Augmented Performance in Dance and Theater

Flavia Sparacino *Christopher Wren* *Glorianna Davenport* *Alex Pentland*
MIT Media Lab
{flavia, cwren, gid, sandy}@media.mit.edu

Abstract. This paper describes motivations and techniques to extend the expressive grammar of dance and theatrical performances. We first give an outline of previous work in performance, which has inspired our research, and explain how our technology can contribute along historical directions of exploration. We then present real-time computer vision based body tracking and gesture recognition techniques which is used in conjunction with a Media Actors software architecture to choreograph digital media together with human performers. We show applications to dance and theater which augment the traditional performance stage with images, video, music, text, able to respond to movement and gesture in believable, esthetical, and expressive manners. Finally, we describe a scenario and work in progress, which allow us to apply our artistic and technological advances to street performance.

1. Introduction

We have built an interactive stage for a single performer which allows us to coordinate and synchronize the performer's gestures, body movements, and speech, with projected images, graphics, expressive text, music, and sound. We have developed two cooperating technologies: one does optical, untethered, tracking of the human body in real time, and the other endows digital media with perceptual intelligence, expressive and communicative abilities, similar to those of a human performer (Media Actors).

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 towards musical and graphical augmentation of human movement. We have 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 computer's role 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 their body as the interface. This 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 pre-chosen 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 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, extend the expressive grammar of theater by allowing the director to show more of the character's inner conflicts, contrasting action/thought moments, memories, worries, 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 computer's role, with the use of perceptive and expressive Media Actors, is that of a *co-actor* which collaborates 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 which share the stage with the human actors.

In the following section, we provide a brief historical outline which shows how our technological intervention is defined within modern and contemporary trends of dance and theatrical performance. We then describe the interactive stage, the real-time optical body tracking system, and the Media Actors software architecture. We give artistic and technical details on DanceSpace and Improvisational TheaterSpace. We also describe a scenario of current and future work which involves the integration of our stage technology into a mobile setting, such as street performance. We draw conclusions based on our work in augmented performance carried out at the MIT Media Lab between January 1996 and today.

2. Background and Motivation

The study of human movement and gesture, from its stylized to its more expressive form, is today an endeavor common to both technology and performance [Huxley, 1996]. We are engaged in the search of new or extended expression modalities which are centered on the body as the primary vehicle of communication, with the aid of the latest technological advances. Along the line of exploration initiated by Muybridge in the late 1800s, we are interested in classifying and analyzing human movement with the use of video cameras and computers, in non-invasive and non-encumbering ways. At the same time, we are investigating how to make creative and artistic use of this enhanced and quantitative understanding of gesture and movement to complement and augment the body in performance.

Both efforts, the study and classification of human movements and gesture, and the search of new artistic modalities of expression which extend the language of staged performance, have historical roots in dance and theater. In many situations these efforts are closely related to each other and aim to achieve a more direct and effective communication with the public. In this section, we describe briefly work in modern and contemporary performance which grounds our research within historical trends, as well as work which has inspired our reflection and exploration in technologically augmented performance in dance and theater.

2.1. Dance

At the end of the 19th century, Emile Jacques-Dalcroze, Rudolph Von Laban, and Mary Wigman started a systematic and rigorous study of dance movements. *Dalcroze* created a system of rhythmic gymnastics, aimed at reaching a harmony between static and dynamic forces of the human body. He believed in the creative richness of mixed media and choreographed pieces at the crossroad between dance and opera. *Laban* gave a theoretical foundation to modern dance. He researched extensively the variety of expressive systems of past and contemporary civilizations to find the basis for a universal language of the body. He affirmed the primacy of movement and dance over music and privileged percussion as the most pure form of rhythmic support. He is known for elaborating a system of choreographic notation called "Labanotation." *Wigman* centered modern dance on the solo performer. A pupil of Dalcroze and Laban, she focused on the expression of the body and the movements which convey innerdeep emotions. Decorations, costumes, even music, and a formal sense of beauty became unimportant in her dances. Along the line traced by Dalcroze, Laban, and Wigman, we are interested in classifying and recognizing movements and rhythms of the body with the aid of technology. We believe real-time computer vision is the ideal tool for this purpose: by using only cameras and computers, the performer is not encumbered with special suits and wires.

A pupil of Martha Graham, *Merce Cunningham* is the emblem of contemporary post-modern dance. He has introduced a series of revolutionary innovations in all aspects of dance and choreography. He borrows elements not only from popular dances but also from everyday gestures. He proposes a polycentric organization of the stage in which each dancer becomes the center of his dance. He has developed a choreography based on structured improvisation: each dancer needs to “invent” his or her own space as a function of the other performers’ improvisation. Rhythm is played around repetitive figures, microvariations, ruptures, bifurcations. Cunningham believes that dance and movement should be designed independently of music: music should be subordinate to movement and may be composed after the principal choreography is finished, just as a musical score is created after a feature film has been edited. Likewise Cunningham, we believe that dance and movement should be designed independently of music. However during rehearsal it may be important to be able to use music to better pace the choreography. For this purpose it is desirable to have a tool which can generate music on the fly by observing the dancer's movements. We resonate with Cunningham's notion of structured improvisation. If visuals/text are shown/projected on stage during the performance, their movement, dynamics, and coordination should reflect or acknowledge the dancers as they coordinate an improvised choreography. Structured improvisation is not equivalent to free form. It is instead a refined skill, which allows a performer to choose among his repertoire of gestures/movements/actions what best suits the current development of the dance, in coordination with the other performers. We believe that any visual/musical material presented during the performance should be modeled in an analog way.

After Cunningham, post-modern dancers further investigated in various directions. Of special interest are the choreographers of the Judson Dance Theater. *Trisha Brown* dances with a film projector attached to her back. In other occasions she asks the public to “animate” her body with cries and vocal encouragement. She elaborates on everyday movements like walking, sitting, running, dressing or undressing. Like Trisha Brown, we think modern dance can import and transpose daily gestures into the performance. We think it is artistically compelling to have a tool which recognizes some of these gestures during the performance and augments the dance with visual projections or sounds. *Twila Tharp* proposes a mixed dance of old and new, accompanied by a juxtaposition of different musical styles. She has performed in unconventional stages such as: public gardens, museums, gymnasiums. Similarly, *Meredith Monk* uses for her performances public spaces such as parking lots, art galleries, and roofs, and invites the public to circulate among the dancers. In accordance with Twyla Tharp and Meredith Monk, we believe that most of our technological tools should be able to scale and function in outdoor situations or should be easily portable to open performance spaces, in which the public can be involved more directly in the performance.

