Behavior Enhanced Instant Messaging Using Desktop Robots

Xing Shusong
School of Software
Nankai University
Tianjin, China
xings@nankai.edu.cn

Zhou Huanyun
School of Software
Nankai University
Tianjin, China
zhouhuanyun@mail.nankai.edu.cn

***Abstract*—This paper presents the development of an application that uses the body movements of desktop robots to enhance online conversations on instant messaging. Acting as the avatars of the remote users, the desktop robots represent instant messages with expressive behaviors to the local users. Robot behaviors are built on a set of action modules and associated with instant messages in the formats of text or emoticons. Behavior representation adds humane features to instant messaging. Responses from the public to the desktop robots indicate that this application might become widespread in the near future.**

***Keywords*—instant messaging, desktop robot, behavior**

I. INTRODUCTION

Instant messaging (IM) allows people to make conversations on the Internet through text, voice, and video in real time. MSN messenger, Yahoo Messenger, AIM, QQ, Skype are popular IM systems. Like the role of non-verbal cues to interpersonal communication, representing instant messages with the behaviors of desktop robots may enhance the expressive effects of online conversations. An IM program called Robim is developed to control the desktop robots that can transform information of conversations into expressive body movements. During interaction with people, desktop robots stand on the desks in front of them and take advantages of the computational and networking capabilities of desktop computers. Comparing to text or image displayed on computer screens, behavior representation is more natural and intuitive to the people. The desktop robots applied to IM conversations are a kind of human machine interfaces that add some humane features to online conversations [1].

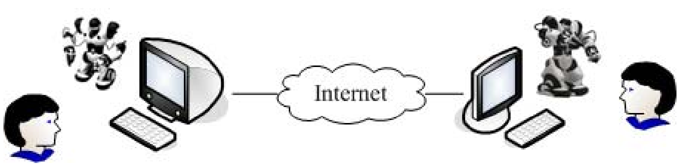


Figure 1. Desktop robots used to enhance instant messaging

The robot on the desk of one side involved in a conversation acts as the physically embodied avatar of the other side (Fig. 1). An avatar is an agent that represents a real person. Different from 3D characters in computers, physically embodied avatars exist in the real world. When a local user input text or emoticons on the user interface of IM software Robim, the text messages or emoticons are transferred on the Internet and transformed into a series of actions on the desktop robot in front of the remoter user.

Figure 2. Behavior representation of desktop robot

The desktop robots used here is a kind of toys whose retail price is less than one hundred dollars. People can describe their emotional states or online status to the remote users by dictating the movements of the robots on the desk of the remote users (Fig. 2).

Robim is a client-side program that implements IM communication using MSN protocol and controls the behaviors of the desktop robots. It uses the services of user authentication and data transfer provided by the MSN servers. A library of expressive behaviors and emoticons is designed for the desktop robots and stored in Robim. The software does not enforce that every side of the same conversation has one robot respectively.

II. BACKGROUND

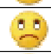
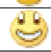
It is estimated that the population of IM users is on the order of 100 millions and continue to increase rapidly [2]. IM systems commonly take client-server architecture: the client side program has a user interface for input and output of conversational contents, while the centralized servers provide services of user authentication and data transfer. People use the client-side program to makes peer-to-peer communication in which both the local user and the remote user have the same privileges.

Emoticons, or smileys, are widely used by IM software to compensate for the limited expressiveness of text messages,

978-1-4244-1676-9/08 /$25.00 ©2008 IEEE RAM 2008

especially in conveying affective cues [3][4]. They take the forms of short sequences of characters or emotional icons that often emulate facial expression or body gestures. With emoticons, people can remember the meanings of body movements. Several emotions whose meanings are easily understandable according to the protocol that has been accepted by the people are listed in the following table (Table 1).

