Proceedings of the 17th IEEE International Symposium on Robot and Human Interactive Communication, Technische Universität München, Munich, Germany, August 1-3, 2008

The Design of a Semi-Autonomous Robot Avatar for Family Communication and Education

Jun Ki Lee, Robert Lopez Toscano, Walter Dan Stiehl, and Cynthia Breazeal
Personal Robots Group, MIT Media Lab

Abstract— **Robots as an embodied, multi-modal technology have great potential to be used as a new type of communication device. In this paper we outline our development of the Huggable robot as a semi-autonomous robot avatar for two specific types of remote interaction – family communication and education. Through our discussion we highlight how we have applied six important elements in our system to allow for the robot to function as a richly embodied communication channel.**

I. INTRODUCTION

AS robotic technologies become a part of our lives and enter into our homes and offices, entirely new avenues for communication and interaction emerge. We believe that robotic technologies are the next logical step in the development of new methods of communication. Robots allow for the combination of many different modalities of sensing. A robot can see through its vision system, hear through a set of microphones, and even sense touch through a sensitive skin system. Additionally, the physical embodiment in the real human world allows for a very unique experience, one in which the robot's own illusion of life is greatly compelling to the person interacting with the robot. If the robot is connected to the Internet with appropriate software systems, entirely new forms of communication can emerge with the robot acting as the communication channel between two people. To help illustrate this we propose two examples of improving this communication channel.

A. *Scenarios of Robot Use to Improve Communication*

In the first scenario, the robot is used for social communication between family members who are separated by great distances, such as grandparents who live far from their grandchildren, a child who is sick in a hospital far from family members, or even a parent who is away on a business trip. In the latter scenario, the robot is in the child's home and connected to the Internet through a wireless connection. The parent accesses the child's robot through a secure website that displays a live video feed of the child's face through the eyes of the robot. A live audio feed from the robot's microphones allows the parent to hear his or her child through the ears of the robot. Additionally, other sensor information is displayed to the parent to allow them to understand how the child is interacting with the robot, such as how they are holding the robot or where they are touching it. The parent is also able to remotely control the robot and have a conversation with their child through the embodied robot. This scenario expands upon the strictly teleoperated scenario posed by Brooks in his book Flesh and Machines [1]. In our scenario, the robot is not just fully teleoperated for surveillance, but is a semi-autonomous robot capable of directing attention to objects in the room or even sharing the attention of the child, and thus can be much more engaging for both parent and child. For example, the parent can read a story to the child through the robot, and the robot could point to different images in the page, allowing for both parent and child to share attention.

The second scenario is one focused on education. Here the robot sits next to the child either in a school setting or at home. Both the child and the robot are in front of a computer screen, upon which the educational lesson is presented. A remote teacher, anywhere in the world, teaches the child through the robot via the website interface. A child interacting with a robot in this way may show similar behavior to those children who read to dogs in the Reading with Rover program [2]. This program has shown that children perform better when they read to a dog than to an adult stranger because the child feels less anxiety. Perhaps a smaller, non-threatening robot character may provoke a similar response. Also, because the teacher is in control of the robot at all times, the teacher is able to respond to a child veering from the lesson or asking a spontaneous question. This is different from a fully autonomous robot where state-of-the-art systems are still not capable of dealing with the full complexity of human behavior.

B. *Past Research*

In this section we provide a brief overview of current

Manuscript received February 17, 2008. (Write the date on which you submitted your paper for review.) This work is partially supported by a Microsoft iCampus grant and the MIT Media Lab Things That Think and Digital Life Consortia.

J. K. Lee is a MS student in the Personal Robots Group at the MIT Media Lab, Cambridge, MA, 02139 USA (e-mail: joon@mit.edu phone:1-617-252-5612 fax:1-617-258-6264)

R. L. Toscano is a M.Eng. student in the Personal Robots Group at the MIT Media Lab, Cambridge, MA, 02139 USA (e-mail: rtoscano@mit.edu)

W. D. Stiehl is a PhD student in the Personal Robots Group at the MIT Media Lab, Cambridge, MA, 02139 USA (e-mail: wdstiehl@mit.edu)

C. Breazeal is an Associate Professor of Media Arts and Sciences, the LG Group Career Development Professor, and the Director of the Personal Robots Group at the MIT Media Lab, Cambridge, MA, 02139 USA (e-mail: cynthiab@media.edu)

978-1-4244-2213-5/08/$25.00 ©2008 IEEE

robotic technologies related to communication applications. The core idea for robotics in communication is telepresence. Telepresence (telexistence) consists of technologies enabling a user to feel his/her presence in a different location from his/her true location.

