Video Content Production Support System with Speech-Driven Embodied Entrainment Character by Speech and Hand Motion Inputs

Michiya Yamamoto[1], Kouzi Osaki[2], and Tomio Watanabe[1,3]

[1] Faculty of Computer Science and System Engineering, Okayama Prefectural University,
111 Kuboki, Soja, Okayama 719-1197, Japan
{yamamoto, watanabe}@cse.oka-pu.ac.jp
[2] Graduate School of Systems Engineering, Okayama Prefectural University,
111 Kuboki, Soja, Okayama 719-1197, Japan
osaki@hint.cse.oka-pu.ac.jp
[3] CREST of Japan Science and Technology Agency

Abstract. InterActor is a speech-input-driven CG-embodied interaction character that can generate communicative movements and actions for entrained interaction. InterPuppet, on the other hand, is an embodied interaction character that is driven by both speech input, like the InterActor, and hand motion input, like a puppet. In this study, we apply InterPuppet to video content production and construct a system to evaluate the content production. Self-evaluation of long-term (5-day) video content production demonstrates the effectiveness of the developed system.

Keywords: Human communication, human interaction, embodied interaction, embodied communication, video content.

1 Introduction

Today, video content that employs CG characters is becoming increasingly popular, and such CG characters are given as much importance and preference as actors or stuffed toys. The possibility of using CG characters will increase with an increase in the availability of televisions with multiple channels. Moreover, there has been an increase in the use of streaming content on networks or websites such as YouTube where users can broadcast content. Many studies have been conducted on the movement generation of CG characters and on video content that uses CG characters [1], [2]. However, the importance of CG characters will increase further if we can make the CG characters in TV programs move in real time.

In human face-to-face communication, not only verbal messages but also nonverbal behaviors such as nodding and body movements are rhythmically related and mutually synchronized between the communicating humans. This synchrony of embodied rhythms, termed entrainment, in communication generates the sharing of embodiment in human interaction, which plays an important role in human interaction and communication [3]. We have already developed a speech-driven CG-embodied

J.A. Jacko (Ed.): Human-Computer Interaction, Part III, HCII 2009, LNCS 5612, pp. 358–367, 2009.

character called InterActor, which performs the functions of both the speaker and the listener by coherently generating expressive actions and movements in accordance with a speech input. We have demonstrated that this system can be effective in supporting human interaction and communication between remote individuals [4]. In addition, we have developed another embodied interactive character called InterPuppet, which permits input in the form of hand motion. The effectiveness of InterPuppet in providing communication support has been demonstrated [5]. In video content, the entrainment movements and actions of CG characters are very important because character motion often appears with voice. If the mechanism of InterPuppet, which involves both entrainment motion and intentional motion, is introduced to a video content production support system, InterPuppet can be widely used in applications such as news programs and live broadcasts.

In this study, a video content production support system using InterPuppet is developed. Further, the effectiveness of InterPuppet is evaluated from the viewpoint of content creators by producing video content for a long period.

2 Video Content Production Support System

2.1 Outline of Content Production

Figure 1 shows the overview of content production using InterPuppet. In our previous researches, we developed a conversation support system using InterPuppet [5]. This system facilitates communication via the CG character, and it is important that there are users behind the CG character. The CG character plays an important role in video content production because the production work is done by the content creator.

The application of InterPuppet to video content production will result in the production of attractive content that involves entrainment. In addition, for any content, this system maintains the entrainment of the body's rhythm and the vividness and natural movement of a CG character such as InterActor. In particular, the viewer can be entrained to the content that presents and explains information when the CG character talks to the viewer as a speaker. Intentional body movements necessary for the production of video content are also added by using hand movements in a manner similar to the manipulation of a puppet. Necessary motion such as pointing is expressed in real time, while automatically generating the communicator's motion from the creator's voice.

