Media in performance: Interactive spaces for dance, theater, circus, and museum exhibits

by F. Sparacino
G. Davenport
A. Pentland

The future of artistic and expressive communication in the varied forms of film, theater, dance, and narrative tends toward a blend of real and imaginary worlds in which moving images, graphics, and text cooperate with humans and among themselves in the transmission of a message. We have developed a "media actors" software architecture used in conjunction with real-time computer-vision-based body tracking and gesture recognition techniques to choreograph digital media together with human performers or museum visitors. We endow media objects with coordinated perceptual intelligence, behaviors, personality, and intentionality. Such media actors are able to engage the public in an encounter with virtual characters that express themselves through one or more of these agents. We show applications to dance, theater, and the circus, which augment the traditional performance stage with images, video, music, and text, and are able to respond to movement and gesture in believable, aesthetical, and expressive manners. We also describe applications to interactive museum exhibit design that exploit the media actors' perceptual abilities while they interact with the public.

Our society's modalities of communication are rapidly changing. Large panel displays and screens are being installed in many public spaces, ranging from open plazas, to shopping malls, to private houses, to theater stages and museums. In parallel, wearable computers are transforming our technological landscape by reshaping the heavy, bulky desktop computer into a lightweight, portable device that is accessible at any time. Such small computers, accompanied by a high-resolution private eye for display and by an input device, have already added an additional dimension to our five senses since they allow us to wear technology just as an element of our everyday clothing. Transmission of information through these media requires new authoring tools that are able to respond reliably not just to mouse point-and-click or drag-and-drop actions, but also to more natural full-body movements, hand gestures, facial expressions, object detection, and location.

Interactive experiences in general benefit from natural interactions, compelling communication, and ease of implementation. We show, according to these principles, how interactive media architectures can be categorized as scripted, responsive, learning, behavioral, or intentional. We have developed a "media actors" authoring technique: We endow media objects—expressive text, photographs, movie clips, audio, and sound clips—with coordinated perceptual intelligence, behaviors, personality, and intentionality. Such media actors are able to engage the public in an encounter with virtual characters that express themselves through one or more of these agents. They are an example of intentional architectures of media modeling for interactive environments.

Since 1995, we have constructed a variety of interactive spaces,[1] and each of them has inspired a revision and improvement of our authoring techniques.

In this paper we focus primarily on our performance and interactive museum work. For dance and theater, we have built an interactive stage for a single performer that allows us to coordinate and synchronize the performer's gestures, body movements, and speech with projected images, graphics, expressive text, music, and sound. For the circus, we created a networked environment in which participants from remote locations meet under a virtual circus big top and interact with the performers in a game of transformations. For museums, we have authored an interactive room as well as a personalized exhibit documentary played in the private eye of a wearable computer, using media actors.

In the following section, we first introduce our media modeling approach by doing a critical analysis of existing interactive media authoring techniques. We then present our media modeling taxonomy and explain in more detail the intentional architecture. In the remainder of the paper we describe our work in interactive dance, theater, circus, and museum exhibit design.

Authoring techniques

Recently, in the field of computer graphics, progress has been made in the creation of lifelike characters or autonomous agents. Autonomous agents are software systems with a set of time-dependent goals or motivations that the agents try to satisfy in a complex dynamic environment (such as the real world).[2] 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. In computer graphics, autonomous agents are lifelike characters driven by autonomous goals. They can sense the environment through real or virtual sensors and respond to the user's input or to environmental changes by modifying their behavior in accordance with their goals.[3–5] Although this approach, called "behavior-based," has proven to be successful for a variety of computer graphics problems, it has not yet been fully understood or exploited for multimedia presentations, digital storytelling, interactive performance, or new forms of interactive art.

It is not uncommon to read about interactive multimedia or artistic experiences that advertise themselves as behavior-based, whereas in reality they would be better described as simply responsive or reactive. These experiences are systems in which short scripts group together a small number of actions on the content displayed. These microscripts are then triggered by sensors, which directly map predetermined actions of the public to an appropriate response of the system. Nevertheless, the fact that these scripts often apply to a human character portrayed in the experience leads to erroneous descriptions of them as behavior-based. The scripts do actually show segments of human behavior; however, the term is not used correctly to describe the internal architecture of the system.

To date there is some confusion and a discrepancy in the way the computer graphics community and the multimedia and electronic art community think of behavior-based modeling. We would like to clarify some important issues in this respect and delineate a direction of work that describes what multimedia and interactive art can learn from computer graphics and how the new behavior-based modeling techniques can be extended and creatively applied to a variety of artistic domains. In this section we explain how the behavior-based approach differs from the scripted or reasoning approach and describe the origin, purpose, and advantages of behavior-based modeling.

Scripted versus behavior-based approaches, and beyond. In the field of multimedia and electronic art, the term "behavior-based" is often ingenuously used to contrast media modeling from more traditional architectures that separate the content on one end and the routines for orchestrating media for the public at the other end. This split architecture leads to complicated control programs that have to do an accounting of all the available content, where it is located on the display, and what needs to happen when or if or unless. These systems rigidly define the interaction modality with the public as a consequence of their internal architecture. Often, these programs need to carefully list all the combinatorics of all possible interactions and then introduce temporal or content-based constraints for the presentation. Having to plan an interactive art piece according to this methodology can be a daunting task, and the technology in place seems to somehow complicate and slow down the creative process rather than enhance or expand it.

There is certainly an analogy between these systems and the more traditional artificial intelligence (AI) applications, which were based on reasoning and symbol manipulation. Both share the idea of a centralized "brain" that directs the system on the basis of operations on symbols derived from sensory or

direct input. When used for authoring interactive media or electronic art, we consider this approach similar to that of having an orchestra director who conducts a number of musicians following a given score. This approach leaves very little room for interactivity, 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. Typical examples of this approach are the many CD-ROM titles that simulate the presentation of an interactive story or game by careful planning of plot bifurcations and multiple-choice menus.

In our view, interactive art and digital media presentations should not be limited to the construction of environments where nothing happens until the participant "clicks on the right spot." In most interactive experiences, 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. This environment induces the public to adopt an *exploratory* type of behavior that tries to exhaust the combinatorics of all possible interactions with the system. The main drawback of this type of experience is that it penalizes the transmission of the message: A story, an emotion, is somehow reduced to exploring the tricks of the gadgetry employed to convey the message. The focus, in many cases, turns out to be more on the interface than the content itself. *Navigation* is, as a matter of fact, the term that better describes the public's activity, the type of interface, and the kind of experience offered by systems authored according to this approach. These systems are often called *scripted*.

Behavior-based is a term originating in artificial intelligence and is often synonymous with "autonomous agent research." It describes control architectures, originally intended for robots, that provide fast reactions to a dynamically changing environment. Brooks[6] is one of the pioneers and advocates of this new approach. Maes[2] has extended the autonomous agent approach to a larger class of problems, including software agents, interface agents, and helpers that provide assistance to a human involved in complex activities, such as selecting information from a large database or exchanging stock. Behavior-based architectures are defined in contrast to function-based architectures as well as to reasoning-based or knowledge-based systems. The latter approach, corresponding to "traditional AI," emphasizes operations on symbolic structures that replicate aspects of human reasoning or expertise. It produces "brains," applying syntactic rules to symbols representing data from the external world, or given knowledge, and which generate plans. These systems work under a closed-world assumption to eliminate the problem of unexpected events in the world. The closed-world assumption declares that all facts relevant to the functioning of the system are stored in the system, so that any statement that is true about the actual world can be deduced from facts in the system. This assumption is useful for designing logic programming environments, but it is untenable in the real world.

Brooks has highlighted the limitations of this centralized and closed-world approach. He has successfully demonstrated the validity of the behavior-based approach by building a number of mobile robots that execute a variety of tasks by choosing an appropriate action from a hierarchical layering of behavior systems (subsumption architecture). Maes has developed an action-selection approach in which individual behaviors have associated an activation level for run-time arbitration,[7] instead of choosing from a predefined selection mechanism as does Brooks. Blumberg[5] has adopted an ethological approach to model a virtual dog able to interact with humans as well as with other behavior-based synthetic creatures. Together with Zeltzer[3] and Johnson,[4] Blumberg has provided an example of how the behavior-based approach can be effective in producing lifelike computer graphics animat creatures (animats = animal + automats) that are able to find their bearings in virtual worlds, and at the same time can perceive commands from a human through the use of real-time computer-vision sensors.

The advantage of behavior-based modeling techniques is that the designer of the experience does not have to think of all the possible sequences and branchings in defining the interaction between the public and the virtual creatures. It suffices to specify the high-order behaviors, their layering structure, and the goals of the creature to produce a realistic and compelling interaction. Blumberg also introduces an abstraction to allow the hard task of coordinating the kinematic motion of the articulated dog to be split from the specification of the high-level behavior system. He describes an improved action-selection mechanism that allows arbitration among commands given by a human and the autonomous drive of the creature determined by its goals and internal motivations.[8] Perlin and Goldberg[9] have applied a similar approach to animating humans in virtual environments.

Although behavior-based computer graphics has proven successful in producing an effective interactive experience between the public and a variety of synthetic animal creatures, it is important to understand and discuss how it can be translated and applied to different application domains. In multimedia, performance, and the electronic arts, the designer of the experience and the public are often involved in more complex forms of interactions and communication that require a revision of the current behavior-based model.

A taxonomy of authoring techniques. The behavior-based approach has proven to be successful when applied to mobile robots and to real-time animation of articulated synthetic creatures. In this context, behavior is given a narrow interpretation derived from behavioral psychology.[10] For animats, behavior is a stimulus-response association, and the action-selection mechanism assigning weights to the layered behaviors can be seen as a result of operant conditioning[10] on the creature. Behavior-based AI has often been criticized for being "reflex-based," because it controls navigation and task execution through short control loops between perception and action.

In our view, Skinner's reductive notion of behavior is insufficient to model many real-life human interactions as well as simulated interactions through the computer. Multimedia, entertainment, and interactive art applications all deal with an articulated transmission of a message, emotions, and encounters, rather than navigation and task execution. As we model human interaction through computer-based media, we need to be able to interpret a person's gestures, movements, and voice, not simply as commands to virtual creatures but as cues that regulate the dynamics of an encounter, or the elements of a conversation.

Researchers have used a variety of approaches to give lifelike qualities to their characters and to give them the ability to interact and respond to a user's commands. Blumberg and Galyean[5] use an ethological model to build behavior-based graphical creatures capable of autonomous action, and who can arbitrate response to external control and autonomy. They introduce the term "directability" to describe this quality. Hayes-Roth[11] uses the notion of directed improvisation to achieve a compromise between "directability" and lifelike qualities. Her research aims at building individual characters that can take directions from the user or the environment and act according to those directions in ways that are consistent with their unique emotions, moods, and personalities (improvisation). Magnenat-Thalmann and Thalmann[12] have built a variety of examples of virtual humans equipped with virtual visual, tactile, and auditory sensors to interact with other virtual or real (suited or tethered) humans.[13] Perlin and Goldberg describe an authoring system for movement and action of graphical characters.[9] The system consists of a behavior engine that uses a simple scripting language to control how actors communicate and make decisions, and an animation engine that translates programmed canonical motions into natural noisy movement. Terzopoulos provided a fascinating example of behavior-based graphical fishes endowed with synthetic vision and able to learn complex motor skills.[14,15] Tosa has built characters that can understand and respond to human emotion using a combination of speech and gesture recognition.[16] Bates and the Oz group have modeled a world inhabited by "woggles" with internal needs and emotions and capable of complex interactions with the user.[17] The users' interface, though, is mainly limited to mouse and keyboard input.

Before discussing extensions or alternatives to the behavior-based approach, we need to analyze the type of problems with which we are faced when authoring systems in our field of research. In this subsection we provide a taxonomy of interactive systems based on the human-machine and human-content interaction modality and the system architecture. This taxonomy does not pretend to be exhaustive. It provides, however, a focus in defining a set of basic requirements, features, and architectures of current interactive media applications. We suggest classifying interactive systems as: *scripted, responsive, behavioral, learning*, and *intentional*. We consider our field to encompass multimedia communications—the world of digital text, photographs, movie clips, sounds, and audio—electronic art, (interactive) performance, and entertainment in general.

- In *scripted systems* a central program coordinates the presentation of visual or audio material to an audience. The interaction modality is often restricted to clicking on a static interface in order to trigger new material to be shown. These systems need careful planning of the sequence of interactions with the public and acquire high complexity when drawing content from a large database. This authoring complexity often limits the experience to a shallow depth of content and a rigid interaction modality. Examples of scripted authoring tech-

nique can be found in Sawhney et al.[18] and in Agamanolis and Bove.[19]

- In *responsive systems* control is distributed over the component modules of the system. As opposed to the previous architectures, these systems are defined by a series of couplings between user input and system responses. The architecture keeps no memory of past interactions, at least explicitly, and is event-driven. Many sensor-based real-time interactive art applications are modeled according to this approach. One-to-one mappings define a geography of responses whose collection shapes the system architecture as well as the public's experience. Although somewhat easier to author, responsive experiences are sometimes repetitive: The same action of the participant always produces the same response by the system. The public still tends to adopt an exploratory strategy when interacting with responsive systems and, after having tried all the interface options provided, is often not attracted back to the piece. Sometimes simple responsive experiences are successful because they provide the participant with a clear understanding of how their input—gestures, posture, motion, voice—determines the response of the system. The prompt timing of the response is a critical factor in being able to engage the public in the experience. Examples of responsive systems are described by Davenport et al.[20] and Paradiso.[21]

- In *behavioral systems* or environments the response of the system is a function of the sensory input *as well as* its own internal state. The internal state is essentially a set of weights on the goals and motivations of the behavioral agent. The values of these weights determine the actual behavior of the agent. Behavioral systems provide a one-to-many type of mapping between the public's input and the response of the system. The response to a particular sensor measurement or input is not always the same: It varies according to the context of the interaction that affects the internal state of the agent. Successful behavioral systems are those allowing the public to develop an understanding of the causal relationships between their input and the behavior of the agent. Ideally, the public should be able to narrate the dynamics of the encounter with a synthetic behavioral agent as they would narrate a story about a short interaction with a living entity, human, or animal. This is one of the reasons why behavioral agents are often called *lifelike* creatures.[8,9]

- *Learning systems* have the ability to learn new behaviors or to modify the existing ones by dynamically modifying parameters of the original behaviors. These systems provide a rich set of interaction modalities and dynamics and offer new interesting venues for interactive media architectures.[14,22]

- *Intentional systems* are modeled according to a new way of thinking about authoring interactive media, which we present briefly in this section and expand on later. We introduce an additional layer in the one-to-many mapping between sensory input and system response, called the perceptual layer. Sensor data are first interpreted by the system as a "percept" and then mapped to an action selected by the behavior system. Both the interpretation and the behavioral mechanisms are influenced by the personality of the agent. The agent generates expectations of the public's behavior and therefore "feels" frustrated or gratified by its experience with people. The intermediate layer of perceptions provides the agent with an interpretation of the interactor's intentions and can be considered as a primitive "user model" of the system.