In the following robotic applications, embodiment or movement of the robot play an important role and provides users with novel experiences, which were not present in typical communication devices. Though not traditional robotic devices, the Hug [3] and the Hug Shirt [4] are novel touch communicative devices and worth mentioning for their emphasis on the tactile modality. The RobotPhone RUI [5] allows the user to modify the shape of connected robotic components and motion played with components can be sent via the Internet to other RobotPhones. PRoP [6] added mobility and embodiment to a traditional video conferencing system. It was not only equipped a camera and a lcd screen on a mobile base, but it also provided a pointing rod with two DOFs for simple arm gestures. Two extensive studies on both operability and network effects have been conducted with this system. iRobot's ConnectR [7] is also a mobile platform for video conferencing or surveillance providing visual and auditory feedback to a remote user. The Logitech's Quickcam Orbit AF has two DOFs to rotate its camera. Logitech's webcam also provides a virtual avatar add-on to existing instant messengers which captures the operator's facial gestures and drives the avatar accordingly.

More advanced robot platforms for telepresence and teleoperation include Robonaut [8], Geminoid [9], Sony's AIBO [10], Quasi [11], and Disney Imagineering's Muppet Mobile Labs [12]. Goza et al. had developed a teleoperation system of Robonaut consisting of VR helmet displays, body posture tracking Polhemus^{TM} sensors, and a finger tracking Cyberglove^{TM}. Although, Goza's system provides a full puppeteering system for the robot, it is inappropriate to use in people's homes. Sakamoto et al.'s Geminoid features many interesting aspects in developing a robot as a communication medium. It has a humanoid form, which resembles the operator's appearance and emulates mouth movement that is captured and sent over the Internet. As with Goza's case, Sakamoto's framework is not designed for home use.

While these are all sophisticated robot systems, their use in a communication scenario relies primarily upon only two senses – vision and audio. We believe that the social communication aspects of these systems can be greatly improved by allowing the operator to see more than just vision and audio, but also understand how the robot is being physically touched, held, or interacted with. Additionally, by adding layers of autonomy on top of the traditionally teleoperated robot we can reduce the cognitive load of the operator while improving the overall interaction experience for the user.

II. SIX DESIGN ELEMENTS NEEDED FOR COMMUNICATION ROBOTS

We believe that there are six design elements needed to create communication robots which are clearly and easily operated by a non-expert person while remaining engaging to the user interacting with the robot. First, the robot must feature systems which allow for the remote operator to direct the attention of the user or be capable of responding to the users' own attempts to direct the robot's attention. Second, both the operator and user should be able to share attention easily, i.e. both user and robot can interact with and focus on the same object. These two features play an important role in our family communication and education scenarios. When the user reads a book together with the robot, either the user or the robot may point at a specific figure or sentence in a book. To enable such features, the embodiment aspect of the robot combined with the operator's ability to directly control its arms and head allow the user to recognize where the robot is gazing and/or pointing at.

Third, the robot must provide the operator with real-time multi-modal sensory information for situational awareness. The data must be presented in a clear, easily understood fashion that allows the operator to be immersed in the interaction. This real-time sensor information may include the physical orientation of the robot, where and how the user is touching the robot, and other descriptive information to improve the interactive experience. Fourth, the robot must be controlled in such a way that reduces the cognitive load of the operator while allowing for rich forms of expression (vocalizations, facial expressions, gestures, etc.). Controlling a robot is still a cumbersome task, especially for elders. Many current control interfaces for robots remain difficult to learn and non-intuitive. For these reasons, making the interface as intuitive as possible by alleviating the cognitive load of the operator is crucial.

Fifth, the robot's expressions and behavior must be readable to the user and convey personality to make the interaction fun, engaging, and personal. This may entail supporting the remote operator's ability to convey his or her own personality through the robot avatar, or to control a robot to convey a consistent character (e.g., a robot that is based on a familiar comic book character). This might includes specific content such as sounds, gestures, and other behavioral elements typical of that character. Finally, the interface between operator and the robot must be widely accessible, ideally from anywhere in the world. For instance, a World Wide Web interface would enable family members or educators to interact with the child at great distances. In the remaining sections of this paper, we describe the implementation of our systems that address each of these six design elements.

