Welbo: An Embodied Conversational Agent Living in Mixed Reality Space

Mahoro Anabuki, Hiroyuki Kakuta, Hiroyuki Yamamoto, and Hideyuki Tamura
Mixed Reality Systems Laboratory Inc.
6-145 Hanasaki-cho
Nishi-ku, Yokohama 220-0022, Japan
{mahoro, kakuta, ymmt, tamura}@mr-system.co.jp
http://www.mr-system.co.jp/

ABSTRACT
This paper introduces a new type of anthropomorphic agent that lives in a 3D space where the real and virtual worlds are seamlessly merged. In this mixed reality (MR) space, people wearing a see-through head-mounted display can interact with both physical and virtual objects in real time. In this type of MR space, an embodied conversational agent, named "Welbo," is implemented to study how agent technology contributes. This agent has several unique features, compared with the conventional desktop agent.

Keywords
Mixed reality, augmented reality, anthropomorphic agent, multi-modal interface, human interface design

INTRODUCTION
Mixture of the real and virtual worlds had been seen only in feature films or TV commercials. Now this situation is changing. A new technology called "Mixed Reality" (MR) makes it possible to merge both worlds seamlessly and augment the physical world with computer-generated information in real time [1]. By wearing a see-through head-mounted display (HMD), people in MR space can see virtual objects that are geometrically and photometrically registered with the real scene.

MR space also allows us to interact with both physical and virtual objects in real time. It is different from physical space and virtual space, where only real and virtual objects are interactive, respectively.

In the field of human-computer interaction, an embodied conversational agent has been studied for years as a potential interface. We believe that this type of agent will become a powerful interface in MR space, once it perceives and acts in the physical scene as well as the virtual world. Thus, this paper introduces an embodied conversational agent into one of our MR systems and describes its design and implementation.

AN AGENT IN MR SPACE
An agent in MR space has the following unique features compared with a desktop agent existing in a rectangular window on a computer monitor.

(1) **Living in an MR 3D space**: Since the real and virtual worlds are registered on one 3D coordinate system, the agent can work in both worlds in the same manner.
(2) **Sharing the MR space with users**: Not only an agent in virtual space, but also users in physical space, interact within the MR space. Thus, the agent can share the MR space with users.
(3) **Embodied with a 3D body**: The stereoscopic images given to the HMD give the agent a 3D body. That is to say, the agent is perceived not only as a symbol, but also as a being, by the users.

Previous agents did not have these features. It is impossible for most of the embodied conversational agents confined to the windows of computer monitors, like "Johnny" [2], to share a 3D space with their users. If an agent is implemented with virtual reality technology, "Steve" [3], for example, it can share a 3D space with users. It shares, however, only a cyberspace and cannot get in touch with the real space. On the other hand, the autonomous agent in the "ALIVE" system [4] can share physical space with its users. This agent, however, shares the space in a large computer monitor similar to a mirror.

DESIGN AND IMPLEMENTATION
A new embodied conversational agent, named "Welbo," is introduced in an MR space. This agent is a robot type interface agent that guides, helps, and serves the users in an "MR Living Room," where the users can visually simulate the location of virtual furniture and articles in the physically half-equipped living room [1] (Figure 1 shows a scene of physical space and an augmented image). This capability is realized with the following functions.

Figure 1: A scene of physical space (left) and an augmented image (right).

(a) **Conversation**: Welbo has the speech synthesis capability to speak some predefined sentences such as sentiments and introductions. It also recognizes the user's instructions from his/her speech. The capability to understand discourse is required to make conversation between the users and the agent smooth.
(b) **Behavior**: The agent acts according to the user's instructions to achieve such feats as movement of furniture and pointing to objects. It also makes some gestures while making conversation and changes its location while waiting for the user's instruction.
(c) **Awareness**: The agent is aware of its status in the cyberspace as well as in the physical space. It also perceives the user's actions and movement.
(d) **Embodiment**: The agent has a human-like robot type body that is familiar with and accepted by the user. The body is rendered with the real-time 3D graphics capability with a sense of reality.