TABLE I. SEVERAL COMMON EMOTICONS

Characters	Icon	Explanation
:-)		Happy, sarcasm
:-(		Sad
:-D		Laughing, very happy
:-O		Surprise
(Y)		Yes
(N)		No

People rely heavily on nonverbal cues to express their meanings and to interpret the communicative activities of others in specific contexts. There is no way to prevent body movements from delivering information during interpersonal communication. Body movements have powerful impact to interpersonal communication because they give cues of omnipresence, form a kind of universal language, provide trusted information, enhance understanding, and sometime cause misunderstanding [5].

In order to make interaction with computers easy and comfortable, people add multimodal communication capabilities, such as speech, intonation, facial expressions, and gestures to them [6]. Human beings have a trend to treat computers and other forms of media in the same way of communicating with other people [7]. Desktop robots having appearances and behaviors similar to human may be used as the avatars of network users. They allow people to treat them as human partners to certain extent. The communications between humans and machines can be performed in a natural and intuitive manner, and humans are no longer constrained to windows, icons, monitors, and pointers as the human machine interfaces.

By applying natural communication modalities to interaction with humans, the robots embed into both the physical space and the social space where the humans live. It is the use of socially acceptable communication in the robots that helps breaking down the barrier between the digital world and the human society, and probably makes the robots become the universal interfaces [1].

A list of robots designed for the purpose of social interaction with human can be found in [8]. To represent the information from digital world composed of networks and virtual environments, three robot marionettes with a rich collection of behaviors were built [9]. A system using two face robots to make face to face interaction between two persons on network is described in [10]. An android that emulates the appearance and gestures of a human in high accuracy is made to study the effects of telepresence [11]. Besides virtual reality, another potential application of haptic devices is to transfer greeting information on IM [12].

III. BEHAVIOR DESIGN

A. *Message classification*

To satisfy people's requirements of online conversations, the messages that can be represented with robot behaviors are grouped into five categories in current implementation: affection, presence, initiation, spatial indication, daily activities. They are organized into tree structure (Fig. 3). The tree leaves are text messages or emoticons corresponding to the robot behaviors.

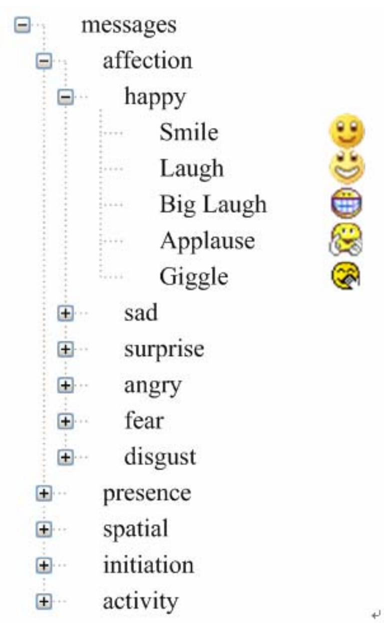


Figure 3. Message classification

The affective messages use keywords and emoticons to describe six emotional states: happiness, anger, disgust, fear, surprise and sadness. Several behaviors are designed to inform people the online status of the remote users. The presence states that can be represented with messages and the corresponding behaviors include login, leave, busy, salutation, etc. The initiation messages include the requests sent to the remote users in order to start a conversation or an activity. The spatial indication messages express the concepts related to the spatial position or direction, like forward, up, there, far away, etc. Daily activities include those activities people do everyday, such as walk, eat, sleep, dance, etc.

People may have requirements to messages and behaviors that can express more specific contents. The information involved in this category is general, while topics of conversations may be specific. For example, people may talk about the feeling of a shopping, a movie, or a party. Extensions to message category and robot behaviors are straightforward under the hierarchical organization of the tree structure.

B. *Emoticons*

One advantage of emoticons is that the meanings they deliver through visual cues are easy for people to grasp. Major IM programs provide many emoticons for expressing affection.

Besides the emoticons of MSN messenger, more emoticons are designed in Robim to represent presence, spatial indication, initiation, and daily activities using the methods of keyframe, metaphor, and definition.