While adopting classical techniques, *William Forsythe* inherits some of the themes of experimentation of the seventies. Fascinated by speed, stress, distortions, unbalance, Forsythe choreographs using collective improvisation, sound effects, dancers' voices, lighting effects. He associates the written word to movement, and has staged beautiful pieces for dancers and projected text. We are inspired by Forsythe’s choreographic experiments which mix typographic design and body movements. We believe technology can provide tools which promote a synergy between the dancer’s body and expressive temporal typography. These tools allow us to synchronize and drive a textual narration of expressive text with body movement.

2.2. Theater: understanding gesture and movement

The systematic study of the expressive resources of the body started in France with *Francois Delsarte*, at the end of the 1800s. Delsarte studied how people gestured in real life and elaborated a lexicon of gestures, each of which, according to the author, had a direct correlation with the psychological state of man. Delsarte observed that for every emotion, of whatever kind, there is a corresponding body movement. He also believed that a perfect reproduction of the outer manifestation of some passion will induce, by reflex, that same passion. Delsarte inspired us to have a lexicon of gestures as working material to start from. By providing automatic and unencumbering gesture recognition, technology offers a tool to study and rehearse theater. It also provides us with tools which augment the actor's action with synchronized digital multimedia presentations.

Delsarte's “laws of expression” spread widely in Europe, Russia, and the United States. At the beginning of the century, *Vsevolod Meyerhold* at the Moscow Art Theater, developed a theatrical approach which moved away from the naturalism of Stanislavski. Meyerhold looked to the techniques of the Commedia dell'Arte, pantomime, the circus, and to the Kabuki and Noh theaters of Japan for inspiration, and created a technique of the actor which he called “Biomechanics.” Meyerhold was fascinated by movement and trained actors to be acrobats, clowns, dancers, singers, and jugglers, capable of rapid transitions from one role to another. He banished virtuosity in scene and costume decoration and focused on the actor's body and his gestural skills to convey the emotions of the moment. By presenting to the public properly executed physical actions and by drawing on their complicity of imagination, Meyerhold aimed at a theater in which spectators would be invited to social and political insights by the strength of the emotional communication of gesture. Meyerhold’s work stimulated us to investigate the relationship between motion and emotion. We are interested not only in recognizing gestures but also in being able to classify them according to how, emotionally, they have been executed. Also, in certain occasions, we want to infer which emotion causes an actor's gesturing.

Parallel to Meyerhold, *Gordon Craig*, imagined an actor’s school which, through rigorous movement training, would give actors the power to improvise. He wanted to free actors from reliance on text and turn them into collaborators in the creation of theatrical artwork. In his essay “The Actor and the Ubermarionette” he did not advocate to replace fallible human actors with giant puppets, as was generally supposed [Rudlin, 1994]. He actually questioned the tradition of western naturalist theater and imagined an actor able to loose his personality, as in certain forms of Asian theater, and take the attributes of a puppet. A puppet however able to improvise. Craig's thinking of an actor as a marionette confirms us in our modeling the body with a stick figure, when analyzing human gestures with the aid of a computer. However, we are not interested in building virtual actors which substitute real ones on stage. We want to augment the stage inhabited by embodied human actors with digital content - visuals/music - in coordination with a real-time analysis of the actor's gestures and movements.

Later in the century *Bertold Brecht* elaborated a theory of acting and staging aimed at jolting the audience out of its uncritical stupor. Performers of his plays used physical gestures to illuminate the characters they played, and maintained a distance between the part and themselves. The search of an ideal gesture which distills the essence of a moment (Gestus) is an essential part of his technique. Brecht wanted actors to explore and heighten the contradictions in a character's behavior. He would invite actors to stop at crucial points in the performance and have them explain to the audience the implications of a character's choice. By doing so he wanted the public to become aware of the social implications of everyone's life choices. Like Brecht, we are interested in performances which produce awakening and reflection in the public rather than uncritical immersion and stupor. We therefore have organized our technology to augment the stage in a way similar to how “Augmented Reality” enhances or completes our view of the real world. This contrasts work on Virtual Reality, or Virtual Theater, or Virtual Actors which aims at replacing the stage and the actor with virtual ones, and to involve the public in an immersive narration similar to an open-eyes dream. Also, like Brecht, we would like to use our augmentation of gestures with the projection of digital text and images, to illustrate contradictions and motivations of a character's choices in the play.

Artaud proposed a Theater of Cruelty which would purge the spectator from antisocial impulses by seeing them played out on stage. He believed in the importance for the actor to make gestural signs which would appeal to the subconscious of the spectator and shock him. Artaud showed that separating the actors from the audience was unnecessary. Actors would walk among the spectators and involve them in the action. He wanted a theater play to be a happening and not a show, with contributions from dance, music, pantomime, and architecture. Artaud's influence on this century's theater has been profound and ranges from the American experimentations of the Living Theater to his European successors. In accordance with Artaud, we want to be able to involve the public in our performances. Many efforts of today's multimedia technologies are aimed at prompting a more active and direct participation of the public in the artwork. This is not new in the history of performance, and we believe we have a lot to learn not only from Artaud, but also from Commedia dell'Arte, street theater, and Improvisational Theater [Johnstone 1979, Viola, 1979]. After Artaud the mime school of *Jacques Lecoq* and the “Theater du Soleil” of *Ariane Mnouchkine* developed a gesture-based improvisational acting strongly inspired by the tradition of Commedia dell'Arte.

More recently, polish director *Jerzy Grotowski* in his Laboratory Theater advocated a “Poor Theater” stripped down to its bare essentials: the space, the actor, the spectators. For Grotowski what is essential to theater is the actor-spectator relationship: lighting, makeup, props, scenery, even the stage, are unnecessary. Through an intense interaction between the actor and the audience, Grotowski was hoping a transformation would take place in both: a holy communion which liberates from falsity and vanity. He designed physical exercises for the body and voice to give actors the athletic strength required to achieve a powerful connection with the public. From Grotowski we borrow focus on an essential theater which includes exclusively the space, the actor, and the spectator. Currently, we are not interested in special costumes, decors, props, makeup, or elaborate sets.

