Physically Interactive Story Environments

Claudio Pinhanez[*], James Davis[†], Stephen Intille[††], Michael Johnson[†],

Andrew Wilson[†], Aaron Bobick[†††], Bruce Blumberg[†]

[*]IBM TJ Watson Research Center, [†]MIT Media Laboratory,
[††]MIT Home of the Future Laboratory, [†††]Georgia Institute of Technology

ABSTRACT

Most interactive stories, such as hypertext narratives and interactive movies, achieve an interactive feel by allowing the user to choose between multiple story paths. This paper argues that in the case of real environments where the users physically interact with a narrative structure, we can substitute the choice mechanisms by creating situations that encourage and permit the user to actively engage his body in movement, provided that the story characters are highly reactive to the user's activity in small, local windows of the story. In particular, we found that compelling interactive narrative story systems can be perceived as highly responsive, engaging, and interactive even when the overall story has a single-path structure, in what we call a "less choice, more responsiveness" approach to the design of story-based interactive environments. We have also observed that unencumbering, rich sensor technology can facilitate local immersion as the story progresses – users can act as they typically would without worrying about manipulating a computer interface. To support these arguments, the paper describes the physical structure, the interactive story, the technology, and the user experience of four projects developed at the MIT Media Laboratory: *"The KidsRoom"* (1996), *"It / I"* (1997), *"Personal Aerobics Trainer"* (1998), and *"Swamped"* (1998).

1. INTRODUCTION

Following the pioneering work of Myron Kruger in the 1980s on creating computer-based interactive spaces [11], the beginning of the 1990s saw an explosion of physically interactive environments for entertainment where the users could explore an environment or interact with a character (for example, [12, 22, 23]). Initially confined to research laboratories, physically interactive environments are now commonly available in arcades and museums. At the same time, video games moved from simple shoot-and-kill scenarios to increasingly complex stories in multi-character stories such as *Myst*, role-playing games such as *Tomb Raider,* or strategic games such as *The Age of the Empires*.

The four projects developed at the MIT Media Laboratory described in this paper created interactive environments that physically engage their users as characters in a story in an attempt to merge the compelling interaction of physically interactive environments with the engagement and suspension of disbelief of complex stories. The *"KidsRoom"* (1996), *"It / I"* (1997), *"Personal Aerobics Trainer"* (1998), and *"Swamped!"* (1998) immerse their users in physical experiences each with beginning, development, and end, exploring the narrative notions of climax and catharsis in the interactive realm. We call such interactive, story-based computer-controlled environments *physically interactive story environments*.

The goal of this paper is to examine and discuss the technological and narrative mechanisms used to combine physical action with interactive stories in these four projects. In particular, these projects do not follow current conventions of interactive story-telling and computer games. First, no cumbersome devices such as joysticks or head-mounted displays are employed. Second, unlike many other VR interactive environments, realistic rendering is not used.

However, the most distinctive characteristic of these four projects is that the participants in those environments have very limited control over the overall development of the story. Although the characters and scene environments are responsive to the actions of the users at any given moment during the story, all users experience the same basic story. Traditional interactive stories, on the contrary, have tried to make the user feel the story responsive by providing him control over the story development through some form of choice among possible story paths [14]. The projects described in this paper do not use the choice mechanism. Nevertheless, our users feel that our environments are interactive. They appear to experience a sense of control over the story. We believe that this responsive feeling is conveyed because (1) our characters respond in compelling ways to small decisions and actions users make while interacting with them, and (2) because the users can engage in natural, pleasing physical movements during the experience. This constitutes what we call the "less choice, more responsiveness" design principle for interactive story environments.

These issues are initially discussed in section 2 of this paper in the context of previous work on physical story-based environments and interactive stories. Next, we describe the four projects we have developed at the Media Laboratory. We conclude by comparing the four interactive experiences, discussing possible reasons for the success of our approach of "less choice, more responsiveness" in designing physically interactive story environments.

2. PHYSICALLY INTERACTIVE STORIES

Since ancient times children have been playing games where they pretend to be characters living real and fantastic stories. Similarly, role-playing has been part of human rituals for millennia in religious ceremonies and theatrical performances (see [21]). Role-playing combines the emotional structure of narrative with the physical activity of the body to create a powerful sense of corporeal immersion in a character, environment, or communal act. The sensation of immersion in such situations is many times enhanced by the use of scenarios, costumes and ritual objects as well as by the presence of "professional" performers portraying some of the characters.

New technologies have been employed throughout the ages to increase the feeling of physical immersion in stories. The advent of computers made it possible to create compelling story-based environments with realistic imagery and sound where computer characters are always available — 24 hours a day — to help play out the user's fantasy.

2.1 Physical Story-Based Environments

Physical realizations of stories have been part of human culture for centuries (e.g. theater and religious rituals). The 19th century's *panoramas* were one of the first examples of capturing stories in environments controlled by machines. A panorama consists of an observation platform surrounded by a painted canvas. To create the illusion of reality, the edges of the canvas are hidden by an umbrella-shaped roof that covers the platform and by a "false terrain" projecting from the observation platform. Panoramas were extremely popular throughout the 19th century. They depicted landscapes, battle scenes, and journeys (see [15] for a history of panoramas).

Although mechanical "haunted houses" have populated carnivals and fairs of the 20th century, the development of *Disneyland* and other theme parks pushed the limits of technology in terms of creating vivid physical renditions of characters and stories with machines. *Disneyland* pioneered the use of *animatronic puppets*, sophisticated robotic devices with life-like movement. These puppets are used to endlessly reproduce a physical rendition of a story where the characters display a fairly complex set of actions.

However, the traditional theme park ride lacks interactivity. Visitors exist in the story space but their actions are never actually reflected in the development of the story or the life of the characters. Although many theme park rides move, shake, and play with the participants' sense of equilibrium, quite commonly the users' physical activity is severely restricted.

About the same time that animatronics became popular, tiny, extremely reactive characters started to populate the screens of video games in arcades. With simple devices such as buttons and joysticks it became possible to interact with such characters in environments that, though responsive, were very limited in their physical immersion: the user's connection to that environment was restricted to a joystick and a video display. Since then the interfaces of arcade games have advanced considerably, enabling full-body action and sensing. Examples include many skiing and cycling games and the recent Japanese “dance” game *Dance Dance Revolution* [1].

There are only a few cases where the full-body arcade interactiveness has been combined with more complex story and characters. A good example is the ride *“Aladdin”*, developed and tested at *Disneyworld* [16], where four users wearing VR helmets loosely control a voyage through a city on a flying carpet. Although most of the user control is restricted to deciding what to see, the presence of the users affects the behavior of the characters in the city.