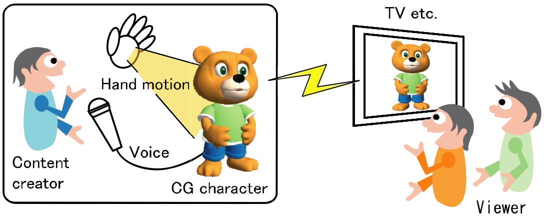

Fig. 1. Outline of content production

By using InterPuppet, we can use a CG character to smoothly produce a TV program, explain the information shown on screen, and interact with other characters. Rich communicative movements and actions based on entrainment body rhythm are generated more easily than those obtained by the method for generating the movement of the character from just the hand motion input, thereby providing support to video content production.

2.2 System Development

We have developed a prototype of the content production system for evaluation and demonstration. Figure 2 shows the system configuration. In order to operate the character like a puppet, the system consists of a data glove (Immersion CyberGlove), a headset for providing voice input/output, a display, and a PC. The PC comprises a video card (DirectX 9.0 support), voice input/output, and a serial port (for connecting the data glove). This system uses an AT-compatible PC with Windows XP. Video content is recorded on the hard disk of the PC and can be replayed freely.

Here, we use an information program that introduces a flowering plant as an example of video content. Figure 3 shows the screen composition. The screen shows the CG character and the background image. An image and a video are set up in the background. If necessary, the position and the size of the video can be changed smoothly in accordance with the composition of the TV screen. This system can produce a wide range of video content by changing the background, movie, and the appearance of the screen.

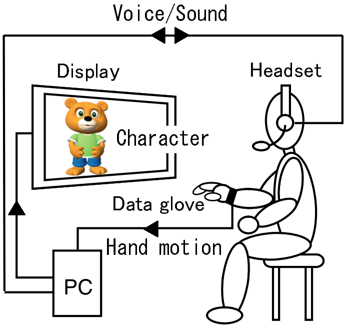

Fig. 2. System configuration

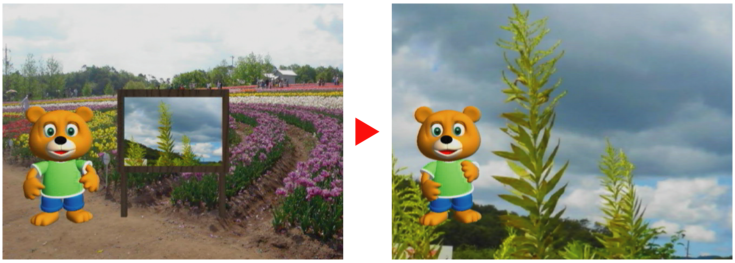

Fig. 3. Screen composition

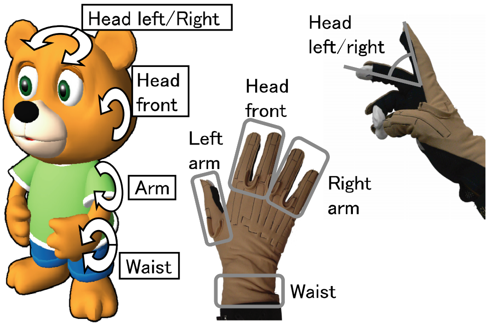

Fig. 4. Data glove operation of InterPuppet

2.3 Operation of a Character

In general, the voice input to InterPuppet is obtained by reading out the script of the program. We have already developed InterCaster; it can produce video content by using the speaker movements of InterActor. InterCaster is used in applications such as television programs developed by InterRobot Ltd. (a laboratory venture). In InterCaster, the effects of communication movements are enhanced when the movements and actions of the listener are added to the movements of the speaker, because the action of nodding is related to the speaker's own voice. Therefore, InterCaster includes both speaker motion and listener motion. In this study, the model that generates the communicative motion of speakers and listeners is the same as that used in our previous research [6].

While generating intentional motion from hand motion, the hand motion input from the data glove is related to the motion of the CG character. By using 18 sensors of the date glove, we can measure hand motion and convert it to the CG character's motion in the form of the motion of a puppet. Figure 4 shows the relationships between character motion and data glove manipulation.

