
Design Considerations of Expressive Bidirectional Telepresence Robots

Ji-Dong Yim

Simon Fraser University
250-13450 102 Ave.
Surrey, BC, Canada V3T 0A3
jdyim@sfu.ca

Chris D. Shaw

Simon Fraser University
250-13450 102 Ave.
Surrey, BC, Canada V3T 0A3
shaw@sfu.ca

Abstract

Telepresence is an emerging market for everyday robotics, while limitations still exist for such robots to be widely used for ordinary people's social communication. In this paper, we present our iterative design approach toward an interactive bidirectional robot intermediaries along with application ideas and design considerations. This study also surveys recent efforts in HCI and HRI that augment multimodal interfaces for computer mediated communication. We conclude by discussing the key lessons we found useful from the system design. The findings for bidirectional telepresence robot interfaces are of: synchronicity, robot's role, intelligence, personalization, and personality construction method.

Keywords

Telepresence, robot-mediated communication, mobile phone, anthropomorphism

ACM Classification Keywords

H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

General Terms

Design, Human Factors, Documentation

Copyright is held by the author/owner(s).
CHI 2011, May 7–12, 2011, Vancouver, BC, Canada.
ACM 978-1-4503-0268-5/11/05.

Introduction

Considering that the personal computer industry once had successful market ancestors (e.g. devices targeted for games and digital printing), it is believed that one or more killer application areas of social robots will be discovered long before the technology advances enough to allow autonomous personal robots around us. Telepresence is an emerging market for everyday robotics. Several companies recently have announced or are already selling a new generation of remote presence systems using mobile robots such as Texai, RP-7i, Tiltr, QB, and Vgo [17]. In most cases, the robots are controllable from afar and capable of transmitting audio-visuals to the operator, in the same way that an unmanned vehicle is used in remote exploration tasks.

For socially interactive robots, especially ones used for bidirectional avatar systems that enable users to telecommunicate to each other, there is a significant difference in the controlling system from tele-operation systems. While the robot interface is the main input/output channels for all users in an avatar-like system configuration, the supervisor of a tele-operation system still has to control the robot through traditional GUIs. A robot system for interpersonal telecommunication must provide a robot-based UI that enables the co-located user to create/edit robot animations, while simultaneously representing the remote user's virtual presence.

A couple of questions arise. In what situation does an interaction technique work well? Which interface modalities are the best for certain applications and what are the alternatives? Our research addresses modality issues of computer mediated communication service. We aim to examine interactive interfaces in the

context of tele-presence artifacts, to explore a new design space of such products, to suggest implementations of such robot systems with possible application areas on inexpensive platforms, and to discuss the implications for user interactions.

In this paper, we present our design approaches toward interactive bidirectional telepresence robots, CALLY and CALLO, along with design implications on characteristics of different kinds of applications. This study also surveys recent efforts in HCI and HRI that augment multimodal interfaces for computer mediated telecommunication. We only provide a brief introduction of the robot's technological aspects in this paper, as we have already described the details of our system specifications, software structure, messaging protocols and robot animation techniques in previous work [33].

Related Work

We see a robot as an interactive tool instead of an artificial intelligent organism. The scope study in this paper thus covers *functionally designed* social robots [7] of which appearance and body movement add value to interpersonal communication, but does not include highly autonomous mobile robots that recognize social cues, make decisions, and learn social skills from human users.

New modalities for Telecommunication

Recent computer-mediated communication tools are augmented with a variety of new services, such as SMS, email, IM (instant messaging), blogs, video call and social networking applications [13][16]. Expressive interaction modalities have not been actively discussed in the field of agent systems. Instead, a number of approaches have been attempted to build interpersonal



Figure 1. CALLY (top; the first generation prototype) and CALLO (bottom; the second prototype).

telecommunication assistants that enhance emotional relationships between remote users, e.g. couples in a long-distance relationship [20].