2.2 Interactive Stories

Most of the academic discussion about interactive stories has been concentrated in the field of literature, which paradoxically uses one of the least interactive media: paper. In [14], Murray examines the characteristics of interactive narrative, particularly in the cases where the story develops as a reaction to the reader’s action. Normally this action consists of choosing the next development in the story from a small set of options proposed by the author.

Many “narrative” video-games, such as *Dragon’s Lair*, are also based on a “choice” structure, although in this case the user’s goal is typically to discover patterns of choice that minimize the traversal of all possible paths. Similarly, there have been attempts to create *interactive movies* where the audience decides, by voting, between different paths for the characters. The most widely known example, *“Mr. Payback”* (*Interfilm*, 1995), was coldly received by the critics and was not commercially successful.

However, most studies of interactive stories (such as [14]) neglect the rich experience of role-playing in dramatic games and human rituals. In spite of all the interaction among role-playing participants, the choices made by the participants tend to keep the game inside of the limitation of the “game reality” (which can encompass a lot of fantasy). Role-players normally avoid the full exploration of the tree of multiple paths (including its uncontrolled dead ends), and seem to extract most of the pleasure from the portions of rewarding mutual interaction and character discovery that happen during the game play through fortuitous encounters.

2.3 Less Choice, More Responsiveness

The two sections above herald the main hypotheses of this paper that in physically interactive stories responsiveness is likely to be more important than choice. We have reached this conclusion based on the fact that we have designed and constructed engaging experiences that feel highly interactive without really providing any real decision power over the story to the participants. These interactive environments, described later in this paper, clearly show that choice is not a prerequisite for interactive stories. It could be argued that it is possible to structure the interactivity in a story-based environment around mechanisms of choice. However, so far we have not seen a successful implementation of a physically interactive story environment based on extensive story choice.

Galyean [8] proposed the *water-skier* model for interactive experiences that is similar to the *less choice, more responsiveness* paradigm proposed here. In the water-skier model, the user is compared to a water-skier who is unable to determine the direction of the pulling boat (the story) but who has some freedom to explore his current situation in the story. This model was employed by Blumberg and Galyean in the *"Dogmatic"* VR experience [4], where the user's avatar encounters a dog, *Lucky*, in a desert town and, ultimately, is led by the dog to her own death. As users look around and decide where to go, they involuntarily became part of the story. Most of the pleasure associated with *"Dogmatic"* seems to be derived from the process of understanding the story. Although the water-skier model advocates limited story choice, it fails to observe that this limitation on story control can be compensated for with an increase in responsiveness and local interaction.

Unlike *"Dogmatic"*, the projects described here focus on creating rewarding experiences in the process of interacting with the characters in the story and in the physical aspects of this interaction. That is, like in many physical rituals and theatrical games, the satisfaction in such interactive stories comes from the pleasure of small, localized interactions with the other characters.

It is possible to keep the users and the characters in the context of a well-structured and interesting story by concentrating story development on local interactions instead of providing multiple story paths. A key problem with multiple-path stories is that some paths are considerably weaker than others. As described by Donald Marinelli [13], choice-based interactive stories are like season tickets to hockey: the experience involves some good games, some boring games, and hopefully a truly remarkable evening that becomes a story to be told to our kids. A great story is a very special, fortunate, and rare conjunction of ideas, events, and characters.

By developing a system that is locally responsive to user actions as the user progresses through a single-threaded story, we can assure that our users always receive the full impact of the best possible story (as hand-crafted by its author) without losing the impression that the story unfolds as a consequence of the participants' actions. To illustrate these ideas, we now examine four projects developed at the MIT Media Laboratory from 1996 to 1999.

3. THE KIDSROOM

"The KidsRoom" project aims to create a child's bedroom where children can enact a simple fantasy story and interact with computer-controlled cartoonish monsters. It is a multi-user experience for children between 6 and 12 years old where the main action happens in the physical space of the room and not "behind the computer screen", as in most video games. Furthermore, the children are not encumbered by sensing devices, so they can actively walk, run, and move their bodies. A detailed description of the project can be found in [5].

3.1 The Physical Setup

"The KidsRoom" is built in a space 24 by 18 feet with a wire-grid ceiling 27 feet high. Two of the bedroom walls resemble real walls of a child's room, complete with real furniture and decoration. The

other two walls are large video projection screens, where images are back-projected from outside of the room. Behind the screens there is a computer cluster with six machines that automatically control the interactive space. Computer-controlled theatrical colored lights on the ceiling illuminate the space. Four speakers, one on each wall, project sound effects and music into the space. Finally, there are three video cameras and one microphone used to sense the children's activity in the room. Figure 1 shows a view of the complete *"KidsRoom"* installation.

The room has five types of output for motivating participants: video, music, recorded voice narration, sound effects, and lighting. Still-frame video animation is projected on the two walls. Voices of a narrator and monsters, as well as other sound effects, are directionally controlled using the four speakers. Colored lighting changes are used to mark important transitions.

3.2 The Interactive Story

"The KidsRoom" story begins in a normal-looking bedroom. Children enter after being told to find out the magic word by "asking" the talking furniture, which speaks when approached. When the children scream the magic word loudly, the room transforms into a mystical forest. There, the children have to stay in a single group and follow a defined path to a river. Along the way, they encounter roaring monsters and, to stop them, they have to quickly hide behind a bed. To guide the participants, the voice of a narrator, speaking in couplets, periodically suggests what the children can do in the current situation.

After some time walking in this forest, the children reach a river and the narrator tell them that the bed is now a magic boat that will take them on an adventure. The children can climb on the "boat" and pretend to paddle to make it "move," while images of the river flowing appear on the screens. To avoid obstacles in the river the children have to collaboratively row on the appropriate side of the bed; if they hit the obstacles, a loud noise is heard. Often, the children add "realism" by pushing each other.

Next, the children reach the monster world. The monsters then appear on the screens and show the kids some dance steps. The children have to learn these dance steps to become friends of the monsters. The monsters then mimic the children as they perform the dance moves. Finally, the kids are commanded to go to bed by an insistent, motherly voice, and the adventure ends with the room transforming back to the normal bedroom.

3.3 The Technology

Three cameras overlooking the *"KidsRoom"* are used for the computer vision analysis of the scene. One of the cameras (marked as the "overhead camera" in fig. 1) is used for tracking the people and the bed in the space. The overhead position of the camera minimizes the possibility of one user or object occluding another. Further, lighting is assumed to remain constant during the time that the tracker is running. Standard background subtraction techniques (described in detail in [6]) are used to segment objects from the background, and the foreground pixels are clustered into 2D blob regions. The algorithm then maps each person known to be in the room to a blob in the incoming image frame. In the scene with the boat, the overhead camera is also used to detect whether the children are rowing and in which direction.

