A Guided Performance Interface for Augmenting Social Experiences with an Interactive Animatronic Character

Seema Patel and **William Bosley** and **David Culyba** and **Sabrina A. Haskell**
Andrew Hosmer and **TJ Jackson** and **Shane J. M. Liesegang** and **Peter Stepniewicz**
James Valenti and **Salim Zayat** and **Brenda Harger**
Entertainment Technology Center
Carnegie Mellon University
700 Technology Drive
Pittsburgh, PA 15219

Abstract

Entertainment animatronics has traditionally been a discipline devoid of interactivity. Previously, we brought interactivity to this field by creating a suite of content authoring tools that allowed entertainment artists to easily develop fully autonomous believable experiences with an animatronic character. The recent development of a Guided Performance Interface (GPI) has allowed us to explore the advantages of non-autonomous control. Our new hybrid approach utilizes an autonomous AI system to control low-level behaviors and idle movements, which are augmented by high-level processes (such as complex conversation) issued by a human operator through the GPI. After observing thousands of interactions between human guests and our animatronic character at SIGGRAPH 2005's Emerging Technologies Exhibition, we strongly feel that both autonomy and guided performance have important roles in interactive, entertainment robotics. Together, the autonomous system and the new Guided Performance Interface allow guests to experience extremely rich, believable, social experiences with robots using technology available *today*.

Introduction

One of the most visible fields in the area of human-robot interaction is that of entertainment robotics. Some of the earliest and most recognized entertainment robots are the elaborate animatronics found in many theme and amusement parks. More recently, the entertainment robotics industry has exploded into the consumer robotics sector. Unlike their traditional, linearly pre-scripted animatronic counterparts, the novelty of personal entertainment robots like Sony's AIBO dog and Wow Wee Toys' Robosapien is in their interactivity. The AIBO, for example, can locate specific objects, recognize patterns, identify its owner's voice, and determine when it's being pet. The robotic dogs can be taught tricks, are capable of "feeling" emotions, and are shipped with a personality that evolves over time as their owners interact with them. It is highly likely that the popularity of these robots stems from their interactive capabilities.

The Interbots Initiative research group at the Entertainment Technology Center (ETC) seeks to leverage the success of household entertainment robotics and introduce interactivity to the field of higher-end entertainment animatronics. Our goals are to: 1) Develop complete, interactive, entertaining, and *believable* social experiences with animatronic characters, and 2) Continue to develop software that allows non-technologists to rapidly design, create, and guide these experiences. Previously, an expressive animatronic character, Quasi, and the kiosk in which he is housed were designed and fabricated. Custom control software was developed allowing Quasi to engage in autonomous interactive social experiences with guests. Intuitive authoring tools were also created, allowing non-technologists to develop complete interactive, entertaining experiences with Quasi in as little as two weeks (Haskell *et al.* 2005). Current work expands upon the previous research in two ways: 1) The design of the animatronic character was extended to include legs, his mechanical system was overhauled, and a new, cable-driven robot was fabricated, and 2) A new piece of control software was created, the Guided Performance Interface (GPI).

We feel that an embodied character is vital to a believable animatronic experience—no matter how impressive the technology, it is impossible to suspend an audience's disbelief and create the illusion of life without an engaging character and personality. Furthermore, the audience cannot distinguish between a character that "actually thinks like a human" and a character that "appears to think like a human." Thus, we are not concerned with replicating human intellect and emotional processes—rather, we wish to emulate them. Instead of focusing on pure artificial intelligence research, we have instead chosen to pursue any and all solutions that maximize believability in social experiences between humans and robots.

The highest standard of success in these experiences is the audience's belief that the creature they are interacting with is alive. The improper use of developing technologies such as speech recognition can lead to frustrating and dissatisfying experiences. With believability as the ultimate goal, it is just as important to decide what technologies *not* to utilize in an experience as it is to be cutting edge—if a piece of technology routinely behaves in a manner that shatters the illusion of life, it will not be used in our platform. We also believe that the most successful entertainment experiences involve the contributions of writers, actors, animators,

---

sculptors, painters, and other artists, hence our focus on creating simple, easy-to-use software tools.

The software in the Interbots Platform allows for Quasi's interactive experiences to range from fully autonomous to completely pre-scripted (Haskell *et al.* 2005). Initially, we envisioned that all interactions with Quasi would be controlled autonomously, and many months were spent developing AI software and autonomous social experiences. Recently, however, a Guided Performance Interface (GPI) was developed for controlling the robot at a live question and answer event. The response it received was stunning; unlike the autonomous interactions which held guests' attention for two to three minutes, the GPI was capable of captivating people for periods of time an order of magnitude longer. Children, especially, never seemed to tire of interacting with the robot. As our goal is to create the *illusion* of life, we fear we may have been too hasty in our initial exclusion of non-autonomous control. In fact, the more we test the GPI, the more strongly we feel that both autonomy and guided performance have valuable roles in the field of human-robot interaction. Following this belief, this paper presents our new hybrid control scheme for entertainment robotics that combines artificial intelligence with human teleoperation.

