Autonomous Robots as Performing Agents

Karl R. Wurst and Robert McCartney
Department of Computer Science and Engineering, U-155
University of Connecticut
Storrs, CT 06269-3155
E-mail: wurst@cse.uconn.edu

Abstract

We are developing autonomous robots whose purpose is to perform for an audience. Computer-based actors have typically been virtual, screen-based characters with no physical presence. Our actors combine techniques from artificial intelligence, robotics and puppetry – a mixture we term *puppotics*. Each of these actors is an autonomous agent with its own processor. They act out their parts, communicating among themselves, with a director and with their audience.

Introduction

Performing agents are designed to entertain an audience through performance. Traditional performing agents are human actors and puppets, but performing agents may be realized in many media – video, audio, text and others, singly or in combination. In most current AI research, these performing agents are on screen or video with no physical presence. Our work is closer to the traditional view of performing agents, having physical puppet actors on stage, and uses techniques and materials from the craft of puppetry as well as the field of autonomous robotics, a mixture that we term *puppotics.* While this research is being done in an engineering school, it draws heavily on the resources and expertise of the University of Connecticut's world-renowned Puppet Arts Program.

Why Physical Performing Agents?

Our agents are physical because we believe that they will be more entertaining and engaging than virtual agents. Audiences have a more immediate rapport with a puppet than they do with an animation because of the physical presence of the puppet.

In addition, we wish to examine the interaction between these agents, the different forms that this interaction may take, and the difficulties involved in implementing this interaction in physical form. By working with physical implementations, we are making explicit many of the assumptions inherent in the software-only research in this area.

Because of the physical nature of our agents communications must be explicit. Software agents are able to communicate with each other with nearly unlimited bandwidth and with few restrictions. This communication may take place through shared memory or interprocess communication. These agents may communicate through some protocol or in the most egregious cases may simply inspect each others data structures. By implementing our agents physically, using a low-bandwidth medium, we are forced to make all of our assumptions about communications explicit.

Communications among Performers

The communications that can take place with performing agents can be characterized as agent⇔agent, director⇔agent, and agent⇔audience. These types of communication are not limited to performing agents, but will manifest themselves in any multi-agent setting (except for possibly the last – agent⇔audience communication.). A team of robots cleaning up an oil spill, for example, would have to communicate to coordinate their efforts, and may receive instructions from a supervisory agent that can see the overall picture.

Agent⇔Agent

Most of the communication is between the agents themselves – the "dialogue" of the performance. This communication may consist of both speech action, and can take the form of point-to-point dialogue between two agents, or a broadcast from one agent to a group. The difference between the two is largely in the intent of the agent. In both speech and action all messages are broadcast to any agent that is within sensory distance. A message may be intended for anyone who can see it, but more often it is intended for the agent that the sender is facing. It is up to the receiver to determine if the message was intended for it or not, and what to do with the information received. One could imagine an agent "eavesdropping" on a conversation between other agents and using the information to change its own internal state in some way – becoming angry perhaps.

Director⇔Agent

Traditionally communication is from the director to the actors only, and takes place before the actual performance. An example of this is what Hayes-Roth calls *directed improvisation*(Hayes-Roth *et al.* 1994). The agents are given high-level, abstract instructions by the director, and improvise the specifics. However, it may be useful to have directions interjected by the director throughout the performance, as a way of keeping the story on track(Kelso, Weyhrauch, & Bates 1992) These directions from the external director would likely add to or change the agent's goals, to force the agent's performance to come into line with the director's overall plan. Other forms of director to agent communication can be imagined, such as a supervisor directing working robots to concentrate their efforts on a particular area of a problem.

Although it is not commonly seen, agent to director communication is also possible. This would allow the agents to inform the director of their internal state (possibly as a response to a query from the director,) or to request new instructions if they get stuck.