The other two cameras (marked as "recognition cameras" in fig. 1) are used to recognize the dance steps performed by the children during the last scene with the monsters. The images from these cameras are processed to segment the silhouette of the child facing the screen. Using motion-template techniques (see [6]) the system distinguishes the occurrence of four different "monster dance steps": crouching, spinning, flapping arms, or making a "Y" figure.

3.4 The Experience

"The KidsRoom" was designed and built in the fall of 1996. The installation was experienced by dozens of children and adults during the three months it remained open (see fig. 2). A new, shortened version of the *"KidsRoom,"* the *"KidsRoom2,"* was built in 1999 in London as part of the Millennium Dome Project and is scheduled to run continuously through the year 2000.

A typical run of the *"KidsRoom"* at the Media Laboratory lasts 12 minutes for children and usually slightly longer for adults. Not surprisingly, we found children to be more willing to engage in the action of the story and to follow the instructions of the narrator. Children are typically very active when they are in the space, running from place to place, dancing, and acting out rowing and exploring fantasies. They interact with each other as much as they do with the virtual objects, and their exploration of the real space and the transformation of real objects (e.g. the bed) enhance the story.

From our observation of the children, there has never been a situation where the children did not understand that they are characters in a story and that they have to act out their parts to make the story flow. Occasionally the children do not understand the instructions and the experience have small breaks in its flow. However, the control software of the *"KidsRoom"* is designed to always push the story forward, so such interruptions are usually overcome quickly.

The story of the *"KidsRoom"* ties the physical space, the participant's actions, and the different output media together into a coherent, rich, and immersive experience. In particular, the existence of a story seems to make people, and especially children, more likely to cooperate with the room than resist it and test its limits. The well-crafted story also seems to make participants more likely to suspend disbelief and more curious and less apprehensive about what will happen next.

In fact, the users of the *"KidsRoom"* have absolutely no control on the overall story development and do not seem concerned at all about that. Some of the best moments of the experience, as judged by enthusiastic reaction of the young users, are connected to simple physical activities (augmented by image and sound) such as rowing on the river, dancing with "live" cartoonish monsters, or simply running in a group towards the bed, away from the monsters, and piling on top of each other.

4. IT / I

"It / I" is a theater play where a computer system plays one of the play's two characters. The computer character, called ***It***, has a non-human body composed of computer graphic (CG) objects projected onto

rear-projection video screens. The objects are used to play with the human character, ***I***, performed by a real actor on a stage. The play is a typical example of *computer theater*, a term (proposed by Pinhanez in [17]) that refers to theatrical experiences involving computers, in a direct analogy to the idea of *computer music*.

"It / I" was created in the context of two main goals. The first is to design an automatic interactive computer character that can co-inhabit the stage with a human performer in front of an audience through the length of a complex story. This creates strong requirements in terms of expressiveness and reliability on the computer actor. The second goal is to create a space where the public can re-enact a story they have watched by taking the place of the human performer, in what Pinhanez calls an *immersive stage* (see [18]). A detailed description of the play and its underlying technology can be found in [18, 19].

4.1 The Physical Setup

Figure 3 depicts a diagram of the different components of the physical setup of *"It / I"*. The sensor system was composed of three cameras rigged in front of the stage. The computers controlled different output devices: two large back-projected screens; speakers connected to a MIDI-synthesizer; and stage lights controlled by a MIDI light board.

4.2 The Interactive Story

"It / I" depicts a story about the relations between mankind and technology. The character played by the computer, called ***It,*** represents the technology surrounding and many times controlling us; that is, in *"It / I",* the computer plays itself. ***It*** is, in fact, quite an unusual creature: it has a "body" composed of CG-objects — representing clocks, cameras, televisions, electrical switches — projected on stage screens. ***It*** can "speak" through large, moving images and videos projected on the screens, through musical sounds played on stage speakers, and through the stage lights.

The play is composed of four scenes, each being a repetition of a basic cycle: ***I*** is lured by ***It***, ***I*** is played with, ***I*** gets frustrated, ***I*** quits, and is ***I*** punished by ***It*** for quitting. For instance, in the second scene a CG-object similar to a photographic camera appears on a small screen and follows ***I*** around. When ***I*** accepts the game and makes a pose for the camera, the camera's shutter opens with a burst of light. Then, on the other screen, a CG-television appears, displaying a slide show composed by silhouette images "taken" by the camera. After some pictures are shown, the camera "calls" ***I*** to take another picture. This cycle is repeated until ***I*** refuses to take yet another picture (that is, the human performer decides it is a good time to finish the cycle), provoking an irate reaction from ***It***, which in response throws CG-blocks at ***I*** while flickering the lights and playing threatening sounds.

4.3 The Technology

The primary sensors in *"It / I"* are three video cameras positioned in front of the stage. Using the information from the three cameras it is possible to segment the human character from the background using a stereo system similar to [9], independently of the stage lighting and changes on the background screens.

The play was written taking into account the sensory limitations of computer vision technology. That is, the actions of ***I*** are restricted to those that the computer can recognize automatically through image processing. In many ways, ***It***'s understanding of the world reflects the state-of-art of real-time automatic vision: the character's reaction is mostly based on tracking ***I***'s movements and position and on the recognition of some specific gestures (using [6]).

Unlike most interactive environments, *"It / I"* portrays a long and complex story that lasts for about 40 minutes. Additionally, the control system of the play has to be extremely robust to cope with the requirement of live performances in front of large audiences. Since at the time when the play was produced there were no story representational languages able to satisfy both requirements (see [20]), it became necessary to develop a special language for representation of interactive stories. In *"It / I"* the control of all the sensor and actuator systems is described in an *interval script*, a paradigm for interaction scripting based on the concept of time intervals and temporal relationships developed by Pinhanez [18], based on previous work with Mase and Bobick [20]. A description of the interval script paradigm is beyond the scope of this paper and can be found in [18].

4.4 The Experience

"It / I" was performed six times at the MIT Media Laboratory for a total audience of about 500 people. The audience clearly understood the computer character's actions and intentions and the play managed to keep the "suspension of disbelief" throughout its 40 minutes. In particular, the sound effects played a key role on creating the illusion that ***It*** was alive and to convey the mood and personality of the character. Each performance was followed by an explanation of the workings of the computer-actor. After that, the audience was invited to go up on stage and play the second scene (as described above) first in front of the audience, and individually afterwards (see fig. 5).

Theater scripts are clear examples of stories where the general story structure and development is fixed but the realization of the character interaction is left to the performers. Although actors usually have no influence on how the story unfolds, they are responsible for discovering and creating the minutia of the moment-by-moment inter-character relations.