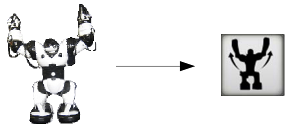
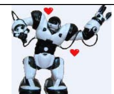


Figure 4. Emoticon created from body movement

Emoticons describing gestures and body movements are generated by making the keyframes of them into icons (Fig. 4). The keyframes are similar to the snapshots that reflect the identifiable features of unique gestures.

TABLE II. SAMPLE ANIMATED EMOTICONS

Icon	Explanation
	Love, like
	Sorry
	Angry
	No

Animated emoticons made of key frames in motion sequences allow the users to preview the body movements of the robots. The animated emoticons are active only when they get input focus from the users, otherwise the flashing images may distract the attention of people. Several animated emoticons that depict the body movements of the robots are listed in (Table 2).

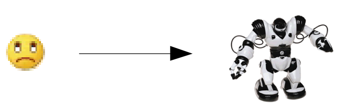


Figure 5. Emotion represented with body movement

Like facial expressions, body movements can also express emotions. Some emoticons are designed according to the visual features of facial expressions. Those emoticons having facial expression on them are mapped to body movements that represent the similar meanings of the facial expressions. For example, the gesture of holding two arms downwards is visually similar to the curved mouth on the Sad icon (Fig. 5). Both the Sad icon and the gesture can be used to express the emotion of sadness.

C. *Motion sequences*

It is expected that the desktop robot has a rich collection of behaviors for the expression tasks. Robot behaviors are built on a set of basic behavior modules and a multilevel behavior representation that organizes the behavior modules into both diversified and complex behaviors. Twenty-five actions from all the movements the toy robots can do are selected as basic behavior modules. Some of them control the movements of single joints, such as the action of rising left arm up. Some generate full body movements, such as the action of walking. Besides the mechanical structure of the toy robots, the preprogrammed behaviors modules form another layer of constraints that must be observed by the behavior design method. Complex behaviors are defined by composing the basic behavior modules into motion sequences. The multilevel structure of behavior representation provides portability and flexibility to behavioral design for the desktop robots with a wide range of mechanical structures.

D. *Behavior design*

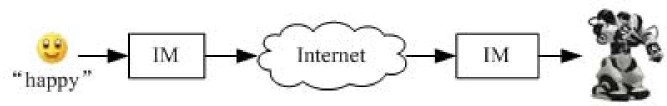


Figure 6. Smile behavior

The IM program Robim keeps a library of behaviors to drive the desktop robots. A behavior is simply defined as the response to a specific stimulus [13]. An emoticon or text message sent out by the remote user will cause the desktop robot in front of the local user to show the corresponding actions (Fig. 6). If the event of sending out an emotion or text message is considered to be a stimulus, then the actions performed by the desktop robot are the responses to it. The whole process from emotion or text message to robot body movements forms a behavior.

The emoticons and robot actions are associated by the designer manually. The actions are designed and examined carefully by the designer. Due to the nonlinear characteristics of robot kinematics, simply combining two actions may produce invalid actions. Those actions whose meanings could not be interpreted by the users are discarded.

The same emotional status can be expressed with different actions. For example, five actions are designed to express happiness: smile, laugh, big laugh, applause, and giggle. Happiness represented with Smile emoticon can be expressed on the robots by the motion sequence of rising two arms and lifting them upwards for seconds. Waving both arms up and down for seconds represents the level of happiness that is equal to laugh. Waving both arms up and down and walk back and forth represents the level of very happy described with the Big laugh emoticon.

The body gestures applied to interpersonal communication are interpreted within a context according to a protocol implicitly accepted by the people involved in a conversation [14]. Robot behaviors designed as the counterparts of human may be more readable to people. To make the robot actions readable when they are performed before the users, the behaviors are designed to have some features of human activities. Many activities of human are so subtle that the robots can not perform. Either designing behaviors manually or learning from human has limited capabilities on creating human readable behaviors on the robots [9].

