# Media Actors:
# Characters in Search of an Author


Flavia Sparacino
*MIT Media Lab*
flavia@media.mit.edu

Glorianna Davenport
*MIT Media Lab*
gid@media.mit.edu

Alex Pentland
*MIT Media Lab*
sandy@media.mit.edu



## Abstract

*Interactive experiences benefit from natural interactions, compelling communication, and ease of implementation. We show how, according to these principles, interactive media architectures can be categorized as scripted, responsive, learning, or behavioral, and give examples of applications in each category. We then propose the perceptive architecture based on Media Actors. We endow media objects - expressive text, photographs, movie clips, audio, and sound clips - with coordinated perceptual intelligence, behaviors, and personality. Such media actors are able to engage the public in an encounter with a virtual character which expresses itself through one or more of these agents. The result is a novel method for interactive media modeling which finds applications in multimedia, electronic art, interactive performance, and entertainment.*


## 1. Introduction

Recently, in the field of Computer Graphics, progress has been made in the creation of life-like characters or autonomous agents. These characters are driven by autonomous goals, can sense the environment through real or virtual sensors, and respond to the user's input or to environmental changes by modifying their behavior according to their goals. While this approach, called behavior-based, has proven to be successful for a variety of computer graphics problems, it has not been yet fully understood or exploited for multimedia presentations, digital storytelling, interactive performance, or new forms of interactive art.

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

There is indeed some confusion and a discrepancy to date in the way the computer graphics community and the multimedia/electronic art community think of behavior-based modeling. We would like to clarify some important issues in this respect and delineate a direction of work which describes what multimedia and interactive art can learn from computer graphics and how the new behavior-based modeling techniques can be extended and creatively applied to a variety of artistic domains. We start by giving a fast overview of the recent literature on behavior-based robotics and animation. We then provide a taxonomy of interactive experiences in the field of multimedia and electronic art, based on the interaction modality and the system architecture. We describe concrete applications according to the given taxonomy, based on our research work, and finally introduce an architecture for effective media modeling for interactive environments.

## 2. Scripted vs Behavior-Based

In this section we explain how the behavior-based approach differs from the scripted or reasoning approach, and describe the origin, purpose, and advantages of behavior-based modeling.

In the field of multimedia and electronic art the term "behavior-based" is often ingenuously used to contrast media modeling from more traditional architectures which separate the content on one end and the routines which orchestrate media for the public at the other end. 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. These systems rigidly define the interaction modality with the public, as a consequence of their internal architecture. Often, these programs need to carefully list all the combinatorics of all possible interactions and then introduce temporal or content-based constraints for the presentation. Having to plan an interactive art piece according to this methodology can be a daunting task, and the technology in place seems to somehow complicate and slow down the creative process rather than enhance or expand it.

There is certainly an analogy between these systems and the more traditional AI applications, which were based on reasoning and symbol manipulation. Both share the idea of a centralized "brain" which directs the system on the basis of operations on symbols derived from sensory or direct input. When used for authoring interactive media or electronic art we consider this approach similar to that of having an orchestra director who conducts a number of musicians following a given score. This leaves very little room for interactivity, and the programmer of the virtual reality experience needs to create break points in the "score" (plot) – somewhat artificially – for the user to be able to participate. Typical examples of this approach are the many CDROM titles which simulate the presentation of an interactive story, or game, by careful planning of plot bifurcations and multiple choice menus.

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

"Behavior-based" is a term which originates from Artificial Intelligence and is often synonymous with "autonomous agent research". It describes control architectures, originally intended for robots, which provide fast reactions to a dynamically changing environment. Brooks [1] is one of the pioneers and advocates of this new approach. Maes [2] has extended the autonomous agent approach to a larger class of problems, including software agents, interface agents, and helpers which provide assistance to a human involved in complex activities – like selecting information from a large database, or exchanging stock. Behavior-based architectures are defined in contrast to function-based architectures as well as to reasoning-based or knowledge-based systems. The latter approach, corresponding to "traditional AI", emphasizes operations on symbolic structures which replicate aspects of human reasoning or expertise. It produces "brains" applying syntactic rules to symbols representing data from the external world, or given knowledge, and which generate plans. These systems work under a closed-world assumption to eliminate the problem of unexpected events in the world.

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

The advantage of Blumberg's modeling technique is that the designer of the experience does not have to think of all the possible sequences and branchings, in defining the interaction between the public and the virtual creatures. It suffices to specify the high order behaviors, their layering structure, and the goals of the creature, to produce a realistic and compelling interaction. Blumberg also introduces an abstraction which allows to split the hard task of coordinating the kinematic motion of the articulated dog from the specification of the high-level behavior system. He describes an action-selection mechanism which allows to arbitrate among commands given by a human and the autonomous drive of the creature determined by its goals and internal motivations [6]. Perlins [7] has applied a similar approach to animating humans in virtual environments.

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

## 3. A Taxonomy of Interactive Experiences

The behavior-based approach has proven to be successful when applied to mobile robots and to real-time animation of articulated synthetic creatures. In this context, “behavior” is given a narrow interpretation derived from behavioral psychology (Skinner). For animats, behavior is a stimulus-response association, and the action-selection mechanism which assigns weights to the layered behaviors can be seen as a result of operant conditioning [8] on the creature. Behavior-based AI has often been criticized for being “reflex-based”, as it controls navigation and task execution through short control loops between perception and action.

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

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

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

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

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

- *Learning systems* have the ability to learn new behaviors or to modify the existing ones by dynamically modifying parameters of the original behaviors. These systems provide a rich set of interaction modalities and dynamics, and offer new interesting venues for interactive media architectures [9].

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

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

# 4. Applications

In this section we provide examples of some of the architectures described above. Our own approach in building interactive experiences has grown from scripted, to responsive, to behavioral, and finally to perceptive. We describe work carried out at MIT since 1995 and show how, at each stage of development, we have learned the advantages and insufficiencies of each of the different architectures. While the literature on entertainment, multimedia, and the electronic arts, provides a large number of examples of approaches and models, we limit our analysis to our own research. We postpone a comparative study of the literature to a later publication, entirely dedicated to this purpose, as its extent would largely exceed the scope of this paper. Some of the projects mentioned in this section have been fully developed. Others have served as experimental testbeds, and are cited as examples which have stimulated our thinking and enriched our experience in building interactive environments.

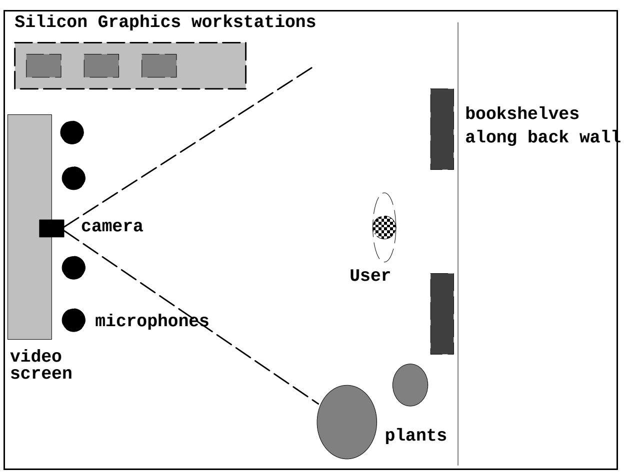


**Figure 1: Example of IVE Space**

Most of our work uses an IVE (Interactive Virtual Environment) setup [10]. IVE is an interactive space developed at the MIT Media Lab, in which the public interacts with visual material presented on a large projection screen which occupies one side of the room [fig. 1]. A downward pointing wide-angle video camera mounted on top of the screen allows the IVE system to track a member of the public. By use of real-time computer vision techniques [11][12][13] we are able to interpret the user'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. The only constraints are a constant lighting and an unmoving background.

IVE was built to enable people to participate in immersive interactive experiences without wearing suits, head-mounted displays, gloves, or other gear. 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 at once. The IVE environment was originally developed for the ALIVE project [14] and has since become our main development platform for interactive experiences.

## 4.1. Scripted Applications

Our first project in the IVE space was an interactive story/museum-exhibit called *Encounters*. A member of the public would meet a 3D humanoid character at a crossroad of a 3D virtual museum-city. S/he would be handed a message and become involved in solving a mystery regarding three contemporary artists. Solving the mystery, brought the participant through a series of chambers, and made him/her become familiar with the work of the artists. The person interacts with the characters, sounds, and images projected on a large screen through simple gestural and voice commands. This project was entirely scripted and – although we managed to author some interesting segments – it soon grew to a size which was very hard to handle for the designers and the participant and yet too simple for the public to be able to enjoy and appreciate.

## 4.2. Responsive Applications

In February 1996 we shifted our attention from storytelling to performance and aimed for a system which would be uncomplicated to understand and use. We created *DanceSpace*: an interactive stage for a single performer [fig. 2] in which music and graphics are generated in the fly by the dancer's movements. In DanceSpace [15] a small set of musical instruments is virtually attached to the dancer's body and generate a melodic soundtrack in tonal accordance with a soft background musical piece. The performer projects graphics on a large back screen using the body as a paintbrush. DanceSpace was a full success as both common users and performers were able to quickly understand the interface and choreograph improvisational pieces influenced by the technological opportunity. It is an ideal example of a responsive experience with one-to-one mappings between sensory input the dancer's output music hands/feet/head/center of body movements and system and graphics.

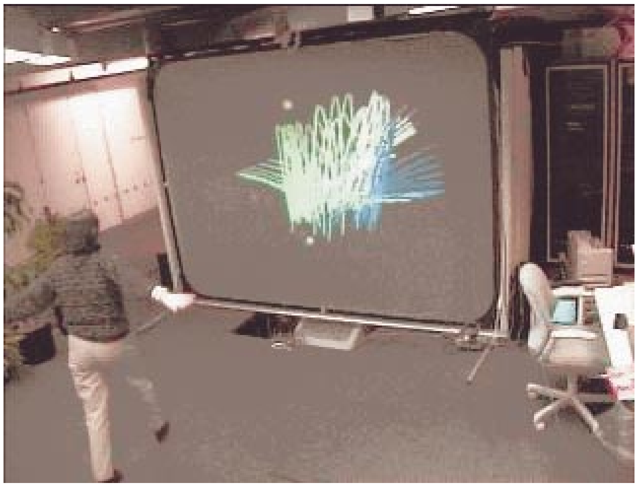

**Figure 2: A Responsive System: DanceSpace**

*Time Window* is another example of successful responsive application we built in May 97. Photographs of the landscape were taken from an office window at regular intervals during the day, and across days in fall season. The most noticeable change in the photographed landscape is given by two large trees which progressively turn from green to red, and finally drop their leaves. As a person walks in the space towards the screen s/he walks through the days from summer to winter. If instead s/he walks transversally across the room, s/he sees the landscape projected on the large screen transform under the light, on one particular day, from sunrise to sunset. A gesture allows the participants to see their image projected on the landscape, as in a magic mirror, and to walk along the roof of the buildings in the back of the picture to meet responsive characters: a virtual cat or a bouncing ball.

We have also developed *Typographic Actor*: a digital design project in which a member of the public can animate expressive typography with both head and mouth movements. S/he can also modulate the appearance of the text by singing to it: the energy and pitch of her voice will be respectively mapped to the size and color of the text.

### 4.3. Behavioral Applications

In the summer of 1995, we created a behavioral experience: the *Hyperplex* [16]. Hyperplex is a virtual movie theater, a building with multiple chambers located at different floors, in which people can view short movie clips representing film trailers. Each floor of the building is associated to a specific movie genre, i.e. adventure, comedy, drama, thriller etc. Movie clips behave as creatures (animats) whose goal is to be seen, and who compete with other movie clips to catch the public's attention. Movie clips belonging to the same genre selected by the public collaborate, whereas the other clips compete to have a central position on the display screen. A member of the public can: "call a movie clip", "grab", "play", "send away", "send to someone else", "take with me", or "ask more info" in the form of sound and text. Both the designers of the experience and the public were fascinated by the idea of being able to browse video on a large screen by using gestural input. However using the body for navigating in a 3 dimensional building was not an easy task for the untrained public.

In April 1996, we decided to combine our experience in building responsive and behavioral environments and created the *Improvisational Theater Space*. 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 [fig. 3]. 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. We performed two improvisational pieces at the Sixth Biennal Symposium for Arts and Technology [17]. This particular research lead us to developing the new perceptive media architecture described in the next section. Through this project we learned that behavior-based media actors are a promising approach to innovative theatrical performances for three main reasons:

1. Behavior-based vs script based theater has room for improvisation, 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 to human error and actually encourages actors to enrich or change the performance according to the reaction of the audience.
3. The system can scale from a performance space to an entertainment space. Behavior-based theater can allow for user participation either during or after the performance without requiring the new users to learn all the script in advance.

The behavioral approach allows for flexible media choreography and contrasts scripted/rule base 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 that both the actors and the director change the script either during rehearsals or even right before or during the final performance [18]. In our view rule based, scripted systems are not able to compensate for human errors or be responsive when some non-planned "magic" between the actors happens on stage. They tend to force human interpreters to rigidly follow a predefined track and therefore impoverishes the quality of the performance.


**Figure 3: Actress Kristin Hall in ImprovTheaterSpace**

*Virtual Studio: Digital Circus* was first constructed in March 1997. It is an immersive behavioral experience in which all objects present in the 3d virtual circus are endowed with behaviors. Advanced real-time computer vision techniques allow to composite and blend a 2d image of the participant inside the 3d world, without the need of blue screens. Individual distant participants can be remotely connected to and share the same virtual world [fig. 4]. Hence such setup can be used at home for collaborative storytelling, visual communication from remote locations, or game playing. In the circus a behavior-based butterfly pet follows the participant around, a cannon fires a cannon woman when the participant virtually presses its "button", an umbrella appears at need. Sitting on a chair causes a gramophone to appear and music to be played, an arm gesture causes the participant to grow taller or become tiny-small, on request. All of these actions/transformations are possible because each object in the virtual space is endowed with an autonomous behavior and it takes care of doing the right thing at the right time.

By grouping our previous work according to the taxonomy described above and by analyzing the interaction modality and experience of the public in each application, we have derived the following conclusions:

- Scripted experiences are difficult to author. They require a careful and detailed planning of the material presented. Complexity burdens both the author and the recipient of the piece. If the interactive experience requires this approach - such as in the case of some storytelling projects - it is important to keep it small and simple.
- Responsive experiences can be successful especially when the system responds in a timely fashion. Also it is important that the input-output mapping becomes clear to the public as soon as possible in the course of the experience. However, due to the fact that the input-output mapping is invariant, responsive applications can become repetitive and eventually obsolete after a few experiences.
- Behavioral experiences allow for more complex forms of interaction. We have described applications in entertainment (cinema, circus) and performance (theater). The behavior architecture allows to distribute the authoring complexity to the various characters/media in the piece. Once the behavior system is built, such systems are much easier to build than the scripted pieces, as they require only specifications about the behavior parameters for the specific application considered.

However, if we want to build experiences which can simulate an encounter, convey emotions, and respond based on content and the history of the interaction, we need to consider alternative architectures which focus more on expression and perception of the human participant's interaction modality.

**Figure 4: Two Connected Participants in Digital Circus**

## 5. Perceptive Systems: Media Actors Architecture

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

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

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

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

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

- a *perceptual layer*: 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.

The importance of having an intermediate layer of sensory representation, and predictions, has also been underlined by Crowley [19]. 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.

The following paragraph describes our perceptive system application.

## 5.1. Perceptive Portraits

A Perceptive Portrait consists of a multiplicity of photographs virtually layered on a high-resolution digital display. The image which is shown at a given time depends on how the viewer approaches and reacts to the portrayed subject. An active computer-controlled camera is placed right above the display. By using real-time computer vision techniques we are able to determine how close/far the viewer is to the portrayed character, her viewing angle, and we can also interpret some of her facial expressions like smile, laughter, surprise, or disappointment [fig. 5].

The viewer's proximity to the image, head movements, and facial expressions elicit dynamic responses from the portrait, driven by the portrait's own set of autonomous behaviors. This type of interaction reproduces an encounter between two people: the viewer and the character portrayed. The experience of an individual viewer with the portrait is unique, because it is based on the dynamics of the encounter rather than on the existence of a unique, ideal portrait of the subject. As the participant observes the portrait he is also being observed by the media actor: the whole notion of "who is watching who" is reversed: the object becomes the subject, the subject is observed.

At ant time during the interaction with a Perceptive Portrait the participant can ask the system to read the story of the encounter from the portrayed character's viewpoint.

**Figure 5: Expressive Behaviors of a Perceptive Portrait**

A first version of this project was shown at ISEA97 [20]. The current research implements the perceptive architecture and shows how Perceptive Portraits interact effectively with people, and are easy to author.

## 6. Conclusions

This work highlights some important technical and creative issues that designers of interactive experiences face in the conception and construction of an interactive piece. It offers an analysis of behavior-based systems, based on work done in robotics and computer graphics. It provides a taxonomy of interactive media applications in order to understand how to extend and improve beyond the behavior-based approach, when designing and authoring a multimedia experience. It gives examples of our research work according to the taxonomy presented and shows the necessity for a new interactive media paradigm. Its contribution to the field of multimedia, entertainment, and electronic arts is to provide a new modeling method which is flexible, scalable and engaging for the public. It also traces a fruitful direction of work and a platform for discussion and improvement in our field.

## References


[1] Brooks R.A., "Intelligence without Reason", *IJCAI-91*, vol 1. pp.569-595.

[2] Maes P., "Modeling Adaptive Autonomous Agents", *Artificial Life* 1: 135-162, 1994.

[3] Blumberg B., Galyean T., "Multi-Level Direction of Autonomous Creatures for Real-Time Virtual Environments", *Computer Graphics* (SIGGRAPH '95 Proceedings), 30(3):47-54, 1995.

[4] 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.

[5] Johnson M., WavesWorld: PhD Thesis, *A Testbed for Three Dimensional Semi-Autonomous Animated Characters*, MIT, 1994.

[6] Blumberg B., "Action-Selection in Hamsterdam: Lessons from Ethology", *Third International Conference on the Simulation of Adaptive Behavior*, Brighton, England, 1994, MIT Press, pp. 108-117.

[7] Perlin K., "Real time responsive animation with personality", *IEEE Transactions on Visualization and Computer Graphics*, 1(1), 1995.

[8] *The Selection of behavior: the operant behaviorism of B.F. Skinner: comments and consequences*, A.C. Catania and S. Harnad eds., Cambridge University Press, 1988.

[9] Jebara T., Pentland A., "Action Reaction Learning: Automatic Visual Analysis and Synthesis of Interactive Behaviour", *ICVS '99*, Las Palmas, Canary Islands, January 1999.

[10] Wren C., Sparacino F., et al., "Perceptive spaces for performance and entertainment: Untethered interaction using computer vision and audition", *Applied Artificial Intelligence*, 11(4):267-284, June 1997.

[11] 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.

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

[13] 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.

[14] Darrell T., Maes P., Blumberg B., and Pentland A., "A Novel Environment for Situated Vision and Behavior", Landy, M., Maloney, L., Pavel, M., eds., *Exploratory Vision: The Active Eye*, Springer-Verlag, 1995.

[15] Paradiso J., Sparacino F., "Optical Tracking for music and dance performance", *Fourth Conference on Optical 3-D Measurement Techniques*, Zurich, Switzerland, September 29-October 2, 1997.

[16] Sparacino F., Wren C., Pentland A., Davenport G., "Hyperplex: a world of 3d interactive digital movies", *IJCAI-95 Workshop on Entertainment and AI/Alife*, 1995.

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

[18] Brook P., *The Open Door*, Theatre Communication Group, New York, New York, 1995.

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

[20] Sparacino F., Oliver N., Pentland A., Davenport G., "Responsive Portraits", *The Eighth International Symposium on Electronic Art*, Chicago, IL, September 22-27, 1997.