[By Cynthia Breazeal, Andrew Brooks, Jesse Gray, Matt Hancher, John McBean, Dan Stiehl, and Joshua Strickon]

~Interactive Theatre

Engaging a human audience through sight, sound, scent, and touch while following a loosely constrained storyline, the Public Anemone and fellow autonomous characters are let loose to entertain—sociably.

Entertainment robotics might be viewed as a relatively new field, but the idea of building lifelike machines that entertain has fascinated humans for at least the past five hundreds of years—certainly as far back as 1515 when Leonardo da Vinci (commissioned by the Medici) built an ingenious self-propelled walking mechanical lion that opened and presented its breast full of lilies to Francis I, King of France, as a token of friendship from the Medici. The practice of creating mechanical automata blossomed in the 16th century when mechanical clock makers in Western Europe began to extend their craft to building mechanical animals. The 18th century craze also in Western Europe for animated objects produced a number of impressive mechanical humanoid automata, including the Jaquet-Droz Writer that emulated a young boy writing a letter at his desk and Joseph Faber's Euphonia, a mechanical talking head an operator could reputedly make speak in several European languages.

Some of the best-known examples of modern entertainment robots include animatronics in theme parks performing fully automated (though fixed and noninteractive) routines and puppeteered animatronics performing with human actors in such Holly-wood movies as *Jurassic Park* (the dinosaurs) and *A.I.* (Teddy). The most sophisticated are designed to look like living creatures and perform highly expressive movements. Less well-known examples include live performance troupes in which remote-controlled robots play a part in theatrical performances—either with human actors, as in Omnicircus (www.omnicircus.com), and with other robots, as in the robotic spectacles of Survival Research Laboratories (www.srl.org)

The field of entertainment robotics continues to find new applications as increasingly sophisticated and life-like autonomous robotic technologies mature. Whereas complex animatronic robots are often not designed to walk, small mobile robots (such as the Ullanta performance robotics [12] and the Carnegie-Mellon robot Improv) with built-in navigational skills have been used in live performances [12]. A number of university projects have explored the use of speech recognition and dialogue systems to allow people to have simple verbal interactions with animatronics [6] and mobile robot performers to have simple "improvisational" dialogues [3]. New tools for managing real-time interactive show control [6, 10], authoring nonlinear narrative, and capturing puppeteer

ROBOT

data have been developed specifically for interactive animatronics [6], often inspired by earlier work in interactive animated characters [1, 4] Meanwhile, new applications for entertainment robots in the home (such as Sony's robot dog Aibo) have sparked toy companies to develop a menagerie of robot toys.

Autonomous Robots that Interact with People

Along with the development of increasingly sophisticated entertainment robots, there is growing interest in the fields of autonomous robotics and humanoid robotics to develop robotic assistants that cooperate with people as partners rather than as mere tools. New areas of inquiry in human-robot interaction and social/sociable robots address

THE PUBLIC ANEMONE ROBOT WITHOUT ITS SILICONE SKIN. THE ARTICULATED STAGES GIVE IT A LIFELIKE, SERPENTINE QUALITY OF MOVEMENT, WHILE THE TENTACLES ARE USED FOR EXPRESSIVE PURPOSES (MIT MEDIA LAB, CAMBRIDGE MA, 2002).

the challenging problems associated with developing robots that interact naturally and appropriately with people, serving as helpers for the elderly, teammates for astronauts, museum docents, and domestic assistants [2, 5].

Sociable robots need to perceive, recognize, and interpret the behavior of humans through multiple modalities, including vision, audition, and touch. Ideally this perceptual awareness and responsiveness could be accomplished without the aid of joysticks, game pads, or other specialized interface devices, given that faces, gesture, and speech are the natural interfaces people use to communicate with one another. While machine perception continues to be a daunting problem, natural interfaces would certainly open new possibilities for interaction with autonomous robots.