IV. DESKTOP ROBOT

Figure 7. Toy robot with infrared remote control

The physically embedded avatars used in the study of behavior-enhanced instant messaging are RoboSapiens (Fig. 7), a kind of toy robots for less than one hundred dollars. Toy robots are suitable for staying and acting on the desks. This robot was designed by robotics expert Mark Tilden [15] to a biomorphic structure and an artful appearance. Alone with the affordable price these factors have led to a sales record of several million items. The robot has a rather simple mechanical structure with seven DOFs, which are enough for it to produce a rich set of expressive gestures and behaviors for the task of enhancing IM communication. Both the right and left arms can be controlled separately but the position and speed of each joint can not be specified individually. Up to 67 gestures and activities are preprogrammed into the onboard circuit in the chest of the robots, such as lifting right arm, going forwards, grasping, throwing, kicking, sweeping, picking up, dancing, whistling, belching, etc.

The toy robots are so simple that they do not have sensors installed to detect the responses of the people on the vicinity. Originally, the movements of the robots are made to be specified with infrared remote controls. Each button on the handheld controls is correspondent to a gesture or activity of the robot. The remote controls send motion instructions encoded with modulated infrared signals unidirectionally to the robots without feedback when the buttons are pressed. When receiving the infrared signals, the onboard circuits of the robot decode the signals and make the robot perform the specified activities. To let a robot make a sequence of activities, the user has to press the buttons one by one.

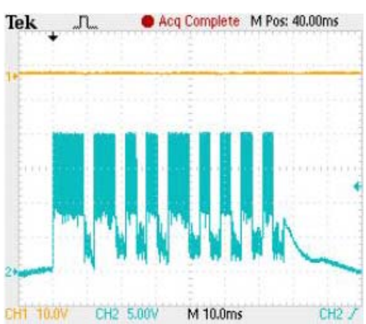


Figure 8. Infrared waveform from the remote control for robot to strike with right arm

In order to act according to IM conversations and generate more behaviors, the robots have to be controlled directly from a PC rather than the remote controls in people's hands. An infrared transceiver with RS-232 interface is connected to the PC on the serial port to replace the remote control and a driver is written to emulate the waveforms sent out from the remote control (Fig. 8). The waveforms from the remote control that can produce all the 67 gestures and activities on the robots are recorded and reproduced by the RS-232 infrared transceiver (Fig. 9). Now the robot can be controlled from PC without modifying the onboard circuit if no single joint was required to be controlled directly. More activities may be built on these pre-programmed moves by composing them into sequences. The communication distance between the robot and infrared transceiver is short because the infrared signals from the transceiver are not modulated and the power supply is not strong.

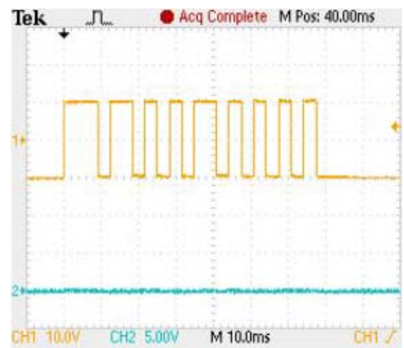


Figure 9. Emulated waveform on RS-232 transceiver for robot to strike with right arm