  The intentional approach allows the system to simulate more closely the dynamics of a human encounter, such as the communication of emotion.

These architectures are not mutually exclusive. They describe the *main* concept, structure, and organization of the system. However, a behavioral system can also learn or eventually scale to be simply responsive, according to the context of the interaction with the participant.

Intentional systems. In this subsection we introduce a new media modeling technique for authoring interactive experiences. We describe "media actors": images, video, sound, speech, and text objects able to respond to humans in a believable, esthetical, expressive, and entertaining manner. We call applications built with media actors intentional systems. In constructing our agents, we have shifted our focus of attention away from a Skinner-like "reflex-based" view of behavior, and we have moved toward building a model of perceptual intelligence of the agent.

Media actors are modeled as software agents whose personality affects not only their internal state (feelings) but also their perception of the public's behavior (intentions) and their expectations about future interactions with their human interactor.

Our media architecture is inspired by a theatrical metaphor. In theater, the director works with the actors with the goal of drawing the public into the story. In a compelling performance, the actors convey more than an appropriate set of actions; rather, they create a convincing interpretation of the story. The audience becomes immersed in the performance, since they are able to identify, project, and empathize with the characters on stage. According to our theatrical metaphor, we see media agents as actors, the programmer or artist as the director of the piece, and the public as a coauthor who, through interaction, gives life to the characters represented by media objects.

We believe that interpretation is the key not only to compelling theater but also to successful interactive media. Media actors are endowed with the ability to interpret sensory data generated by the public—room position, gestures, tone of voice, words, head movements—as intentions of the human interactor. These intentions—friendly, unfriendly, curious, playful, etc.—can be seen as a projection of the media actor's personality onto a map of bare sensory data. The media actor's internal state is given by a corresponding feeling—joy, fear, disappointment—which, in turn, generates the expressive behavior of the agent and its expectations about the future development of the encounter. In this personality model, feelings reflect a variety of emotional and physical states that are easily observed by the public, such as happy, tired, sad, angry, etc., whereas expectations—gratification, frustration, or surprise—stimulate the follow-on action.

Media actors are endowed with wireless sensors to allow natural and unencumbered interactions with the public. Real-time computer-vision and auditory processing allow the interpretation of simple and natural body gestures, head movements, pregiven utterances, and tone of voice. In this type of architecture the sensors are not a peripheral part of the system. On the contrary, the available sensor modalities, as well as their coordination, contribute to the model of the perceptual intelligence of the system.

In line with our theatrical metaphor, media actors are like characters in search of an author as in Pirandello's well-known drama.[23] They are media with a variety of expressive behaviors, personalities whose lifelike responses emerge as a result of interacting with the audience.

At every step of its time cycle a media actor does the following:

- It interprets the external data through its sensory system and generates an internal perception filtered by its own personality.
- It updates its internal state on the basis of the internal perception, the previous states, the expectation generated by the participant's intention, and its own personality profile.
- It selects an appropriate action based on a repertoire of expressive actions: show, move, scale, transform, change color, etc.

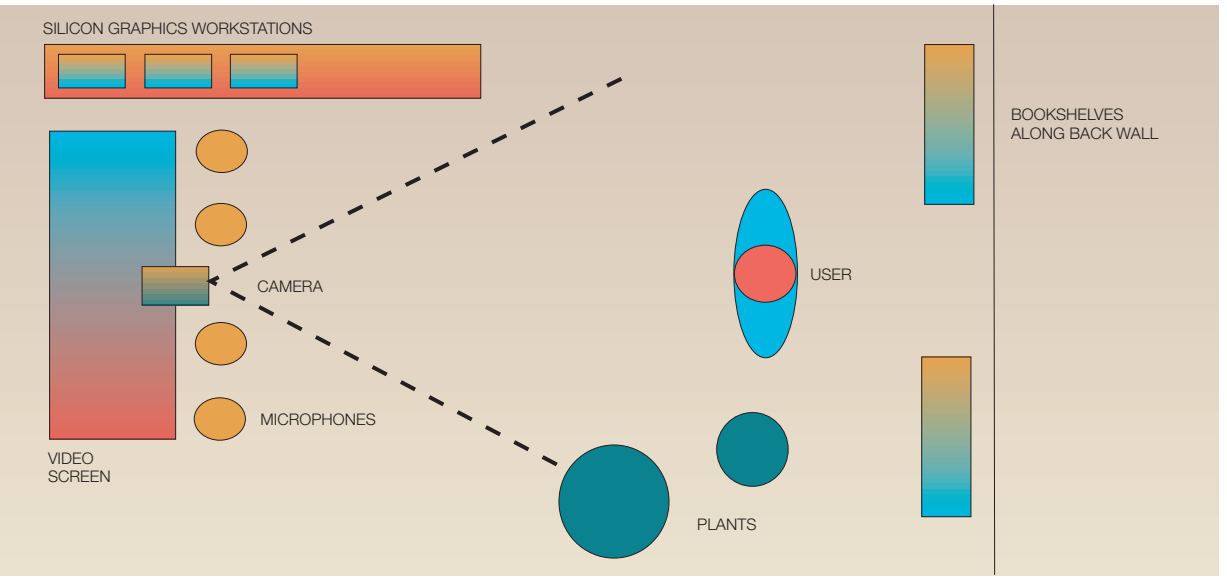
Our model is *sensor-driven*—which explains why we call it perceptual—and *personality-based*, rather than behavior-based. By personality we designate the general patterns of behavior and predispositions that determine how a person will think, feel, and act. We have modeled feelings rather than emotions because we consider emotions to be always in response to some event, whereas feelings can be assimilated to internal states. The internal state of a media actor can then be described with the answer to the question: "How are you feeling today?" or "How are you?"

This type of character modeling for multimedia differs from both scripted and purely behavior-based (animat) approaches. With respect to the classical animat behavior-based approach we introduce:

- A *perceptual layer* in which the sensorial input is translated into a percept that helps define a "user model" as it contributes to the interpretation of the participant's intention
- A notion of *expectation* that the media actor needs to have about the participant's next action so as to model the basic reactions to an encounter such as gratification or frustration
- A notion of *goal* as a desire to communicate, that is, to induce an emotion or to articulate the transmission of a message
- An *internal state* intended as "feeling" that generates an expressive action

The importance of having an intermediate layer of sensory representation, and predictions, has also been underlined by Crowley.[24] However, Crowley's architecture is limited to the construction of reactive visual processes that accomplish visual tasks and does not attempt to orchestrate media to convey a message or to animate a lifelike virtual character.

Figure 1 The IVE stage, schematic view from above



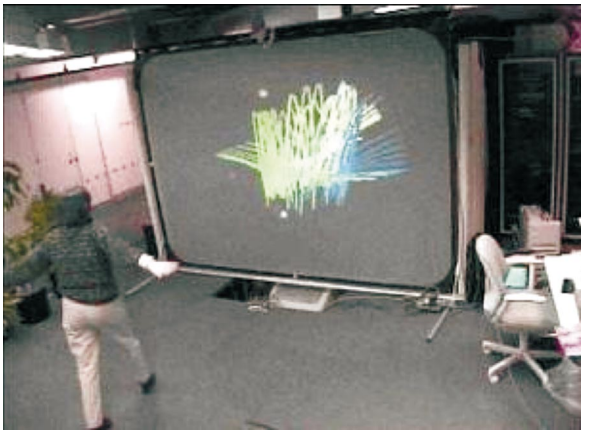
We now describe the interactive spaces we built using media actors and those that have inspired and led us to this authoring method.

The IVE stage and the real-time computer vision input system

Our stage, called "IVE" (Interactive Virtual Environment), is a room-sized area (15 ft × 17 ft) whose only requirements are good, constant lighting and a nonmoving background. A large projection screen (7 ft × 10 ft) occupies one side of the room and is used as the stage backdrop onto which we orchestrate graphical and image projections. A downward-pointing wide-angle video camera mounted on top of the screen allows the IVE system to track a performer (Figures 1 and 2). By use of real-time computer vision techniques,[25–27] we are able to interpret the performer's posture, gestures, identity, and movement. A phased-array microphone is mounted above the display screen for audio pickup and speech processing. A narrow-angle camera housed 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.

The ability to enter the interactive stage just by stepping into the sensing area is very important. The performers do not have to spend time "suiting up," cleaning the apparatus, or untangling wires. IVE was

Figure 2 The IVE stage during rehearsal



built for the more general purpose of enabling people to participate in immersive interactive experiences and performances without wearing suits, head-mounted displays, gloves, or other gear.[1,28] Remote sensing via cameras and microphones allows people to interact naturally and spontaneously with the material shown on the large projection screen. IVE currently supports one active person in the space and many observers on the side. The IVE space is a component of Pentland's Smart Rooms research work, and we often refer to it in the scientific literature as

Figure 3 Pfinder tracking the human body: head, torso, legs, hands, and feet labeled: gray, purple, red, green, orange, and yellow

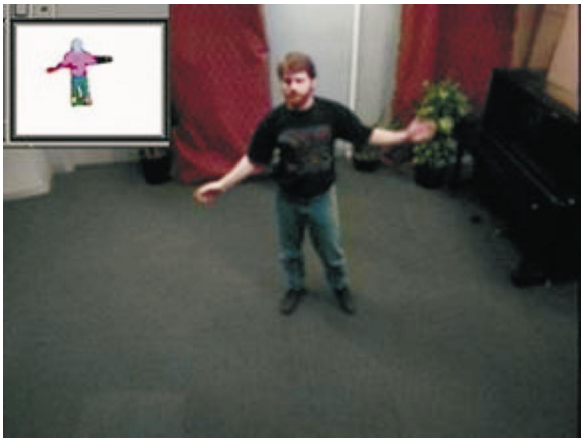
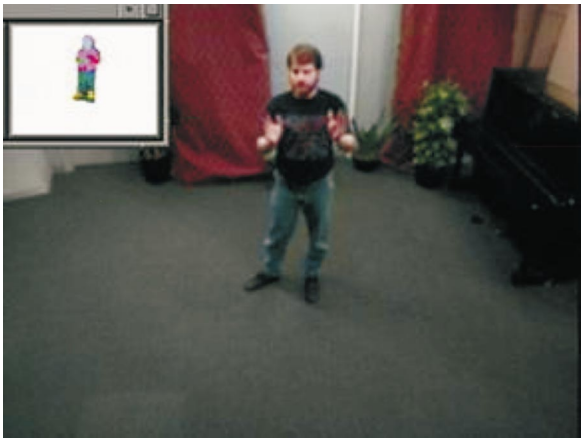


Figure 4 Continued tracking with Pfinder: notice correct hand tracking even when the hands are in front of the body



synonymous with Smart Room[29] or Perceptive Space.[1]

Our real-time computer vision program is called Pfinder, i.e., "person finder."[25] Pfinder is a system for body tracking and interpretation of movement of a single performer. It uses only a wide-angle camera pointing toward the stage and a standard Silicon Graphics O2** computer (Figures 3 and 4). We have recently extended the tracking technology to support many people or performers at once, using infrared tracking of the performers from above as we showed in the "City of News" Emerging Technology Exhibit at SIGGRAPH 99.

Performance

Our work augments the expressive range of possibilities for performers and stretches the grammar of the traditional arts rather than suggesting ways and contexts to replace the embodied performer with a virtual one. Hence, we call our research "augmented performance" by analogy with the term "augmented reality," which contrasts "virtual reality."

In dance, we have conducted research toward musical and graphical augmentation of human movement. We built DanceSpace, a stage in which music and graphics are generated "on the fly" by the dancer's movements. A small set of musical instruments is virtually attached to the dancer's body and generates a melodic soundtrack in tonal accordance with a soft background musical piece. Meanwhile, the performer projects graphics onto a large backscreen using the body as a paint brush. In this context, the role of the computer is that of an *assistant choreographer*: The system is able to improvise a soundtrack and a visual accompaniment while the performer is creating or rehearsing a piece using his or her body as the interface. This role is a great advantage when the choreographer wishes to create a dance performance based on the pure expression of body movements and not by following a prechosen musical score.

In theater, we have done work in gesture, posture, and speech augmentation. In Improvisational TheaterSpace, we create a situation in which the human actor can be seen interacting with his or her own thoughts in the form of animated expressive text projected on stage. The text is just like another actor—able to understand and synchronize its performance to its human partner's gestures, postures, tone of voice, and words. Expressive text, as well as images, extends the expressive grammar of theater by allowing the director to show more of the character's inner conflicts, contrasting action and thought moments, memories, worries, and desires, in a way analogous to cinema. We followed an interaction model inspired by street theater, the mime's world, and the improvisational theater in general, so as to bypass previously scripted and therefore constraining technological interventions. In this context, the role of the computer, with the use of intentional and expressive media actors, is that of a *co-actor* who col-

laborates with the performer in communicating with the public. Media actors can also represent the public's participation in the improvised performance. Many improvisational theater performances require and develop the public's directorial suggestions at specific plot points during the piece. Media actors can be used to incorporate some of these suggestions and become live improvisational characters, sharing the stage with the human actors.

In the following subsections, we give artistic and technical details on DanceSpace and Improvisational TheaterSpace, the virtual studio dance stage, and the Digital Circus.

DanceSpace. DanceSpace is an interactive stage that takes full advantage of the ability of Pfinder to track a dancer's motion in real time. Different parts of the dancer's body (hands, head, feet, torso) can be mapped to different musical instruments constituting a virtual body-driven keyboard. Moreover, the computer can recognize hand and body gestures, which can trigger rhythmic or melodic changes in the music. A graphical output is also generated from the computer vision estimates.[30,31]

It is an ideal example of a responsive experience with one-to-one mappings between sensory input—the dancer's hand, feet, head, or center-of-body movements—and system output (music and graphics).