HCI researchers and designers have suggested expressive and more tangible means of interpersonal communication including emotional icons [24], abstract graphics with animation [6], phonic signals [27], tactile vibrations [31], force feedback [3], and RUI features [26] in zoomorphic or anthropomorphic forms. People are more engaged with a conversational process when they create messages with an interactive user interface [29] and talk to a humanoid robot [25]. Li et al. argue that even a simple robot gesture is able to communicate emotional and semantic content, but knowledge of situational context and facial expressions have much more impact [14].

Robot Animation Techniques

Tele-operation provides a technical basis for interface systems that control remote robots. It has been extensively studied where an unmanned vehicle or a robot plays serious roles for example in a military context and for space exploration [22]. Recent applications show that a wider range of computing devices can now run a robot agent from afar [28][22][11].

Motion tracking techniques, which have a longer history in the film and gaming industries, suggest a convenient interface for robot animations. Timeline based animation techniques support easy editing methods [4]. But some robot platforms do not afford such tracking equipments or large displays [11]. Researchers in HCI and HRI have shown multimodal interface styles such as direct manipulation with/without kinetic memory

[8][23], audio-driven methods [21], vision-based control [15], and the Program by Demonstration [1] with mathematical models as possible methods for incremental refinement [5][9]. Ogawa et al. [21] and Li et al. [15] pointed out an interesting and valuable aspect of robust tracking systems: “quick response and adequate accuracy to the user’s gesture are sometimes more important than precise estimation for avatar-like communication systems.”

Toward Bidirectional Telepresence Robots

Currently available telepresence robots have both strengths and limitations for ordinary person’s social communication. They are meant to deliver the remote operator’s identity to the local users, and are often equipped with a video-conferencing display to render a human face. The abstract and anthropomorphic look-and-feel of such robots has advantages in representing the operator and minimizes the Uncanny Valley problem in part by reducing the local user’s expectations to the robot’s intelligence. The design is also beneficial in terms of commercialization – when a robot needs to resemble the operator’s facial features and expressions, live streaming video will be superior to physically crafted masks. They cost less and are more.

Limitations still exist besides the expense, artificial intelligence and wheel-based mobility for telepresence robots to be widely used. First, there have been only a small number of applications and use scenarios introduced with the robots. While many of the robots are focused on real-time telepresence scenarios, delayed (or asynchronous) telecommunication may be more desirable in some circumstances. Second, the robots mostly depend on verbal and facial cues when

they communicate. Body gestures, especially arm movements are functionally useful and also an important norm for human communication, but not available in existing telepresence robots. Last, the robot interface is inconsistent for the users in different roles. The operator can virtually exist at two different locations by controlling her puppet and allow remote persons to feel her virtual presence throughout the robot avatar. However, as the puppeteer's control unit still consists of conventional GUIs, there is less chance for the remote users to reflect their human affects back to the operator. This type of one-way telepresence fits certain conditions, such as a doctor's round, CEO's meeting, or disabled person's going-out, rather than for everyday situations like ordinary people's social telecommunication.

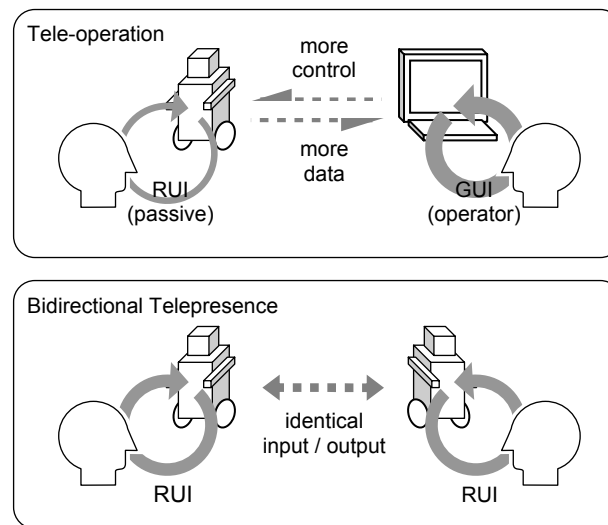


Figure 2. A comparison of interactions between tele-operation (top) and bidirectional telepresence (bottom)

Our robots are designed to address these limitations. As we aim to explore more of believable application scenarios of telepresence robots, our robot prototypes inherit the major advantage of the current system, which is flexible use of flat panel display screens. Our robots are equipped with non-verbal and expressive means of social communication, i.e. anthropomorphic features and body gestures (Figure 1). Regarding the robot interface, we examine two-way interface modalities. We assume that a user would interact with the co-located robot to control the other robot in a remote place, and vice-versa, hence the two telepresence robots should be identical in terms of the interaction scheme as seen in [Figure. 2 (bottom)].