V. SOFTWARE IMPLEMENTATION

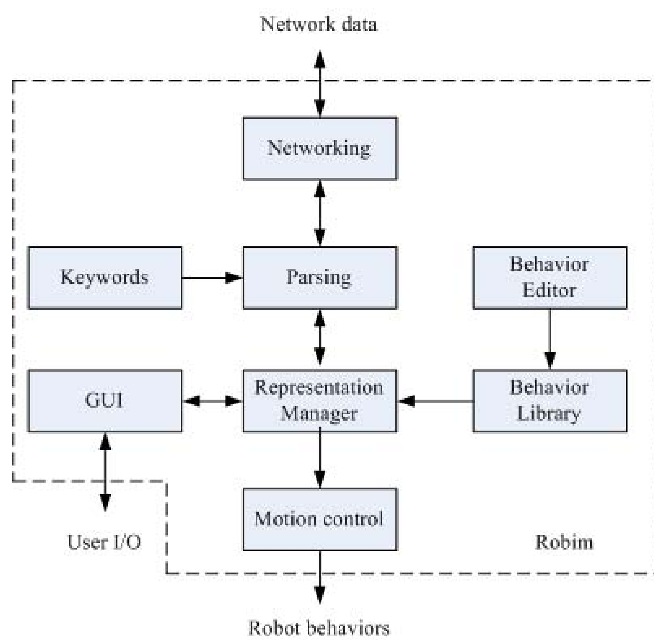


Figure 10. Software composition

Like other IM software, Robim also takes client server architecture. Major developmental work was done on the client side. The client side program consists of eight modules to accomplish the functions including network communication, message parsing, user input and output, robot motion control, behavior editing, storing, and sharing (Fig. 10). The program was developed in C++ and C# with Microsoft Visual Studio 2005.

The networking module implements MSN protocol that IM client programs use to communicate with each other. It packages the conversational contents into network data stream. Besides communication on the Internet, it also has the capability to connect the clients on local area networks where the proprietary servers are setup. It is convenient for students to talk on IM within campus networks isolated from the Internet by firewalls. Besides the common functions of user identity authentication, contact list, data transferring, the servers also store a list of robot behaviors for sharing among users.

The robot motion control module performs low-level motion control to the desktop robots. A list of the infrared waveforms for activating the pre-programmed moves that the robots can perform is coded and saved in a plain text file that records the robot structure. The robot control module reads in the waveforms from the file and writes them to the serial port. The transceiver transforms the waveforms into infrared signals that drive the robots to produce expressive body movements.

The message parsing module examines the received messages and searches for keywords or emoticons that are manually compiled in the tree structure.

The representation manager synchronizes the output of text and image contents to the screen, voices to the sound card, and behaviors to the robot. The text messages and emoticons sent to the users are replaced by robot actions according to the mapping relationships defined in the behavior library.

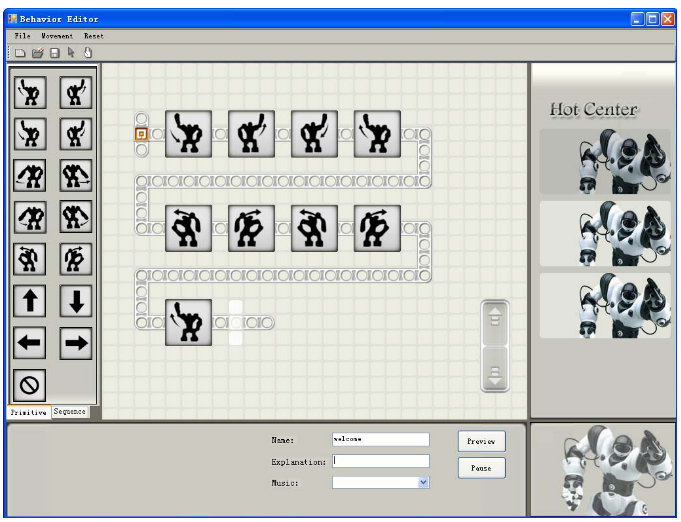


Figure 11. Motion sequence designed with Behavior Editor

Behavior Editor is designed for common users to create customized behaviors (Fig. 11). To minimize the dependency on the specific structure of the robots, this module is designed to have a graphic user interface for operating the available activities but not single joint of the robots in a visualized manner. The parameters for specifying body movements are encapsulated into the behavior modules that are visually represented with emoticons. Each action from the set of behavior modules is represented with an icon and is associated with the constraints that prevent the users from making contradicting combinations with other actions. For example, a walk-forward action may happen with an arm-lifting action at the same time while it conflicts with a walk-backward action if a user tries to make them work simultaneously.