"It / I" follows this traditional structure and therefore, by design, creates an interactive computer environment based on responsiveness. Given the argument above, its similarity to traditional theater makes it clearly a comfortable place for the actor. Beyond that, we observed that the audience also enjoyed engaging in this experience where they have no control of the final outcome of the story but where play-acting is fun.

The lack of story control in *"It / I"* is compensated for by expanding the repertoire of physical interaction. During the play, the actor (and later, the audience) is able not only to explore physical activity but also to use the body to produce and change imagery, sound, and lighting. This is certainly one of the possibilities not present in traditional role-playing that is opened by computer-mediated spaces for physically interactive stories.

5. PERSONAL AEROBIC TRAINER (PAT)

While the two projects described above create stories populated by fantastic characters, the *"Personal Aerobic Trainer"* project, or *"PAT"*, is focused on the quality of the physical activity of the user. The main goal of *"PAT"* is to create a system that helps a user to workout by enthusiastically pushing him through a series of aerobics exercises while monitoring his activity and correcting his movements. A detailed description of the *"PAT"* system can be found in [7].

5.1 The Physical Setup

The silhouetting method for monitoring the user employed in the *"PAT"* system is based on the optical blocking (or eclipsing) of infrared light (IR) rather than the color differences between the person and background (like in *"The KidsRoom"*) or stereo disparity (like in *"It / I"*). This is necessary because to monitor the quality of an aerobic movement it is necessary to have a very precise and sharp silhouette of the user. However, the space in *"PAT"* is carefully engineered to hide the IR and the sensing apparatus.

Figure 6 shows the environment configuration of the *"PAT"* system. It consists of a room where two of the walls are replaced by large screens where video is back-projected. Behind one of the screens there is an array of IR emitters that evenly lights the screen. In front of the opposite wall a camera equipped with an IR filter is positioned facing the IR illuminated screen.

This configuration allows the camera to quickly and effortlessly obtain a high-quality silhouette of the user, in spite of light changes in the room and of the images in the videos projected in the screens. Notice that the infrared light is not visible to the human visual system and thus the user only sees the video projected on the display screens.

5.2 The Interactive Story

The experience in *"PAT"* starts when the user enters the space. The entrance of the user triggers the opening of a video window in the screen with the camera, portraying a virtual instructor (an Army drill sergeant) welcoming the user. The instructor images are obtained from a collection of pre-recorded video clips depicting a multitude of actions, attitudes, and comments spanning a reasonable range of possible reactions for the drill sergeant.

After the brief introduction, the system goes through a sequence of physical exercises. Ideally, that sequence would be personalized according to the identity and physical history of the user. For each exercise, the instructor executes the moves and accompanies the user while he is performing. The drill instructor gives feedback, often humorous, based on how well the user is performing the exercises. During each exercise, background music is synchronized to the user movements (unlike workout videos, which make the user follow the pace of the music). After the workout is complete, the instructor congratulates the user. If the user prematurely leaves the space, the instructor stops and leaves the screen.

5.3 The Technology

The *"PAT"* system employs the same real-time computer vision methods for recognizing large-scale body movements used in the two previously described projects [6]. The method is based on the layering of participant silhouettes over time onto a single template and measuring shape properties of that template to recognize various aerobic exercise (and other) movements in real-time.

Figure 8 presents templates generated from the infrared silhouettes for the movements of left-arm-raise (left-side stretch) and fan-up-both-arms (deep-breathing exercise stretch). The motion of the user is encoded in the varying gray levels within the template. For recognition of these moves, statistical pattern recognition techniques are applied to moment-based feature descriptors of the templates. The system employs user training to get a measure of the variation that arise from different people (see [6] for details on the algorithm).

5.4 The Experience

The first prototype for the *"Personal Aerobics Trainer"* was set up at the Villers Facility of the MIT Media Laboratory in December of 1997 and was experienced by many users in a period of 3 months. Figure 7 displays a model of the interaction of the "PAT" system, where a user is exercising in front of a TV. Users can easily understand the structure of the interaction and become naturally engaged in the "narrative". The comments of the drill sergeant seem to help to make the routine more personal, besides creating a sense of responsibility and accomplishment lacking in traditional workout videotapes.

Although the *"PAT"* system does not have a traditional story line like the three other projects described in this paper, the experience is clearly structured as a narrative. Moreover, it employs a character to interact with the user who, in many ways, tends to assume the role of an Army private.

"PAT" exemplifies a temporally structured human activity where immediate response to physical activity is more important than narrative choice. In fact, the critical aspect of a workout experience is to make the user correctly perform a sequence of physical activities by managing his natural desire to quit. In other words, the system strives *to prevent choice*. Although the adaptation of the system to the user's pace is important, making the user persevere is achieved in *"PAT"* mostly by the personification of the control system, which creates the feeling that the user is being watched and stimulated by another being – the drill sergeant. Moreover, in *"PAT"* the physical activity is the center of the interaction and its healthy benefits constitute the basic source of reward.

6. SWAMPED!

"Swamped!" is an interactive system, developed at the Media Laboratory in 1998, in which the participant assumes the role of a chicken that is trying to protect its eggs from a hungry raccoon in a barnyard setting. Unlike the full body interfaces of the previous projects, in *"Swamped!"* the user

controls a character by manipulating a plush doll representing the character. However, one of the main goals of the project is to examine how a manipulative interface can be used not to explicitly control the character's body movements (like in most shoot-and-kill video-games) but instead to suggest a line of action for a character. A detailed description of the project can be found in [10].

6.1 The Physical Setup

In *"Swamped!"* the user stands in front of a projection screen showing the virtual world and the virtual chicken while holding a plush doll similar to the chicken. She can direct the chicken by making appropriate gestures with the doll. For example, wobbling the doll back and forth makes the virtual chicken walk and flapping the doll's wings will make it fly. The participant's attention is meant to focus on the interactions in the virtual world and not on the doll itself. Figure 9 shows a user in front of the screen holding the "controller" doll while watching the unfolding saga between the chicken and the raccoon.

6.2 The Interactive Story

When the interaction starts in *"Swamped!"*, the user discovers that he is playing the role of a chicken trying to protect its eggs from a raccoon. Figure 10 shows a picture from a typical run. The chicken has various behaviors such as squawking to get the raccoon's attention and make it angry, scratching its own head, kicking the raccoon, and setting a trap for the raccoon. As described above, these behaviors are selected according to the doll's movements and the story context. The raccoon is fully autonomous, choosing what actions to take based on its desires, perceptions, and emotional state [4].