The computer-generated music consists of a richly textured melodic base tune playing 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 generate an improvisational theme above the background track. In the current version of DanceSpace, 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, decreasing the sound as she moves 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. The musical instruments of both hands are played in a continuous mode (i.e., to go from a lower to a higher note, the performer has to play all the intermediate notes). The bells and the drums are, on the contrary, one-shot musical instruments triggered by feet movements. More specific gestures of

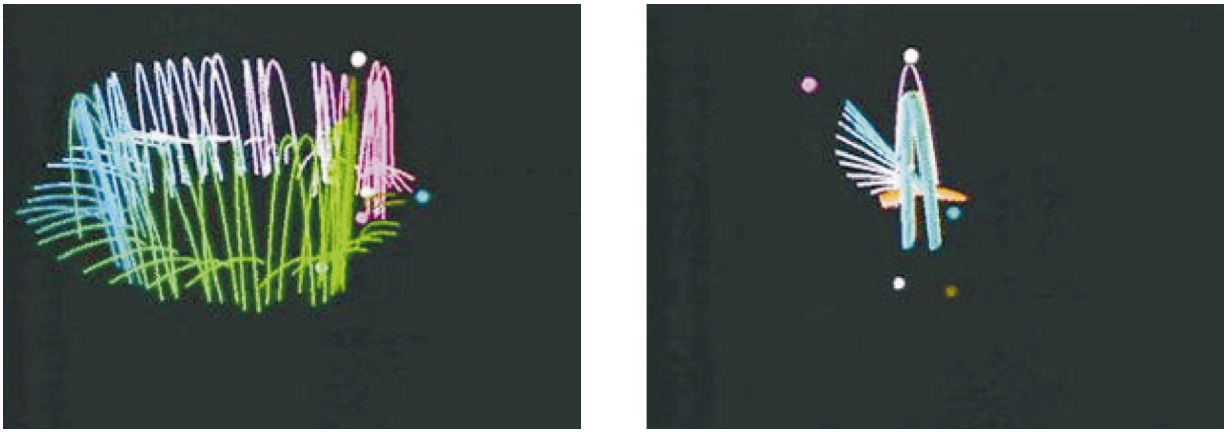
ALGORITHM

Pfinder uses a multiclass statistical model of color and shape to segment a person from a background scene, and then to find and track a person's body parts in a wide range of viewing conditions. It adopts a "Maximum A Posteriori" probability approach to body detection and tracking, using simple 2-D models. It incorporates *a priori* knowledge about people, primarily to bootstrap itself and to recover from errors.

Pfinder builds the scene model by first observing the scene without anyone in it. When a human enters, a large change is detected in the scene, which cues Pfinder to begin constructing a model of that person, built up over time as a dynamic "multi-blob" structure. The model-building process is driven by the distribution of color on the person's body, with blobs added to account for each differently colored region. Separate blobs are generated for the hands, head, feet, shirt, and pants. The building of a blob-model is guided by a 2-D contour shape analysis that recognizes silhouettes in which body parts can be reliably labeled.

The computer vision system is composed of several layers. The lowest layer uses adaptive models to segment the user from the background, enabling the system to track users without the need for chromakey backgrounds or special garments, while identifying color segments within the user's silhouette. This allows the system to track important features (hands) even when they are not discernible from the figure-background segmentation. This added information makes it possible to deduce the general 3-D structure of the user, producing better gesture tracking at the next layer, which 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 humans (heads are usually on top). The highest layer then uses these features, combined with knowledge of the human body, to detect significant gestures and movements. If Pfinder is given a camera model, it also back-projects the 2-D image information to produce 3-D position estimates on the assumption that a planar user is standing perpendicular to a planar floor. Several clients fetching data from Pfinder can be serviced in parallel, and clients can attach and detach without affecting the vision routines.

Figure 5 Graphical output of DanceSpace



both hands or combinations of hands and feet can generate melodic or rhythmic changes in the ambient melody. The dancer can therefore "tune" the music to her own taste throughout the performance. The music that is generated varies widely among different performers in 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.

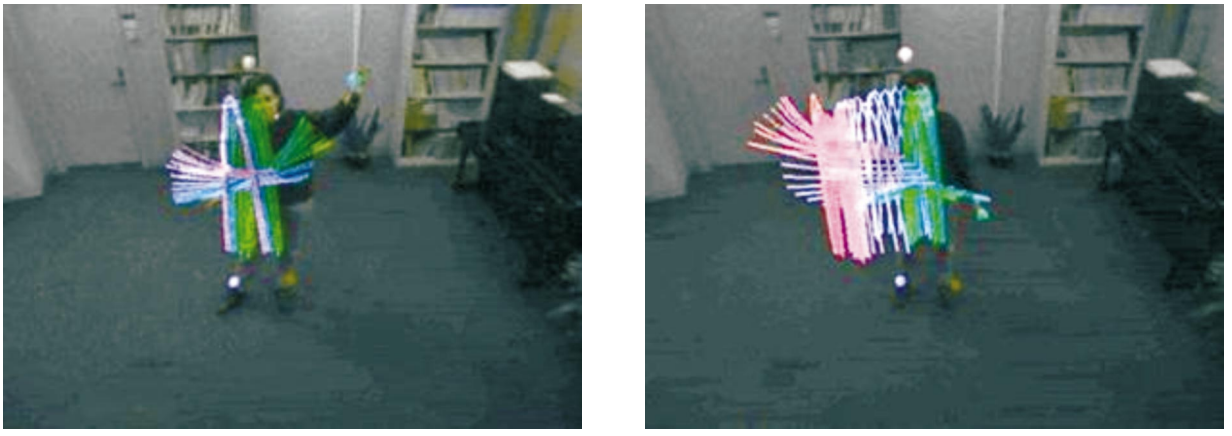
As the dancer moves, her body leaves a multicolored trail across the large wall screen that comprises one side of the performance space. The color of the trail can be selectively mapped to the position of the dancer on stage or to more "expressive" motion cues such as speed. The graphics are generated 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 three-dimensional (3-D) spheres are also drawn to map onto hands, feet, head, and center of the body of the performer. They serve as a reference for the dancer and accentuate 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 30), the dancer can paint more complex abstract figures on the screen (Figures 5 and 6).

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. When concentrating on music more than dance, DanceSpace can be thought of as a "hyperinstrument." Hyperinstruments[32] are musical instruments primarily invented for people who are not musically educated but 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.

The philosophy underlying DanceSpace is inspired by Merce Cunningham's approach to dance and choreography.[33] Cunningham believed that dance and movement should be designed independently of music, which is subordinate to the dancing and may be composed later for performance, much as a musical score is in film.

In the early 1980s, Ann Marion pioneered work on generating graphics through body motion in dance at the Massachusetts Institute of Technology (MIT) by use of polhemus sensors to track ballet steps.[34] Rolf Gehlhaar[35] has built a number of sound spaces where multiple users generate soundscapes through full body motion. Calvert et al.[36,37] have explored a variety of musical and graphical systems for inter-

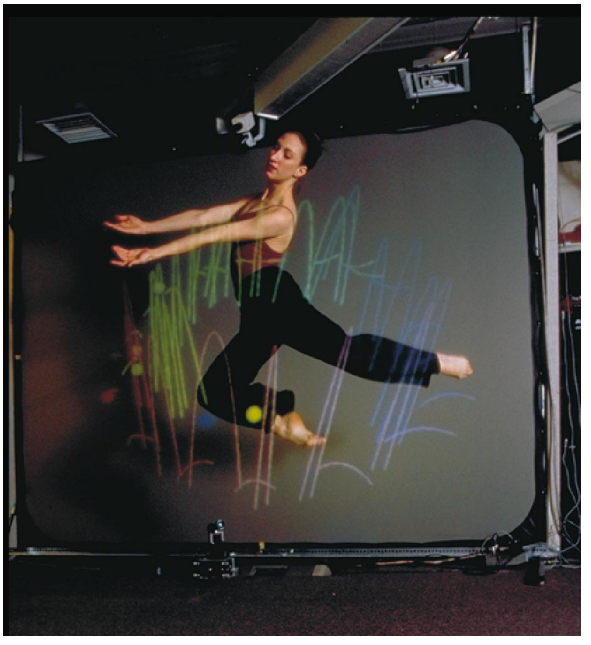
Figure 6 Performer using DanceSpace during rehearsal



active dance composition. Researchers at Georgia Institute of Technology[38] and its 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. DanceSpace differs from the aforementioned examples of interactive dance spaces because the computer-vision tracking system provides real-time information about different body parts of the dancers and not just an estimate 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 or gestures. The principal drawback of DanceSpace is that currently the computer-vision-based sensing technology reliably tracks only one performer at a time.

DanceSpace was first completed in February 1996. It has been tried by a large number of people and performers during several demonstrations at the MIT Media Laboratory, 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 choreographer Erica Drew (Figure 7). 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 so as to better exploit the

Figure 7 Performer Jennifer DePalo in DanceSpace



"comet" effect of the computer graphics trails. Nonperformers who attended the open house seemed to be more interested in exploring the space to obtain a desired musical effect.

We also used DanceSpace to create an original choreography together with choreographer Claire Mallardi at Radcliffe College, Harvard Extension, in the

Figure 8 Performer Diana Aubourg in DanceSpace

spring of 1996 (Figure 8). In this case we had prerecorded the virtual shadows generated by one dancer at the Media Lab in the IVE stage, and then projected them—noninteractively—on the dancers' bodies during the performance (Figure 9). Since 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 occurred on the backdrop of the stage where a parallel performance was taking place: 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 us to further explore this type of mixed-media performance.

Further improvements to be made to DanceSpace include having a large number of different background tunes and instruments available for the dancer to use within the same performance. We are currently expanding DanceSpace to allow for a greater variety of musical mappings and different graphical representations of the dancers.

The virtual studio dance stage. Virtual studios are new forms of TV video productions that combine real foreground images shot in the studio with 3-D computer-generated background scenes. Current virtual studio systems need the following: cameras shooting the foreground action against a blue background, tracking to provide camera position information, rendering to generate images with correct perspective in the virtual scene, and z-mixing to layer background and foreground correctly with respect to depth information.

Research in virtual studios is closely related to the work of building immersive and interactive virtual environments in which people can interact with virtual objects, with animated 3-D creatures, or among themselves. Ultimately these two venues of research converge toward the creation of low-cost, real-time, virtual studios in which the 3-D set is an interactive virtual environment and the compositing of multiple participants in the same set opens up new possibilities for TV and stage production, game playing, collaborative storytelling, or immersive telepresence.

By using the real-time computer vision techniques previously described (Pfinder), we have built a low-cost virtual studio that does not need a blue screen to operate. Hence, such a setup can be used on stage and allows one to perform special effects that are today only possible in the TV studio. In addition, we use computer-vision-based real-time gesture recognition to enable synchronized interaction among the performers and the objects and creatures in the virtual setting.

The type of dance performance we designed is one in which the entire stage becomes a large-scale IVE space with a very large projection screen for compositing. As the dancers move on stage, their image is composited in real time in a virtual 3-D set on the projection screen behind them, and the music is generated simultaneously. We modeled an enchanted forest to be the virtual set of our dance performance (Figures 10 and 11). Although our first application of virtual sets is one with circus performers[39] (shown later in Figures 15 and 16), we conducted an experimental study of a dance performance that virtually takes place in an enchanted forest (Figures 12 to 14). In this type of performance, the public sees the performers dancing inside the virtual 3-D set made by the enchanted forest and also, in conjunction with the dancers' movements, sees virtual butterflies and rabbits following and reacting to the performers, scale transformations, and animated trees dancing along, all of which contribute to create an Alice-in-Wonderland or Disney's "Fantasia" effect on stage.

Figure 9 Performance at Radcliffe College with choreographer Clair Mallardi and dancers: Malysa Monroe and Naomi Housman

