or removed for capturing still shots. A diagram and photos of the modular system are shown on the next page.

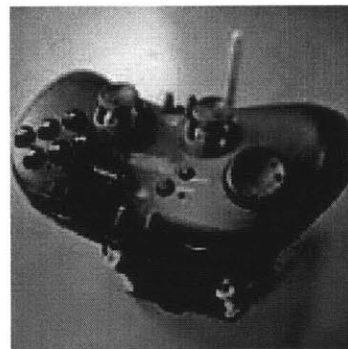


4.6.2 *Diagram of modular camera/control system*

There was still the problem of getting video into the laptop without using a video capture card. The signal from the wireless camera was sent to a small monitor, which was being filmed directly by the laptop's integrated camera. The solution was low-tech, but surprisingly pleasant. The images were softened subtly, darkened in the corners, slightly more choppy, and had a trace of moiré artifact. By adding more technology into the loop, the result wound up looking like a twenty year old home movie shot with an 8mm camera.

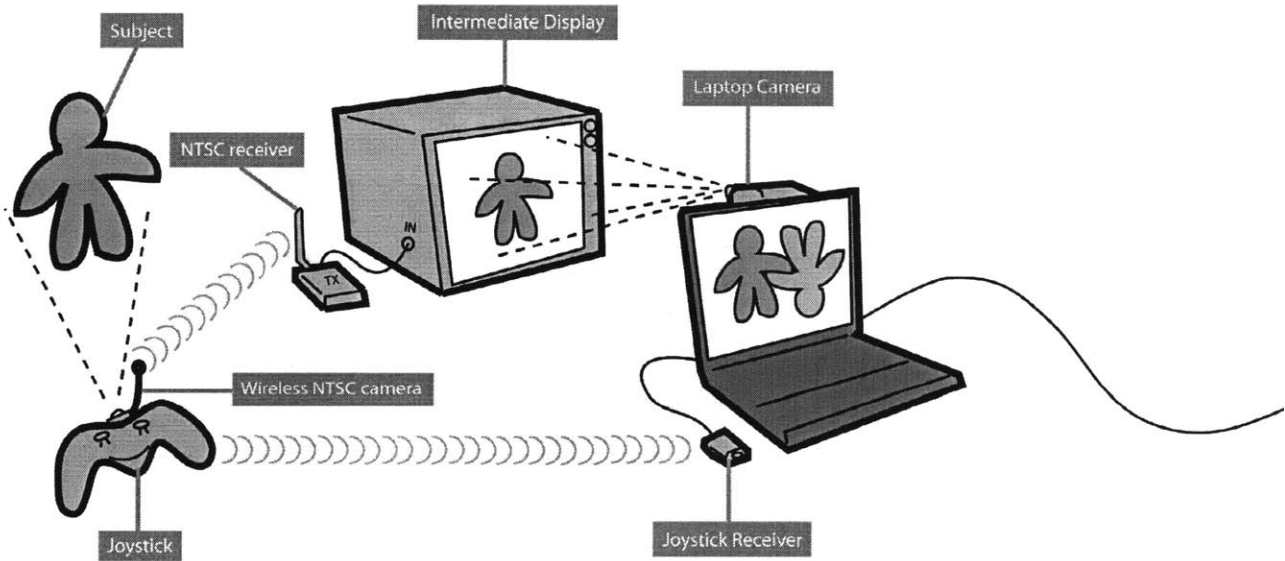


4.6.3 *Wireless Camera*

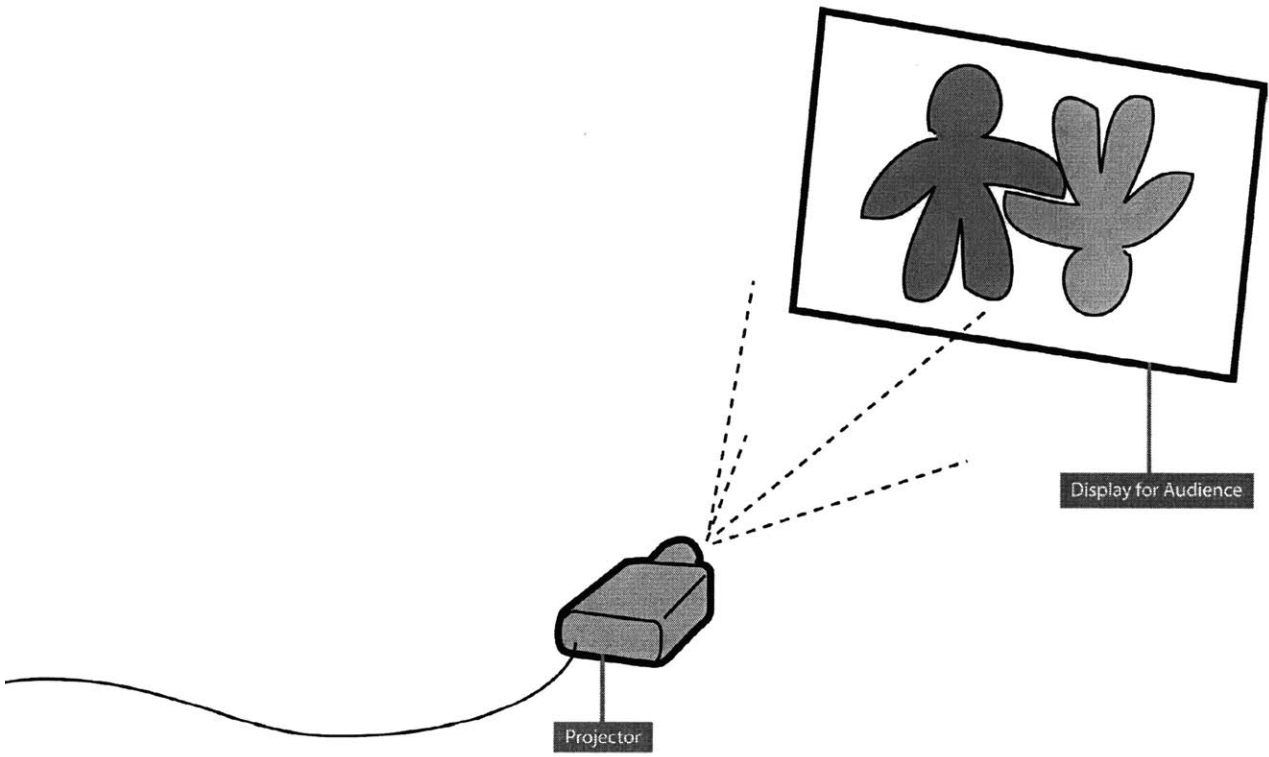


4.6.4 *Wireless Gamepad and Camera*

This large graphic is a diagram of the components used to produce the visuals for the next two performances. Although the projection scenarios varied, the hardware used to capture, control, and manipulate the imagery was identical.