As argued by design guru Don Norman, a good conceptual model is essential for understanding how an entity operates in order to interact with it [8]. With such a model, it is possible to explain and predict what "the other" is about to do, its reasons for doing it, and how to elicit a desired behavior from it. For interacting with socially interactive systems, the ability to recognize and respond to social cues is critical for effective communication and cooperation—skills necessary for envisioned sociable robot applications. In this spirit, designing robots with appealing personalities may help provide people a good (social) model for communicating with them. Personality, according to Norman, is a powerful design tool for helping people form a conceptual model to channel beliefs, behavior, and intentions into a cohesive, consistent set of behaviors. Character animation (both traditional and computer-generated [1, 4]) offers many useful principles, insights, and techniques for personality-rich behavior and movement now being adapted to autonomous robots to serve this purpose.

Robot Theatre as Testbed

To be effective and function coherently in the real world, sociable robots must quickly and successfully handle both physical interactions with the environment and social interactions with people. Each interaction is fundamentally complex, unpredictable, uncertain, and only partially knowable in its own way. For robotics researchers to make progress in the design and development of sociable robots under such difficult circumstances, it is often helpful to explore research

Figure 1. The interactive terrarium, approximately 7 feet wide X 7 feet deep X 10 feet high, integrates eight channels of digital audio and music, 40 color controlled lights, six ultrasonic foggers, one waterfall pump, four tubeworm creatures controlled by one servo controller, and eight drum crystal triggers. The design and control of the intelligent stage is handled by a dedicated software system called Secret Systems developed by Joshua Strickon at the MIT Media Lab [10].
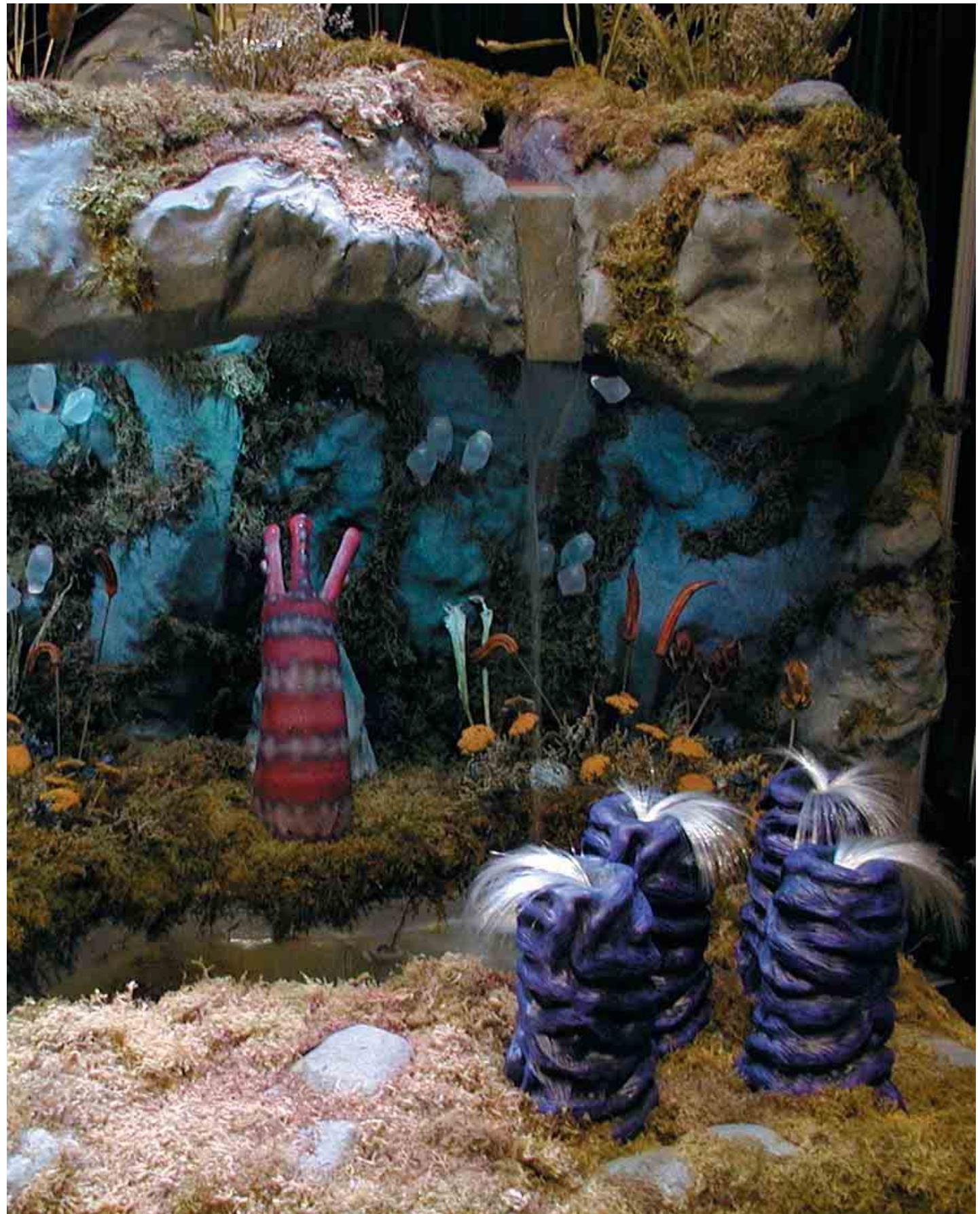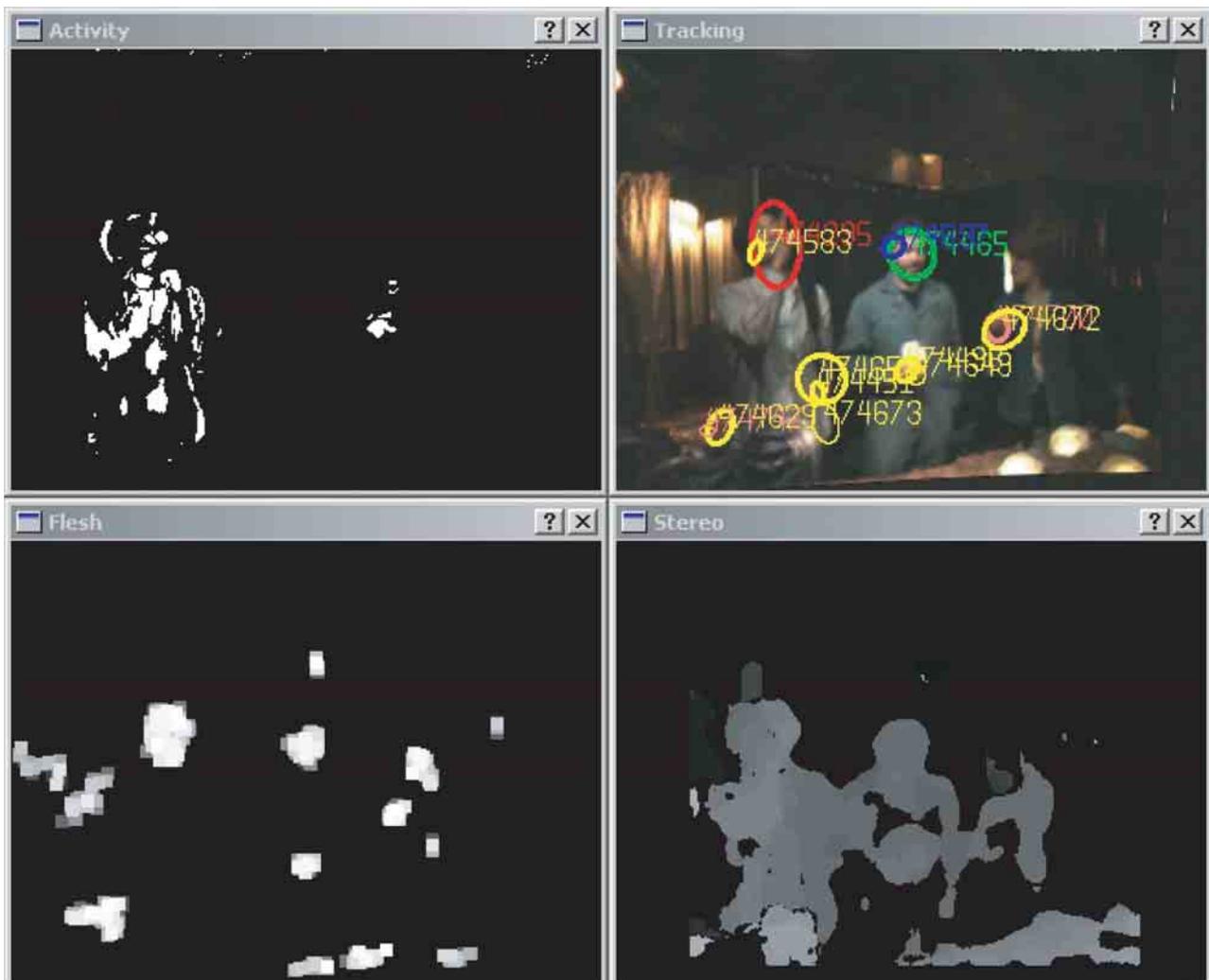
Figure 2. The vision system is embedded in the terrarium wall so the camera faces the audience. Motion is detected (upper-left frame); human skin chromaticity is extracted (lower-left frame); a foreground depth map is computed (lower-right frame); and the faces and hands of audience participants are tracked (upper-right frame). The system was developed in collaboration with the Vision Interfaces Group at the MIT Media Lab.