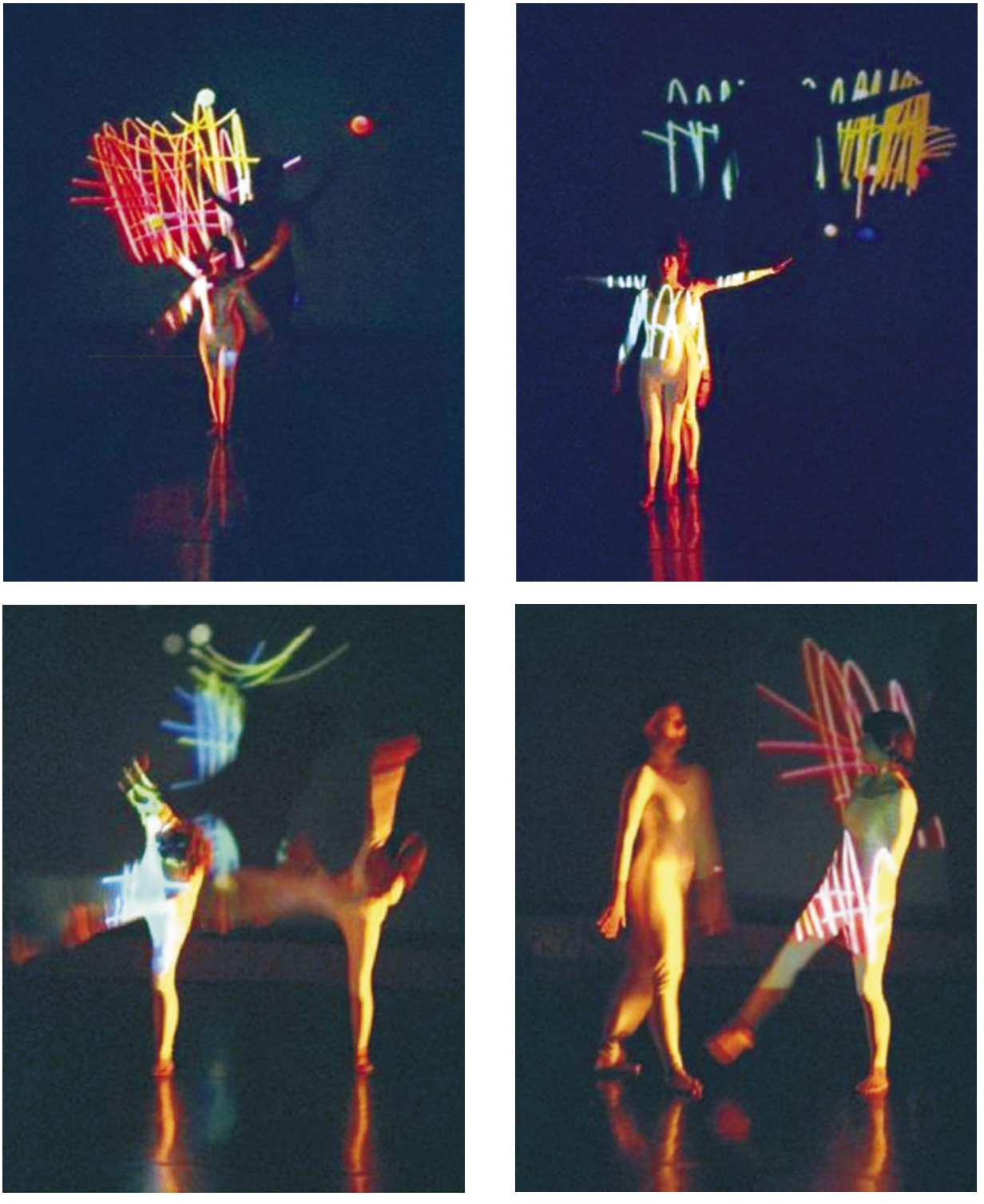


Figure 10 The enchanted forest world view

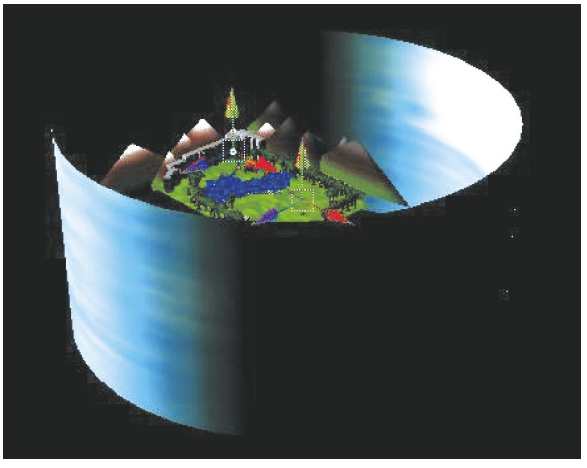
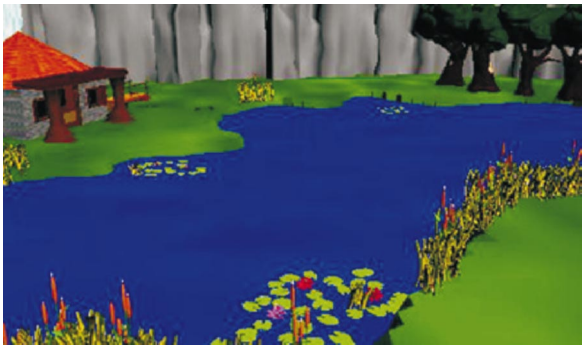


Figure 11 Surroundings of the cabin in the enchanted forest



Although real-time interpretation of the performers' actions is essential for interactivity, authoring a timely, appropriate, and coordinated response by the variety of virtual objects on the set can be a complex task. In order to facilitate authoring and to ensure a timely interaction, we have endowed the virtual objects and creatures that inhabit the 3-D studio set with autonomous responses and behaviors. By using a mixture of purely responsive and behavior-based AI techniques, we are able to distribute the authoring complexity from a central program to the various objects or characters in the set. Objects or creatures in the set can respond appropriately according to the context of interaction without our having to script in advance all the possible combinatorics of time and action sequences on the virtual set.

Our virtual studio dance stage is based on specific features of Pfinder that we now describe. The process of detection and tracking of the human participant is guided by a two-dimensional (2-D) contour shape analysis that recognizes a body silhouette inside which single body parts can be reliably labeled. This silhouette is used to cut out the foreground video image of the person composited in the 3-D set. This process, which achieves the compositing without blue screen, is done in two steps. First, a support graphical silhouette is constructed inside the 3-D environment, and successively the corresponding video image is texture-mapped onto it. Z-mixing is provided by giving Pfinder a camera model of the "home studio." The system back-projects the 2-D image transformation to produce 3-D position estimates using the assumption that a planar user is standing perpendicular to a planar floor. Back-projection ensures correct perspective as well as correct depth layering. Given that our virtual studio uses a fixed camera, we do not address the camera tracking issue.

Networked Digital Circus. Although many interactive experiences and video games tend to focus on exploring new worlds and killing monsters, we have developed a virtual studio application based on the theme of transformations. Participants are engaged in a game of transformations that involves the appearance or disappearance of objects, scaling the image of other participants very large or small, triggering events such as firing a cannon woman from a virtual cannon, or interacting with butterfly creatures that inhabit the set. All of these actions are interpreted and guided by the real-time gesture recognition and compositing features of our computer vision system as well as the responsive and behavioral authoring of the virtual set. We believe it is important to develop, together with the new technology, new genres of collaborative experiences that offer room for exchange, communication, and maybe transformation. Digital Circus moves some steps toward a more educational and artistic genre of networked game playing.[39]

We see the circus as a communication process with its own language. In the circus, information is transmitted mainly in a nonverbal, iconic, and symbolic way. A circus performer can give his or her act in any country, without the need to know much about the language or the culture of that particular country. In this sense he or she is a transcultural performer. Moreover, although the circus is usually connected to childhood, the circus performer usually

Figure 12 Study for a composited dance performance in real time for the virtual studio dance stage

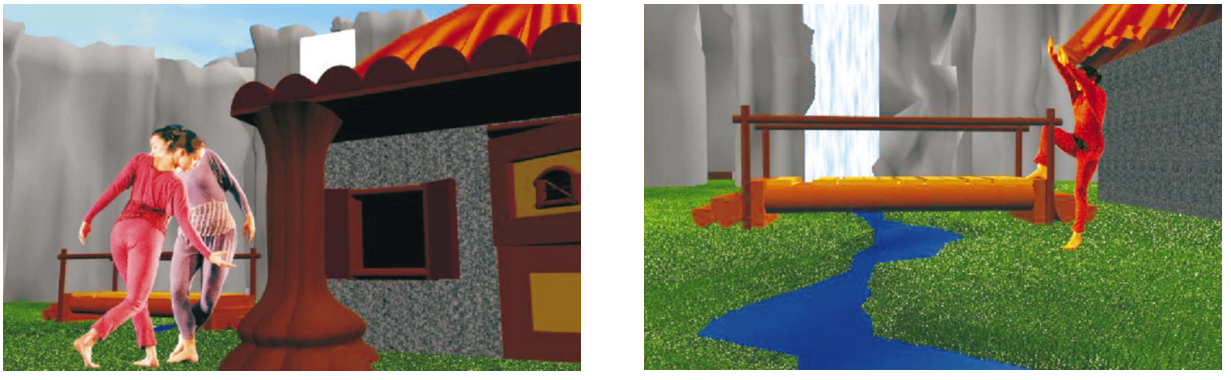


Figure 13 Study for a composited solo performance in the enchanted forest

addresses (and enchants) everyone, without any distinction of age.

The Digital Circus is composed of a basic architectural setting: a big top, a circus cannon, a chair, and a gramophone, together with a series of magic objects that appear or disappear according to the needs of the participant. This digital environment is shown on a large screen that comprises one side of our home virtual studio. The audience experiences the circus by seeing its video image projected onto the virtual environment. All the participants' images are shown in the virtual environment.

We start our game with our basic transformational circus grammar by placing a chair in the real space. This chair projects symmetrically onto a multicolored virtual chair in the Digital Circus. When a person from the audience sits on the real chair, he or

Figure 14 Performers dancing around the magic rock in the virtual enchanted forest

THE LANGUAGE OF THE CIRCUS

The language of the circus is essentially transformational. Objects have a meaning according to their role in the circus act. An elephant might use a phone, play a musical instrument, or eat a meal at the table, like a human. A clown produces sequences of inconsistent behavior. Similarly, the basic rules of balance are challenged or denied. When a circus object appears, it brings a whole set of possible meanings. This set depends on the "encyclopedia" of the audience, whereas its actual meaning is chosen depending on its role in the circus act. An example of this "functional sliding" of the meaning of circus objects is given by the chair. It is a defense tool for a lion tamer; it can be a "character" or a musical instrument for a clown; it is a co-performer for a tightrope walker that places it on a rope and then sits on it. When a chair is presented just as a chair, it is usually for horses and elephants to sit on and it functions as a humanizing factor for the animals.

The circus is a transgressive territory: transgression of laws of ordinary physics by the acrobats (or transgression of the laws of physics for ordinary people), transgression of politeness by the clowns, transgression of perception by the magicians, transgression of roles by the animals. Normally, lions are not mild, donkeys do not do arithmetic, ordinary people do not walk on tightropes, elephants do not eat at the table, objects or people do not disappear or appear or transform in an instant, people do not throw cakes at each other's faces if they disagree. In this respect, the cartoon is a genre close to the circus. There are a specific cartoon physics, cartoon sense of humor, and cartoon description of a character. Cartoon characters are close to humans but not quite humans. They have magic powers, follow laws of physics that violate regular physics, and usually interact in a way that would be judged violent or impolite in our regular way of interacting with other people. This process of transformation, inversion, parody, and caricature of the mainstream culture and its objectives and values produces an ambivalent attitude toward the circus. A strong interest is often blended with a scornful smile. For this reason, the circus is usually mostly appreciated by those individuals that have not been completely integrated within a culture, but instead are marginal, such as children, artists, or poets. These are among the reasons that make the circus an interesting playground for our research: its structural transformational nature and its elected audience. In this respect we believe it can appeal to an audience at least as large as that of the current consumers of videogames, if not larger. Its appeal would also be based on the fact that, along with Bouissac,[40] we see the circus not as a collection of acts but as a set of transformation rules that can generate a potentially infinite series of acts.

Once we have established that the circus communicates through its own *gestural* language, we still need to ask what is the narrative structure of a circus act and what is its relationship with the audience in the process of co-construction of meaning. Most circus acts evolve through successive steps that bear close resemblance to the development of the folktale.[40] They are, with some variance: identification of the hero (performer), qualification trial, principal trial, glorification trial (usually preceded by drum playing), and final recognition from the public. We have so far established a grammar of interaction among participants and among participants and the virtual set. Our current work is aimed at endowing the digital circus with a more robust story architecture.

she will see his or her video avatar sitting on the virtual chair. The system then recognizes the sitting gesture and triggers "magic events." A gramophone suddenly appears and starts to play music. If the person stands up again, the gramophone disappears, and the music stops.

This intrinsic transformational nature of the circus makes the circus story similar to an "open-eyes dream." In particular, the inversion of cause-effect relations, the "functional sliding" of objects, as well as the iconic-symbolic representation of most elements, recalls the "logic of dreams." To convey this point, we have programmed a transformation rule into the system that pops up an umbrella when a member of the audience raises a hand, as though he or she was actually holding an umbrella. Our system also recognizes a flying gesture that virtually lifts the audience up in the clouds, operating a digital plane. This gesture operates literally as a 3-D scene cut that radically transforms the environment and drags the audience into another dream.

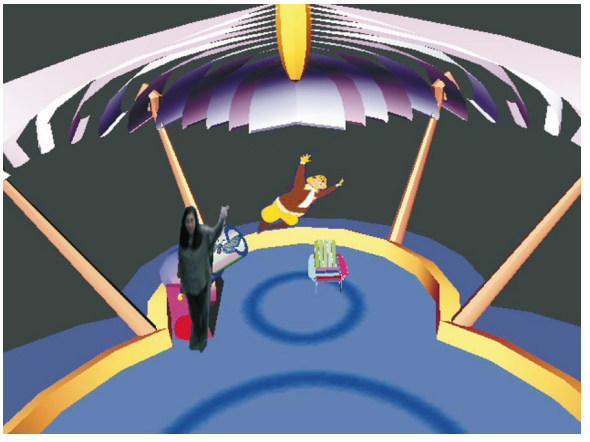
Two people in the circus can "play the clown" with each other (Figure 15). The first one who touches a "magic box" and then lifts or lowers his or her arms causes the other participant to grow very tall or very short. When short, the participant needs to hide so as not to be kicked by the other clown, otherwise he or she is transformed into a ball. A virtual butterfly can "save" the small clown and bring him or her back to original size. If one participant is holding the umbrella and the other touches the magic box, the umbrella lifts the person into the air. Touching the flying butterfly will bring the participant back to the ground. Each participant has his or her own butterfly to help when needed. In order to gain more magic powers or to confuse the other player, a participant can create one or more clones of himself or herself on the set. A clone is an exact replica of the composited image of a participant and moves along with the "main" image of the person, but at a different position in the 3-D set.

As of now, the main character of our circus is a cannon woman (Figure 16). She is fired out of the circus cannon simply by pushing a button situated on the cannon. She allows us to reach a wider audience than the one physically gathered around the home virtual studio. She is fired from the cannon in the Digital Circus and lands on the World Wide Web (WWW) on an applet that can be viewed by other networked participants. Through characters like the cannon woman, who are distributed across the different virtual networked home studios, we hope to gather an active audience of collaborating and engaged participants.

In order to create a compelling experience, it is important not only to design a visually rich 3-D set, but also to give participants the ability to modify and interact with the virtual objects or synthetic creatures inhabiting the set. Authoring an interactive game for multiple participants can be a very complex task, unless an appropriate methodology of modeling interaction is used. Scripted systems are those in which a central program rigidly associates people's actions or inputs with responses of the system by keeping a careful accounting of timing constraints and sequencing. Such systems are in general to be avoided because authoring complexity grows very fast if a variety of interactive features are implemented. In addition, they do not allow for spontaneous and unexpected actions or interactions to happen and are therefore inadequate for many real-life situations such as "live shows" or interactive games.

Figure 15 Two remotely connected participants in the circus with their respective butterflies

Figure 16 The circus virtual studio with participant pointing toward the cannon woman



We have bypassed the complexity of scripted systems by building a responsive application that is defined by a series of couplings between the participant's input and the response of the system. Responsive applications are easier to author because they do not have a central program that takes into account all the sequenced inputs, actions, and outputs. They define instead a geography of one-to-one mappings for which the participant's actions trigger specific system responses. Although easier to author, these systems do not take into account complex story

structures and elaborate timing or sequencing information.