As shown in Fig. 2, the system is composed of the living room subsystem and the agent subsystem. The living room subsystem has three modules. The head tracking module measures the user's viewpoint and view direction with the sensor attached to the HMD. This data is sent to the room data server, which manages the real and virtual environment data and Welbo's location and pose. The data stored in this server are used by the image synthesis and mixing module for synthesizing the computer graphics imagery. This synthesized image is mixed with the physical scene and is displayed on the HMD.

The agent subsystem has the Welbo control module, which is represented as "Welbo" in Fig. 2. This module receives the user's instructions coming from the speech recognition module and perceives the user's movement based on the data in the room data server. Based on this information, the Welbo module renews the data in the server and updates the synthesized image. It also sends the speech data to the speech synthesis module to pronounce the data through a pair of speakers or a headphone.

Figure 3 shows subjective views while a user makes an interior simulation with Welbo.

DISCUSSIONS AND CONCLUSION

Each agent design gives its own impression to the users. In the case of Welbo, the spatial factors of the design have a large impact on the impression. For example, changes in

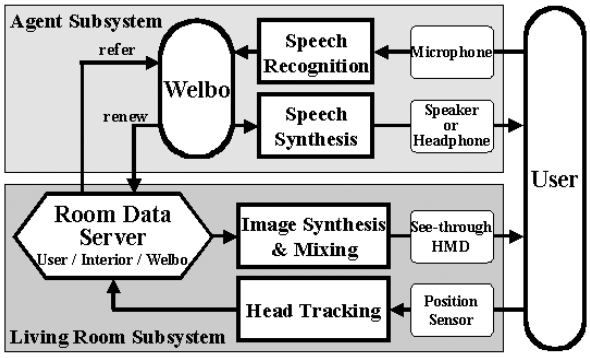

Figure 2: System architecture

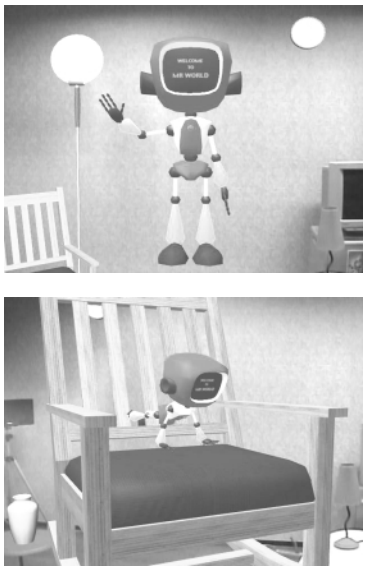

Figure 3: Subjective views while a user makes an interior simulation with Welbo.

the distance from the agent to the user significantly affect Welbo's appearance. The size and location also has similar effects. Therefore, we conducted some simple trials of these factors with Welbo.

Through the experiments, we understood that people preferred a size such that they can see Welbo's whole body in their field of view. Similarly, people like it to stay some distance away from them. As people feel uncomfortable when others look down on them, Welbo gives an unfavorable impression when it floats over them.

Thus, it is sure that spatial factors affect the impression Welbo gives to users. Each person is impressed based on his/her own preference, and we cannot say that one agent design always leads to the same impression. Despite these observations, we hope that it will be possible to find an agent design that tends to make a favorable impression in a certain situation.

ACKNOWLEDGMENTS

The authors thank Ms. Yuko Wakatsuki for her basic planning of this project.

REFERENCES

1. Tamura, H. et al. Steps toward seamless mixed reality, in (Ohta, Y., and Tamura, H. eds.) *Mixed Reality - Merging Real and Virtual Worlds*, Ohmsha & Springer-Verlag, pp.59-84, 1999.
2. Hasegawa, O. et al. Agent oriented multimodal image learning system, in *Proc. of IJCAI-97 Workshop on Intelligent Multimodal Systems*, pp.29-34, 1997.
3. Rickel, J., and Johnson, W.L. Animated agents for procedural training in virtual reality: perception, cognition, and motor control, in *Applied Artificial Intelligence*, Vol.13, pp.343-382, 1999.
4. Maes, P. et al. The ALIVE system: wireless, full-body interaction with autonomous agents, in *ACM Multimedia Systems*, Vol.5, No.2, pp.105-112, 1997.