issues within a constrained yet interesting scenario. For instance, RoboCup Soccer has become a galvanizing test domain for the field of multi-agent robots [7]. The rules of the game and the specifications of the playing field lend enough constraint to make the challenge of having two teams of autonomous robots compete in a soccer match approachable, without sacrificing the interesting problems of team cooperation in the face of dynamic, unpredictable, and adverse conditions.

Live performance with human actors (such as in a theatre) could serve as an equivalent test domain to advance research in autonomous sociable robots. The script places constraints on the dialogue and interaction. The storyline defines concise test scenarios. The stage constrains the environment, especially if it is equipped with special sensing, communication, or computational infrastructure. More important, such an

intelligent stage, with its embedded computing and sensing systems, is a resource that autonomous robotic performers can use to bolster their own ability to perceive and interact with people within the environment.

Good actors often say that half of acting is reacting. Hence, a robot actor must be able to act/react in a convincing and compelling manner to the performance of another entity, whether human or robot, as it unfolds. This requires sophisticated perceptual, behavioral, and expressive capabilities. Creating a robot that is a great character actor with a strong stage presence would certainly address two of the core challenges cited earlier—perception/interpretation and responsiveness/expression. Introducing improvisation or allowing for more audience participation makes the situation that much more unpredictable and unconstrained—approaching open-ended interaction with people. Advances within such a test scenario could help bootstrap the social interactivity of robots in the real world.

Several robotics groups have explored the notion of robot theatre in the context of multiple autonomous robot performers [3, 12]. They are often small mobile robots that resemble vehicles and perform a simple play

by navigating about the stage, perhaps delivering dialogue and reacting to one another's behavior. However, our interest in interactive robot theatre focuses on robot-human interaction. Hence, the work reported here emphasizes the challenges of real-time visual perception of people, natural and expressive movement of creature-like robots, and appropriate responsiveness to people in a loosely constrained storyline. Because our robots do not deliver dialogue, their movement and behavior must be readily apparent and understandable to the audience.

We introduced our interactive robot theatre installation at the 2002 SIGGRAPH Emerging Technology Exhibit (see Figure 1 and robotic.media.mit.edu/projects/anemone.html). The storyline was inspired by the notion of primitive life on an alien world. Audience participants were allowed to interact with the cyber flora and fauna of this fanciful robotic terrarium as it transitioned from day to night. By day, a serpentine, anemone-like creature (called Public Anemone) was awake and carried out its daily "chores," including "watering" nearby plants, "drinking" from the pond, and "bathing" in the waterfall. It perceived human audience members through a real-time stereo vision system, allowing people to compete for its attention and distract it from the chores. It responded by orienting its body toward them and tracking their movements (see the sidebar). If someone got too close, however, it became "frightened" and recoiled defensively, like a rattlesnake.