All the objects present in the circus are embedded with their own set of responses—the cannon, the umbrella, the chair, the gramophone, the cannon woman, and even the set itself—and participate autonomously in the game of transformations, appearances, and disappearances regulated by the participants' actions. We have authored the virtual butterfly character instead by using a behavioral approach. In particular, the butterfly senses the size of each person's image and location on the set.

Improvisational TheaterSpace. In April 1996 we created the Improvisational TheaterSpace.[30] Improvisational TheaterSpace provides us with an ideal playground for an IVE-controlled stage in which embodied human actors and media actors generate an emergent story through interaction among themselves and the public. An emergent story is one that is not strictly tied to a script. It is the analog of a "jam session" in music. Like musicians who play, each with their unique musical personality, competency, and experience, and create a musical experience for which there is no score, a group of media actors and human actors perform a dynamically evolving story. Media actors are used to augment the play by expressing the actor's inner thoughts, memory, or personal imagery, or by playing other segments of the script. Human actors use full body gestures, tone of voice, and simple phrases to interact with media actors. Among the wide variety of theater styles and plays, we have chosen to stage improvisational theater. This genre is entertaining and engaging and allows the audience to drive part of the play. An experimental performance using the IVE setup was given in late February 1997 on the occasion of the Sixth Biennial Symposium on Arts and Technology[41] (Figure 17). We also did a public rehearsal at the MIT Media Lab on the occasion of a Digital Life Open House on March 11, 1997 (Figure 18). Both featured improvisational actress Kristin Hall.

We conceived a theatrical situation in which a human actor could be seen interacting with his or her own thoughts appearing in the form of animated expressive text projected onto a large screen on stage. We modeled the text to be just like another actor, able to understand and synchronize its performance to its human partner's movements, words, tone of voice, and gesture. In a series of very short plays, we showed an actress in the process of interrogating herself in order to make an important decision. A media actor in the form of projected expressive text plays her "alter ego" and leads her to a final decision. The text actor has sensing abilities: It can follow the user around on stage, it can sense a set of basic gestures, and it understands simple words and sentences. Its expressive abilities include showing basic moods through typographic behaviors, such as being happy, sad, angry, or excited.

The stage has a large projection screen along one side and a small color camera aimed at the entire acting space. Using the video image from the camera, the screen is transformed into a mirror that also contains a superimposed expressive text actor. A camera and a microphone are used by the typographic actor as sensors for the real world and to interpret the performer's actions. The audience gathers around the space. The video image of the actor and typographic actor can eventually be broadcast to a larger audience. Any member of the audience can step onto the stage and take the place of the human performer to interact with the typographic actor at any time during the performance. Other scenarios can be envisioned in which performers and public play improvisational theater games with the digital actors.[42]

The idea of conveying emotions through typographic attributes of the text has been largely explored in graphic design.[43–45] The idea of introducing text as a character on stage was inspired by readings of Bakhtin[46] and Vygotsky.[47] Both authors have underlined the role of a dialogical consciousness. For Bakhtin our utterances are "inhabited" by the voices of others: We construct our utterance in the anticipation of the other's active responsive understanding. Vygotsky viewed egocentric and inner speech as being dialogic. We have staged theater work in which the text takes the active role of representing dialogically the character's inner thoughts while he or she is constructing an action or making a decision. Also, the text can represent the different layers of the anticipated response of the audience and, in special cases, the audience itself. By taking this approach, we investigate further along the direction traced by Brecht who, at crucial points of the performance, asked the audience to comment about a character's decisions and motivations. The text actor can also be thought of as a modern equivalent of the chorus in ancient Greek tragedy.[48] Just as in the ancient Greek theater, the chorus, or text actor, engages in a dialog with the "hero," represents the public's collective soul or thoughts, and comments,

Figure 17 Actress Kristin Hall during a performance at the Sixth Biennial Symposium on Arts and Technology

Figure 18 Improvisational performer Kristin Hall in the IVE stage, at the MIT Media Lab, during the 1997 Digital Life Open House

interprets, and highlights the actor's choices in the plot.[49]

Early examples of introducing text as part of a theatrical performance can be found in the Futurist theater. In 1914, Giacomo Balla wrote a futuristic performance piece called "Printing Press" (Macchina Tipografica).[50] In this theater piece, each of the 12 performers became part of a printing press machine by repeating a particular sound. Their movement would also reproduce the movement of the 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.

The software architecture that models the typographic actor has two main layers. One layer groups the basic skills of the actor, such as simple unitary actions it can perform. The above layer groups high-level behaviors that make coordinate use of the low-level skills to accomplish a goal. This type of architecture, which separates low- and high-level skills for animating synthetic creatures, was first introduced by Zeltzer[3] and Blumberg.[5]

The typographic actor also has 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 performing basic skills. Emotional states affect the manner of the pre-

THE TYPOGRAPHIC ACTOR

Following are high-level behaviors for the typographic actor: (1) say typical phrase, (2) attract attention, (3) show off, (4) entertain, (5) daydream, (6) tell a story or explain, (7) suggest what to do next, (8) follow actor (or any of his or her parts: head, hands, feet, center of body).

Behaviors coordinate the execution of basic skills such as: (1) set string, (2) read text from HTML (Web page), (3) read text from file, (4) set color, (5) set font.

Other basic skills occur through time and can happen in different time scales or forms according to *how* the action is executed, based on what emotion the typographic actor wishes to convey. These are: (6) fade to color, (7) glow, (8) scale, (9) jump, (10) goto, (11) rock.

Also, according to which behavior is acting, the text can read in different ways: (12) read word by word, (13) fade word by word, (14) fade letter by letter.

The actions executed by the different skills can be combined as long as they do not use the same graphical resources; i.e., for the color degree of freedom, "glow" cannot happen at the same time as "fade to color."

Time functions that will affect the way the basic 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 skills will be executed according to these time functions. Examples of emotional states of the text are: (1) happy, (2) sad, (3) angry, (4) scared, (5) disgusted, (6) interested, (7) loving, (8) surprised, (9) no particular emotional state.

For example, if the text is "surprised," the active time function will be "oscillating up." In the case of "happy," the biquadratic, exponential, and sinusoidal time function will be chosen during execution of time-dependent skills.

sentation by picking an appropriate time function for execution (which function), and the energy level of the typographic actor determines the speed of execution of the basic skills (how fast).

Examples of use of the text actor are: a word or short sentence rapidly crossing the projection screen at the height of the performer's head to show a thought crossing the performer's mind, a word jumping from one hand to the other of the performer to show a thought wanting to penetrate the performer's consciousness, text following the performer around on stage to show an obsessive thought, rapidly flashing words above the performer's head to show a variety of contrasting simultaneous thoughts, and animated text playing the alter ego of the performer, eventually driven by the public.

Ongoing progress is directed toward a more accurate IVE-based gesture recognition, exploring the challenges and advantages of multimodal interaction and rehearsing a variety of multibranching improvisational plays according to the suggestions of the audience. Inspired by Forsythe's choreographies,[51] we are also exploring modalities of use of the text actor in dance.

Through this project we learned that media actors are a promising approach to innovative theatrical performances for three main reasons:

1. Media actors (versus script-based theater) are a flexible tool, both in the case of the improvisational (or street) theater in general, or for classical scripted theater that the director and the actors need to interpret and, therefore, modify.
2. The system is tolerant of 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 theater can allow public participation either during or after the performance without requiring the participants to learn all the script in advance.

This approach allows the use of flexible media choreography and contrasts scripted or rule-based approaches. The main drawback of scripted media is that the director and the actor have to rigidly follow a script for the system to be able to work. For instance, it is not uncommon in theater for both the actors and the director to change the script either during rehearsals or even right before or during the final performance.[52] In our view, a rule-based, scripted system lacks the responsiveness that creative

artists demand. A scripted system cannot easily compensate for human errors or be responsive when some nonplanned "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.

Interactive museum exhibit design

Applications of technology to museums have so far mainly focused on making extensive and attractive Web sites with catalogs of exhibits. Occasionally these Web sites also present introductory or complementary information with respect to what is shown inside the physical space of the museum. However, unless the public is interested in retrieving specific information about an artist or artwork, they will end up spending time scrolling across photographs and text in static pages and likely will not be involved in an engaging or entertaining experience. Presenting large bodies of information in the form of an electronic catalog usually does not stimulate learning or curiosity.

Museums have recently developed a strong interest in technology since they are more than ever before in the orbit of leisure industries. They are faced with the challenge of designing appealing exhibitions, handling large volumes of visitors, and conserving precious artwork. They look at technology as a possible partner to help achieve a balance between leisure and learning, as well as to help them be more effective in conveying story and meaning.

One of the main challenges that museum exhibit designers are faced with is to give life to the objects on display by telling their story within the context determined by the other objects in the exhibit. Traditional storytelling aids for museums have been panels and labels with text placed along the visitors' path. Yet the majority of visitors express uneasiness with written information.[53] Usually time spent reading labels interrupts the pace of the experience and requires a shift of attention from observing and contemplating to reading and understanding.[53]

Another challenge for museums is that of selecting the right subset of representative objects among the many belonging to the collections available. Usually, a large portion of interesting and relevant material never sees the light because of the physical limitations of the available display surfaces.

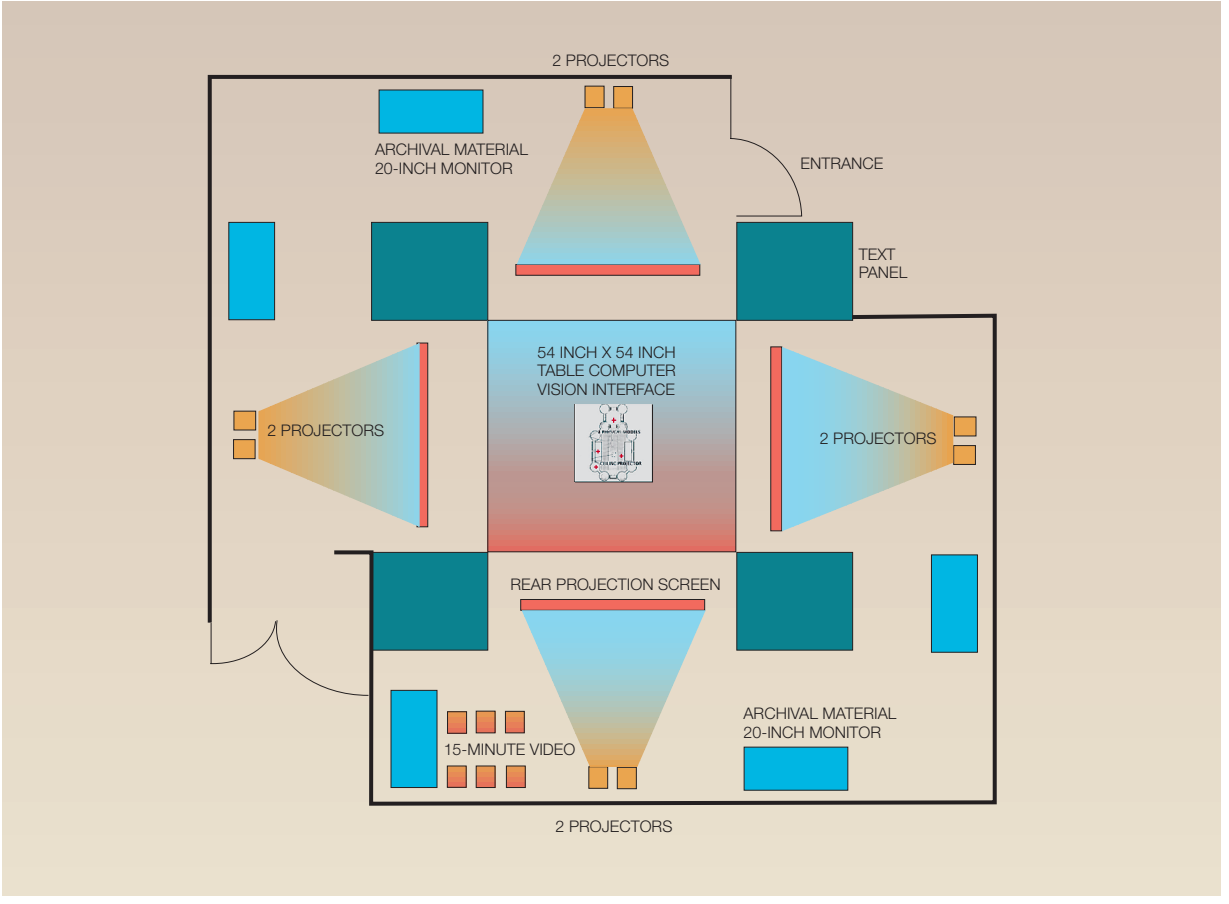
Some science museums have been successfully entertaining their public, mainly facilitated by the nature of the objects they show. They engage the visitor by transforming him or her from a passive viewer into a participant by use of interactive devices. They achieve their intent by installing button-activated demonstrations and touch-sensitive display panels, among other things, to provide supplementary information when requested. They make use of proximity sensors to increase light levels on an object or to activate a process when a visitor is close by.

Other museums—especially those that have large collections of artwork such as paintings, sculptures, and manufactured objects—use audiovisual material to give viewers some background and a coherent narrative of the works they are about to see or that they have just seen. In some cases, they provide audio tours through headphones along the exhibit. In others, they dedicate sections of the exhibit to the projection of short audiovisual documentaries about the displayed material. Often, these movies showing artwork together with a description of their creation and other historical material about the author and his or her times are even more compelling than the exhibit itself. The reason is that the documentary has a narration, and the visuals are well orchestrated and come with music and dialogs. The viewer is then offered a more unified and coherent narration than what would be available in the fragmented experience of the visit. A visit to a museum demands, as a matter of fact, a certain amount of effort, knowledge, concentration, and guidance for the public to leave with a consistent and connected view of the material presented.

On the basis of the above observations, we have identified two areas of technological intervention that contribute to engage the public and enrich its experience during a museum visit. They are: Information Overlay in Smart Rooms[29] (adding technology to the museum space) and Spatialized Interactive Narrative with Smart Clothes[54] (adding technology to the visitor). In this section we describe our museum work in the smart room/IVE space environment. The subsequent section will outline our museum work involving wearable computers (smart clothes).