III. HUGGABLE PLATFORM

For the past three years we have been developing the Huggable robot platform [13]. The Huggable, shown in Figure 1, is a new type of robotic companion designed to function both as a fully autonomous robot as well as a semi-autonomous robot avatar. This paper focuses on the robot avatar mode of operation. In this section, we provide a brief overview of the many hardware and software components of the robot.

A. *Hardware Description*

Underneath its soft plush teddy bear exterior and silicone skin, the Huggable is being designed with a full-body, multi-modal "sensitive skin" [14], two cameras in its eyes – one color and one black and white, a microphone array in its head, an inertial measurement unit in its body [15], a speaker in the mouth, potentiometers to detect joint angle positions, and an embedded PC with wireless networking. The robot has a total of 8 DOFs: a 3 DOF neck (nod, tilt, and rotate), a 2 DOF shoulder motion (up/down and in/out) per arm, and a 1 DOF ear mechanism for expression. The Huggable also uses a hybrid belt-gear mechanical drive system which allows for smooth and quiet motion. Currently, the robot is tethered to a 12V power supply, but ultimately will run under battery power.

B. *Software Description*

Figure 2 shows our current software system implementation for the Huggable platform. The Huggable features a pair of software sub-systems to achieve its interactive behavior. We use the Microsoft Research Robotics Studio (MSRS) to gather data from the various sensors on the robot and process it in real-time. Results of this processing are sent to the other software sub-system: the C5M behavior engine [16]. The behavior engine software processes the sensory data from the user and the operator inputs to make high-level decisions about the robot's behavior, such as where to look or how to move. The custom TCP protocol that MSRS uses, allows us to communicate with computers beyond our local subnet. This is necessary for communication across the Internet between the remote operator and the local user.

IV. APPLICATION OF THE HUGGABLE AS A SEMI-AUTONOMOUS ROBOT AVATAR FOR COMMUNICATION

We have developed several software and hardware based technologies to achieve the design objectives for our interface outlined in Section II. These systems can be divided into technologies that run locally (without the aid of the operator) and technologies that are run remotely (controlled by the operator).

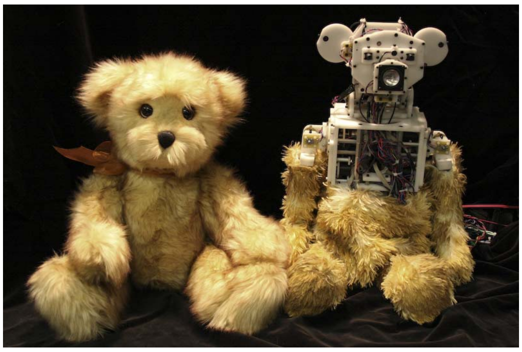


Fig. 1. The Current Huggable V3.0 Prototype (in development) (right) and the Concept Plush (left). In the current prototype only the underlying mechanics of the robot are shown. The sensitive skin system, soft silicone rubber beneath the fur, and final cosmetic fur exterior are not shown in this photo. When fully finished it will look like the concept plush at left.

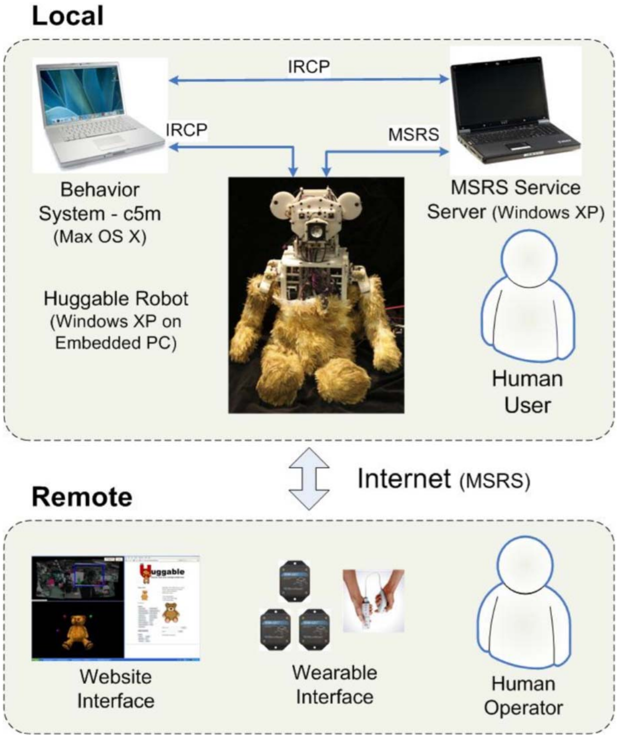


Fig. 2. The Huggable System. On the local side, all the visual, auditory, and somatosensory input from the human user is processed on the embedded PC inside the Huggable and sent to the MSRS service server that visualizes high level sensor information through the web interface. The c5m system receives inputs from the MSRS server, directs the autonomous behavior of the robot, and sends motor control information to the embedded PC. The human operator on the remote side controls the robot via the remote website. The operator may also wear orientation sensors (3DM—GX1) and hold Wii Remotes to control the Huggable via his or her own gestures

A. *Local Technologies*

Face Detection

We use OpenCV to detect upright and frontal faces in a video feed. The robot can use the location of a face in the image to move its head so that the face appears in the center of the video feed, effectively making the robot look-at and track faces.