Agent⇔Audience

The goal of performance is communication between the actors and the audience, with the performer trying to get his message across. There must be a common language that the audience and the performer both understand with conventions for certain concepts. This is one of the areas where we are drawing on the expertise of the puppetry field, where the issues of communication with the audience, often with non-verbal agents with limited physical expression, is explored.

Communication also occurs from the audience to the performer and the importance of this communication varies with the type of the performance. In performances from a fixed text, audience reaction may only be used by the actor at a subconscious level, while in a improvisational performance, audience participation may make or break the performance. Possible uses for this type of communication would be to allow the agents to take a bow to applause, respond to a heckler, or even to adjust its timing to audience response.

The Woggles

Our agents are based on the Woggles of the Oz Project at CMU(Loyall & Bates 1993). The Woggles are video actors, depicted as spheres, that exist in a video landscape. They move through this landscape and interact with each other and with a human-controlled woggle. They communicate with each other by gesturing. The Woggles are also in use at Stanford in the Virtual Theater project(Hayes-Roth 1995).

The Woggles choose their actions based on their internal emotional and physical states, their personalities, and interaction with each other. Each woggle has a distinct personality that affects how it will go about attempting to satisfying its goals. In the CMU Woggle world, *The Edge of Intention,* there are three woggles – Wolf, Shrimp and Bear. Wolf and Shrimp have aggressive and shy personalities respectively. Each may have the goal of amusing himself, but the differences in their personalities will cause them to satisfy this goal differently. Shrimp may decide to go dance around by himself to amuse himself, while Wolf may decide to go pick on Shrimp.

Their performances are unstructured, that is, there is no script or plot that they follow in their performance, and are entirely interaction-driven. They communicate entirely by making physical gestures and visually observing the gestures of others. Their repertoire of gestures includes "Say Hey" – a greeting gesture, squash down, puff up, threaten and spin. This limited number of gestures leads to simple interactions – woggles can gesture at each other, and move toward or away from other woggles. There is no external direction of their actions at this time.

Our immediate goal for our physical agents, the RoboWoggles, is to reproduce the video version of the Woggles as closely as possible with physical devices. We wish to duplicate the behaviors as well as mimic the visual aspects of the original Woggles as much as possible.

The fact that there is a behavioral model already developed for the Woggles, with appropriate gestures for communication makes them attractive as a starting point for our research. Their simplicity, both in processing and in physical design and range of motion, makes them suitable for physical agents.

Implementing Woggles

Our physical implementation matches the ideal quite well as far as gestures are concerned, and adds individual, internal control to each woggle. The body is a ten inch diameter flexible sphere that is mechanically deformable. This allows the characteristic woggle gestures of puff up, squash down and so on. This is accomplished with two servomotors connected to an armature that can deform the body at the top and sides (see Figure 2.)

The woggle is controlled by a system based on the Motorola 68HC11 microcontroller. The system is

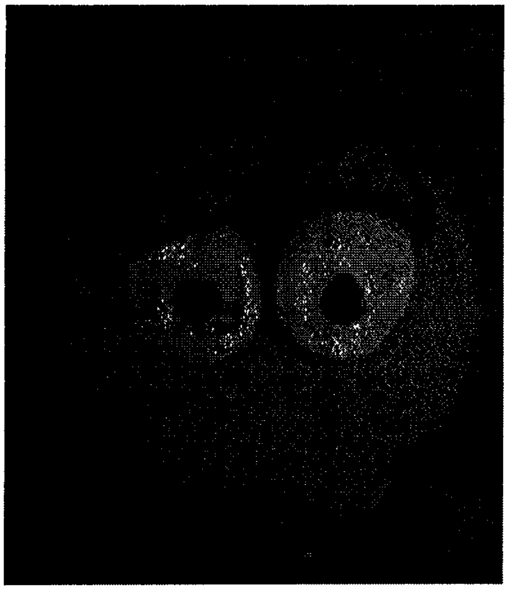

Figure 1: Prototype RoboWoggle

based on Marvin Green's BOTBoard, modified to provide an additional 32K of memory. It can control 4 servomotors, and has a number of analog and digital inputs, and digital outputs.