Louis Kahn interactive: Unbuilt Ruins. By using a variety of wireless sensors (primarily cameras) and by placing audiovisual devices (projectors and speakers) in the museum area, we can use technology to virtually enlarge and augment the exhibit surface. It is then possible to select and show more objects from the ones available in the collection—in the form of images that can be virtually layered on one another

Figure 19 Schematic diagram of the Unbuilt Ruins exhibit



and selected by meaningful interaction with the public.

In January 1999, we created a museum experience to satisfy both the needs of the curator, who needs to be able to feature a great quantity of material in a limited physical space, and those of the viewer, who benefits from a coherent narration of the spatially fragmented objects on display. Prompted by architect Kent Larson and designer Ron MacNeil, we designed an exhibit space, called Unbuilt Ruins, to show a variety of architectural designs by the influential 20th-century American architect, Louis Kahn. The exhibit interactively features radiosity-based, hyperrealistic computer graphics renderings of eight unbuilt masterworks by Louis Kahn. It is housed in a large room, approximately 50 ft $\times$ 50 ft, and in the center contains a square table above which are mounted a camera and a projector pointing downwards. The table is surrounded by four large back-projection screens, parallel to the four sides and situated a few feet away to provide an ideal viewing surface. The table surface contains the projection of the image of the schematic floor plan of one of the eight projects and 3-D printed physical models of each of the projects, on bases 3 in $\times$ 3 in, located two to a side, at the perimeter of the table (Figure 19). Each model contains a digital tag that identifies it.

The schematic plan projected on the table contains graphics symbols indicating camera positions for the selected projects (hot spots). A standard color camera aimed at the table, linked to a computer running a real-time computer vision system, tracks a color-tagged cylindrical object (active cursor). A view is selected when the viewer places the active cursor

over a camera position symbol. When a view is selected, the side screens show a rendering of what a visitor would see if he or she were standing inside the physical construction in the location determined by the position of the symbol on the schematic plan and looking toward the direction indicated in the position symbols (Figures 20, 21, and 22). If no view is selected within one minute, the system automatically scrolls through all views of the selected project until a new view or architectural plan is chosen.

With the eight unbuilt projects presented in the exhibit, Kahn developed and tested ideas quite different from the modern architecture of his time: a configuration of space as discrete volumes, complex ambient light and deep shadow, a celebration of mass, the use of materials with both modernist and archaic qualities, monumental openings uncompromised by frames, and a concept of "ruins wrapped around buildings." By looking in depth at projects left unbuilt, the exhibit attempted to shed light on how Kahn came to develop the architecture he did.

Since these eight projects were incomplete and schematic, the design of each had to be resolved by extrapolating from archival material at the Kahn collection of the Architectural Archives at the University of Pennsylvania. A series of three-dimensional digital stage sets were created, and photographs from Kahn's Salk Institute, Exeter Library, and Yale Center for British Art were used to develop high-resolution texture map montages for each unique surface plane. Essential to the accurate interpretation of these projects was the physically accurate treatment of light. Images were rendered with the Lightscape** visualization system, a radiosity-based renderer, to capture the complex inter-reflections of ambient light in space.

The exhibit was open for one month at the Compton Gallery at MIT in February 1999, and later in the year for another month at the University of Pennsylvania's gallery exhibit space. It received enthusiastic feedback from the visiting public, architects, curators, and classrooms.

Unbuilt Ruins was also used as a model for a Museum of Modern Art (MoMA) exhibit called "The Unprivate House" curated by Terence Riley and shown at MoMA in the summer of 1999. We built an initial prototype in which we adapted the original setup we had at the MIT Compton Gallery to display architectural plans of houses projected on a table. A custom tag system, developed at MIT, was used to select which architectural plan would be explored by the visitor at one time. The tags were embedded inside postcards of the displayed houses, and when placed on the table, they would be detected, and the corresponding floor plan would be projected. The computer vision system tracked three objects in the form of small dolls. When visitors placed an object over a hot spot on the plan, they triggered the system to display photographs showing what would be seen from that point in the house, as well as text that would function as a guide for the exhibit. The three dolls represented the point of view of the curator, the architect of the displayed house, and the owner. Each of these dolls placed on a hot spot told the story of the exhibit from a slightly different viewpoint. The communication between the tag reader, the computer vision software, and the graphical display software was networked such that more communicating tables with displays would be able to exchange information, or be aware of the behavior or exploration of the other visitors to this exhibit (Figures 23 and 24).

Figure 20 Projected views from one hot spot from the architectural plan of the Meeting House of the Salk Institute, La Jolla, California, 1959–1965

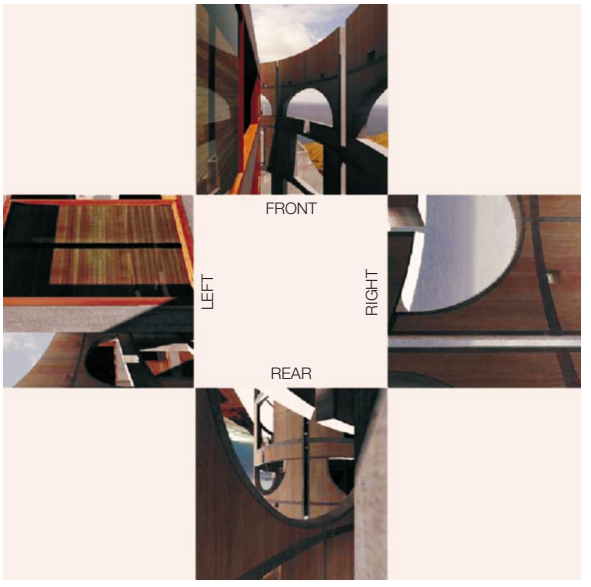


Figure 21 Images of the Unbuilt Ruins exhibit taken at the Compton Gallery at MIT

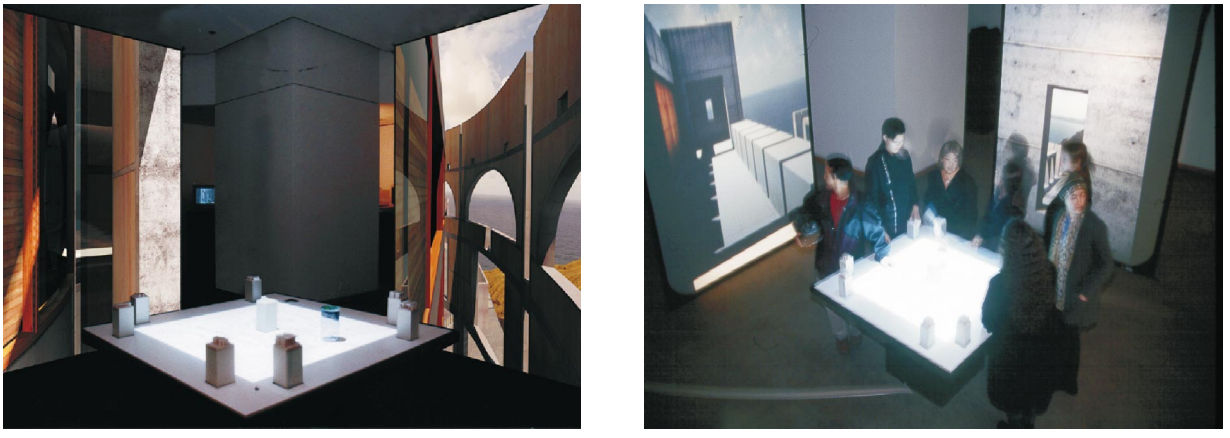


Figure 22 Visitors placing the active cursor on a hot spot on the map and discussing the displayed views

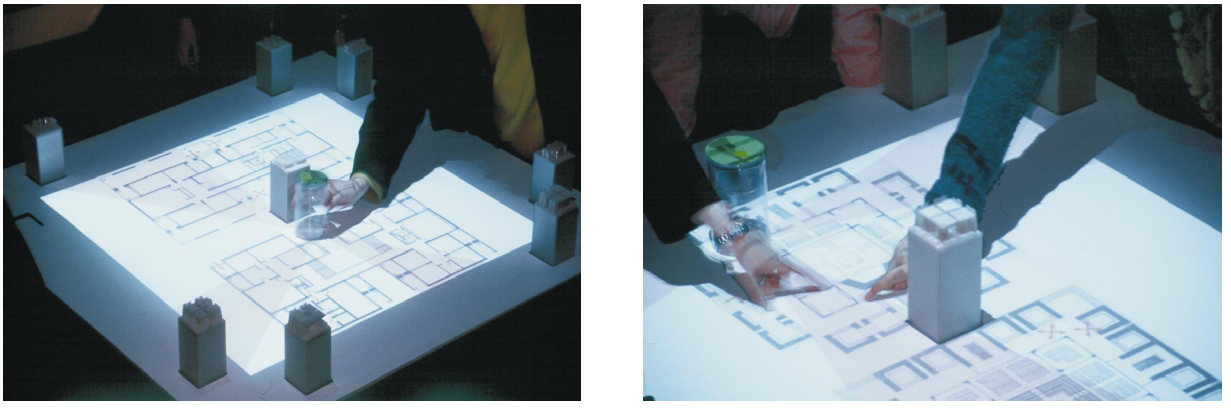


Figure 23 Views of one of the chosen houses for the prototype of "The Un-private House" interactive exhibit at MoMA

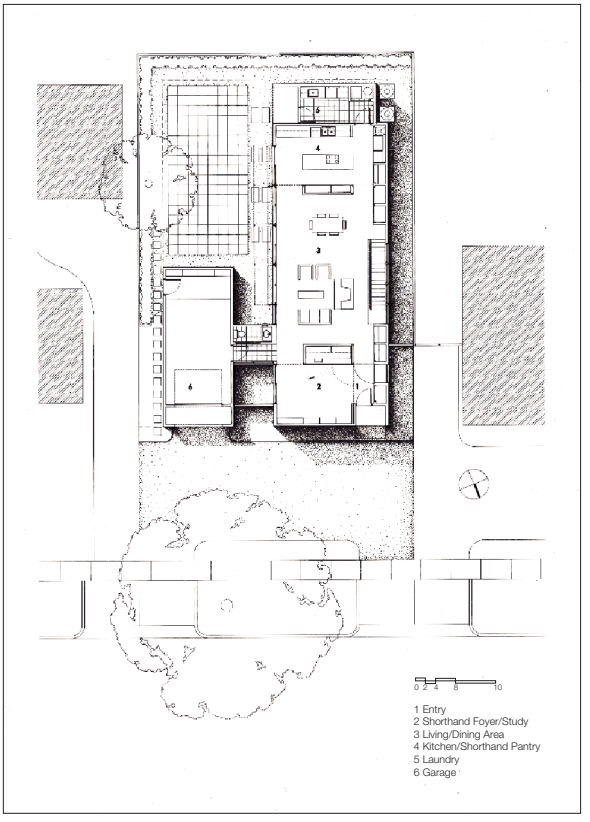
Wearables for museums and performance

A short time and technological leap separates today's thin laptops from Walkman**-sized computers that can be easily carried around by individuals at all times. Such small computers, accompanied by a high-resolution private eye for display and an input device, are already reshaping our technological landscape as they allow us to wear technology just as an element of our everyday clothing. Yet the consequences of offering an uninterrupted connection to the Internet, a TV or computer monitor embedded in eyeglasses, as well as a variety of input, sensing, and recording devices constantly active and attached to our body, is likely to have a profound impact on the way we perceive and relate to the world. Together with, and in addition to, our five senses we will soon have access to an additional dimension that will make us see, remember, access information, monitor our health and time schedule, and communicate in ways in which we had not thought possible before. Starner,[55] Pentland,[54] and Mann[56] describe early work in wearable computing and anticipate some future trends.

In this section we describe how "wearables" can be used in enriching the museum visit and enhancing the expressive capabilities of theater performers. We start by illustrating Wearable City, the precursor of all our wearable work. We then describe Wearable Cinema, which generates an interactive audiovisual narration driven by the physical path of the wearer in a museum space. Wearable Performance illustrates the wearable version of Improvisational Theater Space. All this work benefits from the media actors authoring technique described earlier.

Wearable City. Wearable City is the mobile version of a 3-D WWW browser we created, called "City of News." City of News[57] fetches and displays URLs (uniform resource locators) so as to form skyscrapers and alleys of text and images that users can visit as if they were exploring an urban landscape of information (Figure 25). The system is given a map of a known city at the start. When browsing, the text and images from Web pages are remapped into the facades of the virtual buildings growing from their footprint on the city map. The city is organized in districts, which provide territorial regrouping of urban activities. Similar to some major contemporary cities, there is a financial district, an entertainment district, and a shopping district. One could think of these districts as urban quarters associated with the different conceptual areas of one of the many currently available search engines on the Internet.