4.6.5 Schematic of laptop system with wireless camera and controls.



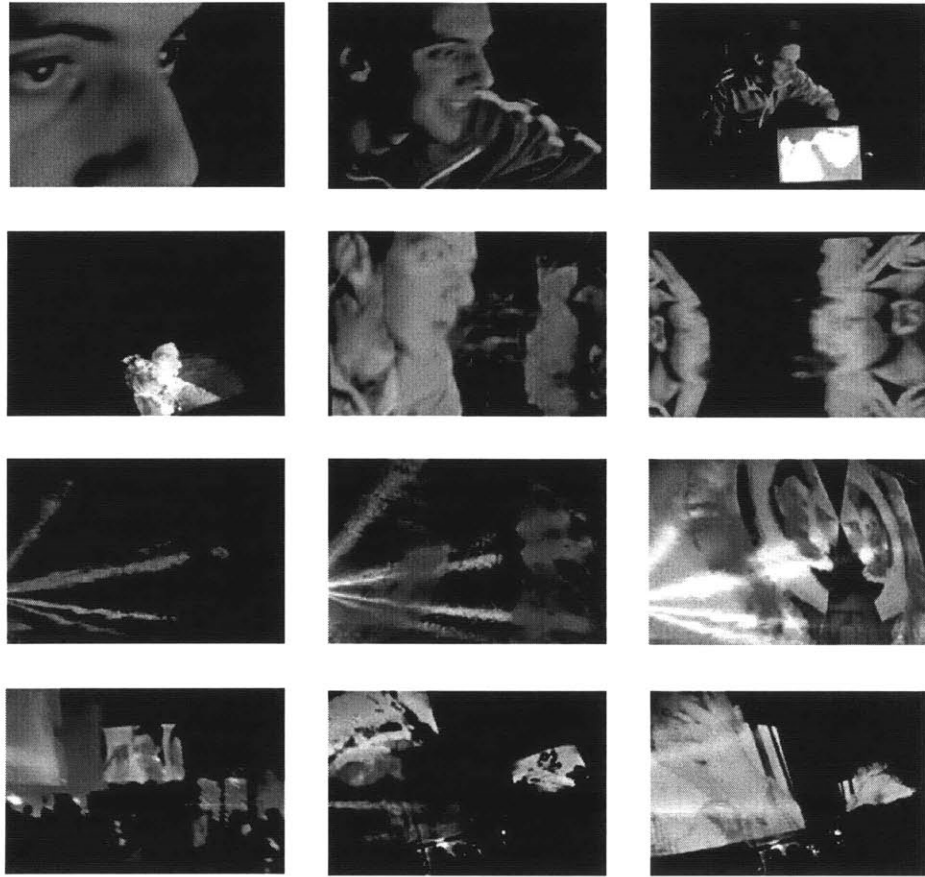
#### 4.7 Octophonic SoundExperience - Enormous Room February 14, 2003

The *Octophonic* event was so named because the audio was delivered to the audience through eight separate channels. Eric Gunther, AKA Cephalopod, built a custom eight track, eight channel sequencer in Max/MSP for the show. Up to eight audio loops could be played simultaneously, and the location and velocity of their source in the room could be controlled in the software. The program would interpolate between speakers to smoothly produce the illusion of travelling sound sources.



4.7.1 Flyer for the *Octophonic* concert

I provided all the visuals for the event which were delivered by two projectors mounted at right angles near the center of the venue, to maximize the number of people who could clearly see them. The projection screens were custom built by Timon Botez out of gossamer and dowel rods. One was placed flush against the wall and the other was suspended half-way between the projector and wall. This created three effective display surfaces because the projected light illuminated the suspended fabric and passed through it onto the wall about three meters behind it.

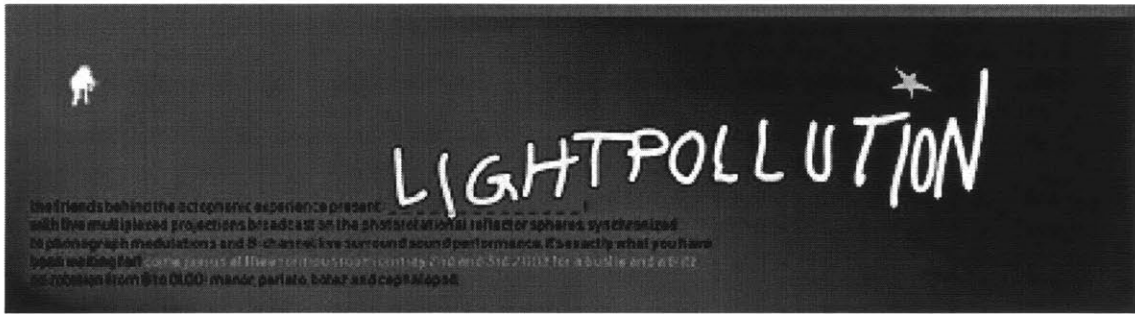


4.7.2 Screenshot from video taken at Octophonic Event. Read from left to right, top to bottom.

The quantity of light being produced allowed for effective use of the wireless camera to legibly image Eric and Timon creating the music as well as members of the audience throughout the space. Attendees usually figured out that I was controlling the visuals when they saw the joystick in my hands. Many people asked to try out the controller. Because it is a common recreational consumer device, it is not intimidating to most people. Concert goers will very rarely ask musicians to hand over their guitars, turntables, or laptops for a quick test run, but lots of people felt comfortable requesting a turn with my system.

#### 4.8 Light Pollution - Enormous Room - May 3, 2003

The most recent event relative to the publication of this document was *Light Pollution*, held once again at the Enormous Room and in collaboration with Eric Gunther and Timon Botez. The stated goal was to cover every available surface in the space with projected imagery. The gossamer screens used in the Octophonic show added so much character to visuals by layering the video and twisting organically in the breeze that it was decided to push the spatial presentation even farther away from the standard viewing rectangle associated with television and movie theaters.



##### 4.8.1 Flyer designed by Timon Botez

A gossamer sheet was hung from the ceiling that spanned nearly the entire length of the room. Three projectors were suspended and one angled so that its foreshortened projection illuminated the whole length of the sheet. Two rotating disco balls modified by Botez were placed in front of the other projectors. He fastened several hand mirrors to the exterior of the disco balls to provide continuous reflection surfaces larger than the tiny mirrors already there.