English director *Peter Brook*, a remarkable contemporary, has accomplished a creative synthesis of the century's quest for a novel theory and practice of acting. Brook started his career directing “traditional” Shakespearean plays and later moved his stage and theatrical experimentation to hospital, churches, and African tribes. He has explored audience involvement and influence on the play, preparation vs. spontaneity of acting, the relationship between physical and emotional energy, and the usage of space as a tool for communication. His work, centered on sound, voice, gestures, and movement, has been a constant source of inspiration to many contemporaries, together with his thought provoking theories on theatrical research and discovery. We admire Brook's research for meaning and its representation in theater. In particular we would like to follow his path in bringing theater out of the traditional stage and perform closer to people, in a variety of public and cultural settings.

2.3. Visual Theater

The idea of creating performances which involve both human actors and objects as characters can be tracked back to the beginning of the century. The avant-garde Italian *Futurist movement* inspired a new form of theater called Synthetic Theater which presented “dynamic, fragmentary symphonies of gestures, words, noises, and lights” [Kirby, 1971]. Emphasis was given to a more direct and non contemplative involvement of the spectators by using techniques such as compression of the text and simultaneity of scenes. In 1909 Marinetti, the founder of Italian Futurism, wrote a theater play called “Electric Puppets” in which human actors acted together with life-like puppets. Other theater works of interest by the futurists include Balla's “Printing Press”, in which large typographic characters became part of the performance, and Podrecca's “Plastic Dances.” The Futurists have provided experimental examples of the use of text on stage. They have also suggested small, granular, provocative narrative examples in opposition to linear narrative forms.

The theater of the *Bauhaus* also experimented with a non-verbal, dance-inspired theater of objects with living actors stylized as geometric shapes [Schlemmer, 1971]. Schlemmer created an experimental stage workshop in which he explored relationships between the stage, the performer, and the visual elements on stage. He investigated the interplay between the actor's body, voice, gestures, movements, and the stage seen visually as form and color in kaleidoscopic motion. We share with the theater of the Bauhaus a view of the stage as a designer's grid or a painter's canvas. *What motivates our understanding of human movement and gestures with the aid of technology is the possibility of transforming the stage in an organic interplay of form, color, text, images, movements, expressive gestures, in which the human is one among the “actors” on the scene*.

More recently *Bob Wilson* has created refined performances which harmonize elements from theater, dance, and opera. His technique owes much to modern technology, and imports cinematographic elements such as the freeze-frame, slow motion, and playback. He uses the stage as a three dimensional canvas in which he mixes sound, gesture, movement, light, and time, to produce epic performances staging relevant social events of our time. Wilson has stimulated us to look at the narrative elements of cinema and to build technological tools which allow to import and reproduce cinematographic effects into the theater.

3. The IVE Stage

Our stage - called “IVE” (Interactive Virtual Environment) - is a room sized area (15'x17') whose only requirements are good, constant lighting and an unmoving background. A large projection screen (7'x10') 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 1,2]. By use of real-time computer vision techniques [Wren, 1997][Darrell, 1996][Oliver, 1997] 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.

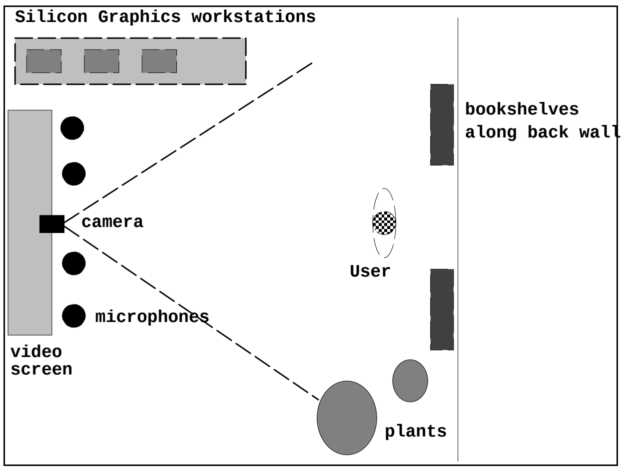


Figure 1. The IVE stage, view from above.

