A Proposal of Video Communication System in Which Talker's Avatar is Superimposed for a Virtual Face-to-Face Scene

Yutaka Ishii [*1], Tomohiro Takada[*2], and Tomio Watanabe[*1]
*1 Faculty of Computer Science and System Engineering, Okayama Prefectural University, 111 Kuboki, Soja, Okayama 719-1197, JAPAN,
*2 Graduate School of Systems Engineering, Okayama Prefectural University,
ishii@cse.oka-pu.ac.jp, t_takada@hint.cse.oka-pu.ac.jp, watanabe@cse.oka-pu.ac.jp

Abstract

We developed a voice-driven embodied CG character in which the user's head motion is detected by a headset-type motion capture device with an acceleration sensor and gyro sensor. We propose the concept of an embodied video chat system in which the CG character is superimposed on the other talker's video image in a face-to-face scene and we describe a prototype system called "E-VChat." Talkers can communicate smoothly, with nonverbal information using video images, and share mutual interaction without detailed adjustment of a video camera. The system provides a new level of interaction awareness for human communication.

1. Introduction

In common real-time communication systems such as telephone or chat, the information is transmitted through limited communication channel. However, communicable information is not much through only voice or text in human communication. In recent years, communication systems which used two or more channels such as a TV phone and CG character are developed, and many researches on multi-modal communication are carried out. However, remote talkers are hard to get interaction awareness because they cannot share the same communication space. Originally, humans communicate smoothly with nonverbal behavior such as gesture and facial expression in face-to-face communication. That is, the sharing of embodiment is very important in communication.

We developed the embodied virtual communication system for human interaction analysis by synthesis using the human type avatar called "VirtualActor" which represents upper body motion of talker, and the effectiveness of the system was demonstrated. By sharing of embodiment through avatars, the scene for sensing mutual interactive rhythms is preferred in the remote communication[1]. More effective communication support system would be developed, if talkers can communicate observing partner's nonverbal information such as facial expression/color as well as embodied interaction awareness.

In this study, we propose the concept of an embodied video communication system in which talker's own video image is superimposed on the other's one. However, this method required detailed adjustments for the natural projection of talkers, such as talkers' sizes and seating arrangements. To address these problems, we develop an embodied video communication system in which the VirtualActor, which represents interactive behavior, is superimposed on the other speech partner's video image in a virtual face-to-face scene. Further, we developed a headset-type motion-capture device that reads the talker's head movements directly, using an acceleration sensor to detect vertical motion and a gyro sensor to detect rotation, and we employ a CG character which moves based on talker's own motion and generates motion automatically based on the on-off pattern of talker's voice.

This paper proposes the concept of an embodied video communication system in which this CG character is superimposed on the other talker's video image in a face-to-face scene, and develops a prototype called "E-VChat" to solve the lack of sensor portability and a feeling of restraint caused by the sensors. Finally, we develop a multiple-character E-VChat system using an audience that nods in response to the talker's voice for more effective video communication.

2. Superimposing of talker's own image in video communication

In order to support smooth interaction between remote talkers, we propose a virtual face-to-face video communication method in which talker's own video image is superimposed on the other talker's video image from the diagonal view as shown in Figure 1. Instead of talker's own video image, talker's avatar is suggested to employ for sharing the embodied interaction as shown in Figure 2. As an avatar, the VirtualActor moves in response to the motion of the talker's head, waist, and both wrists, measured by four magnetic sensors. The talker's own VirtualActor is superimposed on the other talker's video image, using the Chroma-key effect implemented by the video editor. In this way, talkers can observe nonverbal information with mutual embodied rhythms perceived from the video image.

CollabTech 2012 , August 27-29, 2012, Hokkaido, Japan.

Fig.1: Own video image is superimposed.

Fig.2: Own VirtualActor is superimposed.

Although the system solved the problem of detailed adjustments, some issues remained, such as the lack of sensor portability and a feeling of restraint caused by the sensors. Accordingly, we developed an embodied video communication system, called “E-VChat”, in which the motion of the CG character is based on the talker’s own head motion, measured by a lightweight device combined with body movements that are based on the on-off pattern of the talker’s voice[2]. Talker’s own head motion is measured by a headset-type motion-capture device, which reflects the talker’s head movements directly using an acceleration sensor and gyro sensor, in order to obscure the device (Figure 3).

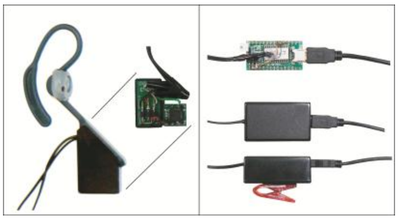

Fig.3: Headset-type motion-capture device.

3. E-VChat system

As shown in Figure 4, the E-VChat system displays the other talker’s video image and the CG character in the form of a VirtualActor. In the E-VChat system, the CG character has two communication functions, as an avatar and as an agent: the avatar has a presence in the remote communication, while the agent acts as a communication support character using interactive motion. The CG character mimics the talker’s head movements detected by the simple motion-capture device and displays interactive motions based on the talker’s voice rhythms. The upper body motion is generated automatically by a moving-average interactive response model, based on the on-off pattern of the talker’s voice[3]. This CG character could have various communicative effects that depend on the generated motion, such as affirmative or negative reaction. During serious communication, such as negotiations or discussion, it may be necessary to show the talker’s intention directly, apart from automatically generated movements. Therefore, we could construct an effective communication system using head movements which play an important role in communication.

This system is unique that each talker sees only his or her own substitute character, and not the other talker’s CG character. On the screen, the partner looks to be able to observe the talker’s own character directly, however, the partner cannot observe and understand it actually. In other words, the character has an effect for only a talker as a direct user of the system.

The proposed system is designed to be used for remote communication, including video conferencing or job interviews, as well as for free conversation between friends. Going one step further, we propose a multiple-character E-VChat system that displays not only avatar of the talker, but also communicative agents around the other talker, as shown in Figure 5. This creates an effective communication space in which multiple characters are superimposed on the other talker’s video image. In stressful communication situations, affirmative support agents could provide the talker with reassurance and improve the comfort level.

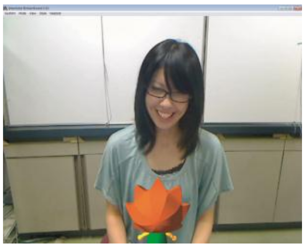

Fig.4: E-VChat.

Fig.5: Multiple characters.

4. Conclusion

In this paper, we propose a video communication system called E-VChat, in which talkers can communicate smoothly, using nonverbal information combined with video images, and can share a mutual interaction without detailed adjustments of a video camera. Furthermore, the prototype systems are developed using the talker’s avatar and multiple agents.

References

[1] Ishii, Y. and Watanabe, T., “Evaluation of an Embodied Virtual Communication System for Human Interaction Analysis by Synthesis”, *In proceedings of RO-MAN 2000*, pp.29-34 (2000).

[2] Yamamoto, M., Osaki, K., Matsune, S., and Watanabe, T., “An Embodied Entrainment Character Cell Phone by Speech and Head Motion Inputs,” *Proc. of the 19th IEEE International Symposium on Robot and Human Interactive Communication*, pp.318-323 (2010).

[3] Watanabe, T., Okubo, M., Nakashige, M., and Danbara, R., “InterActor: Speech-Driven Embodied Interactive Actor; *International Journal of Human-Computer Interaction*,” Vol.17, No.1, pp.43-60 (2004).