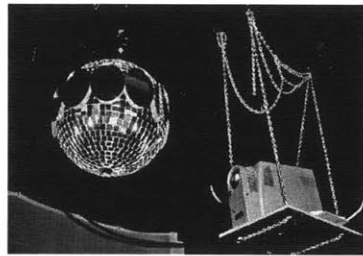
Once the projectors and balls were turned on, the room became filled with moving light of all shapes and sizes. The large sheet was always filled with the unfractured but very stretched images from the angled projection. Large ellipses containing portions of the video signal reflected off the spinning mirrors traveled around the room,



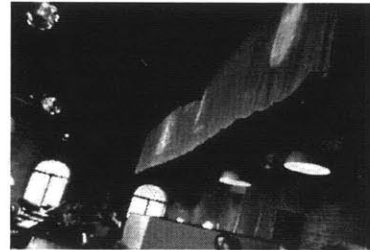
sometimes containing a face or a portion of a car. Tiny flecks of color were everywhere, bounced off the small square mirrors of the disco balls. Stranger shapes were the result of double reflections off both balls or occlusions.

The effect was very different from a light show that might accompany an arena concert or seen inside a night club. Every packet of light moving throughout the space was a portion of a full video feed. The two balls were rotating in opposite directions, causing collision after collision of image pairs.

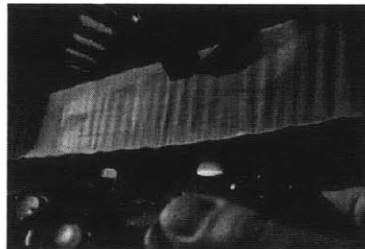
On this page are several photographs of the setup. The sound was all produced by Gunther and Botez, using the Octophonic system once again. On the next two pages are a series of frame grabs from a video shot at the event.



4.8.2 *disco ball projection setup*  
*Screen during daylight*



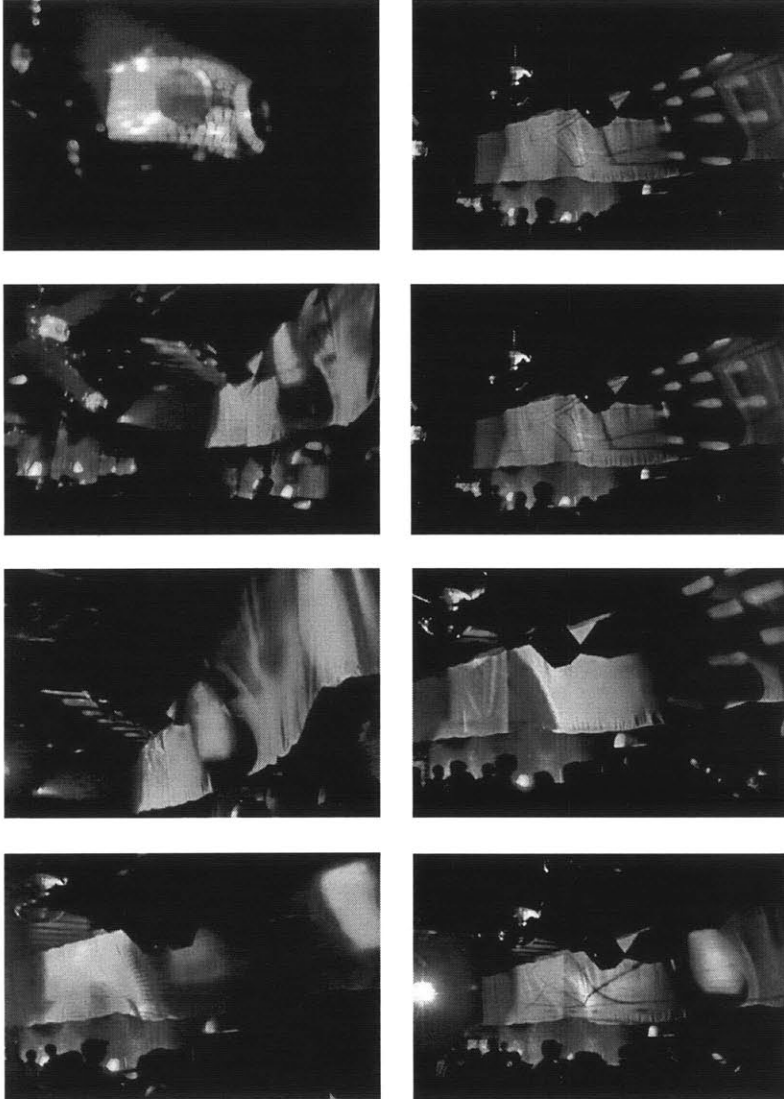
4.8.3 *Rotating elliptical projections*  
*on screen*



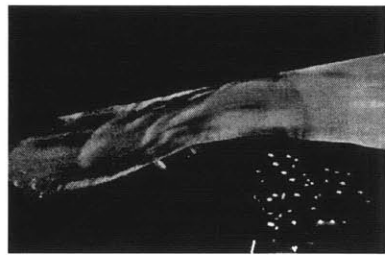
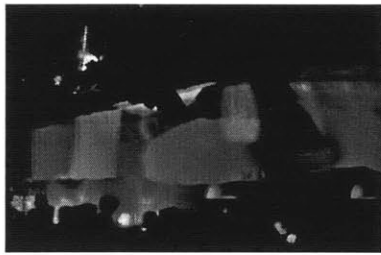
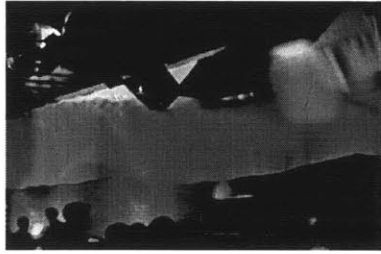
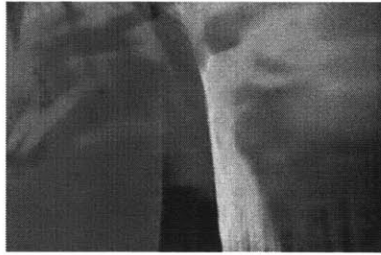
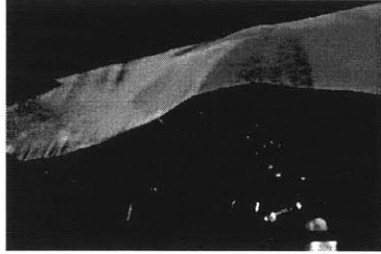
4.8.4 *Me controlling the visuals*



4.8.5 *Eric & Timon controlling*  
*sound*



4.8.6 Screenshots from video taken at Light Pollution. Read top to bottom, left to right.





## 5 Futuristic Interfaces in Cinema

Millions of people spend all day in front of a computer and there is no indication that this trend will reverse anytime soon. Watching others use computers can be a fairly undynamic experience, and when captured on film the results can be downright boring. This may be why so many software interfaces seen in movies are excessively colorful and flashy. Films set in the future are not as restricted as those depicting the present, and can sometimes deliver interesting views of technology and its users that are not yet possible. Directors set aside current limitations of money, practicality and often common sense to bring the audience a glimpse of distant possibilities. There are ideas worth noting in the good movies as well as the bad.