In a normal interaction, the raccoon's attempts to get the eggs are blocked by the user-manipulated chicken. However, the raccoon eventually gets one egg and then runs away with it. Next, it stops to examine it on a giant bulls-eye painted on the ground. Guess what? When the raccoon looks up, a heavy weight descends from the sky and smashes it…

6.3 The Technology

The physical doll used to control the chicken character is fabricated to match the virtual character. An armature made of plastic, brass tubing and wire holds a sensor package and provides an articulated structure (see fig. 11). The sensor package inside the doll includes an array of 13 sensors: two pitch and roll sensors; one gyroscope sensing roll velocity; three orthogonally mounted magnetometers sensing orientation with respect to magnetic north; two flexion (FSR) sensors for wing position; three squeeze (PVDF) sensors embedded in the body and beak, and one potentiometer to sense head rotation about the neck.

Raw data from the doll is processed in real-time on the host computer to recognize gestures that are taught to the system in a learning phase. The system can detect a variety of actions of the doll under user control, such as *walk*, *run*, *fly*, *squeeze-belly*, *hop*, *kick* and *back flip*. Each of these action primitives is learned off-line and recognized using hidden Markov models *(HMMs)* [24].

In *"Swamped!",* the chicken's behavior system treats the output of each HMM as a sensory input to a corresponding consummatory behavior, using a reactive behavior system similar to [3]. For example, when the user flaps the chicken's wings the HMM for the flying gesture surpasses its threshold and stimulates the flying behavior. If this is the most appropriate behavior at the time, the flying behavior becomes active, which causes the virtual chicken to begin flying.

6.4 The Experience

Over 400 users interacted with the *Swamped!* installation in the *Enhanced Realities* exhibit at SIGGRAPH'98. Users were told the cartoon scenario and that the goal was to keep the raccoon busy so that it would not eat any eggs by engaging in the chicken's various behaviors.

In general, three categories of users are encountered: *teachable*, *ideal*, and *skeptical* (in order of approximate group size). The *ideal* users are often children who would pick up the doll, start manipulating it, and immediately understand the concept of the interface. The *teachable* users are by far the largest group. The typical member of this group picks up the doll and tries to manipulate one part, such as one wing or a foot, expecting a direct mapping. After a walking gesture and the "voodoo doll" metaphor is demonstrated, many of these users can quickly learn to use the doll and enjoy the experience. Several users, however, never understand how the doll controls the character and are even *skeptical* about connections between the doll's moves and the character's behavior.

Although the story line of *"Swamped!"* is quite simple, it sets up clear goals to guide the user interaction. Moreover, the raccoon character fulfils the dramatic role of pushing the story ahead, culminating, pathetically, in its own smashing. However, there is no doubt that involving the user in a story has been fundamental in terms of setting up clear objectives for the manipulation of the doll.

Although it can be argued that the user is making choices in terms of story whenever she decides for a particular line of action for the chicken, it is important to recognize that most of the pleasure of the installation seems to come from the manipulation of the doll per se. By combining motor activity and character behavior, *"Swamped!"* is able to produce a novel kind of user immersion that we doubt could have happened if, for instance, the user was shouting to the chicken what to do.

7. DISCUSSION

The main argument of this paper, based on our experiences in the four projects described above, is that in physically interactive stories, immersion and satisfaction can be achieved by creating very responsive physical interactions between the users and the characters and without the use of choice among multiple story paths. This conclusion was reached by observing users while they were experiencing the environments described above. We have not conducted formal studies on user satisfaction mostly because of logistical issues but also because of the lack of established, widely accepted methodologies to measure user pleasure or engagement in a story. User testing has been centered on measuring satisfaction towards the completion of tasks and there is little literature on how to determine how engaging a story is or how pleasing a physical interaction is. Simple methods such as counting laughter during the sessions or making the users answer questionnaires were discarded since they seem to be able to capture only small fractions of the meaning of "having fun" or "being emotionally engaged".

In the projects described above, we experimented with three types of story participation: the user can pretend to be a character, as in the *"KidsRoom,"* in *"PAT,"* or when the audience played with *"It / I;"* the user can be the performer of a character, like the actor in *"It / I;"* and finally, the user can control (puppeteer) a character, as in *"Swamped!"*.

We have not worked on projects where the user becomes the master of the story and is able to control the many different characters. We have also not worked on projects where the user participates in a story as the audience. An example of the latter case is the interaction between the audience and the computer graphics character presented at the SIGGRAPH'99 electronic theater in 1999 (see [2]).

Although it is not clear whether our argument holds in those "story master" situations, it probably holds in the case of audience participation. In fact, audience participation in theater has been mostly successful exactly when the performers manage the audience responses in order to keep the story going in a particular direction. The many attempts in theater, especially in the 60s, of letting the audience control the development and the direction of a play did not prove popular with audiences.

Given that, it is important to examine possible reasons why physical interaction and responsiveness is able to complement the experience in such a way that control of the story seems unimportant. A possible explanation relates to the kinesthetic pleasure associated with moving the body (like the pleasure of dancing). Our conjecture is that by making the body an interface, we can extract aesthetic pleasure and engagement from the variety of responses coming from the muscle actuators and skin sensors to the point that they subsume the need of intellectual pleasure related to determining the developments of a story.

It is important to differentiate this notion of bodily pleasure with the achievement of high-level hand-eye coordination, the foundation of many video games. In our projects, we never had to resort to skill mastery as a way to reward the participants (except, in a different way, in the *"PAT"* project). In physically interactive stories, the participant's goal should be to immerse herself as much as possible in the story, i.e., to "live" the story.

The projects described here show how physical immersion can be greatly enhanced by using non-encumbering sensing mechanisms. The spontaneity of the movements of the children in the *"KidsRoom"* could not be achieved if they had wires or head-mounted displays attached to them. Devices can become a constant (and heavy) reminder that the story is just an illusion, making the "suspension of disbelief" harder to achieve. Furthermore, the gear can interfere with the pleasure of moving which, in our opinion, is a major source of reward in those experiences Of course, given the precarious nature of both computer vision and movement detection technology, it is necessary to carefully craft the stories so that they can accommodate the limitations of current sensing systems. In fact, we found this to be a prevailing issue when developing the four projects described in this paper.

Although we believe that responsiveness is probably a more important factor than story control in physically interactive environments, we found framing our experiences within a story to be extremely positive. Stories seem to give a sense of purpose and discovery to the participants, considerably enriching their interactive experiences. In particular, we found a good story resolution, after a story climax, to be very rewarding for the users. Unlike the "game over" messages in video games, which thrive on frustrating the user out of the story structure, we found it important to keep the participant inside the story up to the last moment and to make the end a natural development of the experience.

8. CONCLUSION

"Stories do not require us to do anything except to pay attention as they are told." (Murray [14], pg. 140). In this paper, on the contrary, we examined four projects in physically interactive stories designed and built to be experienced by people. Based on those experiences, we made a set of

observations that, besides being useful to the design of future interactive stories, seem to run against some common beliefs in the literature of interactive narratives.