IMU Stabilization of Video

Figure 3 illustrates our video stabilization software that keeps the video feed to the remote operator in an upright orientation even as the robot is being picked up and rotated. Our multi-modal technology makes use of both the camera feed and the inertial measurement unit (IMU). Every video frame from the video camera of the robot is coupled with the roll position of the robot given by the on-board IMU. The video frame is then rotated by the negative of the roll value. This counter rotation has the effect of keeping objects upright in the video frame instead of being rotated with the robot (which would be very disorienting for the remote operator).

IMU Motion Classification

Whole body gesture classification also runs locally on the robot. Feature based detection algorithms using frequency, jerk, and other relative sensor measures process the data from the IMU in order to recognize interactions such as when the robot is being picked up, bounced, or rocked.

3D Model

A 3D virtual model of the robot is used to convey the body configuration of the robot's limbs. Potentiometers in each of the robot's DOFs record their current position and subsequently update the joints of the virtual model, and vice-versa. By virtue of this coupling, moving a joint on the real robot moves the joint of the virtual one, and vice-versa. This allows for the remote operator to easily see how much range of motion is left on each joint as the robot's limbs move around. In effect, this provides a "virtual mirror" of the Huggable back to the remote operator so that the operator can see how Huggable's behaviors appear to the user interacting locally with the robot.

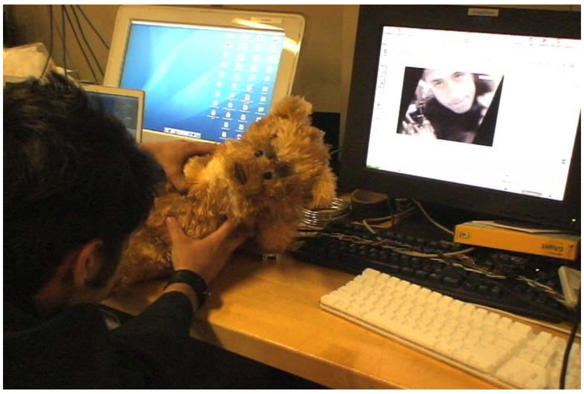


Fig. 3. The IMU Stabilization of the Video Stream. As shown in the figure, the robot is being rolled to its side, yet the image of the face on the screen remains relatively upright because of the counter-rotation. The space where no image data is available (because of the counter-rotation) is filled in with black.

Skin Technology

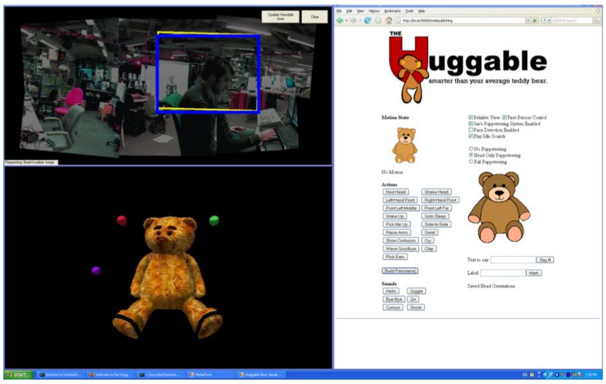
The robot features a sensitive skin technology that makes use of hundreds of sensing elements of three types (pressure, capacitive, and temperature) that shall cover the entire exterior of the robot. We have developed off-line pattern recognition algorithms to classify social-emotional categories of touch (e.g., is the Huggable being tickled, stroked, hugged, grabbed, slapped, etc.). More information on the skin technology can be found in [17]. A real-time version of this technology is currently in development for the 3rd generation Huggable robot.

B. *Remote Puppeteering Technologies*

In this section we describe four systems that suport the remote operator's ability to puppeteer the robot.