Fictional gestural interfaces found in three science fiction movies are examined in this section. I was inspired by these imaginary devices because they were intended to 'wow' the audience with novel uses of technology to perform actions that are currently uninteresting to watch.

## 5.1 Phantom Planet

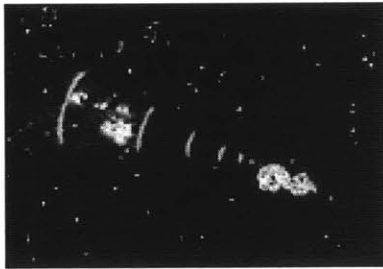
This extremely camp science fiction opera was produced in 1961 and takes place at an unspecified time in the future. Two astronauts are in space investigating the disappearance of several American rockets when they are themselves abducted by the *Phantom Planet* and its tiny inhabitants.

It turns out that the abductors are just little humans who are quite friendly, but in danger of being conquered by the evil monster-race of Solarites. Luckily, the astronauts discover that they are superb pilots of the *Phantom Planet* and are able to fight off the attacks of the Solarites. The controls of the planet are a series of objects resembling oversized champagne glasses. As the pilot's hands come near them the characteristic warble of the theremin is heard, and the viewer suddenly realizes he's experiencing pure mid-20th century entertainment.

The planet/asteroid/spaceship is steered through space using the theremin. The interface is also used to activate the ship's "gravity curtain", which is a powerful deathray that destroys the Solarite fleet once and for all. The *Phantom Planet's* use of a theremin controller speaks of a future where all machine interfaces provide infinite mechanical advantage from our inputs. Users will no longer need to get their hands dirty with buttons, levers, and keyboards because they will only need to wave their hands through the air to get the job done.



5.1.1 *Theremin-like control used to fire weapons.*



5.1.2 *Resulting deathray that saves the day*

## 5.2 Johnny Mnemonic

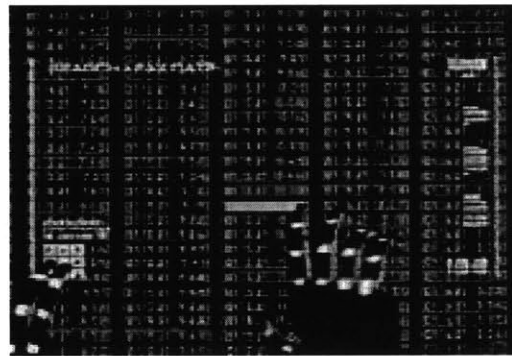
William's Gibson's cyberspace novel *Johnny Mnemonic* was adapted to the big screen in 1995. The plot is set in 2021, in a world run by multinational corporations, where information is the most valuable commodity. Because of hackers, the most precious data is not transmitted via Internet, but by human couriers with memory chips implanted in their heads. One of those couriers is Johnny (played by Keanu Reeves), whose greed and arrogance propel him to double the capacity of his brain-drive, risking a total cranial meltdown.

To access the Internet in 2021, users don virtual reality gloves and goggles much like the ones that received so much hype in the 1990's. Surfers grabbed information 'shapes' and rotated them in three dimensional space to view their resident data. Security portals resembled Rubik's Cubes and similar hand puzzles that had to be spatially reconfigured to gain access to sensitive data.

The director told the digital effects crew that they should forget about where Virtual Reality technology is right now and visualize it as a rich, high resolution, superfast medium. He claimed that technology will follow fiction, and as Virtual Reality is perceived on film, so may it very well come to be.



5.2.1 *Johnny in VR gear*



5.2.2 *Gestural Interface for web surfing*

### 5.3 Minority Report

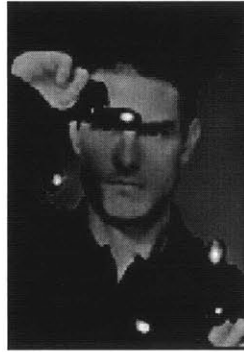
In Spielberg's adaptation of Philip K. Dick's short story, *Minority Report*, cops have learned to see the future and stop murders before they happen. The police detectives analyze the precognitive visions of psychics to search for clues that will lead them to crime scenes.

To navigate the vast quantities of video clips extracted from the psychics, the lead investigator, played by Tom Cruise, would first don a pair of special gloves with lights embedded into the fingertips. By waving his arms in front of a large display, he could sort and seek through many sources of information simultaneously. Clips and pictures could be "grabbed" and manipulated. Zooming, panning, and scrubbing actions were all performed with simple hand gestures.

The system was designed for the movie by Media Lab alumnus John Underkoffler. Spielberg instructed him to come up with an interface that would make its user look like an orchestra conductor. The gestural language used to control the system was so communicative, yet simple, and the technology so plausible that it became a major inspiration for my own work. So much so that I wound up borrowing many of the ideas from this movie for my final project, Cinema Fabriqu e.