At night, when the lights dimmed, it went to sleep, and a variety of nocturnal creatures and special effects emerged, including glowing fiber-optic tubeworms, musical drum crystals, luminescent wall crystals, and a sparkling pond enveloped by a gentle mist. The audience interacted with these creatures through touch, eliciting light and musical responses. The tubeworms detected the nearby proximity of people through capacitive sensing, causing them to react musically, optically, and mechanically by retreating into their shells. The drum crystals allowed participants to create rhythm sequences based on how forcefully they were tapped, synchronizing with the glowing crystal wall theme. Together, these elements created a physically interactive, ever-changing, multisensory experience that engaged the audience through sight, sound, scent, and touch. Meanwhile, the performance followed a narrative, but the interaction of the stage and robot performers with the audience made the experience different each time.

Vision System

The nature of the interaction environment at SIGGRAPH presented special challenges for using video cameras as the anemone's principle sensor for perceiving people. The vision algorithms had to be robust enough to cope with unknown numbers of people simultaneously as they entered and left the robot's field of view at any time. Moreover, audience participants ranged widely in biometric characteristics (such as height and race) while wearing wildly differing apparel and attempting to interact with the robot in inconsistent and unpredictable ways. In addition, visual processing had a strict real-time constraint, since any delay in the robot's reactions would immediately condemn the interaction as unconvincing. We designed the vision system to process no fewer than 10 frames per second for a maximum reaction time of around 0.10 seconds. Given these constraints, we concentrated on two human body features: hands and faces; each is rel-

BECAUSE OUR ROBOTS do not deliver dialogue, their movement and behavior must be readily apparent and understandable to the audience.

atively constant in form among individuals, is unlikely to be obscured by clothing, and would be used in deliberate attempts to attract the robot's attention.

The anemone responded to audience members either by orienting its body toward them and tracking the movement of their hands or faces or by recoiling from them if their reach came too close, or about two feet from the robot. Absolute 3D spatial position was not required, but the robot needed to perceive from a "front-on" perspective to accurately orient itself toward features of interest. It also had to determine the nearness of these objects so it could react to invasions of its "personal space," or again anything closer than about two feet. To address these issues, we designed the real-time vision system to include a pair of fixed-baseline stereo cameras, one mounted behind the robot facing the audience, the other mounted overhead aiming down at the terrarium.