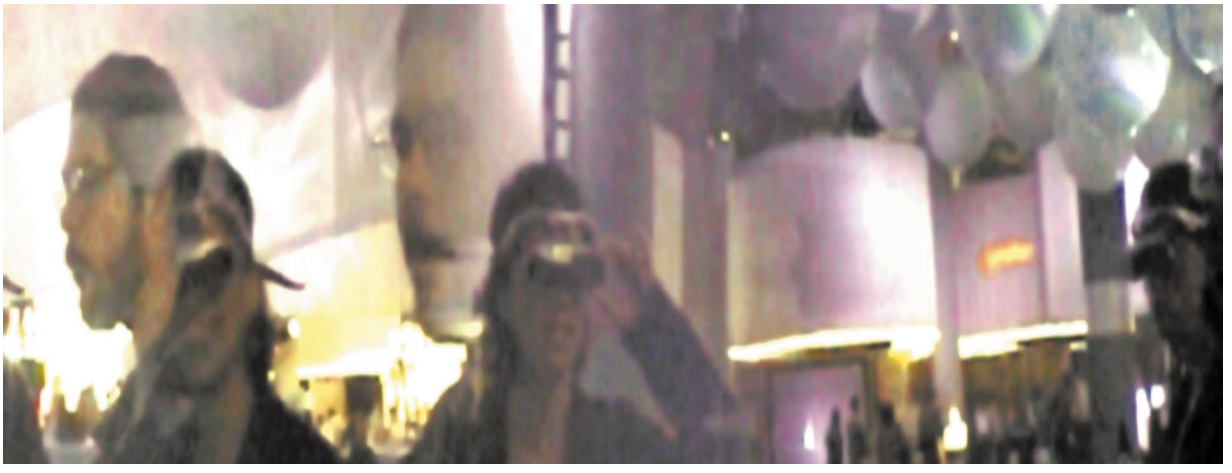
Figure 24 Floor plan of chosen house



To date we have grown cities of news from maps of Boston, New York, Stuttgart in Germany, the SIGGRAPH 99 floor map, and a few imaginary locations. The 3-D browser operates in three stages. At the start, the system analyzes the geometry of the given city map and saves it in memory. Then, when browsing, a parser fetches all the data from the target URL, using a TCP/IP (Transmission Control Protocol/Internet Protocol) socket and extracts all the available page layout information. The parser recognizes a large set of HTML (HyperText Markup Language) tags and builds a meta-representation of the page that contains the text and image data in one field and its formatting information in the associated field. This meta-representation is passed on to the graphical engine. The engine maps the text and images into graphics and textures, and reformats this information according to the real estate available

Figure 25 City of News, a 3-D Web browser that dynamically builds an urban landscape of information

Figure 26 Browsing using a "memory map"—the wearable plays movie clips interactively according to the wearer's path at the Millennium Motel of SIGGRAPH 99



from the assigned location on the map. The location is chosen automatically by an order defined on the map or manually by the user who makes a meaningful association with the given architecture. The software is written with the Open Inventor** graphics toolkit and C++ and runs under the Linux** operating system. A wireless network connection ensures a constant link to the WWW within its range.

We showed an early version of Wearable City at the SIGGRAPH 99 Millennium Motel demonstration floor[58] (Figure 26). A large number of people tried the system during the one-week exhibit and expressed considerable interest and curiosity.

The wearable is a jacket that has an embedded CPU, a sensing system for location identification, high-resolution color head-mounted display glasses, and a touch-sensitive threaded keypad as input (Figure 27). The jacket is a commercial denim long-sleeved jacket, decorated with an illuminated pattern on the back as embellishment.

In order to run high-end graphics on the wearable in 3-D, we have chosen to use an off-the-shelf CPU rather than a custom-made one. We selected a Toshiba Pentium** II, 366-MHz ultra-thin laptop and detached the liquid crystal display (LCD) screen and all unnecessary parts to minimize its weight. We

designed a custom lining to attach the CPU to the jacket, internally, on the back. We also added two side pockets to the lining to carry the batteries for the display and other devices when necessary.

The input system is a compact keyboard (Figure 28), sewn on one sleeve and made with touch-sensitive conductive thread. The keyboard is designed for general use with wearables, and contains as keys the numbers 0–9 and symbols representing VCR (videocassette recorder) controls, which can all be easily remapped for use in a variety of applications.

The display consists of commercial lightweight SVGA (super video graphics array) color head-mounted glasses (Figure 29), sold as non-see-through for immersive viewing. We modified the display for augmented reality applications by manually removing all the electronics from one eye (Figure 30). When wearing the display, after a few seconds of adaptation, the user's brain assembles the world's image of one eye with the display's image seen by the other eye into a fused augmented reality viewing.

The location system is made by a network of tiny infrared devices that transmit a location identification code to the receiver worn by the user and attached to the jacket. The transmitters are approximately coin-sized and are powered by small lithium batteries that last for about a week. They are built around a PIC** microcontroller, and their signal can be detected as far away as about 14 feet within a cone range of approximately 30 degrees.

Wearable Cinema. We envision and work toward making the museum visit indistinguishable from watching a movie or even a theater play. This movie is slightly different from the ones we are accustomed to: It unfolds a story for us as we wander around the museum space but is as engaging and immersive as traditional movies. Interactive movies have been a largely explored topic of research in multimedia and entertainment since the last decade. Davenport did early work that influenced this research field.[59,60] Yet many problems, such as the type of input device, the choice of breakpoints for interaction, and the fragmented and therefore nonimmersive experience that results from interaction, are still unsolved. In some cases, a multithreaded plot appears unjustified and unsatisfactory, especially for fiction, because people expect to see a "meaningful conclusion"—the moral of the story—and a well-edited visual narrative, both of which are hard to do interactively.

Figure 27 Wearable computer: jacket and display

Figure 28 Touch-sensitive keypad

The museum context provides a great platform of experimentation for interactive documentaries, since the interactive input is "naturally" and seamlessly provided by the visitor's path inside its aisles. Thus, it does not require pauses, breakpoints, or loops in the content presentation. It is therefore logical to fuse together the audiovisual documentary that illustrates and extends an exhibit with the visitor's path inside that exhibit, using a wearable computer. We create a new type of experience making the visit indistinguishable from seeing a movie from inside the movie set. This immersive experience cannot be achieved by a simple associative coupling between inputs and outputs. It requires a "perceptive layer" that constantly monitors the input and creates an output having a model of the user, its goals and an "understanding" of the content itself, the logical con-

Figure 29 SVGA color head-mounted display screen, with a view of the SIGGRAPH 99 Wearable City demo

Figure 30 Modified Sony SVGA color Glasstron™ display

nection among the parts, and their emotional impact on the user. We previously described perceptive media modeling in a June 1999 paper.[61]

Our current work transforms our research lab into a museum space. We have gathered a variety of historical footage and authored an interactive presentation for a wearable computer using a Wearable City 3-D graphics presentation to situate the user in the space. The audiovisual presentation of the footage and its description are authored using Macromedia's Flash** authoring environment. A perceptive media modeling of the content unfolds the wearable cinema as the visitor walks around the space, and the camera attached to the wearable recognizes its presence in specific locations or relevant objects.

Oliver et al.[62] developed a wearable computer with a visual input as a visual memory aid for a variety of tasks, including medical, training, or education. This system allows small chunks of video to be recorded and associates them with triggering objects. When the objects are seen again at a later moment, the video is played back. Wearable Cinema differs from this application in many ways. Its scope is to create a cinematic experience and to situate it within the containing architecture. Its focus is to orchestrate content and to carefully construct an immersive experience guided by the perception of the senses of the wearable, using perceptive media modeling. As opposed to the cited application, Wearable Cinema is not a simulation running on a desktop computer connected to a head-mounted display. It actually runs on a wearable especially designed for it, and the computer vision runs in real time on the wearable CPU.

The main distinctive characteristic of this wearable setup is that it uses real-time computer vision as input for easier and faster location finding. This characteristic has the additional advantage that no time needs to be spent distributing the infrared location emitters around the space and substituting the batteries when necessary. A quick training on the locations or objects to be recognized is the only setup needed for the computer vision system.

The wearable is made by sandwiching two CPUs together (Figure 31). One is dedicated to processing the input and the other to producing the output shown on the wearable display. We use a thin ungeared Pentium II Toshiba laptop, stripped down to its motherboard, connected to a super-thin Sony VAIO** Pentium II on a local network. The Toshiba runs the Linux operating system and is used for real-time computer-vision sensing. The VAIO runs the Microsoft Windows** 98 operating system and uses a combination of Flash animations and Open Inventor 3-D graphics to generate the cinematic experience.

These two very thin and lightweight computers are hosted inside a stylized backpack. The wearable is connected to a small wide-angle camera worn on the user's shoulder and to a high-resolution SVGA display.

The computer vision system uses a combination of color histograms and shape analysis to identify objects or locations. Segmentation of regions to be analyzed is achieved by Gaussian modeling of the target color, as in Wren et al.[25] Shape analysis is based on contour extraction and calculation of centralized moments on the contour points. The connection between the input and output software is done using the remote procedure call (RPC) protocol. The vision program runs as a server and communicates the result of its processing to a Java** client through RPC. On the other CPU, a Java server broadcasts this information to the Flash program via JavaScript. The Open Inventor graphics connects to the vision program on the other CPU directly with RPC.

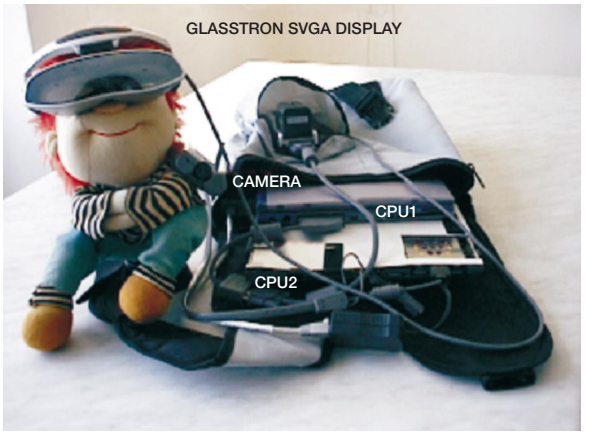
Wearable performance. Wearable computers will determine a migration of the computational engine from the house or the laboratory onto the users themselves. An analog to this transformation can be found in the transition from the drama played in the theater building to street theater.

Street and outdoor performance has a long historical tradition. Its recent form is motivated by the need to bring performance art to the people rather than people to the theater. We have found it therefore natural to try to merge the world of the street performers with the one of the wearable computer and to explore synergies between them. Based on the observation that many street performers are actually skilled craftsmen of their own props, and that some have good technological skills or are at least attracted by the potential offered by technology,[63] we have investigated how some street performers could benefit from the use of an affordable wearable computer.

We believe that wearable computing can contribute to street performance in three ways: (1) It can reduce the amount of "stuff" that the performer needs to carry around by creating "virtual props" or virtual no-weight musical instruments. (2) It can augment and enrich the performance by adding digital actors to collaborate with the performer in the piece. (3) It can allow new types of street performances that were not possible before the introduction and spread of wearable computers.

In our street performance application, we are interested in exploring how point of view transforms our perception of reality. The wearable acts as a "semantic lens" and offers to members of the audience a new, transformed interpretation of the story told by the performer. This lens provides a view-dependent reality augmentation, such that different people observing the performer from various angles are led to a different interpretation of the story told by the actor.

Figure 31 Wearable Cinema setup, showing the backpack, two CPUs, camera, and lightweight color SVGA Glasstron display



We distribute wearable computers among the public (or assume that in a not-so-distant future everyone will carry his or her own wearable), whereas the performer is technology-free. The wearable computer is made up of a lightweight CPU, an augmented reality private eye display, and a camera aiming at the performer (Figure 31). The private eye shows video and graphics superimposed on the user's real-surround view (Figures 29 and 30), just as in Improvisational Theater Space (Figure 18), except that now there is no more need for a stage and a projection screen. By presenting the viewer with the computer-generated images and graphics on one eye and by leaving the view of the other eye unobstructed, we allow the human brain to fuse the real world and the graphical augmentation into a unified augmented reality. A very small and thin camera, attached to the private eye of the wearable analyzes the performer's body movements using the same processing described earlier in the paper. The public sees an expressive text-and-images augmentation of the actor, leading to a different interpretation of the play or improvisational theater, based on the wearer's position relative to the performer. Registration of the two images is provided by the computer vision program running on board.

In Sparacino et al.[64] we also describe a variety of scenarios using wearables in performance. These include the "one man orchestra" which is a wearable version of DanceSpace, the Augmented Mime, and the Networked News Teller.

By customizing a networked, multimedia computer that can be worn as clothing or is built into the performer's or the audience's clothes, we offer the street performer new and powerful tools for expression. We hope that the creation of a community of wearable augmented performers and public with a common set of experiences and needs will also serve as a push toward future improvements of this new technology. Media actors and body-tracking technologies can scale and enrich this new venue of research and technological endeavor.

Conclusions

The future of artistic and expressive communication is tied to our ability to perceive the space and people around us with sensors capable of understanding natural full body movements, hand gestures, and facial expressions, and to perform object detection and location identification. In parallel, to construct interactive experiences that benefit from natural interactions, compelling communication, and ease of implementation, we need authoring techniques that can support more than a simple mapping between inputs and outputs. Media modeling requires a "perceptive layer" to constantly monitor the input and create an output having a model of the user and its goals, as well as an "understanding" of the content itself, the logical connection among the parts, and their emotional impact on the user. We have developed a "media actors" authoring technique: We endow media objects—expressive text, photographs, movie clips, audio, and sound clips—with coordinated perceptual intelligence, behaviors, personality, and intentionality. Such media actors are able to engage the public in an encounter with virtual characters that express themselves through one or more of these agents. They are an example of intentional architectures of media modeling for interactive environments. Using this type of sensing and authoring technique, we have built applications for dance, theater, and the circus, which augment the traditional performance stage with images, video, graphics, music, and text and are able to respond to movement and gesture in believable, aesthetical, and expressive manners. We also discussed work in interactive museum exhibit design and described applications that use either a Smart Space (IVE space) or Smart Clothes (wearable computers) technological framework.

Acknowledgments

Christopher Wren, Ali Azarbayejani, Trevor Darrell, and Thad Starner have worked on Pfinder at different stages of development. The author would like to thank Kristin Hall for her wonderful Improvisational Theater interpretations and for her many hours of rehearsal, discussion, useful suggestions, and encouragement to our work. Many thanks also to choreographers Claire Mallardi and Erica Drew and to dancers Malysa Monroe, Jennifer DePalo, Chung-fu Chang, Erna Greene, Naomi Housman, and Diana Aubourg for their collaboration, interest, and encouragement. MIT undergraduates Teresa Hernandez and Jeffrey Bender did the 3-D modeling work of the enchanted forest using Alias|Wavefront's Maya**. We thank Suguru Ishizaki for his suggestions of the use of temporal typography in performance. Akira Kotani provided his SGI MIDI (Musical Instrument Digital Interface) library, and Chloe Chao contributed to the computer graphics of DanceSpace. Edith Ackerman suggested readings on Bakhtin. Dimitri Negroponte helped direct and film the first rehearsals of Improvisational TheaterSpace. Tom Minka provided gracious support, useful comments, and suggestions.