With behavior editor, users can design new moves to satisfy their personalized requirements by combining the prepared activities from the set of both pre-programmed action modules and activities designed by others into motion sequences. Without prior knowledge about robot motion control, the users can compose new behaviors by operating the pick-and-place components.

In current behavior library, robots' activities are manually associated with text and emoticons stored in tree structure by the users. They are synchronized with short pieces of music and intonations.

The software will not increase the workload of networks. Different from industrial applications, network delay is not a critical factor to behavior representation.

VI. EVELUATION

A conversation between two friends on Robim may happen like this (Fig. 12):

Local user says: Hi

Remote user says: Oh hi! How's it going?

Local user says: Its going great

Local user says: How about your trip to Chengdu?

Remote user says: Terrific

Local user says: Show me the photos

Remote user says: At once

Figure 12. Sample conversation

There are totally five behaviors shown on the robots of both sides during the talk. When the remoter user logs in, a Wakeup behavior is activated on the desk of the local user to indicate the presence of the remote user. The local user describes the mood to the remote user by sending out a Smile emoticon which produces a Smile behavior of lifting up both arms on the robot at the remote side. The word 'Terrific' from the remote user is transformed into a Dance behavior on the local user's robot. The local user expresses the eagerness of seeing the photos by sending out an initiation emoticon. The remote user informs the ongoing action of transferring photos to the local user with the Giving behavior. The robot behaviors accompanying the text messages and images are visually observable and hence enrich the expressiveness of the conversation.

The software and robots were exhibited on the Challenge Cup Competition attended by the students from hundreds of universities. The booth was always crowded by the visitors and 1500 pieces of brochures were sent out within two days. Most visitors expressed their interests to this application and the expectation of having a robot on their desks. Many visitors who have tried the software answered the questionnaire and proposed their suggestions.

The questionnaire consists of a series of questions such as: do you use MSN, Yahoo Messenger, QQ, AIM to talk with others? Do you like to have a robot on your desk? To whom will you use the software and robots to talk with? Would it be helpful to your conversations? Are the meanings of the robot behaviors are clear enough? How do you attracted by the robots? And other questions following the replies of the visitors.

The feedback from visitors indicates that most students prefer QQ than MSN and Yahoo Messenger. One of the reasons for many persons to select QQ to chat online is that their friends use QQ. QQ has an attracting user interface and many emoticons while MSN looks a little conservative. The size of the toy robots is too big for it to be accommodated on the desk. And its appearance should be made more engaging. Male students accept the muscular character while female students prefer cute ones. They also worry about the diversity of behaviors. If there is only a small set of behaviors available for information representation, after the first fever the users may feel boring with repeated movements.

To have better readability, the behaviors of the robots need not to replicate human behaviors precisely. People can understand the conversations in the context constructed by synchronized text messages, emoticons, and robot behaviors. People are not sensitive to every action of the robots. The meanings of some robot actions are ambiguous to people who use Robim for the first time. If the users feel confused with an action, commonly they will enquiry the remote users again in text messages for an explanation. The information represented with robot actions is understood within the context of the conversations. The meanings of the actions depend on the formation of a culture or protocol which is waiting to be setup by the designers. The designers have the freedom to select from a set of actions to express the same messages. For example, many actions can be used to express the situation that a user is online again, the Wakeup action of rising two arms seems more readable. Overall, more work is required to make the robots a usable and comfortable human machine interface.

VII. Conclusion

The preliminary work introduced in this paper focuses on implementation of fundamental aspects of the IM software Robim that can enhance the expressive effects of instant messaging with body movements of desktop robots. The software allows people to use the robots as their avatars in delivering their emotional or status information during online conversations to the remote users. New behaviors can be designed by composing the modules from a set of existing moves. Behavior representation with desktop robots has advantages over plain text, voice, and video in representing online status, special, daily activities. Their vivid actions make online conversation nature and intuitive. The affordable price of toy robots and the responses from the public indicate that this application might become widespread in the next few years. Several problems that became apparent during the progress of software development will be addressed in the future.