It would have been difficult if not impossible to design a computationally tractable model for such an open-ended interaction environment. However, we were able to implement relatively cheap and fast algorithms for performing certain kinds of model-free visual feature extractions (such as proximity, human
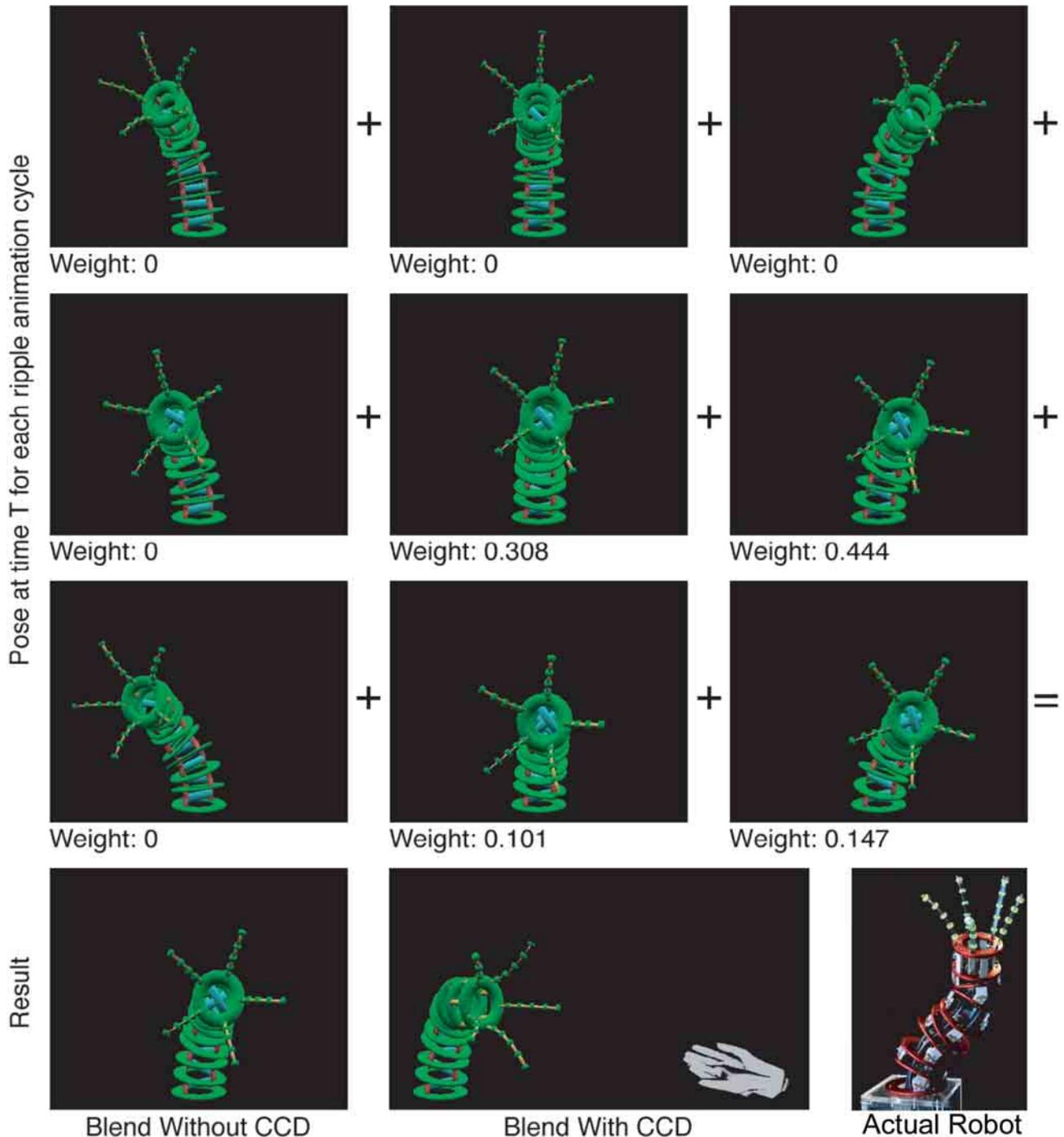
Figure 3. The images in the top three rows are single frames of each of the nine orientation animations for active holds provided by the animators. The images in the bottom row are the result of the weighted blend of nearest neighbors (left) with inverse kinematics applied to the "head" stages (center), and the actual Public Anemone robot (right) without its silicone skin.

skin color, and motion activity). Although individually limited, each was robust within its own narrow scope of performance.

We deliberately had these low-level functions reinforce one another by searching for conjunctions of multiple features. The software consisted of sev-

eral modules performing low-level feature detection, along with a tracking system (see Figure 2). A stereo correlation engine compared the two images for stereo correspondence, computing a 3D depth, or disparity map, at about 15 frames per second. The foreground depth map and the skin probability map were then filtered and combined (by the region extractor), and the optimal bounding ellipse was computed for each region. A stereo correlation engine compared the two images for stereo correspondence, computing a 3D depth, or disparity map, at about 15

frames per second. The depths in this map were compared at each point to a corresponding estimate of the background depth to produce a foreground depth map. The color images were simultaneously normalized and analyzed with a probabilistic model of human skin chromaticity to segment out areas corresponding to human skin. The foreground depth map and the skin probability map were then filtered and combined to extract regions present in both feature spaces. An optimal bounding ellipse was computed for each positive region. For the camera facing the audience, a Viola-Jones face detector [11] ran on each bounding ellipse to determine whether the region corresponded to a face. The regions were then tracked over time, based on their position, size, orientation, and velocity. Connected components were examined to match hands and faces to a single owner. All this information was transmitted to the anemone's behavior system located on a separate computer.