First, all the environments were built based on complex sensory mechanisms designed to make the interaction as natural as possible, completely avoiding cumbersome sensing apparatus such as head-mounted displays or body-tracking suits. By doing so, it is possible to explore using kinesthetic pleasure as an element to reward the participant.

Second, all the environments do not rely on realistic computer graphics or simulations. Most of our characters are cartoonish and in some cases use either low image refresh rates or non-human bodies. However, the characters seem to have been perceived as responsive, intentional, and goal-oriented, largely as the result of the combination of their responsiveness and the context of the story.

Third, unlike most hypertext interactive stories, the feeling of interaction is not based on explicit mechanisms of choice but on making the environments and the characters that inhabited them extremely responsive. The sensation of immersion is, in fact, mostly created by the extensive physical activity of the user.

Finally, the four projects realize complete narratives that take the participants through a clear path with an introduction, character and story development, and a climatic end. Our experience suggests that a well-structured story has the power to engage the users effectively in meaningful interaction.

ACKNOWLEDGEMENTS

All four projects described in this paper were sponsored by the M.I.T. Media Laboratory. *"The KidsRoom"* was developed by Aaron Bobick, Stephen Intille, James Davis, Freedom Baird, Claudio Pinhanez, Lee Campbell, Yuri Ivanov, Arjan Schutte, and Andrew Wilson. *"It / I"* was written and directed by Claudio Pinhanez, and produced by Aaron Bobick; the crew was composed by John Liu, Chris Bentzel, Raquel Coelho, Leslie Bondaryk, Freedom Baird, Richard Marcus, Monica Pinhanez, Nathalie van Bockstaele, and the actor Joshua Pritchard. *"PAT"* was developed by James Davis and

Aaron Bobick, with actor Andrew Lippman. "*Swamped!*" was developed by Bruce Blumberg, Michael P. Johnson, Michal Hlavac, Christopher Kline, Ken Russell, Bill Tomlinson, Song-Yee Yoon, Andrew Wilson, Teresa Marrin, Aaron Bobick, Joe Paradiso, Jed Wahl, Zoe Teegarden, and Dan Stiehl.

Claudio Pinhanez was partially supported by a scholarship from CNPq, process number 20.3117/89.1.

REFERENCES

[1] http://www.konami.co.jp/am/ddr.1999.

[2] B. Blau, M. Butler, and A. Goldberg. "M.C. Leon", *SIGGRAPH'99 Electronic Art and Animation Catalog*, pp. 108. 1999.

[3] B. Blumberg. *Old Tricks, New Dogs: Ethology and Interactive Creatures*. Ph.D. Thesis. Media Arts and Sciences Program: Massachusetts Institute of Technology, Cambridge, Massachusetts. 1996.

[4] B. M. Blumberg and T. A. Galyean. "Multi-Level Direction of Autonomous Agents for Real-Time Virtual Environments", *Proc. of SIGGRAPH'95*. 1995.

[5] A. Bobick, S. Intille, J. Davis, F. Baird, C. Pinhanez, L. Campbell, Y. Ivanov, A. Schutte, and A. Wilson. "The KidsRoom: A Perceptually-Based Interactive Immersive Story Environment", *PRESENCE: Teleoperators and Virtual Environments,* vol. 8 (4), pp. 367-391. 1999.

[6] J. W. Davis and A. Bobick. "The Representation and Recognition of Human Movement Using Temporal Templates", *Proc. of CVPR'97*, pp. 928-934. June. 1997.

[7] J. W. Davis and A. F. Bobick. "Virtual PAT: a Virtual Personal Aerobics Trainer", *Proc. of Workshop on Perceptual User Interfaces (PUI'98)*, San Francisco, California, pp. 13-18. November. 1998.

[8] T. A. Galyean. *Narrative Guidance of Interactivity*. Ph.D. Thesis. Media Arts and Sciences Program: Massachusetts Institute of Technology, Cambridge, Massachusetts. 1995.

[9] Y. Ivanov, A. Bobick, and J. Liu. "Fast Lighting Independent Background Subtraction", *Proc. of the IEEE Workshop on Visual Surveillance (VS'98)*, Bombay, India, pp. 49-55. January. 1998.

[10] M. Johnson, A. Wilson, C. Kline, B. Blumberg, and A. Bobick. "Sympathetic Interfaces: Using a Plush Toy to Direct Synthetic Characters", *Proc. of CHI'99*, Pittsburgh, Pennsylvania. May. 1999.

[11] M. W. Krueger. *Artificial Reality II*. Addison-Wesley. 1990.

[12] P. Maes, T. Darrell, B. Blumberg, and A. Pentland. "The ALIVE System: Full-Body Interaction with Autonomous Agents", *Proc. of the Computer Animation'95 Conference*, Geneva, Switzerland. April. 1995.

[13] D. Marinelli. in *ACM Multimedia'98 Workshop on Technologies for Interactive Movies*, Bristol, England. 1998.

[14] J. Murray. *Hamlet on the Holodeck: the Future of Narrative in Cyberspace*. The Free Press, Simon & Schuster, New York, New York. 1997.

[15] S. Oettermann. *The Panorama: History of a Mass Medium*. Zone Books, New York. 407 pages. 1997.

[16] R. Pausch, J. Snoddy, R. Taylor, S. Watson, and E. Haseltine. "Disney's Alladin: First Steps Toward Storytelling in Virtual Reality", *Proc. of SIGGRAPH'96*, pp. 193-203. August. 1996.

[17] C. S. Pinhanez. "Computer Theater", *Proc. of the Eighth International Symposium on Electronic Arts (ISEA'97)*, Chicago, Illinois. September. 1997.

[18] C. S. Pinhanez. *Representation and Recognition of Action in Interactive Spaces*. Ph.D. Thesis. Media Arts and Sciences Program: Massachusetts Institute of Technology. 1999.

[19] C. S. Pinhanez and A. F. Bobick. "'It/I': A Theater Play Featuring an Autonomous Computer Graphics Character", *Proc. of the ACM Multimedia'98 Workshop on Technologies for Interactive Movies*, Bristol, England, pp. 22-29. September. 1998.

[20] C. S. Pinhanez, K. Mase, and A. F. Bobick. "Interval Scripts: A Design Paradigm for Story-Based Interactive Systems", *Proc. of CHI'97*, Atlanta, Georgia, pp. 287-294. March. 1997.

[21] R. Schechner. *Performance Theory*. Routledge, London, England. 1988.

[22] C. Sommerer and L. Mignonneau. "Art as a Living System", *Leonardo,* vol. 30 (5). 1997.