5.3.1 Cruise using lightgloves



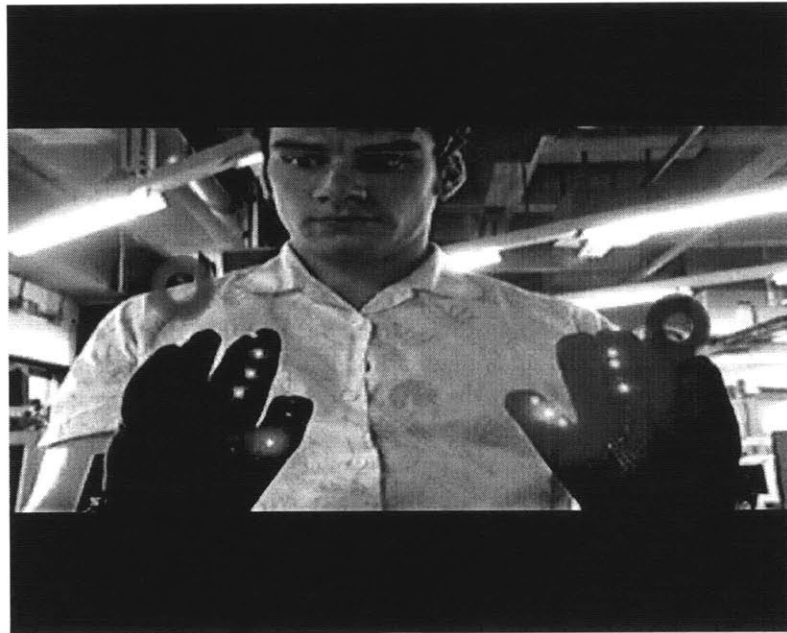
5.3.2 Close-up of gloves







## 6 Cinema Fabriqu 



6.1 Cinema Fabriqu  in action

A gesture is a motion of the body that contains information. Waving goodbye is a gesture. Pressing a key on a keyboard is not a gesture because the motion of a finger on its way to hitting the key is neither observed nor significant. All that matters is which key was pressed. Using your hand to show the motion of a falling leaf is a gesture. Flailing away on a video game joystick is not gesturing but rather an operation of a controller that senses in which of eight possible directions a stick is being pushed. Beckoning with your index finger is a gesture. Handwriting is not a gesture because the motion of the hand expresses nothing; it is only the resultant words that convey the information. The same words could have been typed; the hand motion would not be the same but the meaning conveyed would be. After deciding that all previous interfaces that I designed poorly communicated how I was controlling the sights and sounds that were delivered to the audience I began work on a new system that would hopefully be more transparent in its presentation and more accessible.

## 6.1 Thoughts on Gesture Controllers

The common approach to gesture control design is to develop a language of user gestures and recognition algorithms that interpret hand movements. The sub problems are :

- choosing a set of primitives for the gestural language which is appropriate for the given task domain.
- given a particular input medium (dataglove, video, etc), designing syntactic recognition algorithms to appropriately identify user gestures in terms of the language chosen.
- once particular gestural statements have been recognized, performing an unambiguous mapping from these statements to system actions. [Allport96]

According to Noam Chomsky, human language is separate from and more complicated than other kinds of cognitive systems; languages are so complicated that it is impossible for people to infer all of the rules of sentence production simply by being exposed to samples of grammatical sentences [Chomsky86].

Having an interface as flexible and broadband as Sanskrit or other obscure language would be interesting, but would mean very little to spectators that endured a performance of reasonable length. What was needed was a transparent interaction or metaphor that took very little time for an audience to figure out. The implementation of a symbolic dictionary of hand configurations, akin to sign language, was rejected as a control mechanism because it was felt that it was asking too much of the audience to decipher and remember a collection of hand-words.

I have not implemented a “language” as such, but instead have mapped continuous movements of the user’s hands in three dimensions onto continuous movements of virtual objects that control the scene, like a steering wheel or a record player.

The semantic mapping problem is therefore reduced to apparently trivial questions like “If I move my hands to the left, should the objects on the screen move left or right?” However, these choices is not trivial. For each dimension we sense, there are two choices of direction or orientation of mapping, and the further choice in each dimension of whether to map this onto translational or rotational movement in 3D space. In the domain of video playback and creation, the environment is actually four dimensional, giving us nearly 2 dozen such choices. Then there are the additional choices of how to conceptually model movement in the domain, in terms of movement of self, media objects, or some intermediary.

## 6.2 Visible and Invisible Metaphors

As Bill Gaver points out; graphical computer interface metaphors such as windows differ considerably from everyday metaphors. Instead of highlighting similarities between existing objects in the world, they create new graphical objects. Where the *conceptual* mapping of the interface metaphor differs from the perceptual mapping used to express it [Gaver95].

Since the Cinema Fabriqu e system is controlled with relatively large scale hand gestures, proprioception plays a large role in the experience of the user. Humans have a strong internal sense of what it feels like to make movements of their limbs about in space, and are accustomed to associating these feelings with the motions of their bodies through the world. The control mappings programmed in the video tracking software take advantage of this mind-body link to suggest invisible metaphors, which arise out of the “experience of interaction itself” [Allport96], and are a result of the way certain gestural actions feel.

Many of the effects available in the Key Grip system were ported to the Cinema Fabriqu e software, but the manner in which they are controlled was mapped onto invisible structures that were adjusted with hand motions. Three dimensional scenes are navigated by turning and pushing an imaginary steering yoke. Time based effects are executed in much the same way as records are scratched and pushed on a turntable platter.

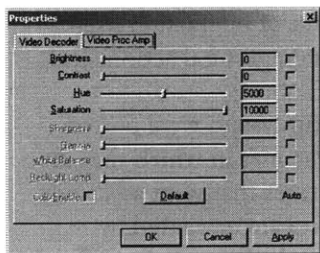
### 6.3 Computer Vision Techniques

*The retinal image produced by the hand of a gesticulating speaker is never the same from moment to moment, yet the brain must consistently categorize it as a hand [Zeki72].*

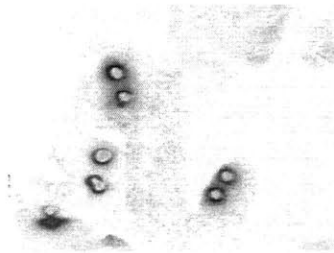
While the human mind has little difficulty categorizing and tracking objects in motion, it is quite difficult to train computers to usefully analyze dynamic scenes. Unless applications are run in extremely controlled environments, issues of occlusion and lighting can confound even the most sophisticated vision systems. Even in laboratory conditions, object recognition and tracking using popular methods of template matching and segmentation rarely produce perfect results.

Simplicity was embraced in the design of the glove tracking system for the Cinema Fabriqu e environment. No part of the scene is ever recognized or tracked per-se. Image moments, developed by Horn [Horn, '85], were employed to estimate the average position and orientation of visible light. Image moments are fast to compute and provide global summaries of a scene's brightness and spatial configuration.

By adjusting the pre-amp settings of the video capture board I was able to filter out most of the normal world that was visible to the analog camera. Setting the brightness and contrast sliders to their minimum value and the saturation to its maximum, only the brightest light sources appear in the video feed. Human forms and background



6.3.1 *Dialog to adjust internal parameter on video capture board*



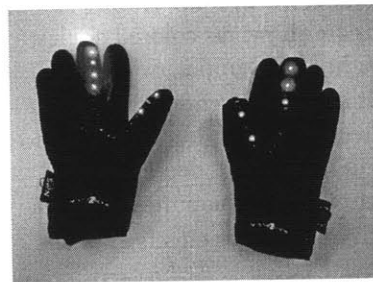
6.3.2 *Light gloves as seen by software. Inverted for clarity.*

objects do not register in the system. LEDs and incandescent lights are readily visible as ring-like forms. Since colored LEDs emit light in a very narrow and predictable wavelength band, their presence is easy to detect. By putting different colored LEDs on the left and right hands, they are easily distinguishable in software. Image moments of each of the red, green, and blue channels of video are calculated separately, yielding average light location and rotation for each. With this method, each hand is effectively tracked through space independently.