Designing Robots and the Interface

The robot phones, CALLY and CALLO, are prototypes developed in our research. They are designed to use robotic social abilities to add anthropomorphic value to telecommunication services. Each robot consists of a cell phone head and a robot body. The cell phone device shapes the robot's face and acts as the robot's brain. When the robot part receives commands from the phone, it allows the system physical abilities such as spatial mobility and body gestures. This robot configuration is especially beneficial for testing realistic scenarios, as it enables the system to be tested on real world telecommunication networks (e.g. during telephony talk) and to be involved with detailed user interactions (e.g. on phone ringing, a user may pick up, wait, or hang up).

In the following subsections, we introduce our iterative design process along with robot applications we developed at each design phase.



Figure 3. One of the paper masks we used for participatory design session.

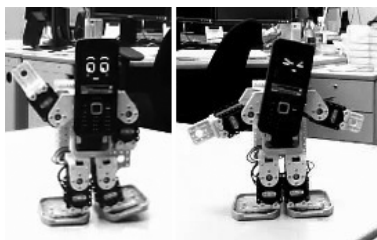


Figure 4. CALLO's call indicator actions; "lover's dance" (left), and "feeling lazy when called from work" (right). The phone application handles telephony routines and controls the robot body.

PHASE 1. Ideation and low-fi prototypes

Our approaches in the first phase were mostly oriented traditional user-centered design methods. Application ideas were generated through brainstorming and detailed by drawing idea sketches. The first prototype, CALLY, was not fully programmed in mobile phones but partly controlled by a PC. The prototype enabled us to examine basic form factors, mobility limitations and robot expressions. However, the primitive robot only allowed us to implement a few number of static functionalities such as facial expressions and simple pre-programmed behaviors in an alarm clock situation.

In spite of the robot's limited functionality, it is confirmed that low-fi prototypes facilitate design collaborations as a participatory design tool. A participatory design method using paper masks was useful for us to describe users' expectations toward the system for each given use context; for example, participants did not press or touch a button on the robot, but just talked to it, say, "Robot, make a call to Greg." or "I'm busy. Ignore the call." [32]

PHASE 2. Simple robots that connect people

One of the lessons we learnt from the previous design exercise was the importance of non-verbal social cues that a robot could possibly utilize by using face expressions and body gestures. So the second generation robot, CALLO, had fewer motors, but equipped with more body gestures. The major improvement of the system was the software structure; CALLO independently runs on its mobile phone brain and became capable of handling telephony services. In the application scenario, the robot was responsible for indicating incoming calls and the caller's identity by

executing different facial/body expressions according to the user's address book information.

PHASE 3. Giving control to users - robot animation interface and gesture messaging

CALLO's call indicator scenario evoked a question: how would one program the robot's gesture animation when setting up the phone book? Considering that people assign different ring-tones for special friends and that some of skillful users make customized fun animations for instant messaging, the question was regarded as

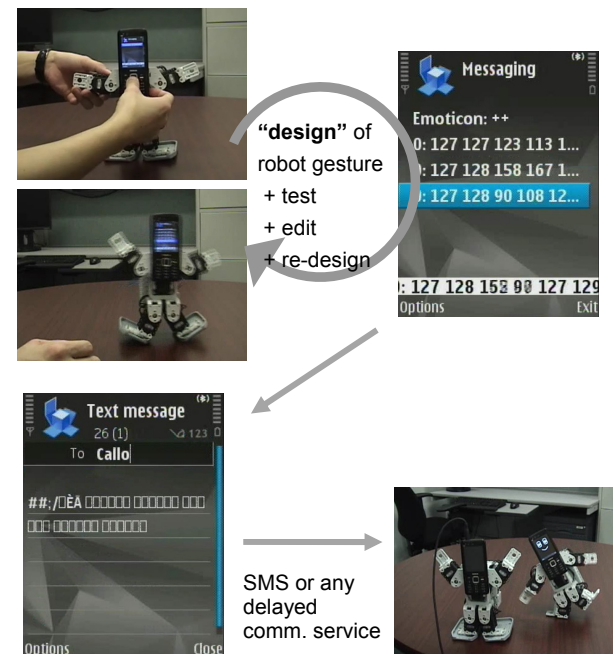


Figure 5. Sending a robot gesture over SMS; a user can design the robot's gesture animation

