Between Real-World and Virtual Agents: The Disembodied Robot

Thibault Voisin
Keio University
3-14-1 Hiyoshi, Kohoku-ku
Yokohama, Japan
tibo@ayu.ics.keio.ac.jp

Hirotaka Osawa
Japan Science and
Technology Agency
4-1-8, Honcho
Kawaguchi, Japan
osawa@nii.ac.jp

Seiji Yamada
2-1-2 Hitotsubashi,
Chiyoda-ku
Tokyo, Japan
seiji@nii.ac.jp

Michita Imai
Keio University
3-14-1 Hiyoshi, Kohoku-ku
Yokohama, Japan
michita@ayu.ics.keio.ac.jp

ABSTRACT

In this study, we propose a disembodied real-world agent and the study of the influence of this disembodiment on the social separation between the user and the agent. In order to give a clue to the user about the presence of the robot and to make possible a visual feedback, we decide to use independent robotic body parts that mimic human hands and eyes. This robot is also able to share real-world space with the user, and react to his presence, through 3d detection and oral communication. Thus, we can obtain an agent with an important presence while keeping good space efficiency, and as a result ban any existing social barrier.

Categories and Subject Descriptors

I.2.9 [**Robotics**]: Operator Interface

General Terms

Design

Keywords

Human-Agent Interaction, Disembodiment, Social Presence, Real-World Agent, Virtual Agent

1. INTRODUCTION

Real-world agents like robots are able to create a social interaction thanks to their body presence and ability to perceive user's actions, when compared to the virtual screen agents. For example, Wainer et al. [8] proposed to study the differences in the perception by the user between virtual and real-world agents. As a matter of facts, the users enjoyed the most embodied agents, and considered them as the ones that were 'watching them the most closely'. A second study, proposed by Lee et al. [6], enforced these results, but showed also that physical embodiment without touch-input capability causes negative effects.

HRI'11, March 6–9, 2011, Lausanne, Switzerland.
ACM 978-1-4503-0561-7/11/03.

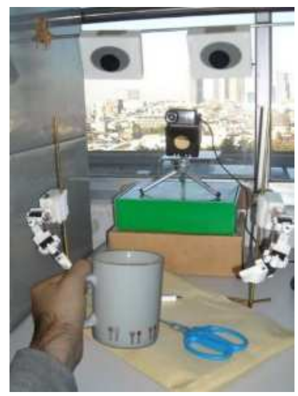

Figure 1: The Disembodied Real-World Agent

These studies clearly showed the advantage of embodied robots, that the physical embodiment enforced the social presence of the robot, and that the effects of physical embodiment could become highly positive when users are able to fully interact with the agent by touching it. However, those real-world agents are not widely used either though, because of their generally high space occupancy, leading to a high spatial cost.

In this paper, we thus present a disembodied agent in the real world, as shown in figure 1, and propose to study the influence of the disembodiment of a real-world agent on its social presence. The agent we consider here consists in body parts that mimic human hands and eyes, in order to give to the user a visual feedback, and is able to react to human contact through 3d detection, thus giving the possibility for the user to interact through touching with the agent.

On the other hand, we allow the user to define the contours of the agent's virtual body: this space will be considered by the agent as being its body. By defining this space, we thus keep the advantage of having a body, while keeping a low spatial cost.

This system has the advantages of real-world robots and virtual agents, thus guarantying a high social presence without sacrificing to the space efficiency.

2. MODEL

In order for the user to get a visual feedback from the disembodied robot, we decided to use the anthropomorphized virtual body parts [7]. These body parts will perform moves in answer to the user's behaviors: the eyes follow the user's hand, and the arms will move accordingly, as a real-world agent would do. An oral feedback will also be performed.

To recall to the user the presence of the agent in the real world, we give to him the possibility to interact with the 'body' of the agent : the user will be able to define the contours of this body in the initialization phase, and to modify it later if he desires, simply by 'pushing' on the agent's body. To reinforce this presence, we chose to define two different spaces: 'inside' and 'outside' the agent's body. These definitions will be used to determine object's position relatively to the agent's body, and will enforce the social presence of the agent as perceived by the user.

3. IMPLEMENTATION

The implementation of this project is mainly based on different parallel data analysis: the detection of the hand of the user, the answer to user's commands, and the treatment of these different data in order for the agent to decide which action to perform.