Cited references

1. C. R. Wren, F. Sparacino, A. J. Azarbayejani, T. J. Darrell, T. E. Starner, A. Kotani, C. M. Chao, M. Hlavac, K. B. Russell, and A. P. Pentland, "Perceptive Spaces for Performance and Entertainment: Untethered Interaction Using Computer Vision and Audition," *Applied Artificial Intelligence (AAI) Journal* **11**, No. 4, 267–284 (June 1996).
2. P. Maes, "Modeling Adaptive Autonomous Agents," *Artificial Life* **1**, 135–162 (1994).
3. D. Zeltzer, "Task Level Graphical Simulation: Abstraction, Representation and Control," *Making Them Move*, N. Badler, B. Barsky, and D. Zeltzer, Editors, Morgan Kaufmann Publishers, San Francisco (1991).
4. M. Johnson, *A Testbed for Three Dimensional Semi-Autonomous Animated Characters*, Ph.D. thesis, MIT, Cambridge, MA (1994).
5. B. Blumberg and T. Galyean, "Multi-Level Direction of Autonomous Creatures for Real-Time Virtual Environments," *Computer Graphics, SIGGRAPH 95 Proceedings* **30**, No. 3, 47–54 (1995).
6. R. A. Brooks, "Intelligence Without Reason," *IJCAI-91*, Vol. 1 (1991), pp. 569–595.
7. P. Maes, "Situated Agents Can Have Goals," *Designing Au-*

tonomous Agents: Theory and Practice from Biology to Engineering and Back, P. Maes, Editor, MIT Press/Bradford Books, Cambridge, MA (1990).
8. B. Blumberg, "Action-Selection in Hamsterdam: Lessons from Ethology," *Third International Conference on the Simulation of Adaptive Behavior*, Brighton, England, MIT Press, Cambridge, MA (1994), pp. 108–117.
9. K. Perlin and A. Goldberg, "Improv: A System for Scripting Interactive Actors in Virtual Worlds," *Computer Graphics, SIGGRAPH 96 Proceedings*, ACM (1996), pp. 205–216.
10. *The Selection of Behavior: The Operant Behaviorism of B. F. Skinner: Comments and Consequences*, A. C. Catania and S. Harnad, Editors, Cambridge University Press, Cambridge, UK (1988).
11. B. Hayes-Roth and R. van Gent, "Improvisational Puppets, Actors, and Avatars," *Proceedings of the Computer Game Developers' Conference*, Santa Clara, CA (1996).
12. N. Magnenat-Thalmann and D. Thalmann, "The Artificial Life of Synthetic Actors," invited paper, *IEICE Transactions*, Japan **J76-D-II**, No. 8, 1506–1514 (August 1993).
13. L. Emering, R. Boulic, R. Balcisoy, and D. Thalmann, "Multi-Level Modeling and Recognition of Human Actions Involving Full Body Motion," *Proceedings of the International Conference on Autonomous Agents 97*, ACM Press (1997), pp. 504–505.
14. D. Terzopoulos, X. Tu, and R. Grzeszczuk, "Artificial Fishes: Autonomous Locomotion, Perception, Behavior, and Learning in a Simulated Physical World," *Artificial Life* **1**, No. 4, 327–351 (December 1994).
15. D. Terzopoulos, "Artificial Life for Computer Graphics," *Communications of the ACM* **42**, No. 8, 32–42 (August 1999).
16. N. Tosa and R. Nakatsu, "Life-like Communication Agent—Emotion Sensing Character 'MIC' and Feeling Session Character 'MUSE'," *Proceedings of the International Conference on Multimedia Computing and Systems* (1996), pp. 12–19.
17. J. Bates et al., "An Architecture for Action, Emotion, and Social Behavior," *Proceedings of the Fourth European Workshop on Modeling Autonomous Agents in a Multi-Agent World* (1992).
18. N. Sawhney, D. Balcom, and I. Smith, "Authoring and Navigating Video in Space and Time: An Approach Towards Hypervideo," *IEEE Multimedia* **4**, No. 4, 30–39 (October–December 1997).
19. S. Agamanolis and M. Bove, "Multilevel Scripting for Responsive Multimedia," *IEEE Multimedia* **4**, No. 4, 40–50 (October–December 1997).
20. G. Davenport et al., "Encounters in Dreamworld: A Work in Progress," *CaiiA*, University of Wales College, Wales (July 1997).
21. J. A. Paradiso, "The Brain Opera Technology: New Instruments and Gestural Sensors for Musical Interaction and Performance," *Journal of New Music Research* **28**, No. 2, 130–149 (1999).
22. T. Jebara and A. Pentland, "Action Reaction Learning: Automatic Visual Analysis and Synthesis of Interactive Behaviour," *Computer Vision Systems, First International Conference*, *ICVS '99 Proceedings*, Las Palmas, Canary Islands (January 1999), pp. 273–292.
23. L. Pirandello, "Six Characters in Search of an Author," *Naked Masks*, E. Bentley, Editor, E. Storer, Translator, E. P. Dutton & Co., New York (1952).
24. J. L. Crowley and J. M. Bedrune, "Integration and Control of Reactive Visual Processes," *1994 European Conference on Computer Vision* (ECCV-'94), Stockholm (May 1994).
25. C. Wren, A. Azarbayejani, T. Darrell, and A. Pentland, "Pfinder: Real-Time Tracking of the Human Body," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **19**, No. 7, 780–785 (July 1997).
26. T. Darrell, B. Moghaddam, and A. P. Pentland, "Active Face Tracking and Pose Estimation in an Interactive Room," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition 96* (*CVPR 96*) (1996), pp. 67–72.
27. N. Oliver, A. Pentland, and F. Berard, "LAFTER: A Real-Time Lips and Face Tracker with Facial Expression Recognition," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition 97* (*CVPR 97*), San Juan, Puerto Rico (June 1997), pp. 123–129.
28. T. Darrell, P. Maes, B. Blumberg, and A. Pentland, "A Novel Environment for Situated Vision and Behavior," *Proceedings of IEEE Workshop for Visual Behaviors*, IEEE Computer Society Press, Los Alamitos, CA (1994).
29. A. Pentland, "Smart Rooms," *Scientific American* **274**, No. 4, 68–76 (April 1996).
30. F. Sparacino, *Choreographing Media for Interactive Virtual Environments*, master's thesis, MIT Media Laboratory, Cambridge, MA (October 1996), submitted for approval in January 1996.
31. J. A. Paradiso and F. Sparacino, "Optical Tracking for Music and Dance Performance," *Fourth Conference on Optical 3-D Measurement Techniques*, Zurich, Switzerland (September 29–October 2, 1997).
32. T. Machover, *HyperInstruments: A Composer's Approach to the Evolution of Intelligent Musical Instruments*, CyberArts, William Freeman, San Francisco (1992), pp. 67–76.
33. J. Klosty, *Merce Cunningham: Dancing in Space and Time*, Saturday Review Press, New York (1975).
34. A. Marion, *Images of Human Motion: Changing Representation of Human Identity*, M.S. thesis, MIT, Cambridge, MA (1982).
35. R. Gehlhaar, "SOUND = SPACE: An Interactive Musical Environment," *Contemporary Music Review* **6**, No. 1, 59–72 (1991).
36. T. Calvert, C. Lee, G. Ridsdale, S. Hewitt, and V. Tso, "The Interactive Composition of Scores for Dance," *Dance Notation Journal* **42**, No. 2, 35–40 (1986).
37. T. Calvert and S. Mah, "Choreographers as Animators: Systems to Support Composition for Dance," *Interactive Computer Animation*, N. Magnenat-Thalmann and D. Thalmann, Editors, Prentice-Hall, Inc., Upper Saddle River, NJ (1996), pp. 100–126.
38. K. Sukel, G. Brostow, and I. Essa, *Towards an Interactive Computer-Based Dance Tutor*, Concept Paper, Georgia Institute of Technology, Atlanta (1998).
39. F. Sparacino, K. Hall, C. Wren, G. Davenport, and A. Pentland, "Digital Circus: A Computer-Vision-Based Interactive Virtual Studio," *IMAGINA*, Monte Carlo, Monaco (January 18–20, 1999).
40. P. Bouissac, *Circus and Culture: A Semiotic Approach*, Indiana University Press, Bloomington, IN (1976).
41. F. Sparacino, K. Hall, C. Wren, G. Davenport, and A. Pentland, "Improvisational Theater Space," *The Sixth Biennial Symposium on Arts and Technology*, Connecticut College, New London, CT (February 27–March 2, 1997).
42. S. Viola, *Improvisation for the Theater*, Northwestern University Press, Evanston, IL (1979).
43. Y. Y. Wong, *Temporal Typography*, master's thesis, MIT, Media Laboratory, Cambridge, MA (1995).
44. E. Spiekermann and E. M. Ginger, *Stop Stealing Sheep and*

Find Out How Type Works, Adobe Press, Mountain View, CA (1993).
45. R. Massin, *Letter and Image*, Van Nostrand Reinhold, New York (1970).
46. M. M. Bakhtin, *The Dialogical Imagination*, M. Holquist, Editor, University of Texas Press, Austin, TX (1981).
47. L. S. Vygotsky, "Thinking and Speech," *The Collected Works of L. S. Vygotsky, Vol. 1: Problems of General Psychology*, R. W. Reiber and A. S. Carton, Editors, N. Minick, Translator, Plenum Publishers, New York (1987).
48. T. B. L. Webster, *The Greek Chorus*, Methuen, London (1970).
49. A. Terzakis, "The Chorus in the Greek Tragedy," *Preuves*, nn. 215–216 (February–March 1969), pp. 8–15.
50. M. Kirby, *Futurist Performance*, E. P. Dutton & Co., New York (1971).
51. H. Gilpin, "William Forsythe: Where Balance Is Lost and the Unfinished Begins," *Parkett* **45**, Parkett Publishers, New York (1995).
52. P. Brook, *The Open Door*, Theatre Communication Group, New York (1995).
53. L. Klein, *Exhibits: Planning and Design*, Madison Square Press, New York (1986), pp. 70–71.
54. A. Pentland, "Smart Clothes," *Scientific American* **274**, No. 4, 73 (April 1996).
55. T. Starner, S. Mann, B. Rhodes, J. Levine, J. Healey, D. Kirsch, R. W. Picard, and A. Pentland, "Augmented Reality Through Wearable Computing," *Presence*, special issue on augmented reality (1997).
56. S. Mann, "Wearable Computing: A First Step Toward Personal Imaging," *Computer* **30**, No. 2, 25–31 (February 1997).
57. F. Sparacino et al., "City of News," *Ars Electronica Festival*, Linz, Austria (September 8–13, 1997).
58. F. Sparacino, G. Davenport, and A. Pentland, "Wearable Cinema/Wearable City: Bridging Physical and Virtual Spaces Through Wearable Computing," *IMAGINA 2000*, Monte Carlo, Monaco (January 31–February 3, 2000).
59. G. Davenport, T. Aguierre Smith, and N. Pincever, "Cinematic Primitives for Multimedia," *IEEE Computer Graphics and Animation*, special issue on multimedia (July 1991), pp. 67–74.
60. G. Davenport, R. Evans, and M. Halliday, "Orchestrating Digital Micromovies," *Leonardo* **26**, No. 4, 283–288 (August 1993).
61. F. Sparacino, G. Davenport, and A. Pentland, "Media Actors: Characters in Search of an Author," *IEEE Multimedia Systems '99, International Conference on Multimedia Computing and Systems* (*IEEE ICMCS'99*), Centro Affari, Firenze, Italy (June 7–11, 1999).
62. N. Oliver, T. Jebara, B. Schiele, and A. Pentland, http://vismod.www.media.mit.edu/~nuria/dypers/dypers.html.
63. B. Mason, *Street Theatre and Other Outdoor Performance*, Routledge Publishers, New York (1992).
64. F. Sparacino, C. Wren, G. Davenport, and A. Pentland, "Augmented Performance in Dance and Theater," *International Dance and Technology 99* (*IDAT99*), Arizona State University, Tempe, AZ (February 25–28, 1999).

Accepted for publication May 15, 2000.

Flavia Sparacino *MIT Media Laboratory, 20 Ames Street, Cambridge, Massachusetts 02139-4307 (electronic mail:flavia@media.mit.edu).* Ms. Sparacino is a Ph.D. candidate at the MIT Media Laboratory. She is designing perceptual intelligence for interactive media, with applications to augmented performance, information architecture, and museum displays. Perceptual intelligence emerges from the integration of multiple sensor modalities and determines the expressive abilities of digital text, photographs, movie clips, music, and audio. Her work was featured at Ars Electronica 97 and 98, ISEA 97, SIGGRAPH 96 and 99, ICMCS 99, and ICHIM99, as well as many other international venues. She holds undergraduate degrees in electrical engineering and robotics, and master's degrees in cognitive sciences and media arts and sciences. Ms. Sparacino received a number of scholarships and awards for her academic work and career, including those from the European Community, the Italian Center for National Research, the French Center for National Research, and Fulbright, among others. She spent some time in film school, theater school, and a small-town circus and has done travel photography in many countries around the world.

Glorianna Davenport *MIT Media Laboratory, 20 Ames Street, Cambridge, Massachusetts 02139-4307 (electronic mail: gid@media.mit.edu).* Ms. Davenport is the director of the Interactive Cinema Group at the MIT Media Laboratory. Trained as a documentary filmmaker, she has achieved international recognition for her work in the new media forms. Her research explores fundamental issues related to the collaborative co-construction of digital media experiences, where the task of narration is split among authors, consumers, and computer mediators. Ms. Davenport's recent work focuses on the creation of customizable, personalizable storyteller systems that dynamically serve and adapt to a widely dispersed society of audience. She has taught, lectured, and published internationally on the subjects of interactive multimedia and story construction. More information can be found at http://www.media.mit.edu/~gid.

Alex (Sandy) Pentland *MIT Media Laboratory, 20 Ames Street, Cambridge, Massachusetts 02139-4307 (electronic mail: sandy@media.mit.edu).* Dr. Pentland is the academic head of the MIT Media Laboratory and cofounder and director of the Center for Future Health and the LINCOS Foundation. His research focus includes both smart rooms (e.g., face, expression, and intention recognition; word learning and acoustic scene analysis) and smart clothes (e.g., augmenting human intelligence and perception by building sensors, displays, and computers into glasses, belts, shoes, etc.). These research areas are described in the April 1996 and November 1998 issues of *Scientific American*. He is one of the 50 most-cited researchers in the field of computer science, and *Newsweek* has recently named him one of the 100 Americans most likely to shape the 21st century. Dr. Pentland has won awards from several academic societies, including the AAAI, IEEE, and Ars Electronica. See http://www.media.mit.edu/~pentland for additional details.