#### 6.4 Glove Design

Women's mild weather gloves were chosen to hold the electronics. Mens' gloves had too much extra baggy material or stylish embellishments. Each glove held eight surface mount LEDs powered off 3 rechargeable AAA batteries. Transformer wire was chosen to carry the current and act as scaffolding for the LED's because it bends easily, is quite durable, and prevents shorts. Transformer wire is coated in non-conductive insulation so that charge is only transmitted through the wire when coiled, instead of across adjacent windings.

The wire was stitched through the glove fabric at regular intervals of a few millimeters. The exposed wire was sanded to remove its insulation in the spots where LEDs were soldered. Current limiting resistors were added in each glove to prevent burnout.

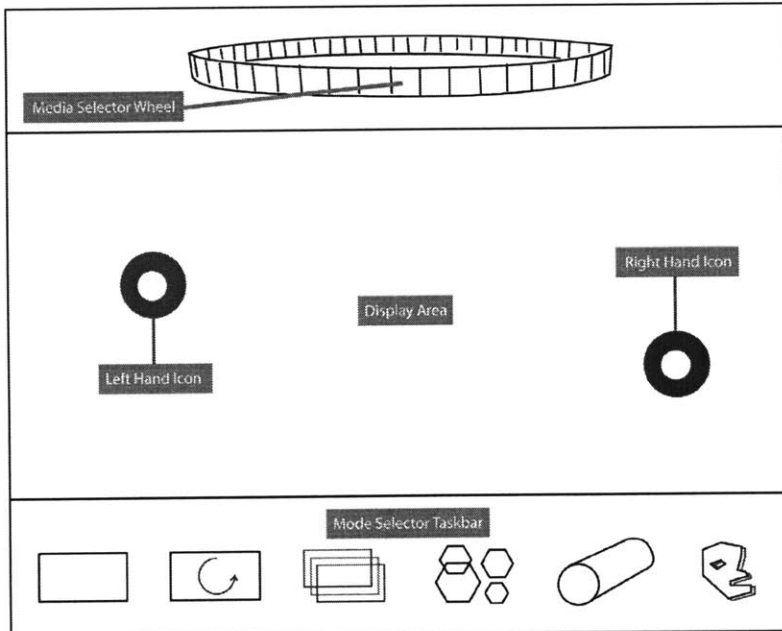


6.4.1 *Light gloves shown palms facing up with LEDs*



6.4.2 *Light gloves show palm down exposing battery pack*

## 6.5 Software Interface



6.5.1 *display layout schematic*

The middle two thirds of the screen is where the visuals are displayed, in widescreen 2:1 format. The top portion of the screen contains all the available media clips and camera feeds, and the different modes and features of the software are accessible at the bottom of the display. The locations of the users hands are shown as colored circles; red for the left hand and blue for the right. All manipulations of visuals and audio occur when the hands are in the 'center stage', and clip transitions or mode changes are affected by bringing the hands to the top or bottom of the screen. In the default mode, depicted as an empty rectangle icon in the mode select region at the bottom of the screen, the clip is played back 'as-is' with no filters or effects applied.



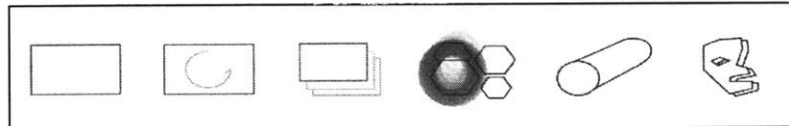
### 6.5.1 Media Track Selection



6.5.2 *video wheel interface*

At the top of the display is a video wheel interface. All available movie clips and camera feeds are shown as a continuous linked ring. The forward most clip is highlighted in blue and is played full size in the large center display area. When the right hand, represented by the blue circle, is brought to the top of the screen, the video wheel fades in from black. At this point, it can be rotated by nudging it with clockwise or counterclockwise motions of the right hand. As the wheel spins, the forward most movie changes, and thus the main output is transitioned.

### 6.5.2 Effect Selection



6.5.3 *effect taskbar shown inverted for clarity*

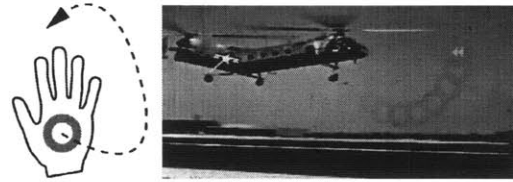
At the bottom of the screen resides the effects 'taskbar'. When the right left hand is brought below the central display area, a taskbar displaying the six different modes of the software appears. By simply moving the hand cursor over an icon, the depicted mode is chosen.

### 6.5.3 Time Based Control - Video Scratching

When the 'scratching' mode is selected (the icon with a circular arrow), the right hand becomes able to seek through the clip being played. Any curved motion will accelerate the playback; clockwise movement results in fast forwarding, counterclockwise movement rewinds. The magnitude of the acceleration is determined by the velocity of the user's hand and the size of the arc that is swept out. Small circles produce slower playback and large circles will speed up the video very fast. Little fast-forward or rewind symbols will appear inside the blue hand icon to inform the user when their actions begin to scratch the video. See the figures below for examples.



6.5.4 clockwise motion fast-forwards the video



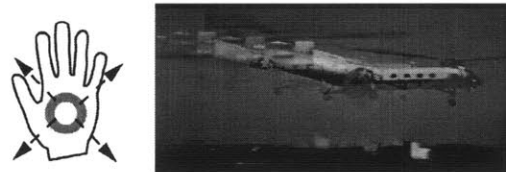
6.5.5 counterclockwise motion rewinds

### 6.5.4 Color Control

In a manner similar to temporally manipulating the clips, the colors can also be 'scratched'. In this mode, denoted by red, green, and blue rectangles, the video can be split into its three component color plates. When the right hand is still, the playback appears normal, but as the users moves around, the red, green and blue portions of the video signal become separated from each other in the plane proportional to the velocity of motion. Examples can be seen below.



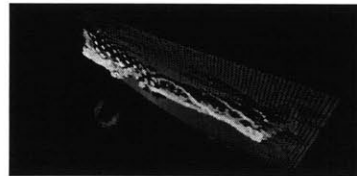
6.5.6 stillness results in relative calm of the colors



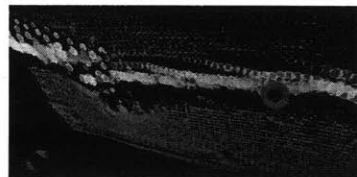
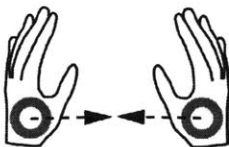
6.5.7 motion created dissonance in the color plates

### 6.5.5 Three Dimensional Navigation

The last three modes involve three dimensional depictions of the video and audio feeds, and are very similar to those in the Key Grip software. To control the orientation of the displayed objects in space, the metaphor of an invisible steering wheel or flight yoke was used. To back away or zoom out of the scene, the user separates her hands in space or pulls them away from the camera. In either case, the software will interpret them as diverging. To zoom in, the user will bring her hands closer together or towards the camera, as if depressing a throttle.