3.1 Synchronization of the cameras

The detection of the hand of the user is based on the use of two cameras: a SwissRanger SR4000 [3] for the 3d acquisition and a Logitech C910 [1] for the hand's recognition. It was in that case necessary to use two cameras, because the 3d camera acquisition is carried out in gray levels, and its resolution was too limited to perform precise hand detection. From the position of the hand on the color camera, we determine the one on the 3d camera, in order to associate a depth to characteristic dots of the hand (top of the fingers, and palm). The synchronization of the cameras is realized thanks to the chessboard calibration provided by OpenCV.

3.1.1 Hand Detection

To perform the detection of user's hand thanks to the color camera, we use in a first time skin color detection, based on the study of Chai et al. [5] to extract the skin color pixels of the image. We then apply a contour detection on the result, and retrieve the contour with the biggest area, from which we retrieve all convexity defects, in order to determine the number of displayed fingers. This feature will then be use for the interactions with the agent's body, like the initialization of its body, or its resizing.

3.1.2 Running Phase

In order to determine the actions the user wants to perform, and how to interpret his moves, the agent needs some additional information. This data are provided through keyboard input now, but will be replaced by oral commands in the future, in order to ease the communication with the agent. The agent can actually understand short sentences, as to trigger the initialization, to resize its body, or to get information about an element added to its body. The robot is also able to follow the face of the user, interpret his expression and react accordingly, through the use of the faceAPI library [2].

At any time, since the agent's body is defined, the user is able to interact directly with the agent by touching its 'body': the agent will react differently depending on the position of the user's hand relatively to its body. The possibilities to scratch and pat the agent's body particularly enforce its social presence as perceived by the user.

4. EVALUATION

During the initialization phase, the agent was able to perceive the user's hand position and set the actual body limit in .8s. This response time is close enough to a human reaction time (between .1s and .4s) [4] to obtain a feeling of spontaneity from the agent. The agent's body resizing suffers also a light 1.2s delay between the end of the user's move and the end of the resizing, which is necessary in order to be sure that this position is the one chosen by the user.

5. CONLUSION AND FUTURE WORKS

In this study, we thus proposed to develop a disembodied agent. It possesses both advantages from real-world agents and virtual ones, gaining space efficiency without losing in social presence.

The next step is now to conduct an experiment to analyze the differences in the user's perception between this disembodied agent, real-world and virtual ones. Also, additional features are still to be implemented, like the understanding of the user's oral commands, and the detection of objects non-equipped with localization sensors. Another improvement could be performed by replacing the couple of cameras by a Kinect device: it would thus avoid the synchronization process, and lead to a gain in the accuracy of the measurements.

6. ACKNOWLEDGEMENT

This work was supported in part by Grant in Aid for the Global Center of Excellence Program for "Center for Education and Research of Symbiotic, Sage and Secure System Design from the Ministry of Education, Culture, Sport, and Technology in Japan".

The second author is supported by PRESTO JST.

7. REFERENCES

[1] Logitech hd pro webcam c910. http://www.logitech.com/.

[2] Seeingmachines faceapi. http://www.seeingmachines.com/product/faceapi/.

[3] Swiss ranger sr4000. http://www.mesa-imaging.ch/prodview4k.php.

[4] S. D. Card, T. P. Moran, and A. Newell. *The Psychology of Human-Computer Interaction*. Lawrence Erlbaum Associates, 1983.

[5] D. Chai and K. N. Ngan. Face segmentation using skin-color map in videophone applications. *IEEE Trans. Circuits and Systems for Video Technology*, 9(4):551, June 1999.

[6] K.M. Lee, Y. Jung, J. Kim, and S.R. Kim. Are physically embodied social agents better than disembodied social agents?: The effects of physical embodiment, tactile interaction, and people's loneliness in human-robot interaction. *International Journal of Man-Machine Studies*, 64(10):962–973, 2006.

[7] H. Osawa, J. Mukai, and M. Imai. Anthropomorphization framework for human-object communication. *JACIII*, 11(8):1007–1014, 2007.

[8] J. Wainer, D. J. Feil-Seifer, D. A. Shell, and M. J. Mataric. The role of physical embodiment in human-robot interaction. *IEEE International Symposium on Robot and Human Interactive Communication*, pages 117–122, 2006.