Web Interface

Figure 4 depicts the web interface we have developed for the remote operator that enables him or her view the state of robot and evoke its behaviors. The web interface is a combination of a website and an application for streaming audio and video to and from the robot. The website includes a diagram that shows which parts of the robot are currently being moved (via the potentiometer sensors), several buttons to execute different actions (movement and sound), an interface to enter text for the robot to speak, and various check-boxes to toggle on and off several of the aforementioned technologies. There are two video streams, one incoming from the robot's video camera, and another stream from the virtual 3D virtual model of the robot. There is also a small animated cartoon of the Huggable that is used to indicate the whole body movement of the robot (whether it is being picked up, rocked, shaken, or bounced). From this application, the operator can also talk to the user through the Huggable's speaker and listen to the user via the Huggable's microphones.



☑ Relative View ☑ First-Person Control
☑ Jun's Puppeteering System Enabled
☐ Face Detection Enabled
☑ Play Idle Sounds

○ No Puppeteering
● Head Only Puppeteering
○ Full Puppeteering

Motion State: No Motion

Fig. 4. A Screen-Shot of our Current Web Interface. The top left portion is the stale panorama.. The bottom left portion is a video stream of the three dimensional virtual model of the robot. The website on the right half of the screen-shot is the interface that the puppeteer uses to control and receive feedback from the robot.

Face Labeling

We have also created a face labeling technology that allows the operator to choose labels for a stationary person or object in its view. These labels can then be referenced later by returning the robot's head to the position that was stored at the time the label was created. The label is defined by the operator and can be entered through the website interface. These labels can be used to mark the locations of objects as well.

Wearable Interface

The wearable interface of the puppeteering system, shown in Figure 5, consists of a set of motion capture devices that aid in capturing the operator's movement such as his/her body posture and gestures. We have two methods of using this system - direct control and gesture based. As seen in Figure 5, the human operator wears a set of orientation sensor units for the head and the arms and holds a Wii Remote and a Nunchuk on both hands. The operator can switch between the different modes of control by pressing a button on the controller and can put down the handheld controllers to focus on the web interface.

In our gesture recognition based system, a Wii Remote and Nunchuk are connected to a computer via Bluetooth. These controllers transmit a total of six acceleration values (3 axes of information per controller). Data streams from multiple people were recorded and used to train six different continuous Hidden Markov Models (HMM) for each gesture. The six gestures are *arms up*, *arms forward*, *bye-bye* (one arm up and waiving), *crying* (wiggling both hands in front of eyes), *ear flickering* (wiggling both hands above the head), and *clapping* (with both arms forward repeating moving both hands left to right in an opposite direction). Recognized gestures are then sent to the behavior engine to evoke animations.

In the direct motion-capture approach, the human operator wears a set of 3DM-GX1 orientation sensors. Each sensor unit consists of three different types of sensors: accelerometers, magnetometers, and gyrometers all in three axes. They are all used to provide stable orientation data in both static and dynamic conditions. The operator wears these sensor units on his or her head and arms to provide joint angles of the operator's posture. This data is then sent to the C5M system where it is mapped to the Huggable's joints and used to modify the joint angles of the robot.

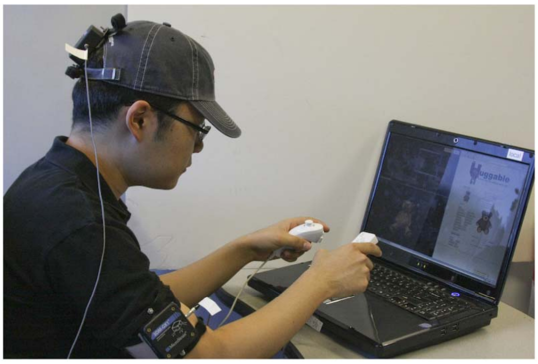


Fig 5. The Wearable Interface. The human operator is holding a Wii Remote and a Nunchuk on both hands and wearing a set of orientation sensors on both arms and the head.

Stale Panorama

One challenge of teleoperation occurs when the robot's video feed presents the operator a relatively narrow field of view (as compared with human peripheral vision). This gives the operator a sense of tunnel vision. To cope with this issue, we have implemented a "stale panorama", shown in Figure 6.

To build the stale panorama, the robot autonomously looks around the room as it captures video frames and stores them with the associated position of the robot's head. The captured frames are then projected to a much larger canvas. The result is a collage of still images that present the remote operator with a panorama of the environment.

The active part of the canvas is the current position of the robot's head, which is a live streaming video feed. The live feed is depicted in Figure 6 as the light colored bounding box.