interesting. However, many traditional animation techniques were not available for cell phone robots, because they often require larger display screens, pointing devices, or tracking equipments.

As a solution for the user to create customized robot gestures, a tangible animation technique – direct manipulation method – was employed in the system. Programming by direct manipulation provided an intuitive interface; the user was able to record robot animations simply by grabbing and moving robot limbs. The interface technique was also advantageous in messaging application scenarios (Figure 5).

PHASE 4. Interface for Real-time Telecommunication

Direct manipulation is an easy-to-learn and robust interface for robot animation tasks. In a synchronous telepresence scenario in which two CALLO robots mediate two remote users through video-call, the interface helped the robots exchange gestural expressions each other [Figure 6]. However, there were significant limitations pointed out by our research team and by the pilot test participants. First, the usability of this sort of interface is mostly determined by the motor system. Some of the pilot subjects reported that they found it hard to move the robot arms and sometimes had their thumb caught in the robot parts. Another problem of direct manipulation method we found more serious is that the interface does not support the human mind handling natural body movements. In other words, a user would not use the interface to control the remote avatar when they are engaged in a video-call, because people do not attentively make gestures when they talk. It is turned out that a direct manipulation interface would work the best for delayed telecommunication scenarios such as SMS, email, and

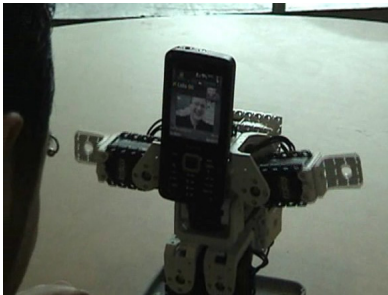


Figure 6. CALLO's video-call; the girl at the other end of the network is controlling the robot

IM, where the users are allowed to “design” the robot gestures to send as they hit the Enter (or Send) key.

PHASE 5. Computer Vision-based Manipulation

In order to address the problems of direct manipulation method, the second animation technique utilized a fast computer vision engine that extracts user's hand and face positions [Figure 7]. The vision engine runs a Gaussian mixture of skin color to distinguish hands and face regions, by using two classification models, for skin and non-skin colors, which are composed of 16 Gaussian kernels respectively based on a Bayesian rule. The vision detector classifies three skin color regions and localizes two hand positions by a face-detection routine.

In this case, the animator controls two motors simultaneously. While it has not statistically been tested yet, vision-based method is expected to be more effective in synchronous communication mode than the direct manipulation method.

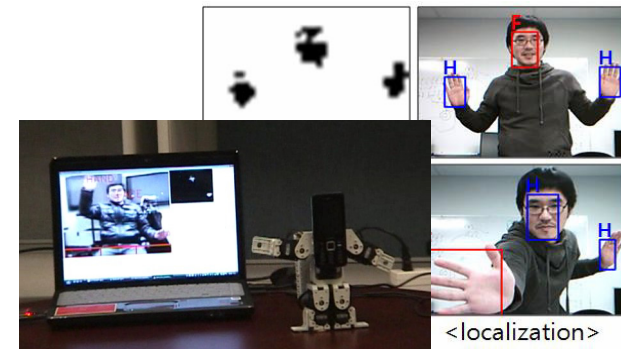


Figure 7. Creating an animation using a live streaming video; the phone's front camera does the work (left); detecting hand regions (right); false result may occur without face detection routine (right-bottom)

Design Implications

In this section, we discuss the implications from our design process and robot implementations. The discussion topics can possibly guide future telepresence application ideas or robot's interface design processes.