The most notable departure from the video woggles is the RoboWoggles' method of propulsion. Rather than attempt the potentially very difficult mechanical engineering problem of producing jumping behavior, we opted instead for having the RoboWoggles roll about on two wheels. The two wheels are differentially driven, allowing the woggles to spin in place as well as move around on the floor.

The woggles' communication is visual: one agent makes a physical gesture, the other observes it. In a "traditional" vision system a camera or similar sensor would capture the complete scene before the agent. A vision processing system would then determine which elements of the scene should be grouped together to form individual objects or agents. Finally, the system must watch the changes in pose of an agent over time to determine what gesture it is performing.

We propose a simpler, but functionally equivalent vision system that eliminates the need for most of the visual input processing. In our "vision-less" vision system, each agent simultaneously gestures and transmits information about the gesture to all agents around him. This vision system should be indistinguishable from a "traditional" vision system from the audience's point of view. It still involves active sensing and processing of visual information to determine what another agent is doing.

Communication between woggles is performed by a line-of-sight, low-bandwidth infrared communication system. Transmission of gestures is broadcast 360° by four transmitters (each covering 90°.) Each transmitter broadcasts the identity of the woggle performing the gesture and the position of the transmitter itself (front, back, right or left) as well as the gesture. This allows a receiving woggle to determine the identity and facing of the transmitting woggle, as well as turning or spinning behavior (by observing the transmitters' signals over time.)

Reception is from the front of the receiver only. In other words, a woggle can only see what another woggle is doing if it is looking at (facing) it, but a woggle can see what other woggles are doing, even from behind them.

State of Implementation

At this time, the prototype microcontroller system is working. Another group in our laboratory has developed the ChickenBoard – a general, expandable, modular robot controller system that will eventually replace the prototype system now in use.

The base mechanics with the drive and steering mechanism are completed as well as the body deformation mechanism. We now beginning to build multiple bases.

A prototype body was made from reticulated foam (a material commonly used for puppet bodies – including the Muppets™.) Unfortunately, it did not have quite the flexing characteristics hoped for, and the amount of sewing necessary to produce it made it prohibitive for making multiple bodies.

We are in the process of producing a new flexible body. The new body will be cast from neoprene, and the casting process will allow multiple bodies to be produced easily.

Future Work

The communication system is still in the design stages. A simple infrared scheme (as in consumer electronics remote control systems) is sufficient to transmit the gesture "vocabulary" needed to mimic vision, but we need to develop communication protocols that can be extended to support broadcast communications and reception of external direction. This extended protocol will need to be able to communicate new goals, change the priorities of goals for an agent, or give explicit instructions, and may implemented using different communication hardware.

We need to develop a communications protocol that can transmit the entire gesture vocabulary needed by the RoboWoggles. Thay protocol must operate over