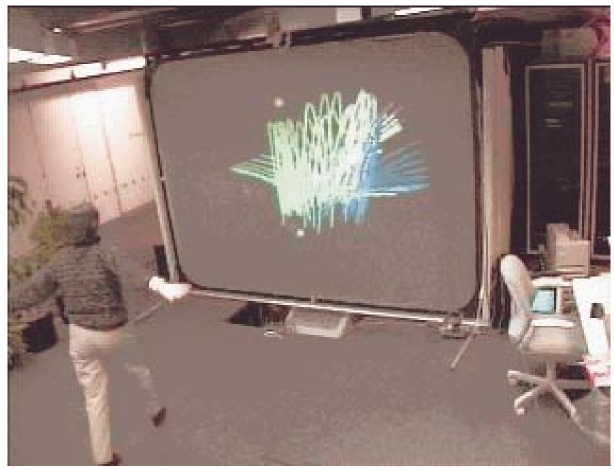


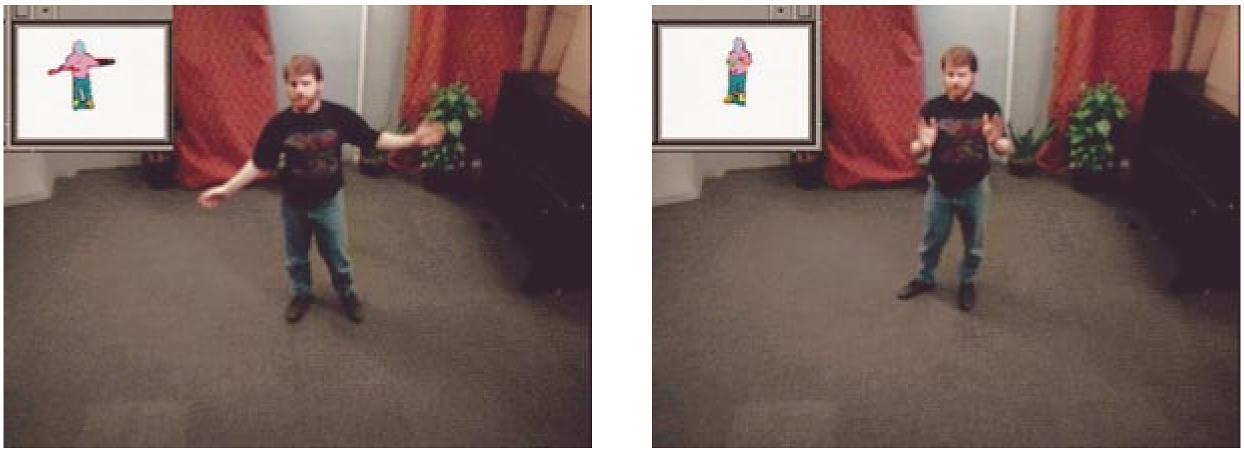
Figure 2. The IVE stage, during rehearsal.

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 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 [Wren, 1996][Darrell, 1994]. 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. We are in the process of extending the tracking technology to support many people/performers at once.

4. The Real-Time Computer Vision Input System

Our real-time computer vision program is called Pfinder, i.e. “person finder” [Wren, 1997]. Pfinder is a system for body tracking and interpretation of movement of a single performer. It uses only a wide angle camera pointing towards the stage and a standard Silicon Graphics O2 computer. The system uses a multi-class statistical model of color and shape to segment a person from a background scene, and then to find and track people's body parts in a wide range of viewing conditions. The system employs a multi-class statistical model of color and shape to segment a person from a background scene, then find and track people'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 two-dimensional 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 people 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 being added to account for each differently colored region [Figures 3,4]. Separate blobs are generated for the person's hands, head, feet, shirt, and pants. The process of building a blob-model is guided by a 2D contour shape analysis that recognizes silhouettes in which body parts can be reliably labeled.



Figures 3, 4. Pfinder tracking the human body: head, torso, legs, hands, feet are labeled: gray, purple, red, green, orange, yellow. In Figure 4, notice correct hand tracking even when the hands are in front of the body.

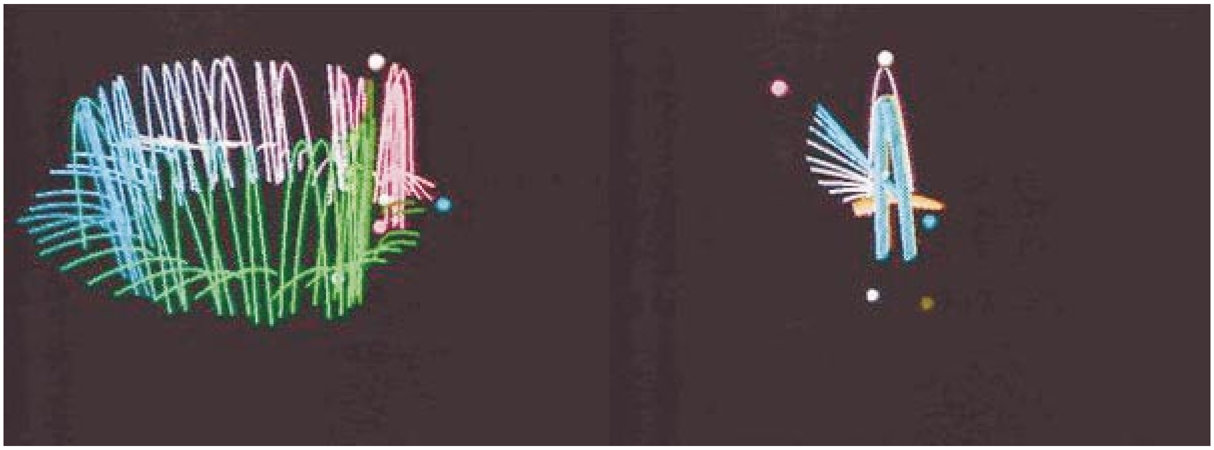
The computer vision system is composed of several layers. The lowest layer uses adaptive models to segment the user from the background, enabling the system to track users without the need for chromakey backgrounds or special garments, while identifying color segments within the user's silhouette. This allows the system to track important features (hands) even when these features aren't discernible from the figure-background segmentation. This added information makes it possible to deduce the general 3D 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 people (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 2D image information to produce 3D position estimates using 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.

5. DanceSpace

DanceSpace is an interactive stage which takes full advantage of Pfinder's ability to track the dancer's motion in real time. Different parts of the dancer's body (hands/head/feet/torso) can be mapped to different musical instruments that constitute 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 [Sparacino, 1996][Paradiso, 1997].

The computer-generated music consists of a richly-textured melodic base tune, which plays in the background for the duration of the performance. As the dancer enters the space, a number of virtual musical instruments are invisibly attached to her body. The dancer then uses her body movements to 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, bringing down the sound as they move closer to the ground. The distance from the dancer's hands to the ground is mapped to the pitch of the note played by the musical instruments attached to the hands. Therefore a higher note will be played when the hands are above the performer's head and a lower note when they are near her waist. Both hands' musical instruments are played in a continuous mode (i.e., to get from a lower to a higher note the performer will have to play all the intermediate notes). The bells and the drums are on the contrary one shot musical instruments that are triggered by feet movements. More specific gestures of 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 of the interactive space. Nevertheless all the music shares the same pleasant rhythm established by the underlying, ambient tune, and a style that ranges from "pentatonic" to "fusion" or "space" music.

As the dancer moves, her body leaves a multicolored trail across the large wall screen that comprises one side of the performance space. The color of the trail can be selectively mapped to the position of the dancer on stage or to more "expressive" motion cues like 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 3D spheres are also drawn to map onto hands, feet, head and center of the body of the performer. This serves as a reference for the dancer and accentuates the stylized representation of the body on the screen. The multicolored trail represents the dancer's virtual shadow which follows her around during the performance. The variable memory of the shadow allows the dancer to adjust the number of trails left by the dancer's body. Hence if the shadow has a long memory of trails (more than thirty) the dancer can paint more complex abstract figures on the screen [Figures 5,6,7,8].