Synchronicity of Communication

Synchronicity is a comparison axis that helps us examine the types of existing communication systems along with conversation patterns from synchronous to asynchronous levels [12]. Face-to-face conversation is a form of synchronous communication which involves all the parties present at the same time, where a letter is a traditional form of asynchronous communication which does not involve the participants at the same place or time. Conference talk and word of mouth are in-between forms with multiple parties in delayed times or difference spaces. The communication patterns have evolved through digital age to electronically mediated forms including telephony, video-conferencing, SMS, IM, Twitter, and so forth.

As we briefly discussed in the earlier section, synchronicity determines whether or not an interface modality is feasible for a computer-mediated system, because the cognitive effort required for gesture animation depends on communication patterns. A user can design the content of a message in most asynchronous and text-based communication systems. However, in a synchronous mode, such as face-to-face or telephone talk, the user will heavily rely on habitual mind to quickly respond to the conversation. It is obvious that each of the two human minds is dominantly involved with each interaction process. IM applications take advantages of both communication systems, with which the sender can hold and edit a

message until she hits the Enter key while all the participants of the talk know that they are communicating via a real time system.

CALLO demonstrated both communication models using SMS and video-call. Its gesture animation interfaces worked differently for synchronous, asynchronous or semi-synchronous applications. It is believed that many more RUI modalities will be discussed with synchronicity. Also, we believe that many of traditional communication patterns, such as word of mouth or pre-recorded conference talks, have not been actively explored in the domain of HRI.

Robot's role

Personal assistant (or secretary) robots deal with information to help the master user, to work independently sometimes, or to interact with other people who want to communicate to the robot's owner. For instance, squirrel robots in [18] takes incoming calls when its master is busy, then the robot automatically retrieves caller information, gives options to users, waits for user input, collects audio signals to recognize the master's availability, and decides a proper behavior such as a subtle indication movement, a silent decline, an audio output, a voice mail mode back to the caller, a user choice, etc. If the system advances enough to decide "proper" behaviors, the artificial secretary will be beneficial because it will enable user to release some social burdens in exchange for a bit of personal information.

Avatar (or surrogate, puppet) robots, on the other hand, are a duplicated physical body of the user at a remote place. Responsibilities of avatar robots are to portray the remote operator's presence to a person (or people)

co-located with it, and to give the operator as much information as it can collect from the remote site. This type of system helps communication by giving more abilities to the user. For example, the operator can virtually exist at two different locations by controlling her puppet; allow remote persons to feel his/her virtual presence throughout the robot avatar; and expand his/her sensing abilities to a place a far.

Intelligence

A social robot can be implemented upon varying complexity of intelligence depending on the robot's role, functionality, behaviors, interaction scenarios and many other aspects. Human-robot interaction is sometimes more specifically described by the characteristics of robots that are of social interface, socially receptive, sociable, socially intelligent, or others [7].

Personal assistant robots are meant to have a sophisticated artificial intelligence. The robot's functionality – e.g. awareness of the master's availability, generation of suggestions, and decision making of a proper behavior – can be highly complicated, and should be carefully designed. The avatar robots demands relatively lower level of intelligence to properly work. Although the telepresence robots are still expensive, the technology is near-commercialization. There are possibilities for the avatar robots to be equipped with higher level technologies, e.g. avoiding cliffs, learning languages.

Personalization

The term personalization is defined as a process that changes the functionality, interface, information content, or distinctiveness of a system to increase its personal relevance to an individual [2]. Personalization,

also often described as customization, is initiated by either a system or a user, or sometime both. An agent system deals with information to help work related personalization and lets the user confirm the change. A user may customize a work environment, use stickers, or change ring-tones for different contact groups in his/her cell phone. Such activities not only give a product an identity and sense of life-likeness, but also make a user feel attachment to a product [30] and to a robot [10]. Thus the personalization activities may build long-term intimacy between the product and the user.

A RUI, if it is well designed, can provide the users with great opportunities of personalization. However, it has not been much discussed as a means of personalization, while screen-based pet applications, different looks-and-feels, and functionalities have been long used in the domain. It is also worthwhile discussing personalization issues along with personality traits to address privacy concerns.