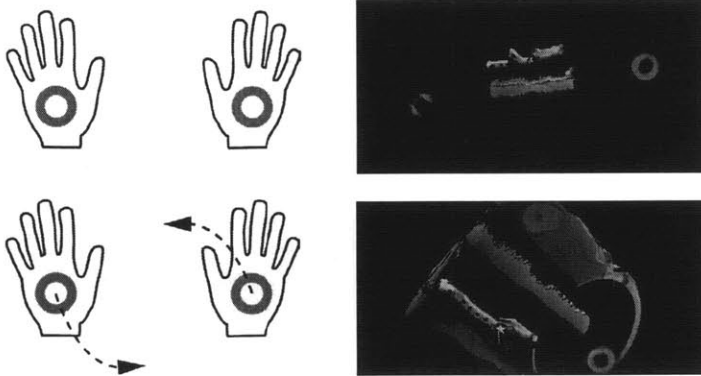


6.5.8 zooming out by pulling hands apart

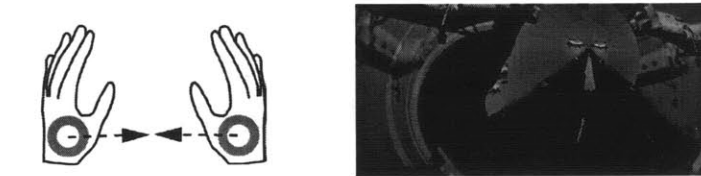


6.5.9 zooming in by bringing hands together

The orientation of the scene is determined by the relative rotations and positions of the hands. Rotation about the Z axis (the line normal to the display) is affected by rotating both hands around the center of the screen, as if turning a steering wheel. Rotations about the X and Y axes are the result of displacement of the centroid of received light of both hands along the Y and X axes respectively. Examples are shown below.



6.5.10 visuals are rotated by rotating hands around center of screen



6.5.11 and then zoomed by bringing hands closer

## 5.6 Cinema Fabriqu  World Premiere River Gods - April 13, 2003

I used the lightglove system to deliver an hour long solo performance at River Gods, a restaurant/bar in Cambridge. This event was the longest continuous set I had ever performed on my own, and I was controlling the audio and visuals the whole time. I performed from an elevated booth overlooking the crowd. Unfortunately I have no photo documentation of the show to share.

I began the set with a video feed of myself wearing the lit gloves, similar to the figure at the very beginning of this chapter. I slowly and deliberately explored the possibilities of each filter applied to the video feed of me and tried to communicate the nature of my control over the system. Then other movie clips and songs were introduced and a switch to a less pedantic style was made.

About a half hour into my set, Rory Keohane, who invited me to perform, gave me some advice to help the audience understand what I was doing. He told me to frequently switch back to the camera feed of myself to reinforce exactly how my hand motions were affecting the audio and visual outputs. He stressed repetitive actions that would drive home the nature of my control. The interface was brand new, and despite my goal to make it communicate its own meaning and function, I should not expect the audience to fill in any blanks that might be caused by flaky LED's, ambiguities in presentation, or performance mishaps.















5.6 *Picture of flyer for event*



## 7 Evaluation

This thesis presented in detail two hardware and software systems for delivering realtime audiovisual entertainment to live audiences. Key Grip and Cinema Fabriqué succeeded, in varying degrees, to satisfy the conditions and goals of this thesis. Their satisfaction of the stated goals and important but unforeseen issues are now detailed.

The discussion will focus on four systems and their relative strengths and weaknesses. The extent to which three variants of the Key Grip family and Cinema Fabriqué managed to fulfill the goals of this thesis is explored. A chart that graphically portrays their merits can be seen below.

system name	picture	weight	setup time	audience comprehension
Key Grip I wired		<b>85 lbs</b>		
Key Grip II wireless control		<b>12 lbs</b>		
Key Grip III wireless camera wireless control		<b>30 lbs</b>		
Cinema Fabriqué		<b>15 lbs</b>		

7.1 Chart of different performance systems' relative merits

## 7.1 Successes

*Portability* - If the extravagant projection systems showcased at the Enormous Room are not taken into account, the amount of hardware necessary to deliver one of my recent performances totals under twenty pounds and can fit into a large backpack. No roadies are needed to help transport the equipment and it can even be taken as carry-on to a flight.

*Fast setup time* - Again assuming I do not need to bring speakers or hang projectors, the hardware and software can be loaded and prepared in under ten minutes if properly tested before arriving to a show.

*Communication of Intentionality* - By exposing the Cinema Fabriqu  interface to the viewers, it was made clear 'what did he do' and 'what did the computer do'. The visual feedback loop of performer input triggering instrument response facilitated audience engagement.

*Audience Participation* - Footage of the audience could be integrated into the performance in both systems. Audience members could also be given the controls of the Key Grip to experiment, which was often more enjoyable for them than watching my performance.

*Use as post production device* - I made several posters and movies by recording the output of the Key Grip system. The system allowed real time manipulation of complex scenes in a smooth, visceral manner. The method of making movies was similar in practice to actors doing several takes to make a scene for a motion picture or television show. It might take me ten tries to create a satisfactory five minute clip, but it is a much faster process than meticulously keyframing dozens of variables in Final Cut or Maya. The final products also communicated a sense of rawness and human touch that would be very time consuming to replicate in other software.

*Speed* - Developing in C++ at the level of graphical primitives and video buffers allows a tremendous amount of flexibility when presenting media in realtime. Ten movies can be played simultaneously or cycled through at a full thirty frames per second.



Features can be extracted from one live camera feed while the user's hands are tracked on another using inexpensive computers. This level of fluid interaction is still not possible with most commercially available systems.

## 7.2 Shortcomings

*Stylistic restrictions* - The visual output of both the Key Grip and Cinema Fabriqu  systems look exactly like themselves. By that, I mean that my style definitely favors three dimensional video manipulations floating in a empty black void, and users are limited to the small set of filters I have loaded at any given time. I have barely explored softer, gentler, visual styles.

*Limited Audio Capabilities* - The audio functionality of the systems is limited to the playback of soundtracks that accompany movie files or MP3s. The audio can be chopped up, looped, and scratched just as the video, but there are no filters or synthesis tools analogous to those available for video signals.

*Lack of haptic feedback* - One of the main problems with video gesture control systems is the absence of physical feedback from the system to communicate its state through touch. Cinema Fabriqu  users cannot take their eyes off the display if they wish to maintain tight control over the output. Even slight drifting of a hand can cause the media clip or effect mode to change.