Figures 5, 6. Graphical output of DanceSpace.

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

[Machover, 1992] are musical instruments primarily invented for non-musical-educated people who nevertheless wish to express themselves through music. The computer that drives the instruments holds the basic layer of musical knowledge needed to generate a musical piece.

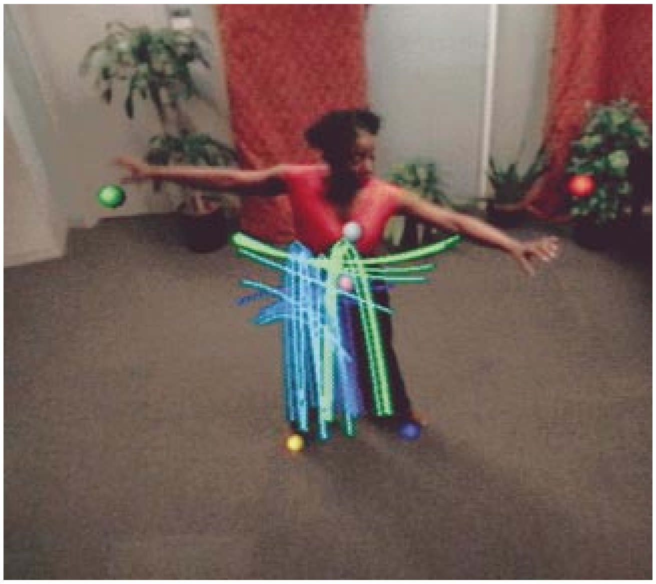
The philosophy underlying DanceSpace is inspired by Merce Cunningham's approach to dance and choreography [Klosty, 1975]. Cunningham believed that dance and movement should be designed independently of music, which is subordinate to the dancing and may be composed later for performance, much as a musical score is in film.

Figures 7, 8. Performer using DanceSpace, during rehearsal.

In the early 1980s, Ann Marion has pioneered work on generating graphics through body motion in dance at MIT by use of use of polhemus sensors to track ballet steps [Marion, 1982]. Rolf Gehlhaar [Gehlhaar, 1991] has built a number of sound spaces where multiple users generate soundscapes through full body motion. Researchers at Georgia Tech University [Sukel, 1998] and Georgia Tech's Center for the Arts together with the Atlanta Ballet are also involved in a Dance and Technology Project. They use active sensing technology by placing sensors on the dancer's body to track the performer's movement on an interactive stage. 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 estimated of gross motion of the performer. Moreover as opposed to other environments it does not require the dancer to wear special clothes or active sensors. Anyone can just walk in the space and generate sound and graphics by body movements/gestures. DanceSpace's principal drawback is that currently the computer vision based sensing technology reliably tracks only one performer a time.

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 Lab, including a one day Open House with people of all ages. Semi-professional dancers from Boston Conservatory have choreographed short pieces for the interactive stage under the supervision of the choreographer Erica Drew [Figure 9]. During these performances the choreographer made an effort to enhance or underline the expressiveness of the human body as opposed to the “coldness” of the musical and graphical output by the computer (her words). The dancers were fascinated by the colored virtual shadow that followed them on stage and soon modified their pieces as to better exploit the “comet” effect of the computer graphics trails. Non-performers who attended the open house seemed to be more interested in exploring the space to obtain the desired musical effect.

We also used DanceSpace to create an original choreography together with choreographer Claire Mallardi at Radcliff College, Harvard Extension, in the spring of 1996 [Figure 10]. 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 – non interactively – on the dancers bodies during the performance [Figures 11,12,13,14]. As the dancers were wearing white unitards their bodies were virtually painted by the projected computer graphics image. As a consequence they changed the choreography of their piece to have more still body poses to better exploit this body-painting effect. Another interesting effect 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 performances.



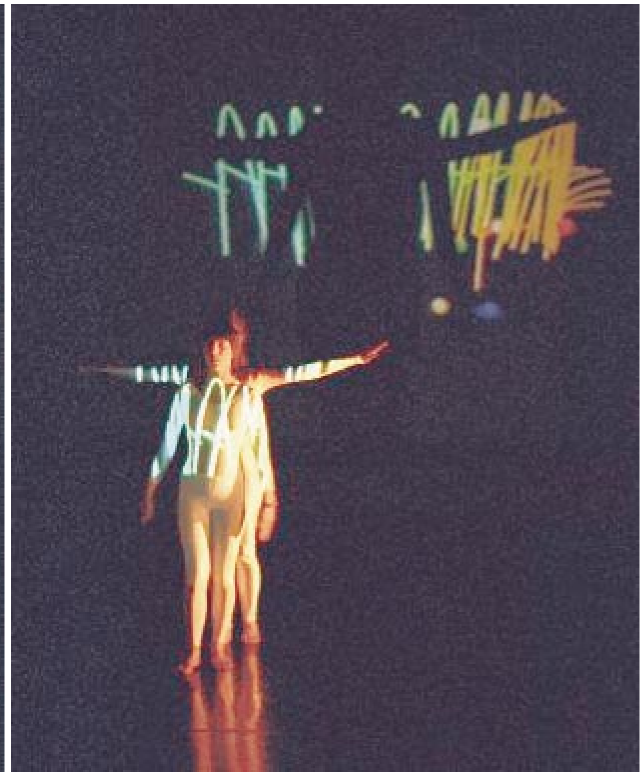
Figures 9, 10. Performers Jennifer DePalo and Diana Aubourg in DanceSpace.

Further improvements 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.

6. Media Actors

In this section we introduce a new media modeling technique for authoring interactive performances. We describe Media Actors: images, video, sound, speech, text objects, which are able to respond to movement and gesture in believable, aesthetic, expressive, and entertaining manners. Media Actors are modeled as software agents whose personality affects not only their internal state (feelings) but also their perception of the performer's behavior (intentions) and their expectations about future interactions with their human interactor. A Media Actor knows whether its content is text, image, a movie clip, sound or graphics and acts accordingly.