[23] N. Tosa, H. Hashimoto, K. Sezaki, Y. Kunii, T. Yamada, K. Sabe, R. Nishino, H. Harashima, and F. Harashima. "Network-Based Neuro-Baby with Robotic Hand", *Proc. of IJCAI'95 Workshop on Entertainment and AI/Alife*, Montreal, Canada. August. 1995.

[24] A. Wilson, A. F. Bobick, and J. Cassell. "Temporal Classification of Natural Gesture and Application to Video Coding", *Proc. of CVPR'97*, Puerto Rico, USA, pp. 948-954. 1997.

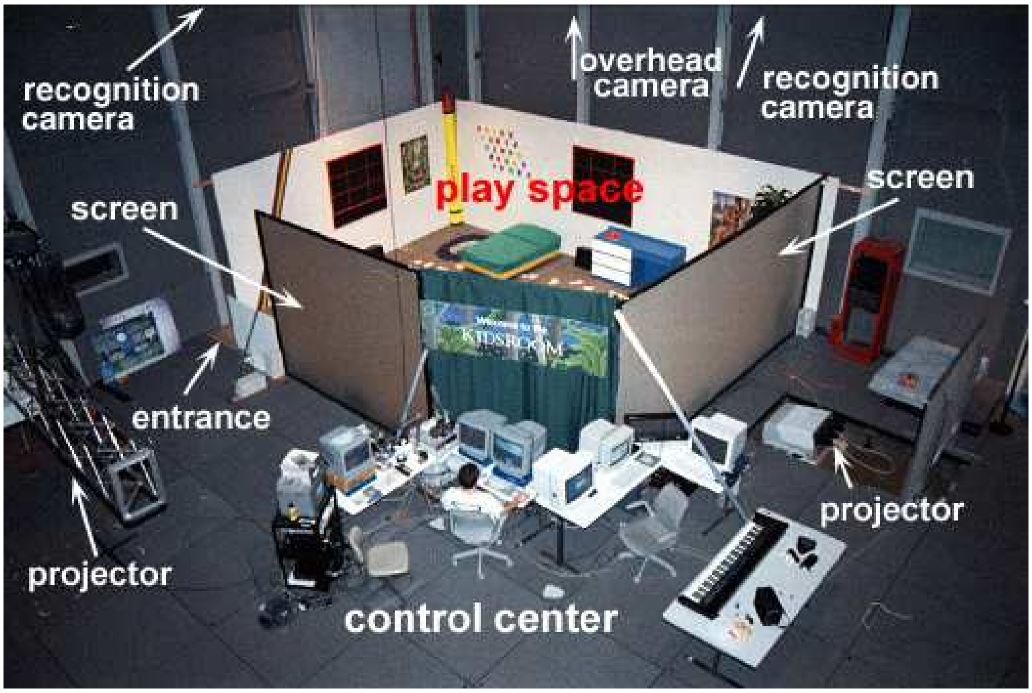

Figure 1. Physical setup of the *"KidsRoom."*

Figure 2. Users experiencing the *"KidsRoom."*

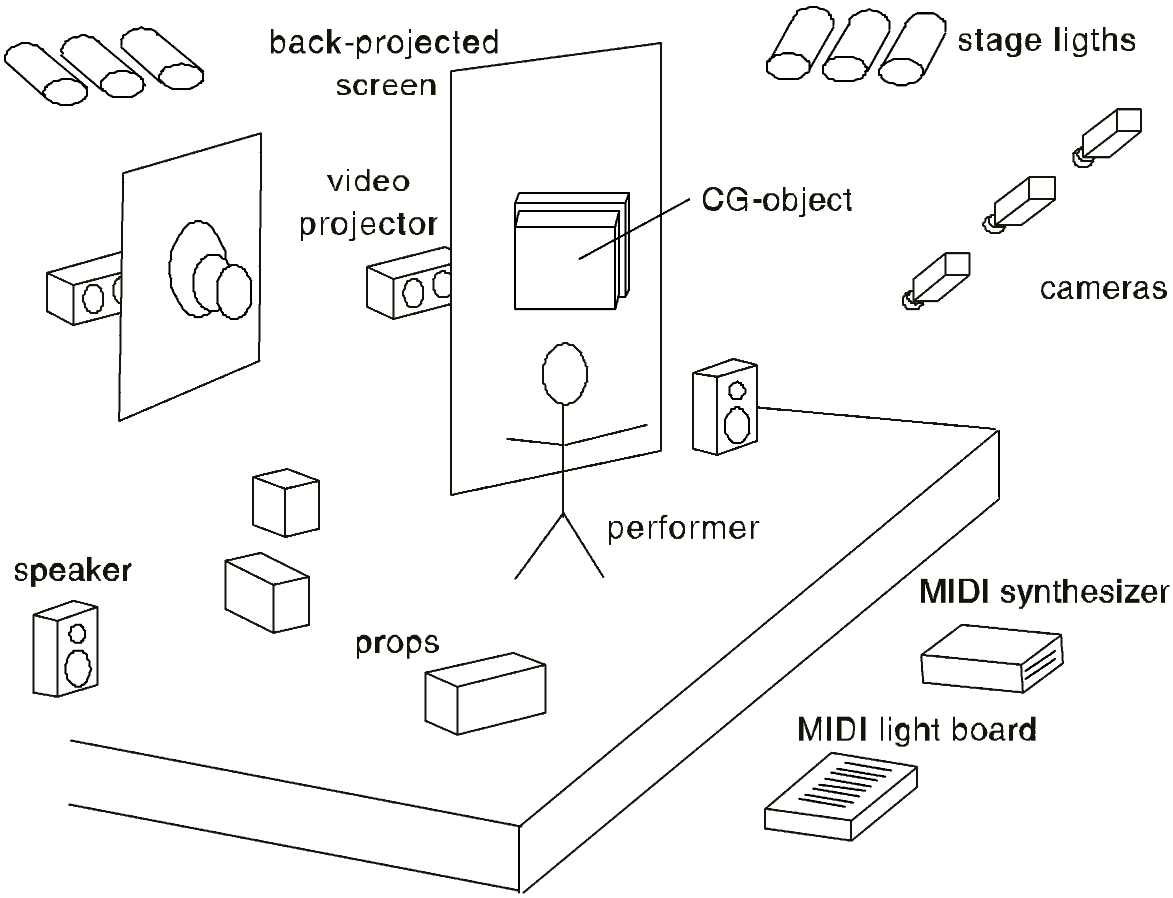

Figure 3. Physical setup of *"It / I."*

Figure 4. The human and the computer characters during scenes of *"It / I"*.