The darker (blue) bounding box is a target box that can be dragged by the operator to control where the robot should look next. Once, the blue bounding box is positioned, the Huggable computes how it must move its head to align its active video feed with that target box.

In addition, the operator can use the stale panorama to control where the robot points its arm to specify a location in space. The operator simply has to click at a location in the stale panorama, and the robot computes the joint angle positions of the shoulder in order to have the arm point in that direction.

The stale panorama technology is no longer relevant if the scene in the panorama drastically changes (i.e. the robot becomes surrounded by people) or if the robot is picked up and moved. The remote operator can discard the stale panorama when this occurs, and build a new one when the robot is stationary again. The IMU sensor offers us the

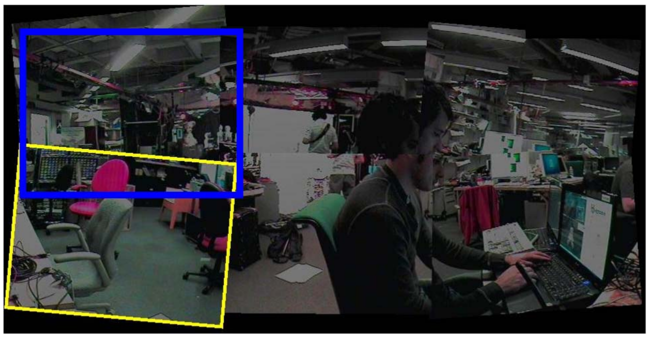


Fig. 6. The Stale Panorama. In this figure, we show the result of the robot building its stale panoramic view of the environment. The bottom box (highlighted in yellow) is where the robot is currently looking and hence not stale, and the top box (in blue) indicates where the puppeteer would like the robot to look at. Once the puppeteer issues the command, the robot's head (and thus the bottom box) move towards the position indicated by the user controller box.

classification of these types of movements to eventually have the Huggable build or disregard the stale panorama on its own.

We also have integrated stitching algorithms to accurately transform captured images so that they align much better and improve the overall clarity of the stale panorama. We use an implementation of the SIFT algorithm [18] to identify scale-invariant keypoints in our frames, and then apply the RANSAC algorithm [19] to find a non-affine transformation matrix. We then use OpenCV's algorithm for warping images to apply the non-affine transforms to the frames in our panorama. Images are collected in the background, and the panorama eventually is constructed and presented to the puppeteer as an asynchronous task.

V. DISCUSSION

In Section II, we outlined six design elements we believe are needed for communication avatar robots. Here we describe how each of the previously described systems can map to one or more of the following six design elements.

A. Directing Attention

The first design element we have addressed is directing attention of the user. This ability is very important in our education or communication scenarios. Many game scenarios for children include pointing and gazing. For example, in a number matching game, a teacher might point to and count two different sets of objects to help children learn the concept of sameness in number.

The remote operator can control the robot's pointing and gazing behavior using two different interfaces. Recall, they can either use the stale panorama or the wearable systems. This allows the operator to chose the control mode that best suits their preference.

B. Sharing Attention

The second design element is to allow the operator to effectively share attention with the user. Sharing attention is crucial to engaging the user and for synchronizing communication. We support this function with the use of our stale panorama technology. The stale panorama allows the operator to quickly change the gaze of the robot—accurately pointing to something in the environment. In addition to the stale panorama, our interface allows the operator to play actions and sounds via the website interface, thus allowing the robot to express forms of acknowledgement and confirmation which also help express sharing of attention. Alternatively, the operator can stream his or her own voice or use the wearable motion capture system if he or she prefers.

C. Real-time Sensor Feedback and Situation Awareness

Our technologies have also made improvements in the area of real-time sensor feedback for situation awareness. One key technology is the virtual 3D model that depicts the state of the robot. The animated image of this classification is shown on the website. This enables the operator to have an idea of how the Huggable is being moved that would otherwise not be conveyed through the 3D Virtual Huggable model or the incoming video feed from the on-board camera of the robot. Another technology that improves on real-time sensor feedback is the use of potentiometers in each of the robot's joints and the visualization of that information in the website interface for the operator. Using this information, the operator is aware of which parts of the robot are being moved. Finally, we also have the ability to display the real time sensitive skin information on the virtual Huggable. We are currently working on adding this visualizer functionality to the website as well.

D. Mitigating Cognitive Load

Some of the technologies that we have developed for the robot help to alleviate the operator's cognitive burden of controlling an 8 DOF robotic avatar.

Again, the stale panorama technology helps with this issue by providing the operator with an easy to use interface for controlling gaze. Our previous implementation of controlling the robot's gaze involved using a game controller. In our own informal internal testing the operator would move the thumb-stick to turn the robot's head, but due to latency and unfamiliarity with this type of input device, operators had a difficult time controlling the robot's gaze. The stale panorama offers a more intuitive interface that controls the robot's gaze at the slide of a viewing window.

Another technology that helps reduce the cognitive load on the operator is the use of IMU data to stabilize the video feed by counter-rotation. This technology helps orient the user in their remote environment, thus helping them understand what they see in the video feed. This type of technology is necessary since the robot is meant to be picked up and carried.

In another scenario, puppeteering the robot becomes tedious if speaking to a user who tends to move around frequently. Keeping the user in the sight of the on-board video camera requires some cognition that might interrupt or at least hinder the intended interaction through the robot. The face detection and look-at technology solves this by automatically centering on the face of someone in the image. If the user moves away from the center of the frame, his or her face is detected and the gaze of the robot turns to center on the user once again. This technology carries the assumption that the user does not leave the video frame completely and that the robot is not already at the edge of its range of motion.

In the scenario where the operator is speaking to more than one user, it would be distracting to move the robot's gaze back and forth from each user. We solve this with our face labeling technology, enabling the operator to quickly move from one face to the other at the click of a button instead of at the movement of a thumbstick or at the sliding of a window.

On the other hand, specific gestures of the robot, such as *bye-bye*, *clapping, and crying*, are evoked and controlled by the wearable puppeteering system as mentioned above. Gestures are captured through sensor devices and used to evoke a set of animations (actions) for the robot. While an operator still controls the robot by directly manipulating each joint, other parts such ears, which cannot be controlled, can be driven using gestures. In addition, the Huggable can also be driven solely by gestures. Thus, it lightens the cognitive load of an operator and enriches the expressiveness of the robot.

E. Conveying Character and Personality

Several features give the robot autonomy and help define the character of the robot. Through the speaker inside the mouth, the robot can play back text-to-speech (TTS) voice outputs and sound effects. It can also play various kinds of animations conveying socially meaningful gestures. Moreover, when the robot is idle, it looks around the room sporadically and plays back a breathing animation, which swings the robot's arms slightly, and slowly raises and lowers its ears. In addition, we provide the operator with the ability to change the idiosyncrasies of the robot by switching between different sets of animations. Although the robot's physical appearance may not change, the change in idiosyncrasies may alter the user's belief of the robot's identity. Alternatively, the operator may choose to stream their own voice and use the wearable motion capture technologies to convey his or her own personality through the Huggable.

F. Global Accessibility

One of our goals with this application was to make sure that the system is accessible via the web. Our website interface helps increase the availability of the puppeteering interface by leveraging current browser technologies. Our use of MSRS allows us to run our software systems across multiple computers using a uniform system. The issue of latency is addressed in our stale panorama technology as well as in other semi-autonomous behaviors such as face auto-centering and the face labeling technology. Because our user base is so diverse, our goal is for our interface to be as accessible as possible.

VI. CONCLUSION AND FUTURE WORK

Robots may be the next natural progression for richly embodied communication technologies. In this paper we have outlined a series of six design elements for a successful communication robot avatar. We have demonstrated how we have implemented each of these elements on the Huggable robot platform and shown some early results of each system. It is important to note that this paper is a purely technical description of our first steps into this realm. Over the next year, we intend to target a specific application in education and customize our interfaces and content to conduct a human subjects study to assess learning efficacy of children with the Huggable.

It is also important to mention that while not discussed in this paper, the use of a communication robot does raise a series of important philosophical questions specifically tied to the issue of identity and privacy. While this topic can easily be the subject of its own paper, it is important to raise a few open questions. As has been shown in the work of Sherry Turkle [20-22], the relationship that people develop with their robots is quite complex. The robot avatar for communication described in this paper also raises some interesting questions. First, if a parent is controlling the robot remotely, does the child understand that the parent is in control, or do they think that their parent is the robot? Different modes of operation will need to be appropriate for children's cognitive and social development. Should we use the parent's own voice or create a standardized voice for the robot that the parent can use to communicate? What is the role of character and autonomy in this relationship for both the child and the parent/teacher? How does one handle issues of privacy? Clearly there are many open research questions about the potential ethical and societal implications that must be carefully studied.

ACKNOWLEDGMENT

The authors would like to thank the other members of the Personal Robots Group at the MIT Media Lab, specifically those undergraduate and graduate students who have assisted with aspects of this work – Allan Maymin, Kristopher Dos Santos, Crystal Chao, and Sigurdur Orn Adalgeirsson. We would also like to specifically thank Daniel Bernhardt of Cambridge University (Graphics and Interaction Group) for his help with the development of the classification system for the IMU and the wearable system for the puppeteering technology. Jun Ki Lee appreciates his support from a Samsung Scholarship. Robert Toscano was funded by a philanthropic gift from D. Steinbrecher. The Huggable has been funded in part by the Media Lab sponsors and Microsoft Research through the iCampus program.

REFERENCES

[1] R. A. Brooks, *Flesh and Machines: How Robots Will Change Us*. New York: Vintage Books, 2002.
[2] "http://www.readingwithrover.org/."
[3] C. DiSalvo, F. Gemperle, J. Forlizzi, and E. Montgomery, "The Hug: An Exploration of Robotic Form for Intimate Communication," in *Ro-Man 2003*, 2003.
[4] CuteCircuit, "The Hug Shirt: http://www.cutecircuit.com/now/projects/wearables/fr-hugs/."

[5] D. Sekiguchi, M. Inami, and S. Tachi, "RobotPhone:RUI for interpersonal communication," in *Extended Abstract of CHI01*, 2001, pp. 277-278.

[6] E. Paulos and J. Canny, "Social Tele-embodiment: Understanding Presence," *Autonomous Robots,* vol. 11, pp. 87-95, 2001.

[7] iRobot, "http://www.irobot.com/sp.cfm?pageid=338."

[8] S. M. Goza, R. O. Ambrose, M. A. Diftler, and I. M. Spain, "Telepresence control of the NASA/DARPA robonaut on a mobility platform," in *Proceedings of the SIGCHI conference on Human factors in computing systems* Vienna, Austria: ACM, 2004, pp. 623-629 %@ 1-58113-702-8.

[9] D. Sakamoto, T. Kanda, T. Ono, H. Ishiguro, and N. Hagita, "Android as a Telecommunication Medium with a Human-Like Presence," in *HRI'07* Arlington, Virginia, USA, 2007, pp. 193-200.

[10] Sony, "Sony AIBO Europe - Official Website - User guides."

[11] Interbots, "http://www.etc.cmu.edu/projects/ibi/."

[12] K. Yoshino, "Disney re-animates theme park with no human in sight," in *The Seattle Times* Seattle, WA, 2007.

[13] W. D. Stiehl, J. Lieberman, C. Breazeal, L. Basel, L. Lalla, and M. Wolf, "Design of a Therapeutic Robotic Companion for Relational, Affective Touch," in *IEEE International Workshop on Robot and Human Interactive Communication (RO-MAN 2005)*, Nashville, TN, 2005.

[14] W. D. Stiehl and C. Breazeal, "A "Sensitive Skin" for Robotic Companions Featuring Temperature, Force, and Electric Field Sensors," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2006)*, Beijing, China, 2006.

[15] S. J. Morris, "A Shoe-Integrated Sensor System for Wireless Gait Analysis and Real-Time Therapeutic Feedback," in *Health Sciences and Technology Sc.D. Thesis* Cambridge: MIT, 2004.

[16] B. Blumberg, M. Downie, Y. Ivanov, M. Berlin, M. P. Johnson, and B. Tomlinson, "Integrated learning for interactive synthetic characters," *Proceedings of the 29th annual conference on Computer graphics and interactive techniques,* pp. 417-426, 2002.

[17] S. Walter Dan and B. Cynthia, "A Sensitive Skin for Robotic Companions Featuring Temperature, Force, and Electric Field Sensors," in *Intelligent Robots and Systems, 2006 IEEE/RSJ International Conference on*, 2006, pp. 1952-1959.

[18] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision,* vol. 60, pp. 91-110, 2004.

[19] M. A. Fischler and R. C. Bolles, "Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography," *Communications of the ACM,* vol. 24, pp. 381-395, 1981.

[20] S. Turkle, *The Second Self: Computers and the Human Spirit 20th Anniversary Edition*. Cambridge, Massachusetts: MIT Press, 2005.

[21] S. Turkle, "Robot as Rorschach: New Complicities for Companionship," in *AAAI Workshop Technical Report WS-06-09*, Boston, MA, 2006, pp. 51-60.

[22] C. Kidd, W. Taggart, and S. Turkle, "A Sociable Robot to Encourage Social Interaction among the Elderly," in *International Conference on Robotics and Automation (ICRA2006)*, 2006.