From the very beginning, it has been realized that there was no standard in the robot world. It is expected that the software was compatible to the robots with different mechanical structures, electrical components, and control methods. As a way to increasing the portability of the software to robots with different numbers of degrees of freedom, the behaviors should be represented in higher abstraction. Thus make the action design be done on behavior modules but not on individual robot joints. To control desktop robots from PCs, it would be better to use radio frequency in wireless communication other than infrared signals which are highly directional. The toy robot is so simple that it does not have a controllable head. Many face robots have been built to convey emotional cues [16]. Having a simpler protocol than Bluetooth, Zigbee could be applied to wireless communication between desktop robots and PCs.

References

[1] C. Breazeal, Designing Sociable Robots. Cambridge, MA: MIT Press, 2002.

[2] R.B. Jennings et al., "A study of Internet instant messaging and chat protocols," IEEE Network, vol. 20, issue 4, pp. 16-21, July-Aug. 2006.

[3] A. K. Amin, "SenseMS: a user-centered approach to enrich the messaging experience for teens by non-verbal means", Proceedings of the 7th International Conference on Human Computer Interaction with Mobile Devices & Services, pp.161-166, Salzburg, Austria, 2005

[4] J. A. Sánchez, etc., "Conveying mood and emotion in instant messaging by using a two-dimensional model for affective states". Proceedings of VII Brazilian Symposium on Human Factors in Computing Systems, pp.66-72, Natal, RN, Brazil, 1996.

[5] J. K. Burgoon, D. B. Buller, W. G. Woodall, Nonverbal Communication: the Unspoken Dialogue, Harper and Row Publishers, New York, 1989.

[6] J. Cassell, J. Sullivan, S. Prevost, E. Churchill, Emobodied Conversational Agents. Cambridge, MA: MIT Press, 2000.

[7] B. Reeves, C. Nass, The Media Equation: How people treat computers, television, and new media like real people and places, Cambridge University Press, Cambridge, UK, 1996.

[8] T. Fong, I. Nourbakhsh, K. Dautenhahn, "A survey of socially interactive robots", Robotics and Autonomous Systems, Special Issue on Socially Interactive Robots, vol. 42, no. 3-4, pp.1-28, 2003.

[9] I. M. Chen et al., "Many strings attached: from conventional to manipulation", IEEE Robotics & Automation Magazine, vol. 12, pp. 59-74, March 2005.

[10] J. Ido et al., "Robotic telecommunication system based on facial information measurement", Proceedings of the 12th International Conference on Intelligent User Interfaces, pp.266-269, Honolulu, Hawaii, January 2007.

[11] D. Sakamoto, "Android as a telecommunication medium with a human-like presence", Proceeding of the ACM/IEEE International Conference on Human-Robot Interaction, pp.193-200, Arlington, Virginia, March 2007.

[12] A.F. Rovers, H.A. van Essen, "HIM: a framework for haptic instant messaging" CHI '04 Extended Abstracts on Human Factors in Computing Systems, pp. 1313-1316, Vienna, Austria, 2004.

[13] R. C. Arkin, Behavior-based Robotics. Cambridge, MA: MIT Press, 1998.

[14] M. Coulson, "Attributing emotion to static body postures: recognition accuracy, confusions, and viewpoint dependence", Journal of Nonverbal Behavior, v.28 no.2, pp.117-139, 2004.

[15] Online contents, www.woowee.com

[16] A. V. Breemen, "Animation engine for believable interactive user-interface robots", Proceedings of 2004 IEEE/RSJ International Conference on Intelligent Robots and System, pp. 2873-2879, Sendai, Japan, Sept. 2004.