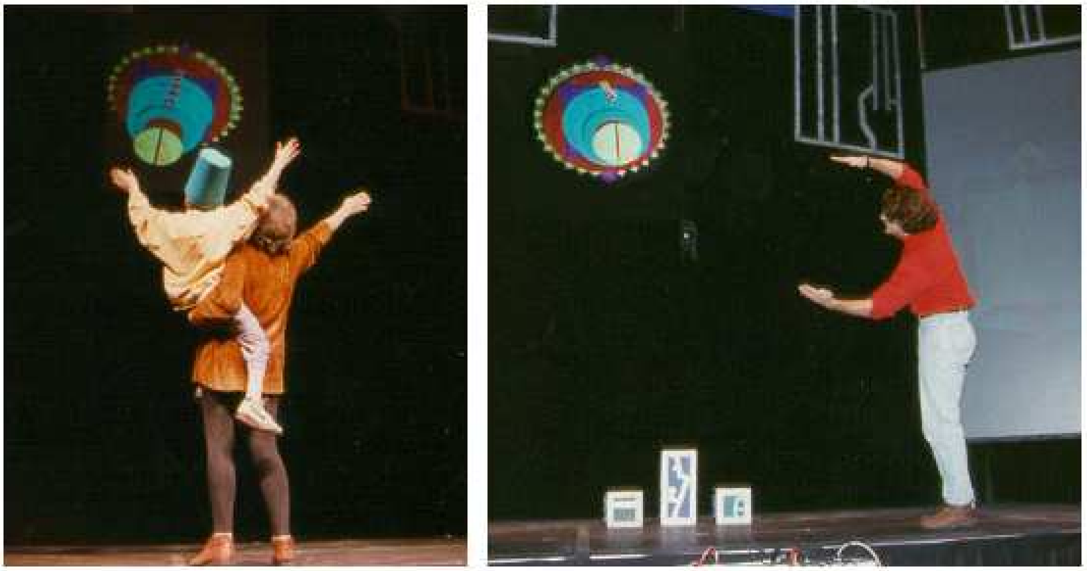

Figure 5. Audience playing in the *immersive stage* of *"It / I"* after the performance of the play.

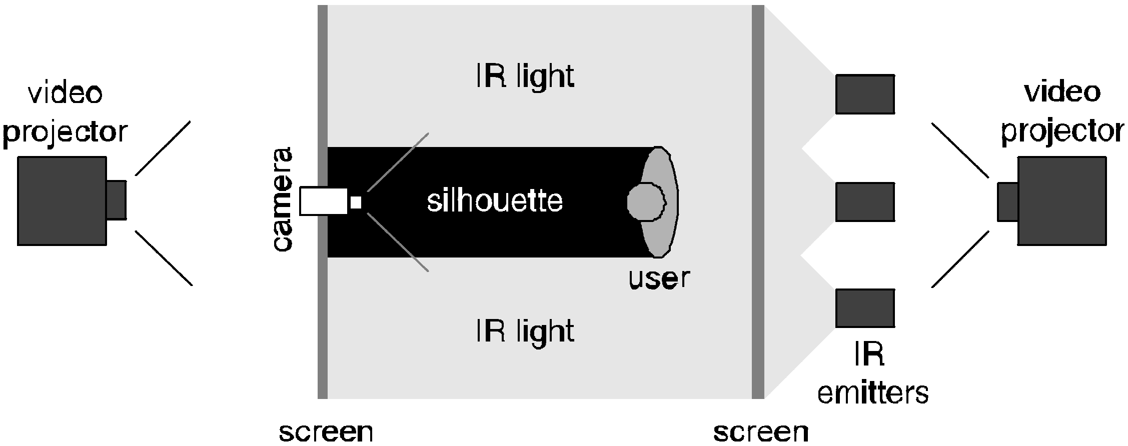

Figure 6. Physical setup of *"PAT."*

Figure 7. Typical interaction of the *"PAT"* system.

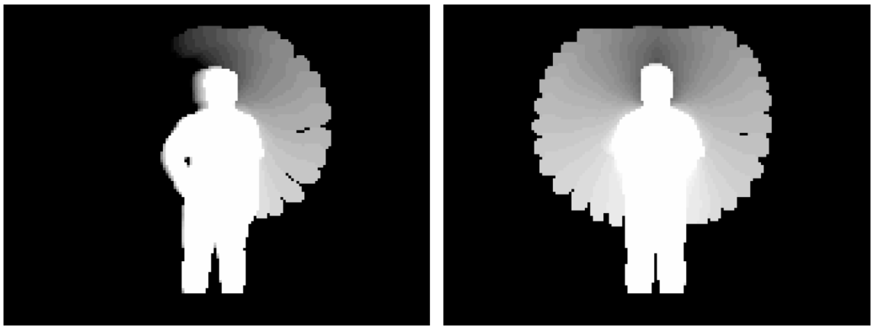

Figure 8. Motion templates used by "PAT" for recognizing "left-arm-stretch" and "deep-breathing-stretch".

Figure 9. User in front of the screen of *"Swamped!"* controlling the computer graphics character through the plush doll.

Figure 10. Snapshot of scene of *"Swamped!"*

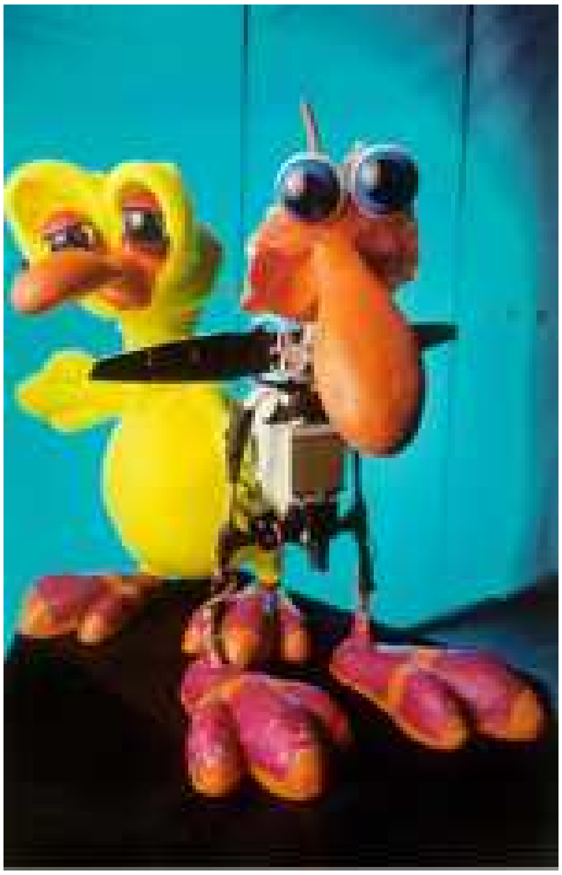

Figure 11. Sensor package installed in the physical doll interface used in *"Swamped!"*.