The typical design of telepresence robot – a wheel-actuated moving base combined with a display screen – has common advantages in co-located human-robot interaction. It helps navigation, and its look-and-feel minimizes the Uncanny Valley problem [19]. When the robots need to resemble the operator's facial features and expressions, live streaming video can quickly help. It is easy to give the operator's personality to avatar robots – the robots are designed for that use. For the secretary robots, however, personalization provokes many challenging questions (e.g. whose personality should a robot resemble? and in which ways can a robot build personality?). Although the technology is not enough to ask the questions for the time being, those questions can potentially become very serious

design problems which involve user's preference, concerns about privacy, situations awareness, and so forth.

Conclusion

In this paper, we introduced the design approaches we took into consideration to create CALLY and CALLO robots. Based on the assumptions about bidirectionally interactive telepresence robots, a design space of HRI has been explored and discussed. We summarized our findings as design implications, each of which asks further research questions in the development of telepresence robot interfaces. More information about CALLY and CALLO are available at our project archives, <http://cally.iat.sfu.ca/>.

Reference

- [1] Billard, A. 2004. Robot learning from demonstration. *Robotics and Autonomous Systems*, 47(2-3), 65-67.
- [2] Blom, J. 2000. Personalization - A Taxonomy. In CHI '00 extended abstracts (The Hague, The Netherlands.). 313.
- [3] Brave, S., Ishii, H., and Dahley, A. 1998. Tangible interfaces for remote collaboration and communication. In Proceedings of the 1998 ACM conference on Computer supported cooperative work - CSCW '98 (Seattle, WA). p. 169-178
- [4] Breemen, A. V. (2004). Bringing Robots To Life: Applying Principles Of Animation To Robots.
- [5] Calinon, S., and Billard, A. (2007). Incremental learning of gestures by imitation in a humanoid robot. In Proceeding of HRI '07 (p. 255). Arlington, Virginia, USA.
- [6] Fagerberg, P., Ståhl, A., and Höök, K. 2003. Designing Gestures for Affective Input: An Analysis of Shape, Effort and Valence. In Proc. of the 2nd

international conference on mobile ubiquitous multimedia (MUM 2003). Sweden.

- [7] Fong, T. 2003. A survey of socially interactive robots. *Robotics and Autonomous Systems*, 42(3-4), 143-166.
- [8] Frei, P., Su, V., Mikhak, B., and Ishii, H. 2000. Curlybot: Designing a New Class of Computational Toys. In Proc. of CHI '00 (pp. 129-136). The Hague, The Netherlands.
- [9] Gribovskaya, E., and Billard, A. 2008. Combining dynamical systems control and programming by demonstration for teaching discrete bimanual coordination tasks to a humanoid robot. In Proc. of HRI '08. p. 33. Amsterdam, The Netherlands.
- [10] Groom, V., Takayama, L., Ochi, P., and Nass, C. 2009. I am my robot. In Proceedings of HRI '09 (p. 31). La Jolla, CA.
- [11] Gutierrez, R., and Craighead, J. 2009. A native iPhone packbot OCU. In Proc. of HRI '09 (p. 193). La Jolla, CA.
- [12] Karahalios, K. 2000. Communication systems: A comparison along a set of major axes. Self-published web paper. <http://web.media.mit.edu/~kkarahal/generals/communication/>
- [13] King, S., and Forlizzi, J. 2007. Slow messaging. In Proc. of Conf. on Designing pleasurable products and interfaces - DPPI '07 (p. 451). Helsinki, Finland.
- [14] Li, J., Chignell, M., Mizobuchi, S., and Yasumura, M. 2009. Emotions and Messages in Simple Robot Gestures. In Proc. of the 13th Conf. on Human-Computer interaction. Part II: Novel interaction Methods and Techniques (Vol. 5611). p. 10. San Diego, CA: Springer-Verlag, Berlin, Heidelberg.
- [15] Li, R., Taskiran, C., and Danielsen, M. 2007. Head pose tracking and gesture detection using block motion vectors on mobile devices. In Proc. of the 4th international conference on mobile technology,

applications, and systems and the 1st international symposium on Computer human interaction in mobile technology - Mobility '07 (p. 572). Singapore.

[16] Lottridge, D., Masson, N., and Mackay, W. 2009. Sharing Empty Moments: Design for Remote Couples. In Proceedings of CHI '09 (p. 2329). Boston, MA, USA.

[17] Markoff, J. 2010. The Boss Is Robotic, and Rolling Up Behind You. *The New York Times*. (Sep. 4, 2010)

[18] Marti, S., and Schmandt, C. 2005. Physical embodiments for mobile communication agents. In Proceedings of the 18th annual ACM symposium on User interface software and technology - UIST '05 (p. 231). Seattle, WA, USA.

[19] Mori, M. (1970). The Uncanny Valley. In Bukimi no tani [The Uncanny Valley] (Vol. 7, pp. 33-35). Energy.

[20] Mueller, F. '., Vetere, F., Gibbs, M. R., Kjeldskov, J., Pedell, S., and Howard, S. 2005. Hug over a distance. In CHI '05 extended abstracts (pp. 1673-1676). Portland, OR, USA.

[21] Ogawa, H., and Watanabe, T. 2000. InterRobot: a speech driven embodied interaction robot. In Proceedings 9th IEEE International Workshop on Robot and Human Interactive Communication. RO-MAN 2000 (p. 322-327). Osaka, Japan.

[22] Podnar, G., Dolan, J., Elfes, A., Bergerman, M., Brown, H. B., and Guisewite, A. D. 2006. Human telesupervision of a fleet of autonomous robots for safe and efficient space exploration. In Proc. of HRI '06 (p. 325). Salt Lake City, UT.

[23] Raffle, H. S., Parkes, A. J., and Ishii, H. 2004. Topobo: A Constructive Assembly System with Kinetic Memory. In Proceedings of CHI '04 (pp. 647-654). Vienna, Austria.

[24] Rivera, K., Cooke, N. J., and Bauhs, J. A. (1996). The effects of emotional icons on remote communication. In Conference companion on Human factors in computing systems common ground - CHI '96 (pp. 99-100). Vancouver, BC, Canada.

[25] Sakamoto, D., and Ono, T. 2006. Sociality of Robots: Do Robots Construct or Collapse Human Relations? In Proceeding of HRI '06 (p. 355). Salt Lake City, Utah, USA.

[26] Sekiguchi, D., Inami, M., and Tachi, S. 2001. RobotPHONE: RUI for Interpersonal Communication. In CHI '01 extended abstracts (p. 277). Seattle, Washington.

[27] Shirazi, A. S., Alt, F., Schmidt, A., Sarjanoja, A., Hynninen, L., Häkkinen, J., and Holleis, P. 2009. Emotion sharing via self-composed melodies on mobile phones. In Proceedings of the 11th International Conference on Human-Computer Interaction with Mobile Devices and Services - MobileHCI '09 (p. 1). Bonn, Germany.

[28] Squire, P., Trafton, G., and Parasuraman, R. 2006. Human control of multiple unmanned vehicles. In Proceeding of HRI '06 (p. 26). Salt Lake City, Utah, USA.

[29] Sundström, P., Ståhl, A., and Höök, K. 2005. eMoto - Affectively Involving both Body and Mind. In CHI '05 extended abstracts (p. 2005). Portland, OR, USA.

[30] Sung, J., Grinter, R. E., and Christensen, H. I. 2009. "Pimp My Roomba": Designing for personalization. In Proc. of CHI '09. Boston, MA.

[31] Werner, J., Wettach, R., and Hornecker, E. 2008. United-pulse: Feeling Your Partner's Pulse. In Proc. of MobileHCI '08 (p. 535). Amsterdam, The Netherlands.

[32] Yim, J. D. and Shaw, C. 2009. Designing CALLY, a Cell-phone Robot. In Proc. of CHI'09, Boston, MA

[33] Yim, J. D., Chun, S. K., Jung, K. C. and Shaw, C. 2010. Development of Communication Model for Social Robots Based on Mobile Service. In Proc. of 2010 IEEE Intl. Conf. on Social Computing, 57-64.