### **7.3 Video, Technology and Performance Conference Brown University April 6-8, 2003**

Todd Winkler of Brown University organized the Video, Technology, and Performance conference in Providence, RI to bring together artists, musicians, and theatre performers who use computers in their work. He invited me to present my projects at a roundtable discussion that was to conclude the weekend of dance performances, lectures, and demos.

Paul Kaiser, who works largely with motion capture technology was clearly not impressed with my work. He referred to my Key Grip project as a “media spigot”, and equated it to “garbage on TV”. His main complaint was the overwhelming variety of clips and sounds available, and that my demonstration was not choreographed in any way. To my credit he disliked everyone else’s work too.

For the most part the Cinema Fabriqu e and Key Grip projects received many compliments and a few people asked me to collaborate with them on future projects. Several people commented that my systems will “make a lot of money”. That remark is actually quite common from people that see my performances. A lot of the attendees complained about theatre and dancing not being lucrative ventures, and how bringing technology into the loop often removed any possible profit margin.

My research has never been monetary motivated. Whether or not my ideas are commercially viable makes no difference to me. I simply wish to entertain people in unique ways. And this I believe I have done. I took what I believed to be the best parts of TV, DJing, and video games, and combined them into a single interface that was accessible and interesting.

## **7.4 Distinctions Received**

I have submitted the Key Grip and Cinema Fabriqu  systems as well as prints and movies created with them to several juried competitions. These are two distinctions I received.

### **7.4.1 PRINT Magazine**

I submitted a portfolio of my projects to PRINT Magazine's Annual New Visual Artist Review. A large portion of the portfolio was documentation of the Key Grip project and music videos and posters created using the system. On the basis of the submitted works I was chosen as one of the twenty up and coming designers under the age of thirty in 2003. Examples of my work and an interview are in the March/April 2003 issue of PRINT.

### **7.4.2 Prix Ars Electronica**

I also sent a short video of the Cinema Fabriqu  project as a submission to the Prix Ars competition for 2003 and received an Honorable Mention in the Interactive Art category. I was also invited by the organizers of the Ars Electronica Festival to be an artist in residence at the Future Lab in Linz, Austria for the summer. I will design an installation form of the project for the Ars museum there and plan a performance to be given to a seated audience of a few hundred people in an opera house during the festival in September.

## **7.5 Future Work**

My systems presented in this thesis are still extraordinary primitive in comparison to most of those examined in the background chapter. Having the gig at the Ars festival will motivate me over the summer to address all of the issues that I have discovered during my research and hopefully discover and deal with many more shortcomings that arise. I look forward to incorporate foot pedals into the Cinema Fabriqu  system to make the system a full body experience and to affect binary changes that are difficult to express and detect through hand gestures.



## 8 References

- [Allport96] Allport, D. Navigating in 3D using the "Flying Fish". *Proceeding of Gesture Workshop '96*. Springer-Verlag. New York. 1996.
- [Armstrong95] Armstrong, David. *Gesture and the Nature of Language*. Cambridge University Press, 1995.
- [Cohen99] Cohen, P.R., McGee, D., Oviatt, S.L., Wu, L., Clow, J., King, R., Julier, S., Rosenblum, L. "Multi modal Interaction for 2D and 3D Environments," *IEEE Computer Graphics and Application*, 1999.
- [Corradini02] Corradini, Andrea. "Multimodal Speech-Gesture Interface for Handfree Painting on a Virtual Paper Using Partial Recurrent Neural Networks as Gesture Recognizer" *Proceedings of the International Joint Conference on Artificial Neural Networks*, Vol III. 2002.
- [Chomsky86] Chomsky, Noam. *Knowledge of Language: its nature, origin, and use*. Praeger, New York, 1986.
- [Cycling74] Cycling 74 company website, makers of MAX/MSP. <http://www.cycling74.com>. 2002.
- [Dancyger02] Dancyger, Ken. *The Technique of Film and Video Editing : Theory and Practice*. Focal Press, Boston, Mass., 2002.
- [Freeman98] Freeman, William et al. "Computer Vision for Interactive Computer Graphics," *IEEE Computer Graphics*, May/June 1998, p. 42-52.
- [Gaver95] Gaver, W. W. Oh what a tangled web we weave: Metaphor and mapping in graphical interfaces. *CHI 95 Human Factors in Computing Systems*, 270-271, ACM press, Denver, 1995.
- [Harling97] Harling, Philip A., Edwards, Alistair D. N. Progress in Gestural Interaction. *Proceedings of Gesture Workshop '96*. Springer-Verlag, London, 1997.
- [Horn86] Horn, Berthold Klaus Paul. *Robot Vision*, MIT Press, Cambridge, Mass., 1986.
- [Kent03] Kent, Justin. EJ company website. <http://www.ejenterprises.tv>. 2003.
- [Kurtenbach90] G. Kurtenbach and E. A. Hulteen. "Gestures in Human-Computer Interaction." *The Art of Human-Computer Interaction*. p. 310. Reading, MA. Addison-Wesley. 1990.

- [NATO03] NATO.0+55+3d software website.  
*http://www.eusocial.com/*. 2002.
- [Palindrome03] Palindrome website. Audiovisual performers.  
*http://www.palindroe.de*.
- [Rokeby02] Rokeby, David. Personal Website : *http://www.interlog.com/~drokeby/home.html*. 2002.
- [Smith02] Smith, Laurel. *Development of an Improved Swept RF Tagging System and its Musical Applications*. MIT Master's Thesis. 2002.
- [Souto82] Souto, Mario. *The Technique of the Motion Picture Camera*: p 151. Focal Press, London.1982.
- [Toulet95] Toulet, Emmanuelle. *Birth of the Motion Picture*: 17, 19. Abrahms, Inc. Publishers. New York. 1995.
- [Meeryman99] Meeryman, Richard. "A Painter of Angels Became the Father of Camouflage," *Smithsonian Magazine*, April 1999. p118.
- [Vo93] Vo, Minh Tue, & Waibel, Alex. "A Multi-Modal Human-Computer Interface: Combination of Gesture and Speech Recognition". *InterCHI Proceedings*, Amsterdam. 1993.
- [Wiener65] Wiener, Norbert. *Cybernetics*. MIT Press, Boston, 1965.
- [Winkler03] Winkler, Todd. Personal website. *http://www.brown.edu/Departments/Music/faculty/winkler*. 2003.
- [Zeki92] Zeki, S. The Visual Image in Mind and Brain. *Scientific American*. 267(3): 68-76. 1992.
- [Zettl90] Zettl, Herbert. *Applied Media Aesthetics*. Wadsworth Publishing, Belmont, CA, 1990.