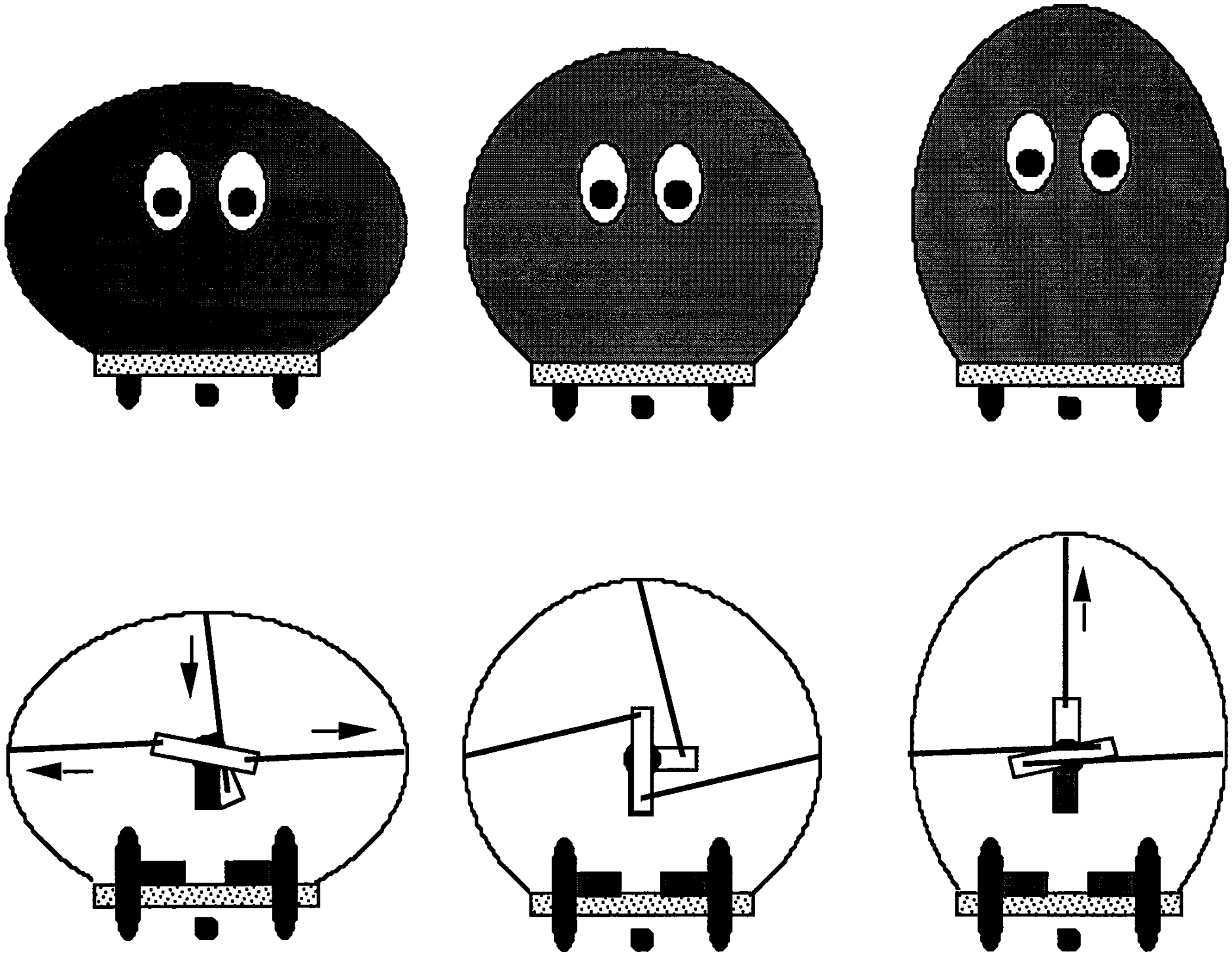

Figure 2: Deformation and locomotion mechanics (squash, normal, and "Say hey")

limited-bandwidth IR channels, and provide the information necessary to mimic vision. The protocol will have to be extendable to include commands from a director, or another communication system could be used for external direction. The nature of these external direction commands needs to be explored further to determine what exactly should be transmitted, and what effect it should have on the internal state of the agents.

We plan to look into structuring the performances by providing some type of plot mechanism. This will require either some sort of script to follow, or the implementation of external direction to keep the improvisations from getting out of control and keep the overall performance on track

Related Work

The Oz Project at CMU(Bates 1992) is developing tools to support virtual worlds with believable agents that humans can interact with. The Woggles and their world, *The Edge of Intention* are just one of the systems that they are building with these tools. They have developed an integrated agent architecture called Tok(Bates, Loyall, & Reilly 1992) which includes a reactive component, Hap(Loyall & Bates 1991), an emotion system, Em(Reilly & Bates 1992), and a natural language system, Glinda. A major thrust of their work is external direction with the goal of being able to immerse a human in a complete dramatic work. They want the system to be able to improvise responses to the human's actions, while still maintaining the overall structure and goals of the plot.