Generating Expressive and Goal-Direction Movement

To enable the anemone to present a lifelike and expressive quality of motion, we based its movements on handmade animations of a scale 3D model of the robot. We imported the model and animations into the C4 system (for Characters, version 4), a behavior-based AI engine for designing and controlling interactive animated characters [4]. Since the 3D model and animations were designed with the same scale and articulation as the physical robot, driving the actual robot was a straightforward translation of rotation information from the model into values for the physical motors.

The target position, velocity, and acceleration commands were each sent via high-speed serial connection to the robot's motor servo boards 50 times per second. These boards ran a proportional-plus-derivative feedback controller (a standard control law that adjusts the commanded movement based on the magnitude of its error from the target position and its derivative at 20KHz). This ensured the robot smoothly followed the prescribed trajectory while performing under varying physical loads (due to gravity) and the configuration of its mechanics.

Within the C4 system, animations were treated as data to be combined in various ways to generate the robot's joint angle trajectories. The simplest type of combination was layering, where different parts of the robot are controlled by multiple independent animations. For instance, a separate animation for moving the tentacles might be combined with another animation controlling the body. Animations controlling the same joints were blended together to produce a new animation combining, on a per-joint basis, these input animations. An arbitrary number of dynamically weighted source animations could be blended together using a quaternion-based algorithm similar to the verb-adverb technique in [9].

The representation of an animation in the system was an ordered list of poses linked together to form a directed graph (called a pose graph) through which the anemone traveled [4]. In general, joint angles were restricted to moving along these paths. However, a more immediate transition strategy was needed for reflexive behaviors (such as the robot's fear response when cowering defensively upon receiving a threaten-

---

THE PUBLIC ANEMONE ROBOT

THE PUBLIC ANEMONE robot was the principle performer. We designed it to have an organic appearance and natural quality of movement. It was highly articulated, with 13 degrees of freedom consisting of eight body stages and five tentacles. Each body stage was rotated 90 degrees relative to the stage below, with servomotors mounted at each stage. This structure gave it a broad range of motion with the ability to precisely control each stage. We equipped the body motors with a combination of optical encoders and potentiometers, enabling both absolute position sensing and precise control via a custom high-density motor control platform. To achieve a smooth quality of motion and avoid the effects of cumulative backlash from the planetary gearboxes, torsional springs at each stage preloaded each gearbox against one side of the backlash region. The robot's skin was a highly elastic silicone rubber (Gl-245 tin condensation cure silicone from Silicone, Inc., custom plasticized for desired softness and elasticity), designed to fit over the mechanics in a way that maximized realism of motion. **C**

ing stimulus). In this case, reflexive behaviors animations had to be interrupted in ways that looked natural. Creating hand-animated transitions from every possible position of the anemone to the "cower" animation would be untenable, so we used a programmatic blend mechanism for simple per-joint interpolation to change the joint angles from their current angle to the positions required for the cower animation.

Whereas a straight programmatic blend from one position to another looked convincingly authentic for a quick movement, it was not sufficient for more prolonged actions (such as orienting toward and smoothly pursuing audience participants); the robot frequently exhibited goal-directed behavior that had to be performed in a lifelike way. Our animators provided a series of nine animations, arranged in a 3 x 3 grid (see Figure 3), each depicting the robot orienting at a particular location while rippling its body segments. We called such behavior an "active hold."

It is important to maintain this style of movement while the anemone engaged with audience participants. Our strategy combined animation blending with the application of inverse kinematics for minor corrections of the top two stages of the robot, or those closest to the tentacles that were interpreted as the robot's head, as in the figure. For the blending component, all but the final two "head" stages were controlled by a programmatic weighted blend of the four animations closest to the desired target orientation. By varying the weights and the animations, the anemone tracked an object smoothly throughout its space while maintaining its characteristic body ripple. For the head stages, we incorporated a cyclic coordinate descent algorithm to produce a more exact orientation of the tentacles to "look at" the target.

Conclusion

We may see more elaborate versions of interactive robot theatre in theme parks, museums, and storefront windows. There may be fanciful robotic characters on Broadway performing with human actors on an intelligent stage. Here, we have sought to address two core challenges: real-time perception of multiple audience members in a highly dynamic and challenging venue and the generation of goal-directed and expressive movement for a highly articulated autonomous robot. In the broader view, interactive robot theatre represents an interesting test scenario for exploring challenging research questions in the development of sociable robots that interact in a natural and socially appropriate manner with humans. These are important skills for robots intended to cooperate with their human counterparts as capable partners, both at home and at work. **C**

REFERENCES

1. Bates, J. The role of emotion in believable characters. *Commun. ACM 37,* 7 (July 1994), 122–125.
2. Breazeal, C. *Designing Sociable Robots.* MIT Press, Cambridge, MA, 2002.
3. Bruce, A., Knight, J., and Nourbakhsh, I. Robot Improv: Using drama to create believable agents. In *AAAI Workshop Technical Report WS-99-15 of the 8th Mobile Robot Competition and Exhibition.* AAAI Press, Menlo Park, CA, 1999, 27–33.
4. Burke, R., Isla, D., Downie, M., Ivanov, Y., and Blumberg, B. Creature-Smarts: The art and architecture of a virtual brain. In *Proceedings of the Game Developers Conference* (San Jose, CA, 2001), 147–166.
5. Fong, T., Nourbakhsh, I., and Dautenhahn, K. A survey of socially interactive robots. *Robot. Auton. Syst. 42,* 3–4 (2003), 143–166.
6. Interactive Animation Initiative. Carnegie-Mellon University Entertainment Technology Center, Pittsburgh, PA; see www.etc.cmu.edu/projects/iai/.
7. Kitano, H., Tambe, M., Stone, P., Veloso, M., Coradeschi, S., Matsubara, H., Noda, I., and Asada, M. The RoboCup 1997 synthetic agents challenge. In *Proceedings of the 1st International Workshop on RoboCup (IJCAI-97)* (Nagoya, Japan). Morgan Kaufmann Publishers, San Francisco, CA, 1997.
8. Norman, D. *How Might Humans Interact with Robots?* Keynote address to the DARPA/NSF Workshop on Human-Robot Interaction (San Luis Obispo, CA, Sept. 29–30, 2001).
9. Rose, C., Cohen, M., and Bodenheimer, B. Verbs and adverbs: Multidimensional motion interpolation. *IEEE Comput. Graph. Applic. 18,* 5 (1998), 32–40.
10. Strickon, J. *Smoke and Mirrors to Modern Computers: Rethinking the Design and Implementation of Interactive, Location-Based Entertainment Experiences.* Ph.D. Thesis, MIT Program of Media Arts and Sciences, Cambridge, MA, 2002.
11. Viola, P. and Jones, M. Rapid object detection using a boosted cascade of simple features. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Kauai, HI). IEEE Computer Society Press, 2001, 511–518.
12. Werger, B. Profile of a winner: Brandeis University and Ullanta performance robotics robotic love triangle. *AI Mag. 19,* 3 (1998), 35–38.

CYNTHIA BREAZEAL (cynthiab@media.mit.edu) is an assistant professor of media arts and sciences and directs the Robotic Life Group at the MIT Media Lab in Cambridge, MA.

ANDREW BROOKS (zoz@mit.edu) is a Ph.D. candidate in the Department of Electrical Engineering and Computer Science at MIT.

JESSE GRAY (jg@media.mit.edu) is a graduate student pursuing an S.M. degree in media arts and sciences at MIT.

MATTHEW HANCHER (mdh@mit.edu) is a graduate student pursuing an M.Eng. degree in the Department of Electrical Engineering and Computer Science at MIT.

JOHN MCBEAN (jmcbean@mit.edu) is a graduate student pursuing an S.M. degree in the Department of Mechanical Engineering at MIT.

WALTER DANIEL STIEHL (wdstiehl@mit.edu) is pursuing a B.S. degree in the Department of Mechanical Engineering at MIT.

JOSHUA STRICKON (strickon@media.mit.edu) is at Apple Computer Corp. in Cupertino, CA. He received a Ph.D. in media arts and sciences from MIT in 2002.