Related Work

Within academia our research is similar to existing work in two areas: social interactions between humans and animatronic robots, and control schemes that combine autonomy with human teleoperation.

Our work within the former category is similar to Leonardo, a fully autonomous animatronic robot that communicates with humans non-verbally through facial expression and body posture (Breazeal *et al.* 2004). Like Leonardo, the autonomous systems of our animatronic robot Quasi are influenced by behavioral, emotional, and environmental factors. Interactions with Leonardo, however, are centered around human-robot collaboration and learning, whereas our focus is on engaging humans in a believable, entertaining experience. Grace (Bruce, Nourbakhsh, & Simmons 2002) and the Nursebot (Pineau *et al.* 2003) are examples of fully autonomous socially interactive robots that, like Quasi, are capable of communicating verbally. However, due to current limitations of language processing software, conversations with these robots are not highly believable like those with Quasi, which are controlled by a human operator.

Our approach is similar to the one behind interactive roboceptionists Valerie and Tank in that our work is also influenced by theatre and focuses largely on the character and personality of the robot. Quasi, Valerie, and Tank all have elaborate character backstories, as well as online presences including e-mail addresses and internet journals. Like the roboceptionists, Quasi can develop autonomous long-term relationships with humans. However, that is not our primary focus; instead we concentrate on mechanisms for creating the richest, highest quality interactive experiences possible.

Some investigation of human responses to teleoperated social robots has been conducted, but these cases differ from our research in several ways. Unlike our work, these robots did not employ any autonomous control. Furthermore, in the case of Sparky, researchers intentionally removed from consideration behaviors that could not be replicated autonomously in the near future, such as natural language comprehension (Scheeff 2000). Outwardly, Tito is similar to Quasi, though his verbal communication is limited to pre-recorded speech (Michaud *et al.* 2005). Currently, Quasi appears to be the only socially interactive robot capable of operating on a sliding scale of three control modes: fully autonomous, pre-scripted, and guided.

Much work has been conducted on adjustable autonomy in a variety of robotic systems from mobility aids for the elderly, to mobile rovers, to free-walking systems (Yu, Spenko, & Dubowsky 2003; Desai & Yanco 2005; Uenohara, Ross, & Friedman 1991). This research, however, has primarily focused on the control of navigational parameters (e.g. speed and direction) and physical orientation in space, whereas our work centers around controlling aspects of social interaction such as behaviors, emotions, gestures, posture, eye gaze, and speech.

Within industry our work is most similar to Lucky, an interactive mobile animatronic dinosaur that frequents the Disney Theme Parks. Our Virtual Show Control System (described in "Control Software") is also very similar to Disney's "Turtle Talk with Crush" exhibit in which an animated 3D turtle is puppeteered live though the use of a digital interface. Not much information is publicly available about these technologies, though it is widely known that the characters are controlled by human operators. In many ways, our research attempts to bridge academia and industry.

Hardware

The hardware in the Interbots Platform consists of three primary components: the animatronic character, the portable kiosk on which he resides, and the show control system.

Quasi the Robot

The most visible (and important) piece of hardware in the Interbots Platform is the custom-built animatronic robot, Quasi. Quasi's stature was modeled after a cartoonish child with a large, round head, large eyes, small body, and short limbs (see Figure 1). This design elicits a friendly, compassionate instinct in those who interact with him.

As Quasi's primary purpose is to interact socially with humans, the first priority when designing him was expressiveness. Eight servos have been dedicated to the movements of his eyelids alone, which were designed to mimic the movements of eyebrows—each eye has an upper and lower eyelid that can raise and lower as well as rotate clockwise and counterclockwise. (These rotations produce a diagonal tilt allowing Quasi to effectively "slant" his eyes, helping him to convey emotions such as anger, suspicion, or sadness.) Quasi has a number of features in addition to his eyelids that facilitate his ability to express emotion, the most prominent of which are LED lighting fixtures for his eyes and antennae. These combine red, green, and blue LEDs to display any color of the spectrum. It was decided that a subset of the LEDs in his eyes would not be dedicated to displaying pupils

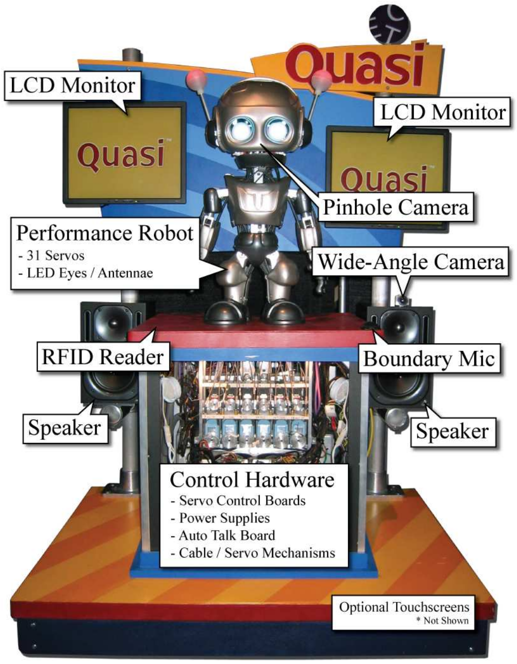

Figure 1: The animatronic character Quasi, his kiosk, and its sensors.

due to the concern that unless his face tracking software was 100% accurate, it would always appear as if he was looking slightly past the guest instead of appearing as if he was making eye contact with the guest. (This decision follows Interbots' philosophy of avoiding technology and techniques that have a high risk of breaking the illusion of life.) Quasi's antennae can rotate both forward and backward as well as tilt in and out, giving them an expressive quality not unlike that of a dog's ears. The physical movements of Quasi's eyelids, antennae, and body posture combined with changing LED colors allow him to effectively communicate emotions and personality even without the use of speech.

Aside from the addition of legs, the biggest difference between the previous design ("Quasi 1.0") and the current iteration ("Quasi 2.0") is the internal mechanical subsystem, which has been completely overhauled and re-designed. In the previous version, all of the hobby servo motors were contained inside the figure's armature. While this allowed for the animatronic figure to be compact and self-contained, it made replacement and repair very difficult and forced the use of smaller, less powerful, less durable hobby servos. In the current design, larger and more powerful industrial grade servos have been placed in an easily accessible "servo bank" underneath the robot and push-pull cables are used to transfer the mechanical forces to the armature. Without the constraint of fitting the servos inside the shell, the number of Quasi 2.0's servos was increased from 27 to 31, allowing for a larger number of movements. The movements themselves also have a much greater range of motion as movement of the armature is not impeded by the internal placement of motors. Inertia of the figure is also reduced, allowing for more precise, faster movements with less mechanical backlash.

Quasi's Kiosk

Interactive experiences with the Interbots Platform involve more than just an animatronic figure alone. Quasi's environment—a kiosk that contains several sensors and a multitude of multimedia hardware—is just as much a part of the Quasi experience as the robot itself. A new, portable kiosk was created for Quasi 2.0 (see Figure 1). Unlike Quasi 1.0's kiosk which houses all of his control hardware and was designed to be a permanent installation in the hallways of the ETC, Quasi 2.0's kiosk was designed with portability as the first priority. Most of the control hardware has been removed to an external roadcase which can be shipped as is. The kiosk itself breaks down into components that can be transported in a single crate. While Quasi 1.0's kiosk has the subdued design and color scheme of a museum installation, Quasi 2.0's environment was designed to reflect the personality of a 12 year old boy and would appear at home in an amusement park.

The kiosk provides information to Quasi's autonomous and guided software through several sensors. A wide-angle camera mounted on the side of the kiosk captures video for the autonomous face-tracking software, allowing Quasi to direct his gaze towards guests. Although it is not technically located on the kiosk, a pinhole camera in Quasi's nose provides a secondary video feed to the Guided Performance Interface, and could also be used as an input to the autonomous software. A sensitive boundary microphone mounted on the front of the kiosk picks up sound for speech recognition software and the GPI. A passive RFID reader allows individuals to be uniquely identified using low-cost tags that can be carried on keychains.

Quasi can enhance his environment with sound; stereo speakers mounted to the side of the kiosk allow him to play speech and music. An LCD touch screen that can be set up in front of the figure allows for intuitive interaction with questions, games, and other active content. Two flat panel monitors mounted behind and on each side of the figure allow for the display of elaborate multimedia content. Although there was concern about people touching Quasi and damaging him, there was a strong sentiment against having him behind glass or any sort of protective barrier. Without the sharing of a physical space, guests would be deterred from emotionally connecting to the figure. Fortunately, these concerns about damage proved largely unnecessary—unless Quasi holds an arm out as if to shake hands, most guests do not try to touch the figure.

In the future, Quasi 2.0's kiosk will be augmented with more sensors to help him better perceive his environment. These may include a motion detector, forward-facing infrared rangefinders, and stereo microphones.

The Control Hardware

In accordance with the goal of keeping the new design as rugged and portable as possible, it was decided to move support equipment to a separate custom-built roadcase that would connect to Quasi via a detachable multi-cable umbilical. Four custom rack-mount PC's serve as Quasi's brains. Gigabit LAN and 802.11g WLAN equipment provides network connectivity between computers. Audio equipment includes microphone pre-amps, a mixer, a sound processor, and wireless audio receivers and transmitters. As Quasi's PCs have development tools installed on them, the road-case also serves as a portable multimedia production facility. These tools may be used to create new software, behaviors, movements, and onscreen content.

Quasi's show control hardware speaks one common language—the DMX protocol. The DMX control line is daisy-chained to multiple pieces of equipment. Power supplies plugged into the chain illuminate the LEDs in his eyes and antennae. Three servo control boards, each supporting sixteen channels, allow up to forty-eight individual motors to be commanded simultaneously. A relay board provides the ability to power Quasi on and off to effectively "put him to sleep".

Software

The following section gives an overview of the autonomous and guided software used in the Interbots Platform. More detailed descriptions of the autonomous control software and a comprehensive overview of the easy-to-use content authoring tools that are used to create autonomous interactions were presented previously (Haskell *et al.* 2005).

Control Software

Character State Control System (CSCS) The backend decision mechanism for a character running on our platform is called the Character State Control System (CSCS). Quasi's autonomous behaviors are stored in a database format as a network of related states and transitions rules that depend on a set of system inputs. At a high level, CSCS reads in the behavioral model as a set of what is called "superstates". These can be thought of as categories of behaviors such as "greeting", "farewell", and "sleep". Inputs to the system are split into internal "factors" (such as the character's current emotional state) and external "facts" (such as sensor readings). Each superstate has a set of "entry requirements" or fact and factor requirements that must be true for the character to enter that superstate. Superstates also have "stay requirements," or requirements that must be true for a character to keep progressing through that superstate. When a character completes a superstate, or when the stay requirements are no longer true, CSCS determines the set of all superstates whose entry requirements are met by the current facts and factors. A superstate is randomly selected from those in this set with the highest priority setting.

Within each superstate is a network of substates which can be thought of as a traditional finite state machine. Substates are explicitly connected via "transitions" and upon entry to a substate two things happen. First, adjustments are made to the character's internal factors. Second, "Actions" take place. Actions are sequences of output that the character executes. This can include events such as executing character animations, lighting changes, playing sounds, displaying something on a video screen, or even explicitly jumping to a different superstate. After a substate is completed, CSCS considers all transitions leading from that substate. Transitions can also be weighted to control their probability of being executed and also have transition requirements much like superstates have entry requirements. If there are no transitions exiting a substate, CSCS exits the current superstate.

Real-Time Show Control System (RSCS) Quasi's physical movements stem from animations created using a virtual replica of the robot in Maya, a popular 3D modeling and animation application. RSCS is the actual software interface to the hardware—it receives messages from CSCS specifying which animation files to load and which channels to play them on, and then sends DMX data to the show control hardware. RSCS can linearly blend animations together, as well as seamlessly transition from one animation to another. A recent development is the Virtual Show Control System (VSCS), which provides the same exact functionality as RSCS, only the end recipient is not a physical character, but rather the virtual 3D animated character running in the open source Panda3D simulation engine. This provides developers with a simple way to test interactions quickly without the need for a physical character. Content creators can construct complete interactive experiences even while the physical figure is being constructed or is out of commission.

Other Control Applications The Interbots Platform utilizes a tool for face tracking called "cVision" which is an implementation of Intel's OpenCV face tracking library. We believe the ability to detect faces to be a crucial form of input for a character designed to interact with humans. When cVision is running it overrides control of Quasi's head, allowing him to appear to be looking directly at the face of a guest. The variation in movement induced by this method also lends an element of realism and believability to the character's movement—in fact numerous feedback indicates that cVision is the difference between interacting with a "machine" and a character that appears alive.

The Babble application is a simple implementation of the open-source Sphinx voice recognition package developed at Carnegie Mellon University, enhanced with the capability to pass messages between applications. Babble works decently for simple voice responses to Quasi's questions in a fairly controlled setting. However, it is very sensitive to local sound conditions and is not reliable in noisy environments, thus we limit its use to controlled research situations.

The Guided Performance Interface

Originally, the Guided Performance Interface (GPI) was developed for unique situations that the autonomous software can not easily accommodate, such as a live question and answer session. Because responses in a live event must be nearly instantaneous, we felt that an operator to robot ratio of one would be optimal. Thus, unlike the animatronics found in the special effects industry which can require 10 or more puppeteers, the GPI allows for a robot to be "guided" by a single actor equipped with a wireless tablet

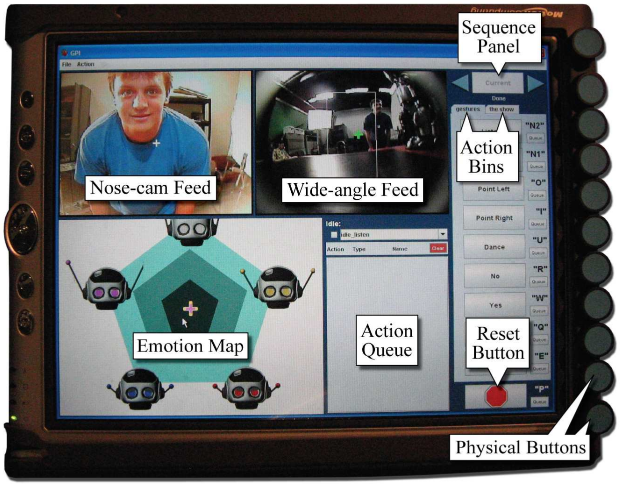

Figure 2: The Guided Performance Interface.

PC, a stylus, and a microphone headset. (The use of the term "actor" instead of "operator" is intentional. All of the members of the Interbots group have had improvisational acting training and strongly believe that the most successful guided interactions are the result of an operator skilled in live entertainment.) Audio from the boundary microphone on Quasi's kiosk is fed into the headset, allowing the actor to hear guests' speech. As the actor speaks into the microphone on the headset, their voice is pitch shifted and sent to an auto-talk board which moves Quasi's jaw based on the amplitude of the sound signal. This allows the actor to speak as Quasi in real-time.

The upper portion of the interface is occupied by two video feeds (see Figure 2). The feed on the left is from a pinhole camera placed in Quasi's nose. It essentially allows the actor to see what Quasi is seeing. The video feed on the right is from a wide-angle camera mounted on the side of Quasi's kiosk. The white rectangle superimposed over this video feed represents Quasi's body. By moving the green crosshairs inside of the white rectangle, the actor directly controls the movement of Quasi's head. If the actor moves the crosshairs to the left or right side of the rectangle, he or she can drag the rectangle across the camera feed. By moving the rectangle the actor can turn Quasi's body to the left and to the right. The end result is a very natural movement in which the head leads the body in a gaze, a traditional animation technique. The actor can also place the crosshairs on the upper corners of the rectangle and drag the corners down, tilting the rectangle. This allows the actor to tilt Quasi's head from side to side. The use of both camera feeds together allows the actor to spot a guest in the wide-angle view, move Quasi so that he faces the guest, and then use the nose-cam view to make fine adjustments to ensure that Quasi is making eye contact with the guest.

The right side of the interface contains the Action Bins. An Action is a set of atomic actions (e.g. animations, sound files, movie files, Macromedia Flash files, etc.) along with a start time, a duration, and a name. The actor can create an Action and add it to an Action Bin by right clicking on the bin. The actor then selects the atomic actions from a library and specifies the start time, duration, and name. The start time is essentially a delay period—a two second start time indicates that the Action will not be executed until two seconds after it is fired. The duration that the actor specifies need not be the exact duration of the Action; it is essentially a period of time that must elapse before another Action can be executed. The actor can fire an Action by either tapping the rectangular button on the interface or by pushing the adjacent physical button on the right edge of the tablet. This allows the actor to control Quasi's movement with the stylus in one hand while simultaneously firing actions by pressing physical buttons with the other hand. The actor can create multiple Action Bins, each of which can be brought to the forefront of the interface by tapping on the corresponding tab at the top of the panel.

Directly below the Action Bins is a special "reset" Action button which, when fired, immediately halts Quasi and returns him to a neutral position. Directly above the Action Bins is the Sequence Panel. By right clicking on the Sequence Panel the actor can set up a series of Actions for execution. Tapping the right arrow causes the next Action in the sequence to fire, while tapping the left arrow causes the previous Action to fire. The name of the current Action appears in box between the arrows, and the current Action can be replayed by tapping the box.

To the left of the Action Bins is the Action Queue. Notice that each Action button in an Action Bin has a smaller "Queue" button to its right. By tapping on the Queue button the Action is added to the Action Queue. If no other Action is in the Action Queue, the Action fires immediately. If there are other Actions in the Action Queue, the Action will not fire until the duration of the Action immediately ahead of it elapses. If an Action is fired through an Action Bin while there are Actions playing in the Action Queue, the Action from the Action Bin will override the Actions in the Action Queue.

The lower left quadrant of the interface is occupied by the Emotion Map, which is quite possibly the most innovative feature on the GPI. Each corner of the pentagon represents a different emotion; clockwise from the top they are: happy, confused, angry, sad, and embarrassed. Quasi physically expresses his emotional state through a combination of eye and antennae color, head position, and body posture (see Figure 3). Moving the pink crosshairs to an emotion icon on the outer pentagon causes Quasi to express all three physical aspects of emotion. If the crosshairs are placed on a point of the middle pentagon, his body remains in the same position, but his head position and color change to reflect the emotion. Placing the crosshairs on a point of the inner pentagon changes only Quasi's eye and antennae color. In this respect, the crosshairs' position on the circumference of the pentagon represents the emotion's valence, while the radius from the center of the pentagon represents intensity.

Changing Quasi's emotional state has the side effect of dynamically changing the upper and lower limits on the ranges of Quasi's movements. Therefore, a "wave" animation executed when Quasi is happy will look decidedly dif-

Figure 3: Quasi's emotional expressions and corresponding eye/antennae colors; clockwise from top left: happy (green), angry (red), confused (yellow), neutral (white), embarrassed (pink), and sad (blue).

ferent than when Quasi is sad. When happy, Quasi's head is high, his torso is upright, and the arm he's waving is held high in the air. When sad, Quasi looks at the ground, his torso is slightly slumped over, he holds his arm low, and he waves more slowly. Thus, instead of having to create a separate animation for each action in every possible valence/intensity combination, only one animation is required. This effectively expands our animation library by at least an order of magnitude.

One completely unexpected result from the Emotion Map was linear blending of the physical expression of emotion. For example, the first time the GPI was being tested the actor moved the crosshairs directly between "angry" and "sad". The immediate response from observers was "Wow, he looks so disappointed!" Further experimentation revealed that the halfway point between "sad" and "embarrassed" resulted in an expression of "self-loathing." Although it was suspected that moving the crosshairs around the map would produce unusual effects, it was never anticipated that the blends would map onto other identifiable emotions. The Emotion Map opens up new avenues of research in the physical composition of emotional expression which we plan to pursue in the future.

Autonomous vs. Guided Control

As noted in "Related Work", the majority of interactive animatronic robots utilize autonomous control. It often seems like there is a sentiment within the social animatronics research community that if it's not autonomous, it's "cheating", or is at least inferior to autonomous control. Thus, in our original approach, it did not occur to us to pursue non-autonomous software. However, after seeing the response to the GPI (particularly from children) at several events, as well as the response from the technological community at SIGGRAPH 2005's Emerging Technologies Exhibition, we have come to realize that guided control has many strengths to offer the field of human-robot interaction, especially when the ultimate goal is entertainment. The biggest advantage of guided control is that it offers the most advanced situational awareness available—that of a human. The actor is capable of natural language processing, can carry on complex conversations, can recognize unique faces (and voices), and can recall and apply relevant information. Many guests at SIGGRAPH were surprised and delighted simply by the fact that the robot they had visited the previous day remembered them.

Situational awareness is directly related to the next advantage of guided control: believability. Recall that our goal is to create *believable* social experiences with robots; it is much easier for a guest to believe that the character they are talking to is alive if the high-level intellect behind the character is powered by a human rather than autonomous software. Guided control is also extremely flexible; voice is generated in real time, Actions can be created as needed, and combinations of Actions can be executed on the fly, all of which must be prepared for in advance in the autonomous mode. The biggest weakness of guided control is that it requires a human operator. Any time guided interactions are desired, an actor must be present and available. In general, the GPI offers the better "bang for the buck"—a high level of intelligence and believability for a relatively small amount of code.

The biggest disadvantage of the GPI is conversely the greatest strength of autonomous control; no humans are required. The autonomous software can run 24 hours a day, seven days a week, if desired. Autonomous control is also superiorly suited for low-level actions and behaviors such as general face tracking, eye-blinking, breathing, idle body movements, etc. Additionally, it is better at exact repetition and executing multiple commands simultaneously. The biggest weakness of autonomous control is that the most advanced artificial intelligence is not currently capable of passing the Turing Test. Similarly, natural language processing is extremely difficult and current technology does not come close to replicating human ability. There will undoubtedly come a time when artificial intelligence and speech generation are sophisticated enough to take the place of a human conversationalist but current technologies generally diminish the believability in character experiences, not increase it. This is not a problem for interactive robotic creatures that do not speak, like Sony's AIBO dog, but for humanoid animatronics, it is a big disadvantage. Autonomous control is also an incredible amount of work with comparatively little payoff in terms of guests' reactions. Our autonomous software is extensive and intricate, providing for deep, but ultimately constrained interactions. Experiences can still be believable, but they are confined to the domain set forth by the software, such as a game of tic-tac-toe, an easter egg hunt, or a karaoke show. Guest reactions to the autonomous experiences are still positive, but not nearly as strong as their reactions to the guided interactions.

With memorable experiences as our ultimate goal, we must ask: in an entertainment venue, is an autonomous agent the best approach to engaging interactions, or is it currently just a novelty? Do guests appreciate autonomous interactions because of the experiences they provide, or for

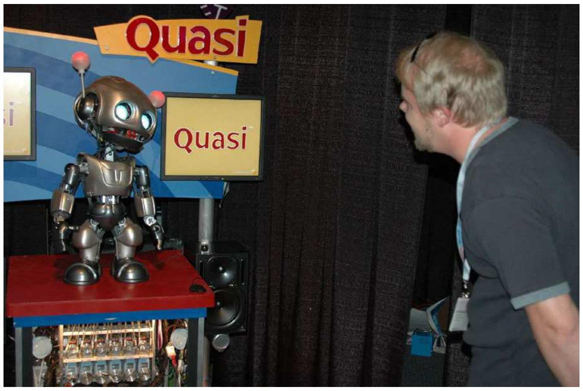

Figure 4: Quasi conversing with a guest at the SIGGRAPH 05 Emerging Technologies Exhibition.

the technological innovation? Is it more effective to use a live actor? What if we combine the strengths of autonomous and guided control? Just as the hierarchical and behavior based robotic architectures eventually merged to form the hybrid architecture, perhaps a current solution in entertaining human-robot interactions lies in a hybrid autonomous/guided control.

Following these thoughts, we created a hybrid control system for use at the SIGGRAPH 2005 Emerging Technologies Exhibition. The autonomous portion of the control system served two purposes: 1) Take the burden of low-level behaviors off the shoulders of the actor, and 2) Quickly demonstrate to guests the platform's autonomous capabilities. For these reasons, the autonomous system consisted of a simple behavioral network, a set of animated responses to user input, and a set of low-level animations. Six RFID tags represented various actions a guest could take: kiss, praise, insult, slap, tickle, and request a dance. Waving a tag in front of the RFID reader caused two things to happen: first, Quasi would fire an animation in response to the action (such as sharply moving his head to the side after a slap), and second, his emotional levels would change. Kiss and praise were positive actions that increased Quasi's overall mood while insult and slap were negative actions that decreased his mood. Too many kisses and praises put Quasi into an "in love" state; his eyes and antennae turned pink, his upper eyelids closed halfway, his antennae drooped down, and an idle animation played in which he slowly moved his head as if dazed. Similarly, too many insults and slaps put Quasi into an "angry" state in which his eyes turned red, his antennae lay flat behind him like the ears of a growling dog, and his eyelids slanted in to create the impression of glaring. When in a good mood, tickling Quasi or asking him to dance resulted in appropriate responses. When in a bad mood, however, all tickles and dance requests were ignored. Although the autonomous software is capable of much more complex interactions, the aforementioned system allowed guests to quickly and clearly see how their actions affected Quasi.

The Guided Performance Interface was used to augment the autonomous interactions. Instead of having to worry about directing idle behaviors or firing animations in response to guests' actions, the actor concentrated on engaging in conversation with guests, directing Quasi's gaze at specific individuals, and firing situational animations, such as a wave goodbye. The autonomous behaviors and animations also gave the actor something to play off of. For example, when a guest gave Quasi enough kisses to make him fall in love, the actor built upon the "in love" idle behavior described above by gazing at the guest and acting distracted, speaking in a slow, happily content voice, and repeatedly flattering and complimenting the guest. This hybrid approach allowed us to exploit the strengths of both autonomous and guided control, providing the best possible interactions for guests.

Although not equipped to collect scientific data, we informally observed thousands of interactions at SIGGRAPH, looking for applications for both autonomous and guided control. The informal conclusion was that each interaction seemed to move through the following four phases:
1. *Greeting Period:* "Hello, my name is..., I'm from..." for both human and robot. This is usually characterized by tentativeness or skepticism on the part of humans, particularly adults. Often, Quasi has to resort to benign jibes to elicit a response from the person (e.g. "Don't humans talk?").
2. *Breaking the Barrier:* Quasi says something undeniably intelligent and relevant, e.g. "I like your purple shirt" (visual understanding), "Your name is Christopher? Like Christopher Columbus? Are you a sailor?" (speech and concept recognition), "I remember you from yesterday" (memory).
3. *Passing a Test:* Human asks Quasi a difficult question ("What's the circumference of a circle with a radius of 2?") or asks Quasi to do something physical like dance, wave, or change emotions. The guest is testing the limits of Quasi's intelligence and Quasi shows himself to be more than they expected.
4. *Conversation:* At this point the guest has accepted Quasi as a legitimate character. They may even know there's a person controlling him but they address all questions to the character and respond accordingly. The illusion of character is complete.

In the future, we would like to explore the expanded use of autonomy in each of these phases. The Babble voice recognition software (described in "Control Software") could allow for phases 1 and 3 to be largely autonomous, but phases 2 and 4 contain elements such as natural language processing that make them better suited to guided control.

We are often asked how guests respond when they discover that the high-level intelligence in the interactions comes from a human actor. (At SIGGRAPH intentional care was taken to ensure that every guest understood the technology behind the experience.) Although guests initially laugh embarrassedly (it is very similar to delivering the punchline on a candid camera show), the overwhelming majority continue to converse with the character. It seems that most people, even those that are technologically savvy, don't care that the intelligence isn't purely artificial. Most people would not walk up to an adult and begin a random conversation, but there's something about the *character* of Quasi that draws guests in for exactly that experience.

Future Work

The Guided Performance Interface opens up several avenues of research we wish to pursue. We are especially interested in determining if autonomous behavior can be derived from guided commands. We wish to ascertain if there are sets of guided commands that actors repeatedly execute, and if any such commands can be replicated through autonomous control. We would also like to conduct more rigorous comparison tests between autonomous and guided control in order to isolate the nuances of believability. In a controlled experiment, would guests find a robot utilizing cVision (autonomous face tracking software) less believable than an actor manually directing the robot's gaze? If so, why, and can we extract the difference and implement it autonomously?

In terms of autonomous software development, we would like to enhance cVision with a feature that would allow an actor to select a specific face in the robot's vision, and have the software automatically track that face. Ultimately our goal is to take as much physical movement off the actor's shoulders as possible, allowing them to focus solely on controlling that "spark of life".

We are also exploring the addition of a virtual representation of the robot to the Guided Performance Interface. Actors have indicated that visual feedback from the robot is helpful, but often their view of the physical robot is blocked. By sending all commands simultaneously to the physical character through RSCS and to a 3D virtual character on the interface running VSCS (described in "Control Software"), the actor would be able to visually verify all of their actions. The GPI has already been used to control the virtual model of Quasi with phenomenal results. This will be investigated further to determine if guests' reactions to the interactive virtual character are the same as their reactions to the physical character. In essence, does physicality matter? Is it possible that interactive virtual characters will open up the door to an entirely new field of Human-Character Interaction?

Conclusions

The Interbots Initiative follows the philosophy of technological transparency: allow technology to support the experience, but not become the experience. Thus, we will not implement a piece of technology if it does not fully support the illusion of life in our animatronic character. Currently, the majority of research in social robotics focuses on autonomous control, a technology that currently makes the development of a believable humanoid personality very difficult. The development of our Guided Performance Interface (GPI) has allowed us to create a hybrid approach that combines the strengths of both autonomous software and human intellect. Low level behaviors (such as idle movements) are controlled by an autonomous system which is augmented by high level processes (such as complex conversation and strong personality) issued by a human actor through the GPI. After observing thousands of interactions between guests and our animatronic character, we strongly feel that both autonomy and guided performance have important roles in interactive, entertainment robotics. Non-autonomous control should not be eliminated from consideration simply because it is seen to be inferior to autonomous control. Together, our autonomous system and Guided Performance Interface allow people to experience extremely rich, believable, social experiences with robots using technology available *today*.

References

Breazeal, C.; Brooks, A.; Gray, J.; Hoffman, G.; Kidd, C.; Lee, H.; Lieberman, J.; Lockerd, A.; and Mulanda, D. 2004. Tutelage and collaboration for humanoid robots. *International Journal of Humanoid Robotics* 1(2):315–348.

Bruce, A.; Nourbakhsh, I.; and Simmons, R. 2002. The role of expressiveness and attention in human-robot interaction. In *Proceedings of the IEEE International Conference on Robotics and Automation. ICRA '02*.

Desai, M., and Yanco, H. A. 2005. Blending human and robot inputs for sliding scale autonomy. In *Proceedings of the 14th IEEE International Workshop on Robot and Human Interactive Communication*.

Haskell, S.; Hosmer, A.; Leu, E.; Stepniewicz, P.; Zayat, S.; Zhu, J.; Patel, S.; and Harger, B. 2005. An extensible platform for interactive, entertaining social experiences with an animatronic character. In *Proceedings of the ACM SIGCHI International Conference on Advances in Computer Entertainment Technology, ACE '05*.

Michaud, F.; Ltourneau, D.; Lepage, P.; Morin, Y.; Gagnon, F.; Gigure, P.; Beaudry, E.; Brosseau, Y.; Ct, C.; Duquette, A.; Laplante, J. F.; Legault, M. A.; Moisan, P.; Ponchon, A.; Raevsky, C.; Roux, M. A.; Salter, T.; Valin, J. M.; Caron, S.; Masson, P.; Kabanza, F.; and Lauria, M. 2005. A brochette of socially interactive robots. In *Proceedings of the Twentieth National Conference on Artificial Intelligence, AAAI '05*.

Pineau, J.; Montemerlo, M.; Pollack, M.; Roy, N.; and Thrun, S. 2003. Towards robotic assistants in nursing homes: Challenges and results. *Robotics and Autonomous Systems: Special issue on Socially Interactive Robots* 42(3–4):271–281.

Scheeff, M. 2000. Experiences with sparky: A social robot. In *Proceedings of the Workshop on Interactive Robotics and Entertainment*.

Uenohara, M.; Ross, W. P.; and Friedman, M. 1991. Torcs: A teleoperated robot control system for the self mobile space manipulator. Technical Report CMU-RI-TR-91-07, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA.

Yu, H.; Spenko, M.; and Dubowsky, S. 2003. An adaptive shared control system for an intelligent mobility aid for the elderly. *Autonomous Robots* 15(1):53–66.