The Virtual Theater project at Stanford(Hayes-

Roth, Brownston, & Sincoff 1995) is also concerned with *directed improvisation,* but as a more general paradigm for human-computer interaction. Their system, the Virtual Theater, uses the Woggles as an environment for children to write and perform plays. The children may write out detailed scripts for the characters to perform, create abstract stories for the characters to improvise, or directly control the characters as video puppets. The children can mix these modes, and in fact, switch between them at any time.

The ALIVE System at the MIT Media Lab(Maes *et al.* 1996) presents a virtual world populated by autonomous agents with their own goals and intentions, modeled with their Hamsterdam(Blumberg 1994) agent toolkit. It has a novel interface that allows use of the user's entire body to interact with the agents. It does this by projecting a video image of the user onto a large projection screen, placing the user into the virtual world. The user sees himself interacting with the agents and objects he sees. The system interprets the user's actions through video processing that identifies the user's gestures.

Conclusions

We are building physical implementations of performing agents whose main purpose is to provide entertainment for an audience. The ROBOWOGGLES are an example of *puppotics* – a combination of artificial intelligence, robotics and puppetry technologies. They communicate with each other by physical gestures, and communicate those gestures over an infrared communication system, while performing the actual gesture for the audience. These simple performing agents provide a good starting point for exploring performances with more complex plots and interactions and the possibility of external direction of the agents. In addition, we are exploring the issues involved in the communication among the agents and between the agents and an external director. The communication issues that must be addressed are not limited to performing agents, but will manifest themselves in any setting where multiple agents are working together.

Acknowledgements

Thanks to Christian Netter for all his help in developing the prototype controller board and the CHICKENBOARD. Thanks to Bartolo P. Roccoberton Jr., head of the University of Connecticut Puppet Arts Program for providing the expertise needed to make the ROBOWOGGLES *real.*

References

Bates, J.; Loyall, A. B.; and Reilly, W. S. 1991. Broad agents. In *Proceedings of the AAAI Spring Symposium on Integrated Intelligent Architectures.* Available in *SIGART Bulletin,* Volume 2, Number 4, August 1991, pp. 38-40.

Bates, J.; Loyall, A. B.; and Reilly, W. S. 1992. Integrating reactivity, goals, and emotion in a broad agent. In *Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society.*

Bates, J. 1992. The nature of characters in interactive worlds and the Oz Project. Technical Report CMU-CS-92-200, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA.

Blumberg, B. 1994. Action-selection in Hamsterdam: lessons from ethology. In *Proceedings of the Third International Conference on the Simulation of Adaptive Behavior.*

Hayes-Roth, B.; Sincoff, E.; Brownston, L.; Huard, R.; and Lent, B. 1994. Directed improvisation. Technical Report KSL-94-61, Knowledge Systems Laboratory, Stanford University, Palo Alto, CA.

Hayes-Roth, B.; Brownston, L.; and Sincoff, E. 1995. Directed improvisation by computer characters. Technical Report KSL-95-04, Knowledge Systems Laboratory, Stanford University, Palo Alto, CA.

Hayes-Roth, B. 1995. Agents on stage: Advancing the state of the art of AI. In *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence.*

Kelso, M. T.; Weyhrauch, P.; and Bates, J. 1992. Dramatic presence. Technical Report CMU-CS-92-195, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA.

Loyall, A. B., and Bates, J. 1991. Hap: A reactive, adaptive architecture for agents. Technical Report CMU-CS-91-147, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA.

Loyall, A. B., and Bates, J. 1993. Real-time control of animated broad agents. In *Proceedings of the Fifteenth Annual Conference of the Cognitive Science Society.*

Maes, P.; Darrell, T.; Blumberg, B.; and Pentland, A. 1996. The ALIVE System: wireless, full-body interaction with autonomous agents. *ACM Multimedia Systems.* To be published.

Reilly, W. S., and Bates, J. 1992. Building emotional agents. Technical Report CMU-CS-92-143, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA.