DirectIVE: Choreographing Media for Interactive Virtual Environments

by
Flavia Sparacino

Diplôme d'Etudes Avancés en Sciences Cognitives
Université de Paris VI, France, 1994

Diplôme d'Ingenieur
Ecole Centrale Paris, France, 1993

Laurea in Ingegneria Elettronica
Politecnico di Milano, Italy, 1993

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

February 1997

Author

Program in Media Arts and Sciences
October 4th, 1996

Certified by

Alex Pentland
Associate Professor of Computers, Communication and Design Technology
Program in Media Arts and Sciences

Accepted by

Steven A. Benton
Chair
Departmental Committee on Graduate Students
Program in Media Arts and Sciences

DirectIVE: Choreographing Media for Interactive Virtual Environments

by Flavia Sparacino

Submitted to the Program in Media Arts and Sciences
School of Architecture and Planning
on October 4th, 1996
in partial fullfillment of the requirements for the degree of
Master of Science in Media Arts and Sciences
at the Massachusetts Institute of Technology

Abstract

This work describes methods for content choreography for performance and entertainment spaces, using remote sensing technology for interaction, so that the user/performer is not encumbered with wires or sensors. Artificial Life programming methods are used to avoid rigid scripting of user and content interaction. This results in the design of active content endowed with intentionality and autonomous behavior (media creatures). By following this approach I created an Improvisational Theatre Space where autonomous media creatures react to the user's movements, speech etc., to create an improvisational dialog between the user/performer and the performance/entertaintment space's content (images, video, sound, speech, text). This space is the last of a series of experiments in the construction of Interactive Virtual Environmnents (IVEs) that include also DanceSpace and NetSpace. Each of these experiments aims at exploring a specific aspect among those that are essential to create an immersive experience.
This research contributes to the field of digital graphic design and content orchestration for virtual reality applications by offering new tools for designing interactive immersive spaces.

Thesis Advisor:
Alex Pentland
Associate Professor of Computers, Communication and Design Technology

DirectIVE: Choreographing Media for Interactive Virtual Environments

by Flavia Sparacino

The following people served as readers for this thesis:

Reader:

Glorianna Davenport
Principal Research Associate
MIT Media Laboratory

Reader:

Michael Hawley
Assistant Professor of Media Arts and Sciences
Program in Media Arts and Sciences

Reader:

Patrick Purcell
Visiting Professor of Human Computer Interaction,
Imperial College, London

Acknowledgements

I would like to thank my advisor Sandy Pentland for having encouraged me to pursue this research even when at an early stage it looked vague and too large to become a Master's Thesis.

I am also very grateful to my readers: Glorianna Davenport, Mike Hawley and Patrick Purcell for having been so generous with good advice and support since I arrived to the Lab, and for their infinite patience in supervising the writing of this document.

Special thanks to Bruce Blumberg who has coached me along the path to learning how the (virtual) dog's brain works, and to Chris Wren, Ali Azarbayejani and Erik Trimble who helped with all the technical difficulties associated to this project.

I have also been fortunate to interact with Suguru Ishizaki: his stimulating conversation and suggestions provided a useful guidance when this project started.

Thanks to all pentlandians past and present: Ali, Baback, Chris, Dave, Deb, Erik, Irfan, Jeff, Ken, Nuria, Stan, Sumit, Thad, Trevor, and Wasi for sharing their good humor in the late working hours.

Special thanks to Dave Becker for having given us such a great proof of courage and wisdom of life.

Special thanks also to my UROPs Chloe Chao and Michal Hlavac who have worked with competence and enthusiasm on some important aspects of this research.

Special thanks to my office-mate Baback Moghaddam who has contributed to this research with some excellent music and movie selections that have added fun along the path.

Thanks to our wonderful assistants Karen Navarro and Laurie Ward and to all vismoders. In particular Tom Minka and Martin Szummer have been a continuous source of inspiration in my long path to becoming a "MIT hacker".

Thanks also to my friends in this part of the world: Martin, Daniele, Isabella, Erik, Obed, and Rob who have put up with my late working hours and busy week ends during these past two years at the Media Lab.

Thanks to Margaret and Perry Kipp who have been my dedicated host family in the United States.

And my special gratitude goes to my parents for their endless support.

Contents

1. Introduction

Whereas film is used to show a reality to an audience, cyberspace is used to give a virtual body, and a role, to everyone in the audience. Print and radio tell; stage and film show; cyberspace embodies ... Whereas the playwright and the filmmaker both try to communicate the idea of an experience, the spacemaker tries to communicate the experience itself. A spacemaker sets up a world for an audience to act directly within, and not just so the audience can imagine they are experiencing an interesting reality, but so they can experience it directly ... Thus the spacemaker can never hope to communicate a particular reality, but only to set up opportunities for certain kinds of realities to emerge. The filmmaker says, "Look, I'll show you." The spacemaker says, "Here, I'll help you discover."

-Randall Walser, 1990
Elements of a Cyberspace Playhouse

1.1. Motivation

To create an engaging experience for the user "*you need to think about intelligent animation - characters who know what they look like and how they move*" [Laurel, 1989]. Recently, in the field of Computer Graphics, progress has been made in the creation of life-like characters or autonomous agents. These characters are driven by autonomous goals, can sense the environment through real or virtual sensors and respond to the user's input or to environmental changes by modifying their behavior according to their goals. While this approach, called behavior-based, has proven to be successful for a variety of computer graphics problems it has not been yet fully exploited for content presentation or digital design.

Content presentation, on the other hand, should not be limited to the construction of environments where nothing happens until the user "clicks on the right spot". In most virtual reality applications visual elements, objects or characters always appear in the same position on the screen and their appearance is triggered by the same reactive event-based action. The static nature of these worlds gives them an alienating and surreal look and induces the user to have an exploratory type of behavior. *Navigation* is, as a matter of fact, the term that better describes the user's main activity while she is interacting with a current virtual reality application.

This thesis aims at establishing some criteria for designing immersive spaces by a series of progressive experiments that combine behavior-based interactive design with an unencumbered man-machine interface. The research carried out merges the worlds of autonomous dynamic graphics and improvisational performance.

1.2. Problem Statement

Behavior-based design endows design elements - like digital text, photographs, movie clips or graphics - with intentionality and autonomous goals. My contribution to this field is that of creating media objects *with a sense of themselves* that I call *media creatures*. Media creatures are autonomous agents with goals, behaviors and sensors. A media creature knows whether its content is text, image, a movie clip, sound or graphics and acts accordingly. It also has a notion of its role and *personality attributes.*

This design approach implies that there is no separation between content and choreography of content. Each design element or media creature embeds some basic choreographic strategies that form its repertoire of actions to interpret content. These choreographic strategies correspond to a variety of behaviors that map sensorial information to design attributes of the media object, like size, form, color or typeface.

Traditional digital content presentation uses passive content and a separate program that coordinates the presentation based on the user's input. The model is analog to that of having an orchestra director who conducts a number of musicians following a given score. This leaves very little room for interactively and the programmer of the virtual reality experience needs to create break points in the 'score' (plot) - somewhat artificially - for the user to be able to participate.
Behavior-based design adopts instead the '*jam session* model, where a number of musicians, each one with its own musical personality and instrument, meet to create a musical experience with no previous program or score. Musicians share a common ground of musical repertoires and techniques that contribute to the creation of harmony and coordination among the players. The improvisational nature of the experience leaves space for interactivity with the public. A similar metaphor could be used to compare traditional theatre with improvisational street theatre or the circus.

In this type of environment, the metaphor for the interaction between the user and the virtual world is not that of an *exploration* but that of an *encounter* with media objects (creatures). By encounter I mean a two way movement: one by the user in search of interesting things to see and the other by the media creatures looking for someone who is interested in their story or performance.

1.3. Research Overview

Using this approach I created three immersive interactive spaces for performance and entertainment. These spaces evolve from a simple responsive environment to a more complex environment that has active content and a richer interface.

DanceSpace is an interactive performance space where dancers can generate music and graphics through their body movements. It is a responsive environment where content (sound and graphics) is seamlessly generated by the performer's movement in space.

NetSpace is an interactive World Wide Web browser in which text and images form a three dimensional landscape that the user builds dynamically through exploration and can then revisit. Here text and images are reactive design elements that respond to the user's gestures and voice commands.

The Improvisational TheatreSpace is an interactive stage on which human actors perform accompanied by virtual actors made of text, images, sound, movie clips, speech or graphics. Digital content is active and autonomous responding to the performer's gestures, speech and tone of voice. This system is an example of an interactive virtual environment which is both engaging and immersive. It has been tested both as a performance space, where the user is a mime/performer and as an entertainment space, where the user is a non-performer. In both cases the user/performer and the active content are involved in an improvisational situation in which they are singing or acting following a loose script.

Each environment uses remote sensing technology to interpret the actions of the user/performer. This work is premised on the assumption that immersive spaces require natural, wireless interaction so that the user/performer is not encumbered with wires or sensors. Most current virtual environments use bulky head-mounted displays, data gloves and body suits with multiple cables. While such systems can be extremely accurate they limit severely the freedom of movement of the user. This results in an unnatural man-machine interaction and it is the first obstacle to creating a truly engaging environment. The immersive environments described in this document use sensing devices which are embedded in the environment, endowing the latter with perceptual intelligence. As people use primarily vision and audition to communicate with each other, these interactive spaces use real time computer vision and audition technology as their source of perceptual information.

In the following chapters I will explain my approach to designing immersive Interactive Virtual Environments (IVEs) and I will describe each of the spaces mentioned above. I will evaluate my work through the feedback of the many people that have experienced these spaces. I will conclude with a section on related works, a section on future developments and some final remarks.

2. Approach

2.1. Inter+Active Design

Richard Buchanan identifies one area of design as that of "the design of complex systems or environments for living, working, playing and learning" [Buchanan, 1995, pg 8]. In this work I have explicitly treated the goal of creating engaging and immersive Interactive Virtual Environments (IVEs) as that of articulating a design problem and discovering solutions as design principles to build such environments.
The relationship between *signs* and *actions* is treated by a "double repositioning of the design problem" [ibid], once from the point of view of the participant and again from the content presentation perspective. On the one hand, *from action to sign*, the participant, user or performer, communicates with the system through body gestures or voice commands that are interpreted by the interface and affect the signs, i.e. the visual (or sound) elements of the presentation. On the other hand, *from signs to actions,* the visual elements take actions to choreograph content presentation by virtue of their behavior-based autonomous structure (figure 1).

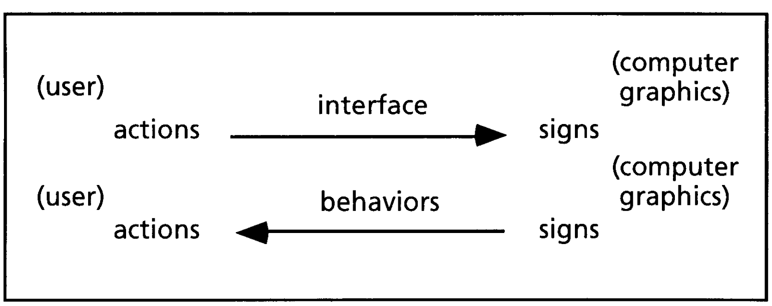


Figure 1: double repositioning of the design problem

My approach to digital (graphic) design is that of taking traditional - passive - visual elements and provide them with their own presentation capabilities and behavioral attributes. This is why I use the term "Inter+Active Design". The inter+active design elements that I created are composed of three layers (figures 2A and 2B):

1. the actual information: text, images, video, sound, speech, graphics.
2. the interface, that defines how the element responds to the user's input.
3. the choreographic component, that is responsible for the autonomous behavior and for choosing the most appropriate choreographic strategy for that element in the given situation.

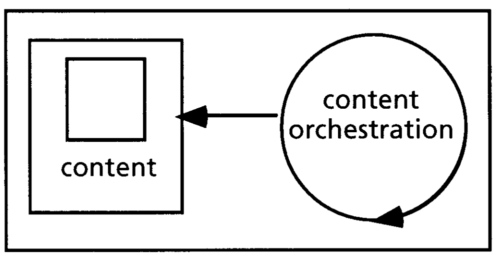


Figure 2A: passive content

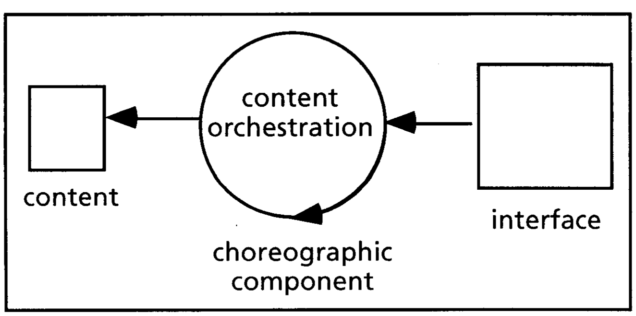


Figure 2B: active content

In this context I define as *interface* a mapping between the space formed by all possible values given by sensory input and the space of all actions that manipulate content (figure 3).

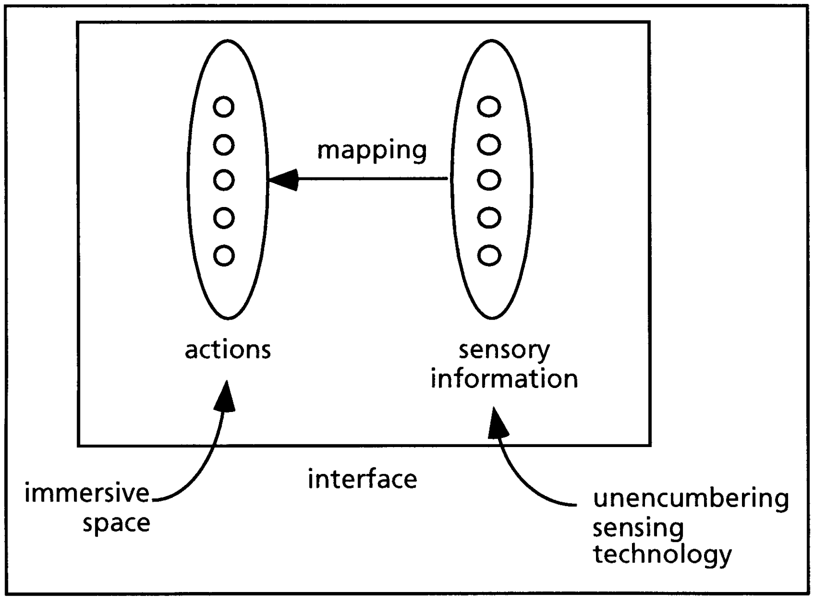


Figure 3: definition of interface

This research was carried out as a series of progressive experiments in building immersive spaces, each one exploring a different interaction and content presentation modality. Inter+active design elements evolve through these experiments from a simple Content + Interface structure to a more complex Content + Behavior (Behavior = Interface + Choreographic Component) structure.

In DanceSpace, content consists of a trail of colored virtual shadows that follow the dancer moving in the space and create colored patterns on the large display screen that forms the backdrop of the interactive stage. In addition music is generated by the dancer's movements in space. The trail of virtual shadows constitutes an interactive design element that incorporates the interface as part of its structure. The interface is part of the element because the shape of the graphical shadows changes in real time according to the dancer's movements. There is a direct mapping between sensory information retrieved from the computer vision system and the graphical output of the application. This is a particular case for which the actions that manipulate content coincide with the content itself.

NetSpace is an entertainment space where people use full body gestures to browse through a virtual city made by web pages. In this application content consists of text and pictures that compose the web pages and it is made responsive to browsing gestures. In this case each web page is a design element that incorporates an interface that has a gesture interpretation module. This module acts as a filter between sensory input and system output (figure 4). In NetSpace content orchestration is still done by the program and it is not included as part of the design element itself (figure 5).

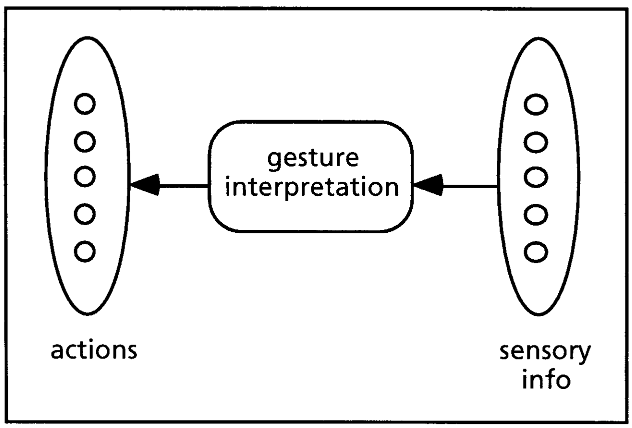


Figure 4: interface with an interpretation module

The Improvisational Theatre Space is an interactive stage in which human actors perform and are accompanied by virtual actors. In this application virtual actors belong to a specific category of digital performers that I call *media creatures.* Media creatures constitute an example of active content that choreographs its own presentation to the user. Media Creatures are inter+active design elements composed of content, interface and a directive component (figure 6). Content can be graphics, text, photographs, movie clips or sound. The choreographic component is responsible for the autonomous behavior of the creature. It also allows for directability by the user when she needs to issue specific commands.

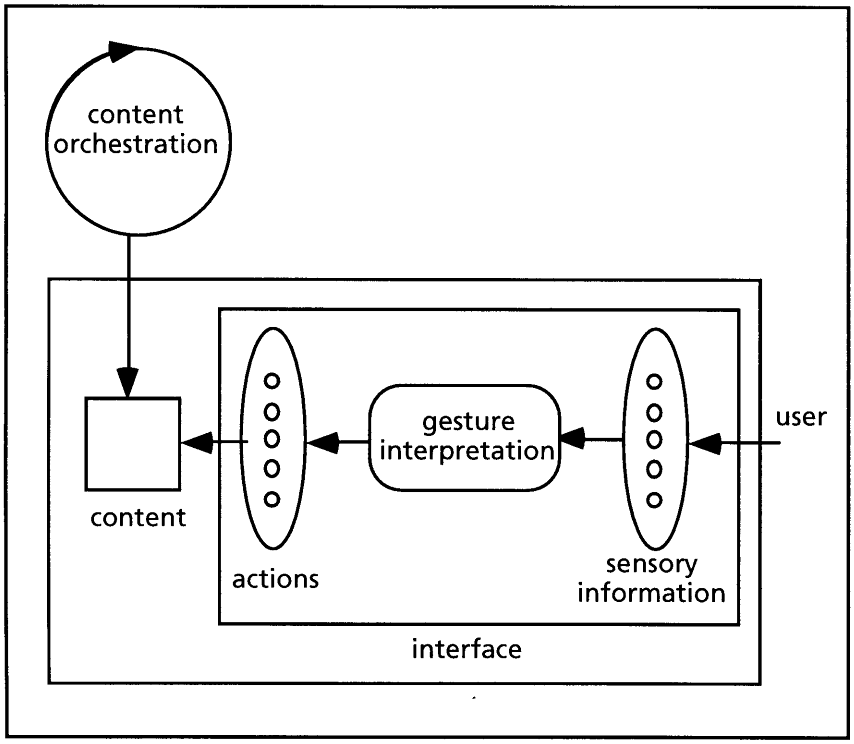


Figure 5: Content Orchestration for NetSpace

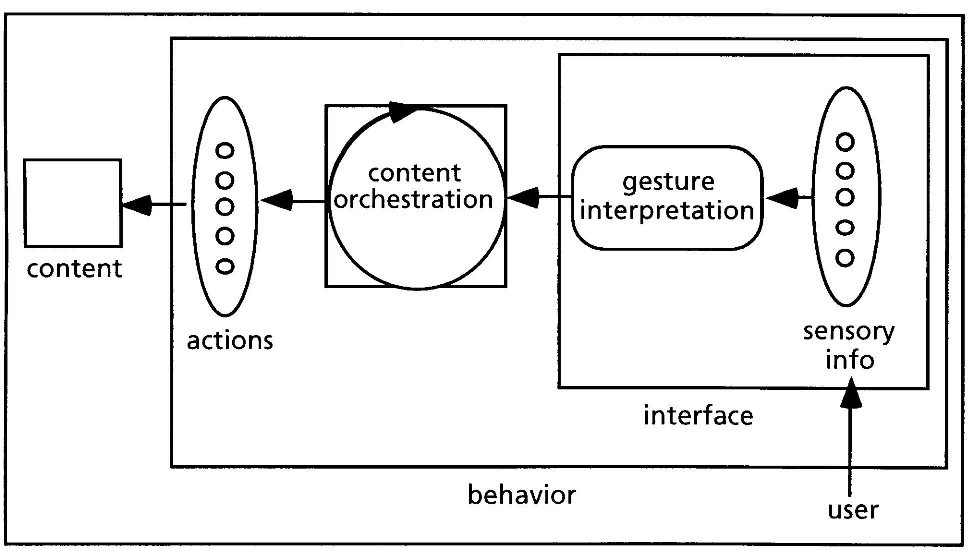


Figure 6: architecture of Media Creatures

2.2. Behavior-Based Design for Choreography of Content

2.2.1. Autonomous Agents

One of the main areas of research in Artificial Life today is concerned with modeling autonomous agents to solve problems for which traditional AI has produced poor or insufficient results. Autonomous Agents are a new field of Artificial Intelligence which is highly inspired by biology and in particular by ethology.

Research in Autonomous Agents is synonymous with behavior-based AI - as opposed to mainstream knowledge-based AI -, with bottom-up AI - versus top-down AI -, and also with "animat" approach. The autonomous agent or behavior-based approach "is appropriate for the class of problems that require a system to autonomously fulfill several goals in a dynamic, unpredictable environment" [Maes, 1994].

Autonomous agents are therefore software systems with a set of time-dependent goals or motivations which they try to satisfy in a complex dynamic environment [ibid]. An agent is autonomous in the sense that it has mechanisms for sensing and interacting in its environment as well as for deciding what actions to take so as to best achieve its goals.

Given my objective of creating active content for interactive performance and entertainment spaces, I used a behavior-based approach to build inter+active design elements and gave them an autonomous agent architecture. The choice of using a behavior-based rather than a knowledge-based approach was determined by the following considerations:

1) in the case of interactive spaces for performance and entertainment, the system is continuously observing the user/performer and changes the presentation according to the user's input. The environment is not only taken into account dynamically, but its characteristics are exploited

to serve the functioning of the system. The domain knowledge of the system is constantly changing and therefore the traditional AI approach, that was developed for closed environments, would not be as efficient as the autonomous agent approach.

2) Traditional AI has developed systems that provide depth rather than width in their competence. In order to model active content I need instead to be able to assign to each inter+active design element a multiplicity of low-level competencies and goals that range from following the user around the space to changing color, font, or showing off, or suggesting to the user what to do next.

3) In addition traditional AI focuses on modeling knowledge for the system as a whole. AI systems have declarative knowledge structures that model aspects of the domain expertise. In contrast, inter+active design elements do not have a world model, nor any information about the world except that which is available via their sensors. Autonomous Agents are completely self-contained, monitor the environment to figure out by themselves what to do next and are able to deal with many conflicting goals simultaneously.

The agent architecture I used for my research is the one proposed and successfully implemented by Blumberg [Blumberg, 1995]. Blumberg's approach is heavily based on ethology [Blumberg, 1994] and uses an action-selection algorithm for animats that draws from computational models of animal behavior. Hence the name *creature* to designate autonomous agents that are animated software objects capable of goal-directed and time-varying behavior and that follow an ethological-based action-selection mechanism [ibid].

This approach was conceived to meet the need to integrate autonomy with *directability*. Directability is a quality of an agent that is not only capable of autonomous action but that can also respond to external control. It allows the user to co-direct, together with the system, an autonomous creature at the motivational level, the task level and the direct motor level.

The structure of a creature consists of three parts - Geometry, Motor Skills and Behavior - with two layers of abstraction between these parts -

Controller and Degrees of Freedom - [Blumberg, 1995]. The Geometry provides the shapes and transforms that are manipulated over time for animation. The Motor Skills provide atomic motion elements which manipulate the geometry in order to produce coordinate motion. Behaviors implement high level capabilities such as: 'present oneself to user' or 'follow the user around' as well as low level capabilities such as 'move to' or 'change size' by issuing the appropriate motor command.

2.2.2. Media Creatures

Media Creatures are autonomous agents with behaviors, sensors and goals (internal motivations). A media creature knows whether its content is text, image, a movie clip, sound or graphics and acts accordingly. It also has a notion of its role and "mood", i.e. it can express basic emotions like happy, sad, angry, scared etc.
Media creatures are modeled according to Blumberg's tool kit for behavior-based animation of autonomous creatures [Blumberg, 1995]. The choreographic component of a media creature specifies its behavior according to the context of the interaction with the performer. As opposed to scripted animation that imposes a pre-defined sequence of actions to a virtual actor, behavior-based animation, according to Blumberg, defines a tree of actions that are driven either by the internal motivation of the creature or by the external actions of the user/performer or by a combination of both. Hence *the model of the interaction between the human and the digital actor consists of a nearly infinite tree of possibilities rather than a linear sequence of consecutive actions* (figure 7).

Behaviors are grouped into sets of mutually excluding behaviors and branch out into more granular and specific activities or low level behaviors as the tree grows in depth. For example, if a media creature wants to entertain it may follow two branches in the behavior tree, starting from the behavior *entertain* at the root of the tree (figure 8). The creature can decide to *say something* or to *wander around* according to what the user is doing.

In both cases the creature will have the choice among a group of mutually excluding behaviors. Only the leaf behaviors, that are at the bottom of the (inversed) tree can execute motor commands for the creature.
The choice of the behavior in which a creature is involved at a certain instant of time is done by an action-selection algorithm that implements a winner-take-all strategy on the tree of behaviors by computing the weighted mean of the value of the sensory information and the internal motivations.

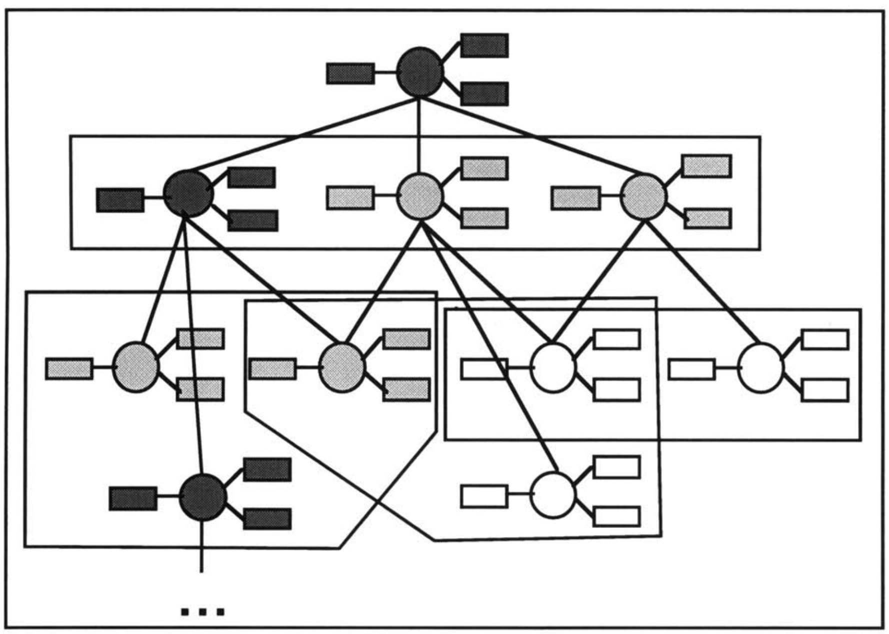


Figure 7: behavior tree

Although only the winning behavior is active at one time it is still possible for less important behaviors to issue suggestions in the form of secondary commands. This is the way in which the emotional state of the creature is expressed. Each of the emotions/moods of the creature will issue the appropriate secondary command for that particular state. These commands will affect the size/form, color and the type of movement of the creature. Moods are set by *internal variables* of the creature whereas emotions depend on things that the user does (gestures/tone of voice). Independently of their source, moods and emotions will map to the following basic states: happy, sad, angry, scared, surprised, disgusted, interested, loving, no-emotions.

Internal Variables are a particular case of autonomous agents' Internal Motivations. Although they do not explicitly favor the choice of a behavior according to the given motivation, they are weighted in the action-selection algorithm against the information coming from the sensors for the choice of the most appropriate behavior.

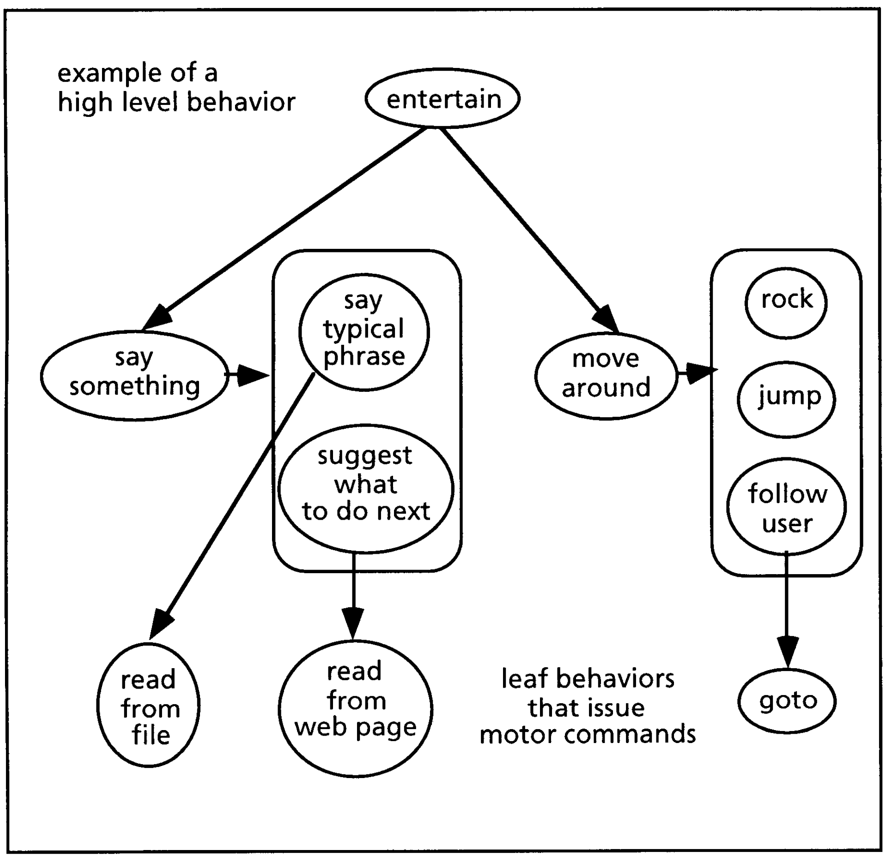


Figure 8: Example of a branch of the behavior tree of a Media Creature

Creatures can have both real and virtual sensors. Creatures use real sensors, like video cameras or microphones, to interact with the user. Virtual sensors are used for communicating among creatures. Virtual sensors can operate in

two ways: either directly via interrogation of other virtual creatures and objects or by use of synthetic vision. In the latter case the virtual scene is rendered from the creature's eye viewpoint and the resulting image is used to generate a potential field that allows the creature to do obstacle avoidance and low-level navigation [Blumberg, 1995]. Direct interrogation of other virtual creatures is based on the features of Open Inventor which is the three-dimensional graphical tool kit used for building autonomous creatures. Open Inventor is a library of objects and methods used for interactive 3D graphics. Direct query on the status of other creatures is possible due to Inventor's ability to provide named access to field variables at run time.
Most of the building blocks that compose a creature, including all of the components of the behavior system and the action-selection algorithm itself, are derived from "Inventor" classes. As a consequence most of the specific characteristics of a creature as well as its behavior system can be defined via a text file using the Inventor file format. This is important for fast prototyping to quickly adjust or tune the creature's behavior during the rehearsals between the human performer and the media creatures.

The information coming from the sensors is evaluated by Releasing Mechanisms [Blumberg, 1995] that identify objects and/or events - i.e. user's gestures - which are relevant to a behavior (stimuli). Releasing Mechanisms output a continuous value which depends on whether the stimuli was found, or on its distance and some measure of its quality. This will account for the creature displaying a specific behavior on the basis of a weak stimulus with a strong motivation or vice versa.

One of the main advantages of using creatures in interactive immersive environments is that each creature can have their own specific interface. This means that the same action of the user/performer can trigger different behaviors in different creatures. Moreover, if the context of the interaction between the human and the digital actor changes the only thing that the programmer of the interactive experience needs to change is the weight of the value of the high level behaviors.

As a consequence of the interaction with the user, creatures can modify their behavior at three levels: the motivational level, the task level and the direct motor level [Blumberg, 1995]. At the highest level the creature would be influenced by changing its current motivation, whereas at the lowest level a creature receives a command that directly changes its geometry (figure 9). By having access to multiple levels of directability to influence the behavior of the creature, the user/performer can choose the appropriate level of control for a given interactive situation.

In the following part of this document I will use the term *media creatures* or *media actors* or *media tellers* as synonymous, according to the particular context in which they are interacting with the user/performer.

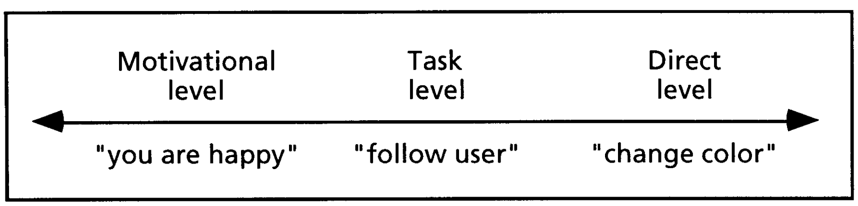


Figure 9: levels of directability of a creature

2.2.3. DirectIVE and ALIVE

ALIVE, an acronym for Artificial Life Interactive Video Environment, is a system which allows wireless full-body interaction between a human participant and a graphical world inhabited by autonomous agents. It has been developed at the Media Lab during the past three years by joint efforts of the Vision and Modeling Group and of the Agents group. ALIVE provided the ideal platform to develop my research project both in terms of the human-machine interaction and of the autonomous agent modeling.

ALIVE is an interactive virtual environment that allows the user to interact in natural and believable ways with autonomous and intelligent agents. These agents include a virtual dog named Silas (figure 10), a puppet named Homi, a hamster, a predator etc. In ALIVE the user experiences the environment through a magic mirror. The interactive space mirrors the real space on the other side of a projection display, and augments that reflected reality with the graphical representation of the agents and their world (including a water dish, partitions and even a fire hydrant). In ALIVE the user communicates with the creatures of the augmented world by pointing gestures, voice commands, and by moving around the space so as to intersect the path of a creature. ALIVE was developed at the Media Lab as a collaborative project between the Vision and Modeling Group and the Agents Group [Darrell, 1994]. Computer Vision techniques [Wren, 95] are used to separate the user's image from the background and to recover the 3D location of the user's head, hand and feet. The creatures in ALIVE respond to the participant in real time on the basis of this information and of their own internal motivations.

Several thousand people have experienced the ALIVE system at SIGGRAPH installations, at a temporary installation at the City Art Museum of Nagoya, Japan and at the Media Lab. The system has proved to be robust and an enjoyable and novel experience for the participants.

Following a tradition that attaches the postfix 'IVE' (Interactive Video Environments) to all projects that are developed in the same setting as ALIVE, i.e. wireless and natural interaction with autonomous creatures on a large space in front of a big screen, I decided to name collectively the interactive space that I developed DirectIVE. The reason behind this choice is that in my systems the user/performer gives directorial suggestions that drive the presentation of the performance or entertainment space.

In DanceSpace the performer drives the musical and graphical output through its body movement.

In NetSpace the user issues commands directly to build and explore interactively a 3D webscape.
In the Improvisational TheatreSpace the user or performer becomes also the director of the experience - like an improvisational actor - because through interaction with the media actors she drives the performance in the direction that she as an artist or the audience in that context requires.

Figure 10: interacting with a virtual dog

3. Building Performance and Entertainment Spaces

3.1. The Unencumbered Interface

While many advances have been made in creating interactive worlds, techniques for human interaction with these worlds lag behind. The ability to enter the virtual environment just by stepping into the sensing area is very important. The users do not have to spend time "suiting up," cleaning the apparatus, or untangling wires. Furthermore, social context is often important when using a virtual environment, whether it be for game playing or designing aircraft. In a head mounted display and glove environment, it is very difficult for a bystander to participate in the environment or offer advice on how to use the environment. With unencumbered interfaces, not only can the user see and hear a bystander, but the bystander can easily take the user's place for a few seconds to illustrate functionality or refine the work that the original user was creating.

Figure 11 demonstrate the basic components of an Interactive Space that occupies an entire room. In this space - called 'IVE' (Interactive Virtual Environment) - the user interacts with the virtual environment in a room sized area (15'x17') whose only requirements are good, constant lighting and an unmoving background. A large projection screen (7'x10') allows the user to see the virtual environment, and a downward pointing wide-angle video camera mounted on top of the projection screen allows the system to track the user [Wren, 1995]. A phased array microphone [Casey, 1995] is mounted above the display screen. A narrow-angle camera mounted on a pan-tilt head is also available for fine visual sensing. One or more Silicon Graphics computers are used to monitor the input devices in real-time.

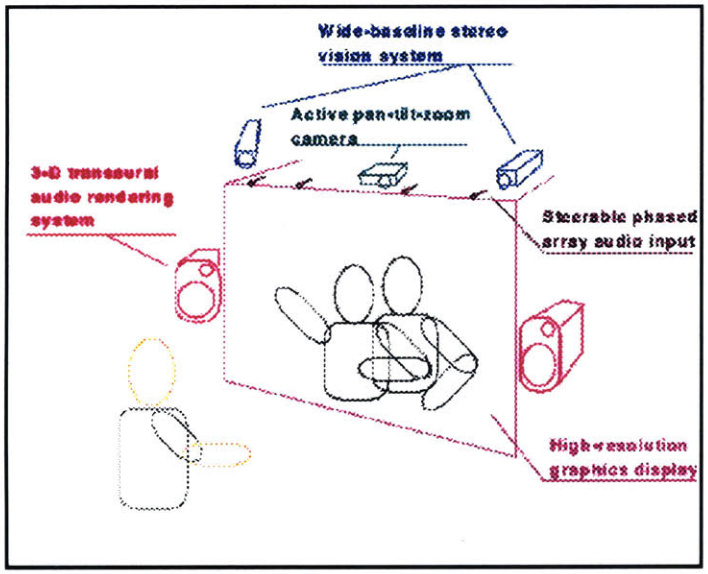


Figure 11: wireless sensors in the ALIVE space

3.1.1. Perceptive Spaces

The interactive environment interface is built to be entirely non-invasive. The use of a computer vision system to measure the user eliminates the need to harness the user with many sensors and wires. A large display format allows an immersive experience without the need for head-mounted displays and opens the environment up to multiple users.

The computer vision system is composed of several layers. The lowest layer uses adaptive models to segment the user from the background. This allows the system to track users without the need for chromakey backgrounds or special garments. The models also identify color segments within the users silhouette. This allows the system to track important features (hands) even when these features aren't discernible from the figure-ground segmentation. This added information may make it possible to deduce general 3D structure of the user allowing better gesture tracking at the next layer.

The next layer uses the information from segmentation and blob classification to identify interesting features: bounding box, head, hands, feet, and centroid. These features can be recognized by their characteristic impact on the silhouette (high edge curvature, occlusion) and (a priori) knowledge about people (heads are usually on top).

The highest layer then uses these features, combined with knowledge of the human body, to detect significant gestures. Audio processing included at the various levels allows the system to use knowledge of hum an dialog to better recognize both audio and visual gestures.

These gestures become the input to the behavioral systems of the agents in the virtual environment. This abstraction allows the environment to react to the user on a higher, more meaningful and inflected level. It can also allow to avoid the distracting lag inherent in many other immersive systems.

3.2. DanceSpace

DanceSpace is an interactive performance space where both professional and non-professional dancers can generate music and graphics through her body movements (figures 12A and 12B).

The music begins with a richly-textured melodic base tune which plays in the background for the duration of the performance. As the dancer enters the space, a number of virtual musical instruments are invisibly attached to her body. The dancer then uses her body movements to magically generate an improvisational theme above the background track.
The dancer has a cello in her right hand, vibes on her left hand, and bells and drums attached to her feet. The dancer's head works as the volume knob, bringing down the sound as they move closer to the ground. The distance from the dancer's hands to the ground is mapped to the pitch of the note played by the musical instruments attached to the hands.

Therefore a higher note will be played when the hands are above the performer's head and a lower note when they are near her waist. Both hands' musical instruments are played in a continuous mode (i.e., to get from a lower to a higher note the performer will have to play all the intermediate notes). The bells and the drums are on the contrary *one shot* musical instruments. When the performer raises her feet more than 15 inches off the ground then either of the bells/drums are triggered, according to which foot is raised.
The music that is generated varies widely among different users of the interactive space. Nevertheless all the music shares the same pleasant rhythm established by the underlying, ambient tune, and a style that ranges from "pentatonic" to "fusion" or "space" music.

As the dancer moves, her body leaves a multicolored trail across the large wall screen that comprises one side of the performance space.
The graphics are generated by drawing two bezier curves to abstractly represent the dancer's body. The first curve is drawn through coordinates representing the performer's left foot, head, and right foot. The second curve is drawn through coordinates representing her left hand, center of her body, and right hand. Small 3D spheres are also drawn to map onto hands, feet, head and center of the body of the performer. This serves as a reference for the dancer and accentuates the stylized representation of the body on the screen. The multicolored trail represents the dancer's virtual shadow which follows her around during the performance. The variable memory of the shadow allows the dancer to adjust the number of trails left by the dancer's body. Hence if the shadow has a long memory of trails (more than thirty) the dancer can paint more complex abstract figures on the screen.

The choreography of the piece varies according to which of the elements in the interactive space the choreographer decides to privilege. In one case the dancer might concentrate on generating the desired musical effect; in another case or in another moment of the performance, the dancer may concentrate on the graphics - i.e. painting with the body - finally the dancer may focus on the dance itself and let DanceSpace generate the accompanying graphics and music autonomously.

DanceSpace is inspired by Merce Cunningham's approach to dance and choreography [Klosty, 1975]. Cunningham believes that dance and movement should be designed independently of music; music should be subordinate to movement and may be composed after the principal choreography is finished just as a musical score is created after a feature film has been edited.

When concentrating on music, more than dance, DanceSpace can be thought of as a "hyperinstrument". Hyperinstruments [Machover, 1992] are musical instruments primarily invented for non-musical-educated people who nevertheless wish to express themselves through music. The computer that drives the instruments holds the basic layer of *musical knowledge* needed to generate a musical piece. DanceSpace can also become a tool for a dancer/mime to act as a street performer who has a number of musical instruments attached to her body. The advantage of DanceSpace over the latter is that the user is unencumbered and wireless; she can be therefore more expressive in other media as well (its own body or computer graphics).

A disadvantage of the current implementation of DanceSpace is that as a music improvisation system it does not easily reproduce well known musical tunes.

Future improvements to DanceSpace include having a number of different background tunes and instruments available for the dancer to use within the same performance. Another important addition will also allow the user to adjust the music's rhythm to her rhythm of movement. I would also like the color of the dancer's graphical shadow to match an expressive or emotional pattern in the dance and become an active element in the choreography of the piece.

I see DanceSpace as a possible installation for indoor public places, for example an airport, where people usually spend long hours waiting, or interactive museums and galleries. DanceSpace could also become part of a

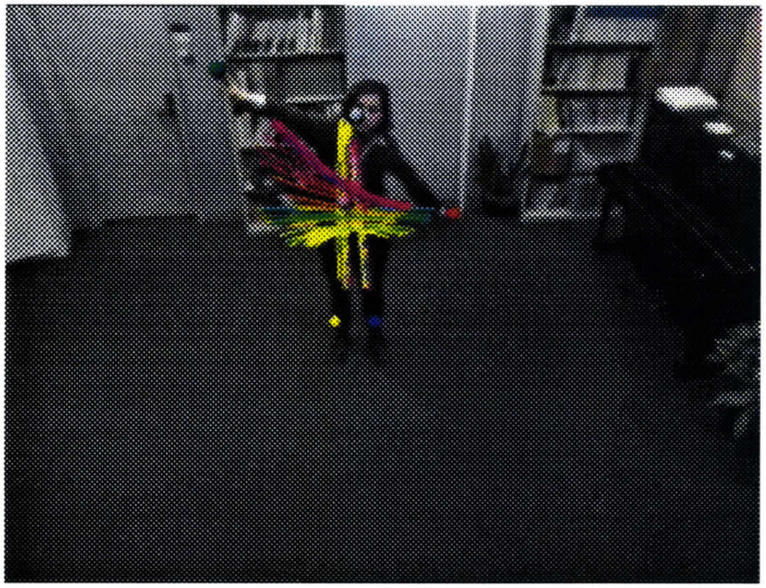


Figure12A: user interacting with her virtual shadow

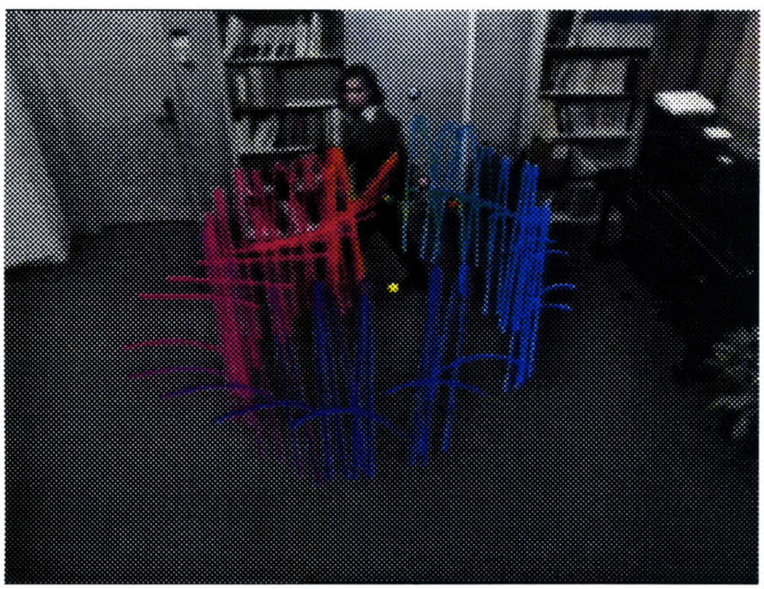


figure 12B: user generating a circle with her virtual shadow performance space (figure 13A and 13B), allowing a dancer to play with her own shadow and generate customized music for every performance.

Figures 13A and B: performers in DanceSpace

3.3. NetSpace

NetSpace is an immersive, interactive web browser that takes advantage of the human ability to remember 3D spatial layout.
For example, most people can easily remember where most of the hundreds of objects in their house are located. In comparison to our spatial memory, our ability to remember lists of information using spatial memory versus lists without, is greatly impoverished. NetSpace capitalizes on this ability by mapping the contents of URLs into a 3D graphical world projected on the large IVE screen. This gives the user a sense the URLs existing in a surrounding 3D environment.

NetSpace was conceived as a natural extension to Hyperplex, our first experiment using IVE as an immersive browser for movies [Sparacino, 1995]. To navigate this virtual 3D environment, users stand in front of the screen and use voice and hand gestures to explore. The user can scroll up and down a page by pointing up and down with either arm. "Previous" and "Forward" browser commands are implemented by pointing intuitively left and right. The user can zoom into a page by moving closer to the screen and can look at the overall landscape that she generated though exploration by moving away from the screen. Pointing with both hands will cause the browser to follow the currently active link. When a new page is loaded, the virtual camera of the 3D graphics world will automatically move to a new position in space that constitutes an ideal viewpoint for the current page.

The URLs are displayed so as to form an urban landscape of text and images (figures 14A,B,C,D,E) through which the user can "surf" [Wren, Sparacino et al, 1996]. Example of voice commands are: *Follow Link*, *Go There*, *Place Bookmark*, *Remove Bookmark*, *Previous, Next*, and *Help.*
A graphical set of looping text and graphical animations helps the user remember and learn the variety of gestural and voice commands offered by this multimodal interface (figure 15).

Figures 14A,B,C,D,E. A variety of webscapes generated using NetScape

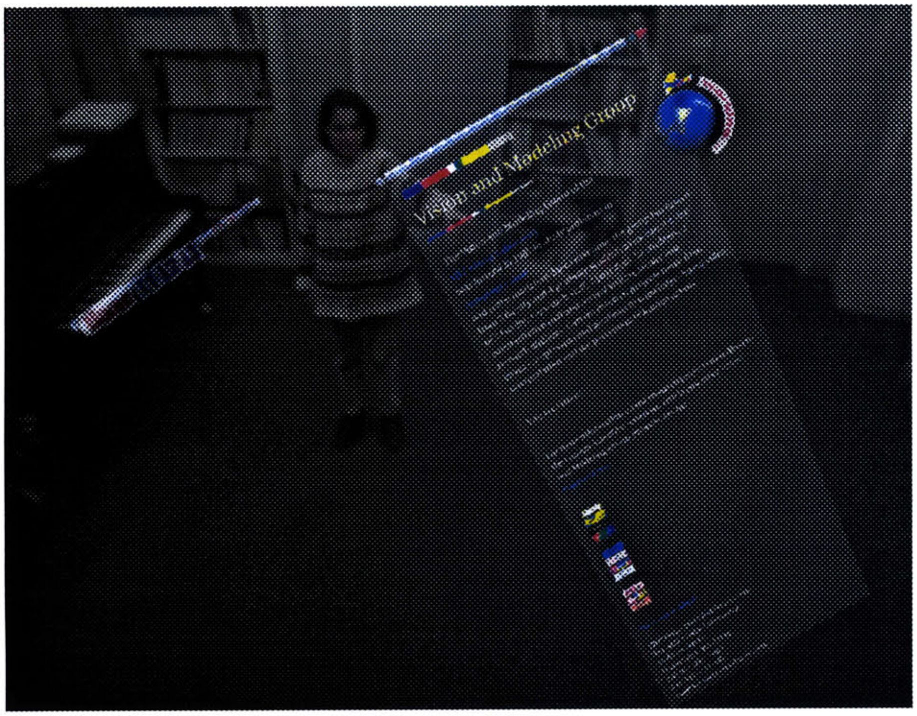


Figure 15: user issuing a pointing gesture to load a web page

In despite of the graphical help in many cases the intervention of a human guide seems necessary to help the user learn all the gestural and vocal commands available for exploration. More research is needed to establish a minimal set of actions that the user can rapidly learn.

An HTML parser, written in C++, reads HTML pages from remote servers and tokenizes the text according to HTML standard formatting tags. This information is then passed on to an Open Inventor based 3D renderer that maps HTML tags into its own customized 3D layout.
The browser currently supports standard HTML with formatted text and pictures. Future extensions include supporting MPEG movies, stereo browsing, with the use of Crystal Eyes glasses, and exploring a variety of web landscape architectures.

3.4. Improvisational Theatre Space

The endeavor to free man from his physical bondage and to heighten his freedom of movement beyond his native potential resulted in substituting for the organism the mechanical human figure: the automaton and the marionette. E.T.A. Hoffmann extolled the first of these, Heinrich Kleist the second. The English stage reformer Gordon Craig demands: "The actor must go, and in his place comes the inanimate figure - the Übermarionette we may call him". And the Russian [Valeri] Bryusov demands that "we replace actors with mechanized dolls, into which a phonograph shall be built".

from Schlemmer, Moholy-Nagy, Molnár, The Theater of the Bauhaus, ed. Walter Gropius, 1971.

The Improvisational Theater Space is an interactive stage where human actors can perform accompanied by virtual actors. This space was created following the model of the street theater, the mimes world and the improvisational theater in general (ie. Viola Spolin, Del Close, Keith Johnstone).

I wanted to create an interactive stage where human and digital actors meet to generate an emergent story through their interaction. An important constraint was to bypass strict scripts of the encounter. I thought of an Improvisational TheatreSpace not just as a performance space but also as a playground for participatory theater or interactive storytelling.
By participatory theater I mean a situation where a user first watches a performer interacting with a virtual actor and then is offered to take the place of the human performer and experience an emergent story from a subjective point of view. In this case it is important to avoid the need for the user to memorize lines of text based on a one time observation of the performance. The user only needs to roughly grasp the *situation* and the *interaction modality* in order to be able to participate. This augmented

storytelling environment responds to the teller's gestures, words, and tone of voice by enriching the presentation with sound, text, images, or graphics.

All of these scenarios need to satisfy the following necessary conditions for the interactive experience to be truly engaging and immersive:

- The experience needs to be scalable both for a common user and an expert performer
- The virtual actors need to be believable characters.
- Interactions between the human user/performer and the digital performer has to happen naturally (for the human) and in real time, i.e. the system needs to be responsive.

To satisfy the first condition, I created TheatreSpace as an improvisational stage. A set of possible situations is set in advance and the human actor is given the choice to change the order of the actions and the mood in which scenes are played.

In order to have believable characters, I modeled the virtual actors as media creatures. The autonomous structure of the media creatures allows them to exhibit a behavior that *makes sense* to the user even when she is not *doing the right thing*. In addition the behavior system and sensors of the creatures enable them to understand situations while interacting with the user/performer and to act as improvisational theater performers.

Finally for the third condition to be met the interactive stage is a perceptive space in the sense described in 3.1.1. Human actors use a full range of body gestures, tone of voice, as well as simple phrases to interact with media creatures.

Other scenarios can be envisioned in which users play improvisational theater games with the digital actors. Examples of these games are [Spolin, 1977; Johnstone, 1979]: 'spit takes', 'offer and accept', '3 words offer and accept', 'one word stories', 'one word scenes' (1).

(1) suggestions to play these improvisational theatre games came from performance artist and improvisational actor Kristin Hall.

3.4.1. The Typographic Actor

Computers are theater. Interactive technology, like drama, provides a platform for representing coherent realities in which agents perform actions with cognitive, emotional, and productive qualities ... Two thousand years of dramatic theory and practice have been devoted to an end which is remarkable similar to that of the fledgling discipline of human-computer interaction design; namely, creating artificial realities in which the potential for action is cognitively, emotionally, and aesthetically enhanced.

-Brenda Laurel, 1991
Computers as Theater

3.4.1.1. Scenarios

In 1914 Giacomo Balla wrote a futuristic performance piece called "Printing Press" (Macchina Tipografica) [Kirby, 1971]. In this theatre piece each of the twelve performers became part of a printing press machine by repeating a particular sound. Their movement would also reproduce the movement of the printing machine. The performance was to take place in front of a drop and wings that spelled out the word "TIPOGRAFIA" (Typography) in large black letters.

Following the example of Balla's work I created two experiences to show the potential and the capabilities of the Improvisational Theater Space. Both experiences involve interaction between a human actor and a media actor that takes the form of text. Hence the name *typographic actor* for this particular media creature.

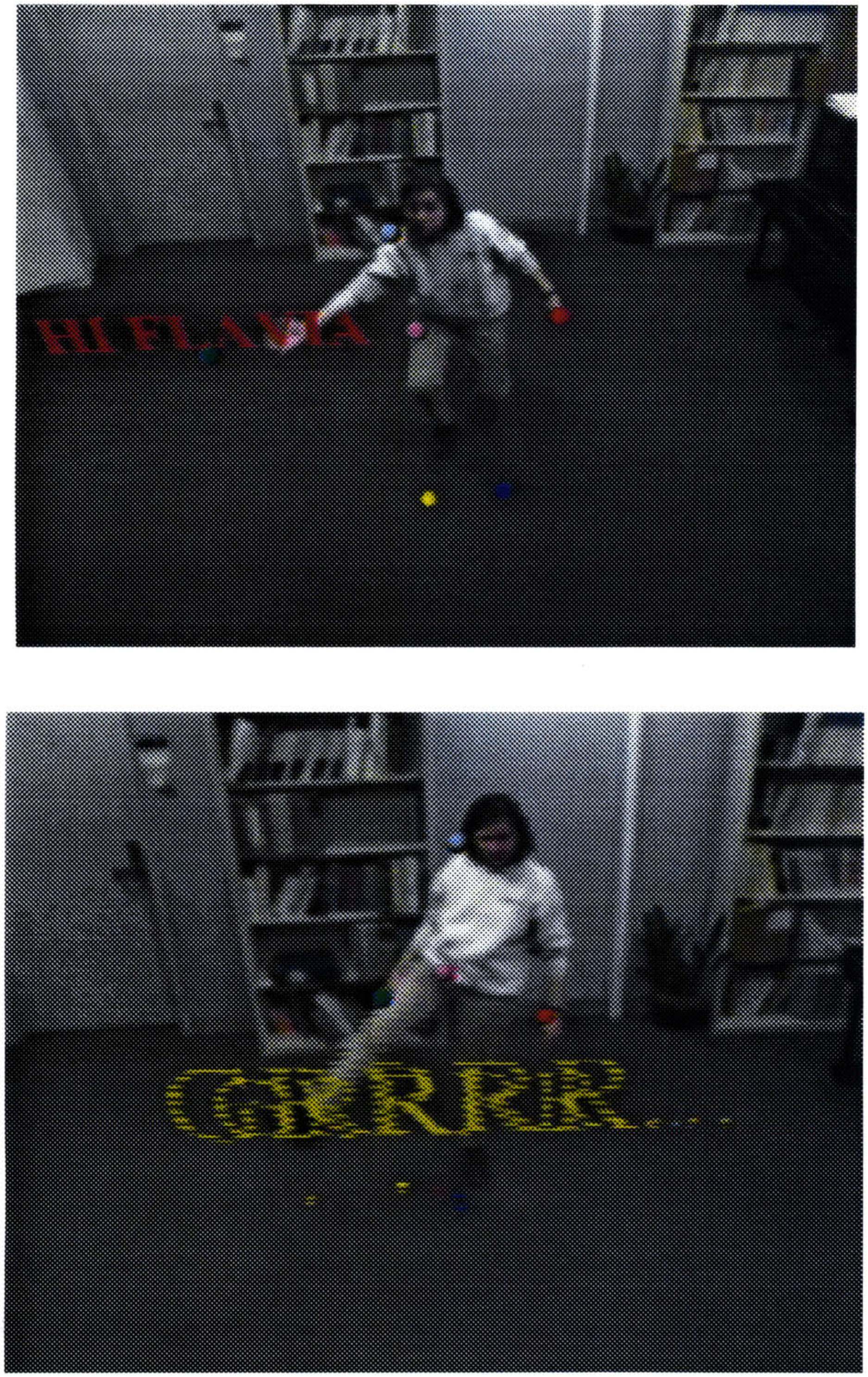
The stage is made of a large space where the actor moves and a large screen that reflects the real image of the actor and contains a graphical representation of the text. A camera and a microphone are used by the typographic actor as sensors for the real world and to interpret the user's actions.
The audience gathers around the space. The video image of the human actor and the typographic actor can eventually be broadcasted to a larger audience. Any member of the surrounding audience can step into the stage and take the place of the human performer to interact with the typographic actor at any time during or after the performance.

Experience 1: The encounter

A) The typographic actor is alone in the space. It is *daydreaming...* (figures 16A and 16B).

The actor walks into the stage:

T: dum-dee-dum-dum *(happy)*
H: Hi there *(happy)*
T: hmmmm *(scared)*
H: I am talking to you !! *(angry)*
T: Hi there! *(scared)*
H: Come say hello *(happy)*
T: Hi. *(happy)*
H: Come here *(normal)*
< text moves to actor's right hand >
H: Come over this side *(normal)*
< text moves to actor's left hand >
H: How are you feeling today? *(happy)*
T: Annoyed *(sad)*
< text moves in front of actor's face. Actor gets upset - She can't see herself anymore in the magic mirror >
H: Stop it! *(angry)*
T: grrrr.... *(angry)*



Figures 16A and B: user interacting with the typographic actor

< text moves to actor's right foot as though it wanted to bite it >

H: Go away! *(scared)*

T: Leave me alone *(sad)*

< text moves away from actor >

H: bye bye *(sad)*

T: dum-dee-dum-dum *(happy)*

B) Scene starts with the human actor on stage, looking at her watch, waiting for something or someone. It could be the bus or a friend (script supervisor: Brenda Cotto Escalera, Assistant Professor, Music and Theatre Department, MIT).

T: Hi Hi Hi *(happy)*

< text appears in different parts of the screen >

H: how are you doing? *(happy)*

T1: Fine *(happy)*

T2: OK *(sad)*

H1,2: I haven't seen you in a long time! *(surprise)* or *(anger)*

T1: I was busy *(normal)*

T2: I've been away *(sad)*

H1,2: What have you been doing ? *(interest)* or *(sad)*

T: [I have been] looking for something *(normal)*

H1,2,3: for what? *(surprise)* or *(disgust)* or *(interest)*

T: Something I've lost *(sad)*

H: <large gesture of surprise>

T: ... a long time ago.

H: What is it?

T: ... the joy of childhood

H: I don't understand <shaking head>

T: Memories beyond the surface of everyday conversation

H1,2,3,4: So what did you find? *(surprise)* or *(disgust)* or *(interest)*

T1: that I am still alive *(surprise)*

T2: a treasure of forgotten feelings *(interest)*

T3: Nothing *(sadness)*

T4: I found you! *(love)*

Experience 2: Sing!

The user sits in front of a table and by a large screen. She is prompted with the words of a popular refrain, and animates the text by singing (figure 17):

Row row row your boat
Gently down the stream
Merrily merrily merrily merrily
Life is but a dream

Figure 17: animating the typographic actor by singing

When a line appears the user will sing it. The loudness and pitch of her voice will match some typographic attribute of the text like color, font, scale, speed, or type of movement (i.e. linear, quadratic, exponential, etc..)

The user's hands rest on a table. The distance between them maps to the size of the typographic actor on the screen.

In version A of Experience 1, the encounter, the human performer has the choice of playing her lines in whatever order. In each case the typographic actor will respond in an appropriate manner, according to its behavior system.

In version B, the human actor can set the mood of the play by using appropriate gestures and tone of voice. The response and content of the text will follow.

The singing experience is a demonstration of directability of the typographic actor as a media creature. The user's way-of-singing will affect directly the behavior of the digital actor and therefore its appearance and choreography on the large screen. The written text will shape so that its typographic attributes will match the *expression* of the singing voice.

3.4.1.2. Building a Character: A repertoire of Actions for a Typographic Actor

The following is a list of high-level behaviors for the typographic actor:

1. say typical phrase
2. attract attention
3. show off
4. entertain
5. daydreaming
6. tell a story or explain

7. suggest what to do next
8. follow creature (user)

Behaviors trigger the execution of Motor Skills. Basic Motor Skills for the typographic actor are: [Dondis, 1973]

1. set String
2. read Text from HTML (web page)
3. read Text from File
4. set Color
5. set Font

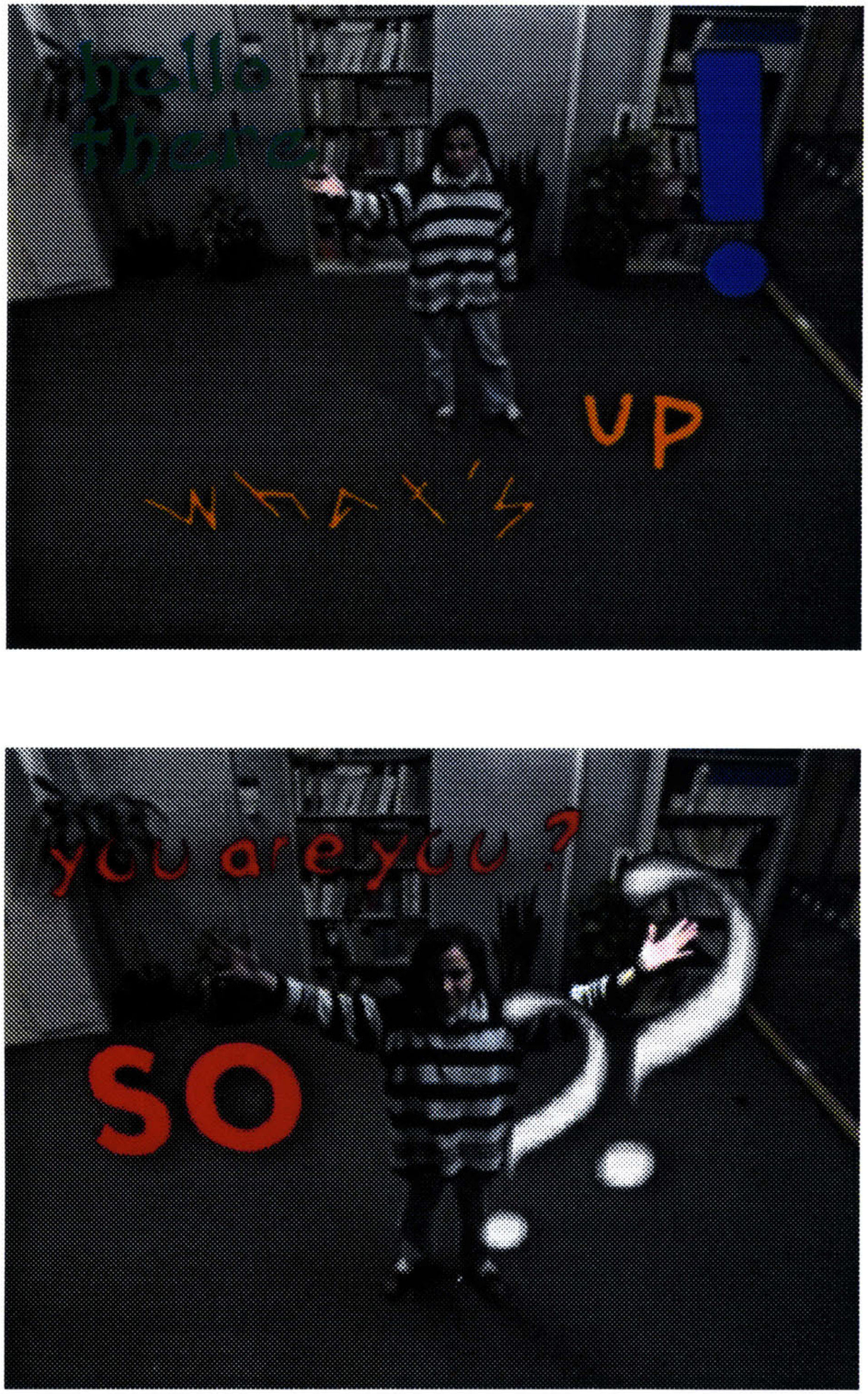
All the previous Motor Skills have immediate execution after they are issued by the behavior system.

Other Motor Skills take place through time and can happen in different time-scales or function according to *how* the action is executed based on what emotion the typographic actor wishes to convey (figures 18A and B). These are:

6. fade to Color
7. glow
8. scale
9. jump
10. goto
11. rock

Also, according to which behavior is active, the typographic actor can read text in different ways:

12. read word by word
13. fade word by word
14. fade letter by letter



Figures 18A and 18B: examples of motor skills of the typographic actor

The actions executed by the different Motor Skills can be combined as long as they do not use the same degree of freedom as the text (i.e. for the color degree of freedom *glow* cannot happen at the same time as *fadeToColor*)

Time functions that will affect the way the Motor Skills are executed are:

1. linear
2. quadratic
3. quadratic decay
4. inverse quadratic
5. inverse biquadratic
6. biquadratic
7. sinusoidal
8. quasi-sigmoid
9. oscillating up
10. exponential
11. logarithmic
12. hyperbole modulated by a sinusoidal

The emotional state of the text will affect how time-dependent motor skills will be executed according to these time functions.

Emotional states of the text are Internal Variables [Blumberg, 1995] of the media creature. The typographic actor can be:

1. happy
2. sad
3. angry
4. scared
5. disgusted
6. interested
7. loving
8. surprised
9. no particular emotional state (normal)

The idea of conveying emotions through typographic attributes of the text has been largely explored in graphic design (figures 19A,B,C) [Wong ,1995; Spiekermann, 1993; Massin, 1970]. In the case of the typographic actor if it is happy, the behavior system will issue a secondary command which affects the way in which the Motor Skill issued by the primary command is executed. In the case of *happy* that will match to *biquadratic, exponential,* or *sinusoidal*. If the text is *surprised,* the active time function will be *oscillating up*.

The typographic actor has also a level of energy that can be high, low, or medium. Text draws energy from the user's speed of movement and loudness of speech. It spends energy while executing motor commands. As well as emotional states affect the presentation, by picking an appropriate time function for execution, the energy level of the typographic actor determines the speed of the execution of a motor command.

Figures 19A,B,C. From [Spiekermann, 1993]: expressivity through type.

4. Performance Evaluation

DanceSpace has been tried by a large number of users and performers during several demonstrations at the MIT Media Lab, including a one day Open House with people of all ages. Semi-professional dancers from Boston Conservatory have choreographed short pieces for the interactive stage under the supervision of the choreographer Erica Drew. During these performances the choreographer made an effort to enhance or underline the expressiveness of the human body as opposed to the "coldness of the musical and graphical output by the computer" (her words).

The dancers were fascinated by the colored virtual shadow that followed them on stage and soon modified their pieces as to better exploit the "comet" effect of the computer graphics trails. Common users (non-performers) who attended the open house seemed to be more interested in exploring the space to obtain the desired musical effect.

I also used DanceSpace to create an original choreography together with choreographer Claire Mallardi at Radcliff College, Harvard Extension, in the spring of 1996 (figures 20A,B,C,D). In this case I had prerecorded the virtual shadows generated by my own dance at the Media Lab and then projected them - non interactively - on the dancers bodies during the performance. As the dancers were wearing white unitards their bodies were virtually painted by the projected computer graphics image. As a consequence they changed the choreography of their piece to have more still body poses to better exploit this body-painting effect. Another interesting effect was also happening on the backdrop of the stage where a parallel performance was happening: the dancers' black shadows were dancing in real time with the colored shadows generated by the virtual dancer (not on stage).

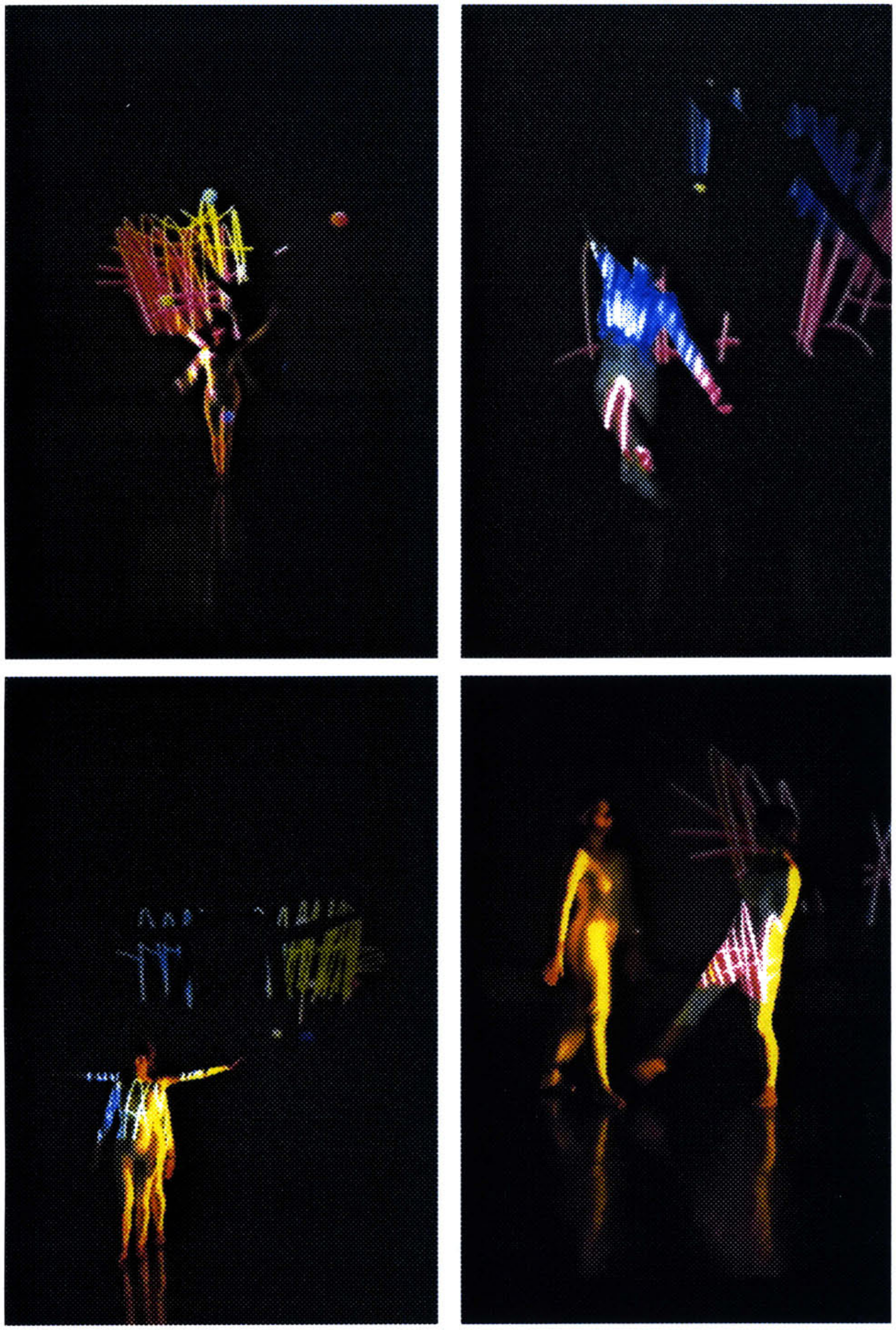
The public reacted positively and encouraged me to further explore this type of mixed-media performances.

It seems necessary now to expand DanceSpace to allow for a variety of musical mappings and graphical representations of the dancers (2).

NetSpace and the Improvisational TheatreSpace were shown in the digital Bayou section at SIGGRAPH 96. Users enjoyed using pointing gestures and voice commands to explore the web in 3D. The main problem I had was that oftentimes I had to instruct the user on the gestures that "worked" and on the voice commands that the system recognized. Even though I built a graphical aid to help the user to learn the interface, the suggestions of a human guide seemed to be necessary for the user to fully exploit the capabilities of the system.

At SIGGRAPH users also animated the typographic actor by singing and enjoyed watching the dynamic change of colors/font/size of the text according to their input. An early version of the typographic actor interacting with a human actor was also informally shown at CHI96 and attracted the attention and received positive comments by the design community attending the conference.

(2) DanceSpace was documented in a Scientific American Frontiers special which was filmed at the MIT Media Lab by the Chedd-Angier production crew in the spring of 96 and by the Austrian TV that broadcasted a piece on DanceSpace during the days of the prize awards for the Linz Ars Electronica Festival 96.



Figures 20 A,B,C,D.
Dancers dancing with DanceSpace's virtual shadow during a public peformance

5. Related Work

5.1. Synthetic Theatre

The idea of creating performances that involve both human actors and inanimate objects as characters can be tracked back to the beginning of the century. The avangard Italian Futurist movement inspired a new form of theatre called *Synthetic Theatre*. In the Futurist Theatre manifesto emphasis was given to a more direct and non contemplative involvement of the spectators with the presentation. This was to be achieved by using techniques such as compression (of the text), simultaneity (of the scenes) and the involvement of the audience in the performance. In 1909 Marinetti, the founder of Italian Futurism wrote a theatre play called *Electric Puppets* in which human actors acted together with life-like puppets. A few years later he wrote a *drama of objects* called "They are Coming" (Vengono) where the main characters are future objects in a room that try to find each other after the servants come in a rearrange the furniture around for a new social gathering. Other theatre works of interest by the futurists include Balla's *Printing Press* (Macchina Tipografica) and Podrecca's *Plastic Dances* (Balli Plastici).
The futurists created a theater of situations to show the diversity of life, a diversity "present for a moment on a train, in a cafe, at a station, and which remains filmed on our minds as dynamic, fragmentary symphonies of gestures, words, noises and lights" [Kirby, 1971].

The theatre of the Bauhaus also experimented with a non-verbal, dance-inspired theatre of objects with living actors stylized as geometric, often cubic shapes, resembling modern marionettes and automatons [Schlemmer, 1971].

Although futurist performances did not have great success with the public they can be seen as early tentatives of staging human actors together with objects in a theatre of situations that required participation from the public.

The futurists' work certainly influenced Pirandello's theatre, which in turn influenced the work discussed in this thesis. In particular, the dramatic structure underlying the encounter metaphor is similar to that of the theater play "Six characters in search of an author" by Pirandello. "Six Characters" was born of Pirandello's contact with the characters of his imagination. The play is organized around a meta-play in which six people find a theater manager and ask him to make them become 'characters' in an artistic theatrical sense. They would like to see their life's daily struggle transformed into eternal artistical expression but they don't feel truly represented by the medium. By the artifice of the *theater-in-the-theater* they end up offering the viewer the plain drama of their little lives and miss their opportunity to become eternal through art. Each of the characters has a story that needs to be told, or a story that must never be told. In some sense, the characters are their story, because we are never shown anything important about them that is also not important to their story.

This type narrative structure has been used by some filmmakers. Two successful examples are: "Sherlock Jr." by Buster Keaton and "The Purple Rose of Cairo" by Woody Allen [Cook, 1990]. In my system the user is the 'author' that through interaction with the media creatures develops a story of which she is also a part.

5.2. Virtual Reality

Both DirectIVE and ALIVE (of which DirectIVE is an extension) were inspired by Myron Krueger's Videoplace system [Krueger, 1990]. The public area of VIDEOPLACE consisted of two or more rooms separated from on another by various geographic distances.

Video cameras, mixers and projectors would make it possible for people in any of the different rooms to interact with the video images of others who were physically located elsewhere.
DirectIVE and VIDEOPLACE differ primarily in three respects. The first difference is that VIDEOPLACE focuses on 2D rather than 3D worlds and interaction. A second difference is DirectIVE's emphasis on autonomous agents. Most of Krueger's worlds allow users to interact with other users, a notable exception being the "critter", a 2D animated sprite. A final difference is that DirectIVE can recognize hand and body gestures as patterns in space and time.

Another system that bears similarities to DirectIVE is the Mandala system [Vincent, 1993] which composits the user's color image with a virtual world, which is sometimes video-based and sometimes computer-animated. A difference with DirectIVE is that the Mandala system only supports 3D; it does not attempt to recognize parts of the user's figure nor does it do any gesture recognition.

An important predecessor of this research is Brenda Laurel's work. Her early projects at Atari in the 80s sought to build 3d graphics worlds with artificial characters that respond to each individual human user like an actor in an improvisation exercise, supported by rule-based systems that included dramatic expertise based on Aristotele's theories of drama.

DirectIVE differs from this approach because it builds autonomous animated characters that are behavior-based instead of rule-based. This gives the system more flexibility and choice for improvising a believable interaction with the user.

Joe Bates and the Oz group have made a Woggle world [Bates, 1993]. Bates' woggles have several internal needs and emotions which result in fairly complex interactions with the user. In contrast with DirectIVE, users interact with the woggles using a mouse, rather than being able to interact directly.

Pinhanez has envisioned a computerized theatre that relies on the knowledge of the script to drive a computerized stage [Pinhanez, 1996]. It proposes to use a scripting technique based on interval logic to recognize the actor's action and to drive special effects on stage. The main drawback of this approach is that it relies on the director and the actor to rigidly follow a script for the system to be able to work. For instance it is not uncommon in theatre that both the actors and the director change the script either during rehearsals or even right before or during the final performance [Brook, 1995]. In my view this type of rule based system will not be able to compensate for human errors or be responsive when some non planned "magic" between the actors happens on stage. It tends to force human interpreters to rigidly follow a predefined track and therefore impoverishes the quality of the performance.

In summary I believe that behavior-based media creatures are a better approach to computerized theatre than the above for three main reasons:

1. Behavior-based versus script based theatre has room for improvisation, both in the case of the improvisational or street theatre in general or for classical scripted theatre that the director and the actors need to interpret and therefore modify.

2. The system is tolerant to human error and actually encourages actors to enrich or change the performance according to the reaction of the audience.

3. The system can scale from a performance space to an entertainment space. Behavior-based theatre can allow for user participation either during or after the performance without requiring the new users to learn all the script in advance.

5.3. Temporal Typography and Information Landscapes

The Media Lab's VLW has done extensive research in the use of digital text in 3D dynamic landscapes. David Small explored the use of dynamic typography in interactive 3D presentations [Small, 1994]. Suguru Ishizaki has developed multiagent models of dynamic design for typographic spaces [Ishizaki, 1996]. Ishizaki offers a conceptual model and a descriptive language that designers can use in the course of creating dynamic design solutions. Wong's research [Wong, 1995] provides a designer's tool for scripting expressive text for message exchange, storytelling and electronic chat.

The main difference between these works and the typographic actor is that the latter is an autonomous creature conceived to interact with users or performers in an interactive environment and it is not a designer's tool. Although the text can be directed by the user to act or behave in a specific manner the degree of control that the user has over it is limited by its autonomous nature. VLW's scripted text has more knobs and features than the typographic actor as it needs to satisfy the artistic expression requirements of the designer.

Also in the VLW, in his Master Thesis "The Mind's Eye", Earl Rennison demonstrates a novel interface approach for visualizing, navigating and accessing information objects in a large body of unstructured information [Rennison 1995]. A World Wide Web browser in 3D space was also build in the VLW by Jeffrey Marshall [Marshall, 1996]. Although it was built along the lines set by VLW's information landscapes, NetSpace differs for two main reasons:

1. it attempts to build dynamically an urban landscape through exploration.
2. the interaction modality takes place with natural gestures and voice commands in a perceptive space.

Relative to point 1, research on NetSpace was actually inspired by Jeffrey Shaw's "Legible City" (figure 21) in which a cyclist virtually wanders through urban alleys made of 3D text.

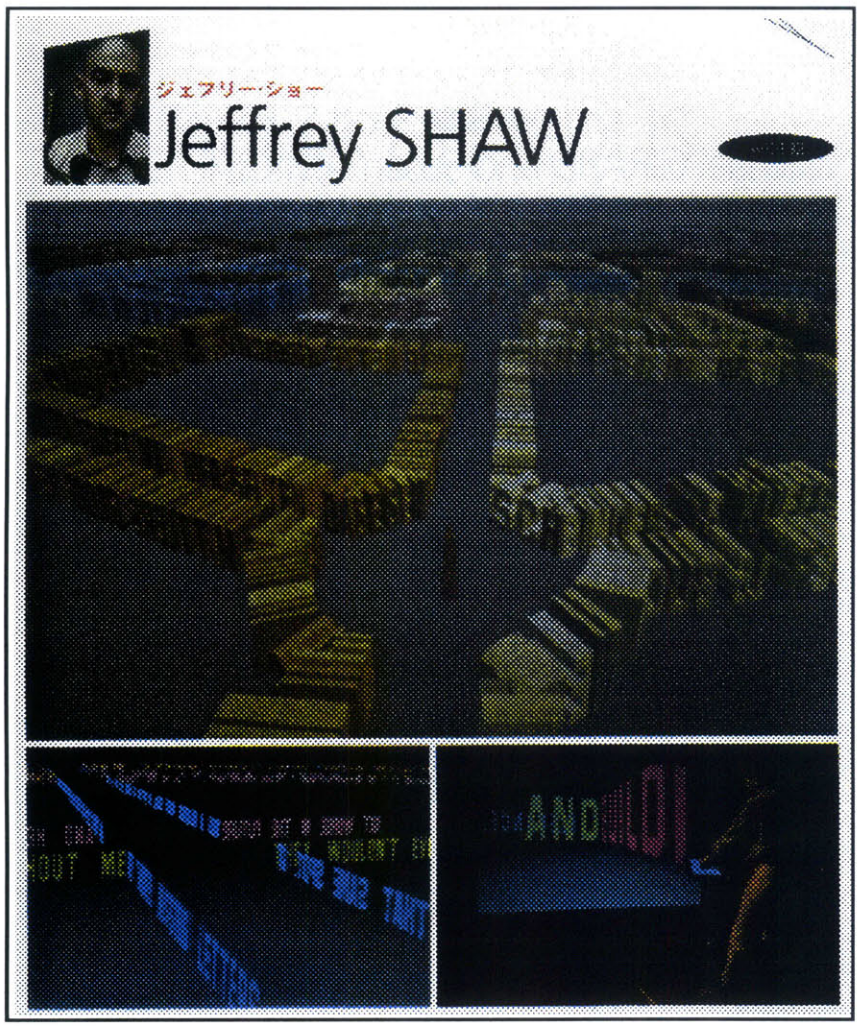


Figure 21. Images by Jeffrey Shaw's "Legible City"

5.4. Interactive Sound and Dance Spaces

Ann Marion has pioneered work on generating graphics through body motion in dance at MIT. In her Master's Thesis at the MIT Architecture Machine Group, Marion describes the use of polhemus sensors to track ballet steps and illustrates a beautiful performance project with the Joffrey Ballet dance company in NY where the computer graphics image generated by the dancers becomes integral an part of a performance [Marion, 1982]. Steve Mann has also done work with dancers generating sound through movements on stage by using radar and ultrasonic technology [Mann, 1991].

Rolf Gehlhaar [Gehlhaar, 1991] has built a number of sound spaces where multiple users generate soundscapes through full body motion. Researchers at Georgia Tech's Center for the Arts together with the Atlanta Ballet are also involved in a Dance and Technology Project. They use active sensing technology by placing sensors on the dancer's body to track the performer's movement on an interactive stage.

Dance Space differs from the afore mentioned examples of interactive dance spaces because the computer vision tracking system provides real time information about different body parts of the dancer and not just an estimated of gross motion of the performer. Moreover as opposed to other environments it does not require the dancer to wear special clothes or active sensors. Anyone can just walk in the space and generate sound and graphics by body movements/gestures. DanceSpace's principal drawback is that currently the computer vision based sensing technology reliably tracks only one performer a time.

6. Future Work

6.1. The Inter-Museum

Immersive digital exhibition spaces may constitute an interesting complementary experience to that of visiting a Museum in person. If however you are looking forward to a *virtual visit* to a remote Art Museum or an Art Exhibition site on the web, you are most likely to be disappointed. You will find mostly collections of photographs and/or text in static pages that can eventually just scroll along the plane of the display monitor. Unless you are interested in retrieving a specific information about an artist or artwork, you are certainly not involved in an entertaining or engaging experience. The nature of this type of experience is closer to that of browsing through an electronic catalogue than that of being in an immersive space.

On the other hand, real Museums are more than ever before into the orbit of leisure industries. They are faced with the challenge of designing appealing exhibitions, handling large volumes of visitors and conserving precious artwork.

One of the main problems they are faced with is to *give life* to the objects on display by telling *their story* within the context determined by the other objects on display.

Interactive Science Museums have partially solved this problem, mostly facilitated by the nature of the objects they exhibit. Their intent is to engage the visitor, transforming him/her from a passive viewer into a participant by use of interactive devices.

They achieve their intent, amongst other things, by installing button-activated demonstrations and touch sensitive display panels that provide supplementary information when requested. They make use of proximity sensors to increase light levels on an object when a visitor is close by and/or to activate a process.

Other Museums offer audiovisual material to give viewers some background and a narrative coherent exposition of the works they are about to see or that they have just seen. Often movies that show artwork together with a description of their conception and other historical material about the author and her times are even more compelling than the exhibition itself. The reason is that the movie has a narration, the visuals are well orchestrated and come with music and dialogue. The movie presents the viewer with a more unified and coherent exposition of the content than the fragmented experience of the visit.
A few historical Museums in the United States have adopted an effective interactive solution to entertain and inform the user. An example is the Plymouth Plantation Museum. In this museum, 17th century New England is brought to life in the 1627 Pilgrim Village. Visitors find themselves transported in time by Pilgrims who are dressed in 17th century clothing, even speaking in period dialects as they perform the task of daily life. In this dynamic exhibit real people create an interactive environment which visitors can use to ask questions and hear first person accounts from the periods' inhabitants.

A truly engaging Virtual Museum should then use active content and media creatures that act as the human storytellers of the Plymouth Plantation Museum. In this case the following scenario could become possible:

The user wants to visit the exhibition on Cézanne, hosted by the web site: http://sunsite.unc.edu/wm/paint/auth/cezanne/. She will see on the large screen the title of the exhibition/web page - Paul Cézanne - and the Logo of the site. The words "cubism" and "impressionism" which are two hot links at the top of the web page will be blinking on the sides. A voice will be reading the introductory part of the text. Then an arrow will appear and invite the user to move forward. She will move closer to the large screen and she will soon be invaded by a flock of flying words each one indicating one of her possible paths in the exhibitions. They are "Biography", "From Impressionism to Classicism and Cubism", "Cézanne's works", "Portraits", "Bathers", "The Still Life gallery", "Landscapes" as on the original site.

These words will move around the user trying to catch her attention in different ways by their movement on the screen. The user chooses to see the portraits by grabbing and holding the word "Portraits" in her hand. All of the previous text disappears. As the voice describing this section starts playing she shuts it off by moving her right finger next to her nose. She is only interested in the images for this part of the visit. Images of the paintings tile together forming a corridor at the right and left of the user. As she starts walking along the corridor the title of the painting she is walking close by appears. She points her arm towards the painting she wants know more about. As she does that the painting will move towards her to introduce itself. The other paintings move away. Music starts playing ...

6.2. Futuristic Performances

Although cinema has used text in the past as a means to communicate the character's inner thoughts instead of voice over - examples include Woody Allen's Annie Hall and Peter Greenway's Prospero's Book - no similar effects have been used in theatre to my knowledge.
The choreographer William Forsythe has experimented using text in his dance performances with the Frankfurt Ballet [Gilpin, 1995].
I am planning to use the typographic actor as an extension to DanceSpace's graphical output to create a type of live performance that narrates poetry using the body to play the text on the stage.

In 1964, Robert Massin designed a special edition of Ionesco's play *The bald Soprano* [Ionesco, 1956] in which he makes a special use of custom designed typography to express the *voice* of the speakers and simultaneity of expression. Characters shouting and talking simultaneously, actors facing away from the audience and other expressive aspects of Ionesco's dramatic techniques were lost when his play script was printed in monotonous lines of type on plain pages. Massin translated the atmosphere, the movement, the speeches and the silences of the play trying at the same time to convey

the idea of duration of time and space on the stage by the simple device of interplay of image and text (figure 22) [Massin, 1970].
Following Massin's example I would like to set up a short performance, one act from Ionesco's Bald Soprano, that exploits the expressivity and interactivity of the typographic actor to augment the director's artistic options on stage.

Figure 22.
A page from Massin's typographic interpretation of *The Bald Soprano*

7. Conclusions

In this work I describe methods for content choreography in live performance and entertainment spaces. I used Artificial Life programming methods and behavior-based design to avoid rigid scripting of user and content interaction. The main result of this research is the construction of animated media creatures endowed with intentionality and autonomous behaviors. Media Creatures allow content to be active and to present itself to the user by dynamically adapting to the context of the interaction. I used Media Creatures to create an engaging Improvisational TheatreSpace where the user/performer is engaged in an improvisational dialogue with a typographic actor. This space is an example of an immersive Interactive Virtual Environment that can be used to create a variety of experiences like augmented Storytelling or Virtual Museum explorations.

Along the path leading to the Improvisational TheatreSpace I have also built DanceSpace and NetSpace. Both of these spaces exploit the advantages of wireless computer vision and audition technology to allow a user/performer to move and interact naturally with content.

This research contributes to the field of digital graphic design and content orchestration for virtual reality applications by offering new tools for designing interactive immersive spaces.

The work carried out in this thesis also defines directions for future research for each of the interactive spaces created. Examples are: the exploration of a variety of mappings of full body gestures to autonomously generated music for DanceSpace, intuitive man-machine interfaces and interactive landscape construction for NetSpace, and learning from experience of previous interaction with the user for the Media Creatures.

Bibliography

Bates J., Altucher J., Hauptman A., Kantrowitz M., Loyall A.B., Murakami K., Olbrich P., Popovic Z., Reilly W.S., Sengers P., Welch W., Weyhrauch P. and Witkin A., *Edge of Intention*, SIGGRAPH-93, Visual Proceedings, Machine Culture, ACM SIGGRAPH, pp. 113-114, 1993.

Blumberg Bruce M., Galyean Tinsley A., *Multi-Level Direction of Autonomous Creatures for Real-Time Virtual Environments*, ACM-0-89791-701-4/95/008, SIGGRAPH 1995.

Blumberg Bruce M., *Action-Selection in Hamsterdam: Lessons from Ethology*, Proceedings of the 3rd International Conference on the Simulation of Adaptive Behavior, Brighton, August 1994.

Brook Peter, *The Open Door*, Theatre Communication Group, New York New York, 1995

Buchanan Richard, *Wicked Problems in Design Thinking*, in The Idea of Design, Margolin and Buchanan Eds, MIT Press, 1995.

Casey Michael A., Gardner William G. and Basu Sumit. *Vision steered beam-forming and transaural rendering for the artificial life interactive video environment (alive)*. In Proceedings of the 99th Convention of the Aud. Eng. Soc. AES, 1995.

Cook, David A., *A History of Narrative Film*, WW Norton & Company, 1990.

Darrell T., Maes P., Blumberg B., Pentland A., *A Novel Environment for Situated Vision and Behavior*, Proc. IEEE Workshop for Visual Behaviors, IEEE Comp. Soc. Press, Los Alamitos, CA, 1994

Dondis Donis, *A Primer of Visual Literacy*, MIT Press, 1973

Gehlhaar Rolf, *SOUND = SPACE: an Interactive Musical Environment*, Contemporary Music Review, 6(1), pp 59-72, 1991.

Gilpin Heidi, *William Forsythe: Where Balance is Lost and the Unfinished Begins,* Parkett, 45, 1995

Ionesco Eugene, *The Bald Soprano*, typography and photography by Massin and Cohen, Grove Press, NY, 1956

Ishizaki Suguru, *Typographic Performance*, PhD Thesis, MIT Media Lab, 1996

Johnstone Keith, *Impro: improvisation and the theatre*, New York, Theatre Art Books, 1979.

Kirby Michael, *Futurist Performance*, New York, E.P. Dutton, 1971

Klosty James, *Merce Cunningham: dancing in space and time*, Saturday Review Press, 1975

Krueger Myron, *Artificial Reality II*, Addison Wesley, 1990

Laurel Brenda, *A Taxonomy for Interactive Movies*, New Media News, Vol. 3 No 1, Winter 1989.

Laurel Brenda, *Computers as Theatre*, Menlo Park, CA, Addison Wesley, 1991

Machover Tod, *Hyperinstruments: A composer's approach to the evolution of intelligent musical instruments*, CyberArts, 1992.

Maes, Pattie, *Modeling Adaptive Autonomous Agents*, Artificial Life 1: 135-162 , 1994.

Mann Steve, *DopplerDance, A Sculpture of Waves*, public performance, Toronto, March 22nd 1991.

Marion Ann, I*mages of Human Motion: Changing Representations of Human Identity*, Master Thesis, MIT 1982

Marshall Jeffrey Barret, *A World Wide Web Browser Utilizing Three-Dimensional Space*, Bachelor Thesis, MIT, 1996.

Massin, Robert. *Letter and Image*, New York, Van Nostrand Reinhold, 1970

Pinhanez Claudio, *Computer Theatre*, Perceptual Computing Technical Report No. 378, MIT Media Lab, 1996

Pirandello Luigi, *Six Characters in Search of an Author*, in Naked Masks, Bentley, E., ed.,Storer, E., trans., E. P. Dutton & Co., 1952.

Rennison Earl, *The Mind's Eye. An Approach to Understanding Large Complex Information-Bases Through Visual Discourse*. Master Thesis, MIT Media Lab, 1995.

Schlemmer Oscar, Moholy-Nagy Laszlo, Molnár Farkas, *The Theatre of the Bauhaus*, ed. Walter Gropius. Middletown, Conn. Wesleyan University Press, 1971

Small David, Ishizaki Suguru, Cooper Muriel. *Typographic Space.* CHI 1994.

Sparacino F., Wren C., Pentland A., Davenport G., *HyperPlex: a World of 3D Interactive Digital Movies,* IJICAI 1995

Spiekermann Erik and Ginger E.M., *Stop Stealing Sheep & find out how type works,* Adobe Press, Mountain View, CA, 1993

Spolin Viola, *Improvisation for the Theater,* Northwestern University Press, 1979.

Vincent V.J., *Mandala: a Virtual Village*, SIGGRAPH-93 Visual Proceedings, Tomorrow's Realities, ACM SIGGRAPH, pp.207, 1993.

Walser Randall, *Elements of a Cyberspace Playhouse*, Proceedings of the National Computer Graphics Association,1990, Anaheim, CA, March 1990

Wong Yin Yin, *Temporal Typography*, Master Thesis, MIT Media Lab, 1995

Wren Christopher R., Sparacino Flavia et al., *Perceptive Spaces for Performance and Entertainment: Untethered Interaction using Computer Vision and Audition*, to be pushished in: Applied Artificial Intelligence (AAI) Journal, March 1996.

Wren Christopher,Azarbayejani Ali, Darrell Trevor, Pentland Alex, *Pfinder: Real-Time Tracking of the Human Body*, Integration Issues in Large Commercial Media Delivery Systems (Conf. 2615)", SPIE, 1995.