In our approach to modeling Media Actors, we provide traditional – passive – visual/sound elements with their own performing skills, behavioral attributes, and perceptive abilities. This approach contrasts with scripted media modeling, which separates the content to be shown and the routines which orchestrate

Figures 11,12,13,14. Performance at Radcliff College, with choreographer Claire Mallardi and dancers: Malysa Monroe and Naomi Housman.

media. This split architecture leads to complicated control programs which have to do an accounting of all the available content, where it is located on the display, and what needs to happen when/if/unless. Scripted 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 performance art piece according to this methodology can be a daunting task, and the technology of scripting inevitably complicates and slows down the creative process rather than enhance or expand it. By distributing the computational burden in the Media Actors, instead of in a central program which does media accounting and timing throughout the performance, we ease the task of both the programmer and the director of the interactive performance piece.

Media Actors are composed of three layers:

1. The actual content: it can be text, images, video, sound, speech, graphics.
2. The perceptual layer, which is responsible for “understanding” and interpreting the performer’s gestures and movement.
3. The choreographic component, which selects the most appropriate expressive behavior in the given context. The choreographic component is responsible for the autonomous behavior of the synthetic actor.

In the Media Actors architecture, there is no separation between content and the choreography of content. The choreographic component of a media Actor specifies its behavior according to the context of the interaction with the performer. As opposed to scripted animation, which imposes a pre-defined sequence of actions to a “virtual actor”, we provide a tree of actions that are driven either by the internal motivation of the digital actor or by the external actions of the performer or by a combination of both. Hence the model of the interaction between the human and the digital actor consists of a nearly infinite tree of possibilities rather than a linear sequence of consecutive actions.

Our software 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 story. We believe that interpretation is the key not only to compelling theater but also to successful mixed-media interactive performances. Media actors are endowed with the ability to interpret sensory-data generated by the performer or 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 which are easily observed by the public such as happy, tired, sad, angry, etc, while expectations – gratification, frustration, or surprise, stimulate the follow-on action.

Media actors are endowed with wireless sensors which allow for natural and unencumbered interactions with the public. Real-time computer-vision and auditory processing allow for interpretation of simple and natural body gestures, head movements, pre-given 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, contributes to model 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 [Pirandello, 1952]. They are media with a variety of expressive behaviors, personalities whose life-like responses emerge as a result of the interaction 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 behaviors: show, move, scale, transform, change color, etc.

Our model is sensor-driven – which explains why we call it perceptual – and personality-based, rather then behavior-based. By personality we designate the general patterns of behavior and predispositions which determine how a person will think, feel, and act. We have modeled feelings rather than emotions, as 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?".

Our original implementation of Media Actors was inspired by Blumberg's behavior-based animation approach [Sparacino, 1996][Blumberg, 1995]. Since then we have developed a media architecture [Figure 15] which focuses more on expression and perception [Sparacino, 1998]. This new type of character modeling for multimedia differs from both scripted and purely behavior-based approaches. With respect to the classical behavior-based approach [Blumberg, 1995] we introduce:

- A perceptual layer: the sensorial input is translated into a percept which helps define a "user-model" as it contributes to interpret the participant's intention.
- A notion of expectation that the media actor needs to have on 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: to induce an emotion or to articulate the transmission of a message.
- An internal state intended as "feeling" which generates an expressive action.

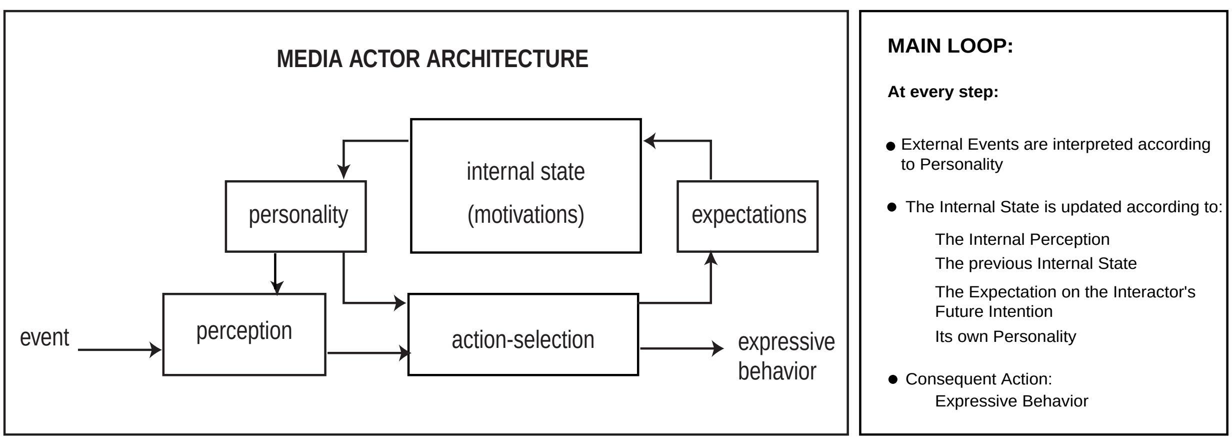


Figure 15. Media Actor architecture.

The importance of having an intermediate layer of sensory representation, and predictions, has also been underlined by Crowley [Crowley, 1994]. However Crowley's architecture is limited to the construction of reactive visual processes which accomplish visual tasks, and does not attempt to orchestrate media to convey a message or to animate a life-like virtual character.

7. Improvisational TheaterSpace

In April 1996 we created the Improvisational TheaterSpace [Sparacino, 1996]. 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 which 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 autonomous agent-based text, images, movie clips, and audio. These 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 is an entertaining and engaging genre which allows the audience to drive part of the play. An experimental performance using the IVE setup was shown in late February 1997 in the occasion of the Sixth Biennial Symposium on Arts and Technology [Sparacino, February 1997]. We also did a public rehearsal at the MIT Media Lab, in the occasion of a Digital Life Open House, on March 11, 1997, both with improvisational actress Kristin Hall.

We conceived a theatrical situation in which a human actor could be seen interacting with his own thoughts appearing in the form of animated expressive text projected onto a large screen on stage. We modeled the text 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 understands simple words and sentences. Its expressive abilities include: showing basic moods through typographic behaviors, like being happy, sad, angry, or excited.

The stage is made of a large space where the actor moves and a large screen that reflects the real image of the actor and contains a graphical representation of the test. A camera and a microphone are used by the typographic actor as sensors for the real world and to interpret the performer's actions. 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 into 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 [Viola, 1979].

The idea of conveying emotions through typographic attributes of the text has been largely explored in graphic design [Wong 1995, Spiekermann 1993, Massin 1970]. The idea of introducing text as a character on stage was inspired by readings of Baktin [Baktin, 1981], and Vygotsky [Vygotsky, 1987]. Both authors have underlined the role of a dialogical consciousness. For Baktin 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/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.

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") [Kirby, 1971]. In this theater piece each of the twelve performers became part of a printing press machine by repeating a particular sound. Their movement would also reproduce the movement of the 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 which 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 which 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 [Zeltzer, 1991] and Blumberg [Blumberg, 1995].

The following is a list of high-level behaviors for the typographic actor: 1. Say typical phrase; 2. Attract attention; 3. Show off; 4. Entertain; 5. Daydream; 6. Tell a story or explain; 7. Suggest what to do next; 8. Follow Actor (or any of his/her body 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 take place 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, or sinusoidal time function will be chosen during execution of time-dependent skills.

The typographic actor has also a level of energy that can be high, low, or medium. Text draws energy from the user's speed of movement and loudness of speech. It spends energy while performing basic skills. Emotional states affect the manner of the presentation 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; animated text playing the alter ego of the performer, eventually driven by the public.

Ongoing progress is directed towards a more accurate IVE-based gesture recognition, exploring the challenges and advantages of multimodal interaction, and rehearsing a variety of multi-branching improvisational plays according to the suggestions of the audience. Inspired by Forsythe's choreographies [Gilpin, 1995], 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 – vs. 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 for public participation either during or after the performance without requiring the participants to learn all the script in advance.

This approach allows for flexible media choreography and contrasts scripted/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 [Brooks, 1996]. In our view, a rule based, scripted system lacks the responsiveness which creative artists demand. A scripted system cannot easily compensate for human errors or be responsive when some non-planned "magic" between the actors happens on stage. It tends to force human interpreters to rigidly follow a predefined track and therefore impoverishes the quality of the performance.

Figures 16, 17. Improvisational performer Kristin Hall in the IVE stage, at the MIT Media Lab, during the Digital Life Open House, on March 11, 1997.

Figures 18, 19. Actress Kristin Hall during a performance at the Sixth Biennial Symposium on Arts and Technology, in February 1997.

8. Current and Future Work: Wearable Performance

In this section we describe a scenario of current and future work aimed at porting our current stage technology to the mobile setting of street performance [also in: Sparacino, October 1997]. This is made possible by the rapid technological advances in Wearable Computing [Starner, 1997]. We describe motivations, modalities, and scenarios which scale and extend our IVE stage-based technology to a more portable context.

Wearable computers are transforming our technological landscape by reshaping the heavy, bulky desktop computer into a lightweight, portable device that is accessible people at any time. Although the computational power of wearable computers is certainly not equivalent to that of some high-end desktop computers, the portability and set of functionalities will nevertheless determine a migration of the computational engine from the house or the lab onto the user itself. An analog to this transformation can be found in the transition from the drama played in the theater building to the street theater.

Street and outdoor performance has a long historical tradition. However 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 and 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 [Mason, 1992], 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 that collaborate with the performer in the piece. 3. It can allow for new types of street performances that were not possible before the introduction and spread of wearable computers.

Bim Mason has carried out an extensive study of street performers [Mason, 1992]. He has defined five categories that group performers according to their motivation and artistic intent. There are: Entertainers, Animators, Provocateurs, Communicators and Performing Artists. We provide a short description of these categories and use them to classify the three types of wearable street performers presented in this section. *Entertainers* are defined as those performers with the simple aim of pleasing the audience, either by making them laugh or by impressing them with skills such as juggling, acrobatics or magic. In contrast, *Animators* play games with the audience. They use audience interaction not just for part of the show but as the main act itself. *Provocateurs* are more concerned with loosening-up society as a whole. They ask questions of society by going to the limits of conventionally acceptable behavior. *Communicators* see themselves as educators who feel they have something to teach to the rest of society or a message to pass on. Finally, *Performing Artists* are mainly interested in showing an artistic work, and their own personal view of art, focusing more on form rather than content. Our investigation in wearable performance describes examples of how some of these street artists’ work can be transformed by the new technology, according to the above mentioned taxonomy. We present three possible scenarios of investigation that we are currently evaluating.

8.1. The Performing Artist: The Augmented Mime

We create a performance that the public is able to enjoy both as is and also as an augmented performance. The augmented performance allows us to operate a semantic transformation of story fragments acted by the mime, through the use of the added computer graphics objects. In the augmented reality display water is turned into fire, simple inanimate objects become dangerously animated, and yawning generates expressive digital typography. The mime is wearing a small wearable computer [Starner, 1997] in his backpack. A flat panel display is connected to the wearable for the audience to use. Any member of the audience can take the display and hold it against the performer. The panel will act as an augmented performance loupe, showing digital props or graphically augmented objects that collaborate with the performer. This is done by attaching a very small camera the other side of the display such that when the viewer holds the display against the performer’s body the camera is also taking a wide angle video image of him. Through this artwork we are interested in exploring how point-of-view transforms our perception of reality. The “semantic lens” carried by the mime offers to members of the audience a new, transformed interpretation of the story told by the mime.

8.2. The Communicator: The Networked News Teller

The Networked News Teller is a street performer that carries a wearable computer with a “private eye”. The wearable runs a program that shows a constant update of news in the private eye of the performer. This program is set to explore a fixed set of news providers’ web pages. It then reconstructs a page that reports the same news from the different point of views of the different news providers. After having chosen a topic, the News Teller reads updated information about it through his private eye and interrogates the passerbyers about their opinion on the subject matter. She can then “perform the news” in the street based on her interaction with the public and the information appearing on her wearable. The performance consists in a re-interpretation or enactment of the different point of views expressed by the public interviewed in the street and the information published by the press. This type of street performance becomes particularly interesting when the news being reported is one that creates expectation and clustering of opposite opinions of the public. Good examples are presidential elections, or waiting for the verdict of the court in a trial. A performer who interprets artistically and “discusses” with the audience on the street themes that have great relevance and impact for our society, represents a social figure of stature.

8.3. The Entertainer: The One-Man Orchestra

Many of us are familiar with street musicians that are strapped with wires and carry a large number of musical instruments that can be triggered by different parts of their body. Usually they carry drums and cymbals that receive input from the feet, a mouth organ fixed in front of their mouth, and a guitar. All this gear is usually heavy to carry around and encumbering while playing. In addition the musical mapping between the musical instruments and the body parts cannot change during the performance. The One-Man Orchestra needs to carry only a lightweight wearable computer in his backpack. The computer is connected to five accelerometers placed on his head, hands and feet. The sensory information is processed in real time and a midi output is sent to a synthesizer. Loudspeakers and amplifiers can be worn or placed nearby. Music is generated by body movements by creating a significant mapping between body parts and musical instruments as well as between movements and musical notes. This mapping can dynamically change during the performance according to the artistic intent of the musician. Sound generated from the surrounding environment or speech coming from the audience can be easily sampled and integrated into the performance.

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

9. Conclusions

We have built an interactive stage for a single performer which allows us to coordinate and synchronize the performer's gestures, body movements, and speech, with projected images, graphics, expressive text, music, and sound. We have developed two cooperating technologies: one does optical, untethered, tracking of the human body in real time, and the other endows digital media with perceptual intelligence, expressive and communicative abilities, similar to those of a human performer (Media Actors). We have described how our technological augmentation of the stage is aligned along a historical direction of work and exploration in dance and theatrical performance. We have shown applications to dance and theater which augment the traditional performance stage with images, video, music, text, able to respond to movement and gesture in believable, esthetical, and expressive manners. Our main contribution to the field is that of introducing Media Actors – a digital media modeling technique which has room for improvisation and coordinated choreography with human performers. This technique enables both performers and programmers to escape the burden of scripted authoring, and offers a flexible and creative tool for the performance artist. We have also described a scenario and work in progress, which allow us to apply our artistic and technological advances to street performance.

Acknowledgements

The authors 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, Erica Drew and dancers: Malysa Monroe, Jennifer DePalo, Chung-fu Chang, Erna Greene, Naomi Housman, and Diana Aubourg, for their collaboration, interest, and encouragement. We thank Suguru Ishizaki for his suggestions of use of temporal typography in performance. Bruce Blumberg provided help and code for the first early implementation of Media Actors. Akira Kotani provided his SGI MIDI library, and Chloe Chao contributed to the computer graphics of DanceSpace. Edith Ackerman suggested readings on Baktin. Dimitri Negroponte helped direct and film the first rehearsals of Improvisational TheaterSpace. Tom Minka provided gracious support, useful comments, and suggestions.

Bibliography

Bakhtin, M. M., *The dialogical imagination.* (M. Holquist, Ed.) Austin: University of Texas Press, 1981.

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

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

Crowley J.L., Bedrune JM., *Integration and Control of Reactive Visual Processes*, 1994 European Conference on Computer Vision, (ECCV-'94), Stockholm, May 1994.

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

Darrell T., Moghaddam B., Pentland A., *Active face tracking and pose estimation in an interactive room.* In CVPR 96. IEEE Computer Society, 1996.

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

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

Huxley M., and Witts N. eds, *The Twentieth Century Performance Reader*, Routledge, 1996.

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

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

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

Machover Tod, *HyperInstruments: A Composer's Approach to the Evolution of Intelligent Musical Instruments*, CyberArts, William Freeman, San Francisco, pp. 67-76, 1992.

Marion, A., *Images of Human Motion: Changing Representation of Human Identity*. MS Thesis, Massachusetts Institute of Technology, Cambridge MA, 1982.

Mason B., *Street theatre and other outdoor performance*, Routledge Publisher, New York, 1992.

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

Oliver N., Pentland A., Berard F., *LAFTER: A real-time Lips and Face Tracker with Facial Expression recognition*, CVPR 97, IEEE Computer Society , San Juan, Puetro Rico, June 1997.

Paradiso J.A. and Sparacino F., *Optical Tracking for Music and Dance Performance*, Fourth Conference on Optical 3-D Measurement Techniques, Zurich Switzerland, September 29-October 2, 1997.

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

Rudlin J., *Commedia dell'Arte*, Routledge, 1994.

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

Sparacino F., DirectIVE: *Choreographing Media for Interactive Virtual Environments*, Master Thesis, MIT Media Lab, 1996.

Sparacino F., Hall K., Wren C., Davenport G., Pentland A., *Improvisational TheaterSpace*, The Sixth Biennal Symposium for Arts and Technology, Connecticut College, New London, CT, February 27 - March 2, 1997.

Sparacino F., Pentland A., Davenport G., *Wearable Performance*, International Symposium on Wearable Computers, Cambridge, Massachusetts, October 13 - 14, 1997.

Sparacino F., Davenport G., Pentland A., *Media Actors: Characters in Search of an Author*, submitted for pubblication, 1998.

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

Starner T., Mann S., Rhodes B., Levine J., Healey J., Kirsch D., Picard R.W. and Pentland A., *Augmented Reality Through Wearable Computing*, Presence, Special Issue on Augmented Reality, 1997.

Sukel K., Brostow G., and Essa I., *Towards and Interactive Computer-Based Dance Tutor*, Concept Paper, 1998.

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

Vygotsky, L. S., *Thinking and speech*. In R. W. Reiber and A. S. Carton (Eds.), (Trans., N. Minick), The collected works of L. S. Vygotsky. Vol. 1: Problems of General Psychology. New York, NY: Plenum, 1987.

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

Wren C., Azarbayeani A., Darrell T., Pentland A., *Pfinder: Real-time tracking of the human body*. IEEE Trans. Pattern Analysis and Machine Intelligence, 19(7):780-785, July 1997.

Wren C., Sparacino F., et al, *Perceptive Spaces for Performance and Entertainment: Untethered Interaction using Computer Vision and Audition*, Applied Artificial Intelligence (AAI) Journal, June 1996.

Zeltzer D., *Task Level Graphical Simulation: Abstraction, Representation and Control*, in: Making Them Move. Ed. Badler, N., Barsky, B., and Zeltzer D., Morgan Kaufmann Publishers, 1991.