3 Content Production Experiment

3.1 Setup

In order to evaluate the effectiveness of the proposed method from the viewpoint of content creators, we performed an evaluation experiment in which a live TV broadcast was assumed (Figure 5). In this experiment, the proposed method was evaluated from the viewpoint of producing video content; the experiment was performed by following a previously established procedure. Moreover, in order to create a realistic work scene, we made a subject rehearse his/her task before recording the video content program.

Figure 6 shows the experimental setup. Dummy broadcasting equipment (Sony, FXE-120) was arranged to create the atmosphere of an actual production site. Two experimenters played the roles of production staff. The experimenters handed out realistic instructions to the subject, operated the dummy broadcasting equipment, and

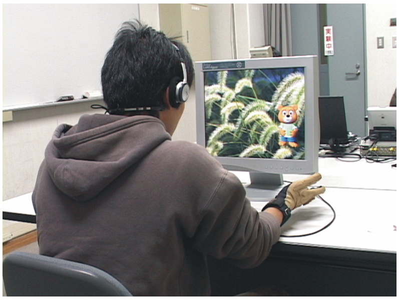

Fig. 5. Experimental scenery

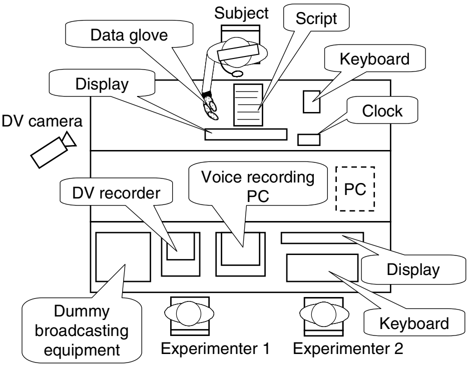

Fig. 6. Experimental Setup

monitored the DV camera. In addition, we assumed that the programs were broadcast to many unspecified viewers. In order to make the subjects perform their tasks with the highest efficiency, we told them that their produced work would be made available to general public for viewing. We prepared a script and video for the production of a one-minute program that introduced a flowering plant. While reading out the script, the subjects generated movement in accordance with the video.

The program was recorded 15 times. In order to examine regular content production, three recordings were made in one day; such recordings were made on five different days in a total span of 16 days. A different video and script for the introduction of the flowering plant were prepared for each program. On average, the scripts had 261.8 moras, and it was possible to read out all the information in approximately 1 min. Moreover, the timing of the video was changed, and the BGM was set to a value equal to this timing.

In the experiment, the following three modes of operation were compared. The program was recorded in each of these modes.

A: Hand motion input only (movements of mouth and eyes are the same as those in mode B)
B: InterActor
C: InterPuppet (A + B)

Before beginning the experiment, we confirmed that the subjects were well acquainted with the system. These subjects were briefed on the character operation method to be followed while using this system. Further, the data glove was calibrated when necessary. Figure 7 shows the procedure of the production of the program. First, the subjects used the system. The three modes were used randomly in a day. Then, these subjects watched the recorded content and evaluated their own work at the end of each trial.

In order to evaluate the mental workload, the subjects generated the character motions as work. We used NASA-TLX for the evaluation of their work. Using NASA-TLX, we asked the subjects to assign weights on the basis of a paired comparison. They were asked to rate six items (mental demand, physical demand, temporal demand, own performance, effort, and frustration level) between 0 and 100. Then, the

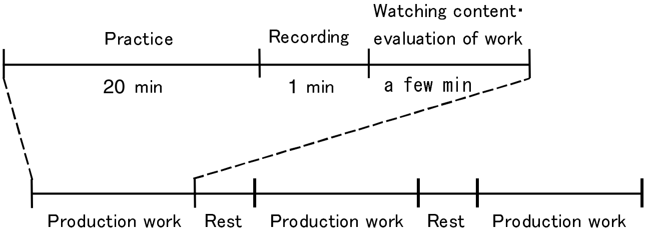

Fig. 7. Time table

mean weight workload (WWL) score, which was an integrated score of the mental workload obtained by a weighted average, was calculated. Moreover, the subjects used a seven-point bipolar rating scale from –3 (not at all) to 3 (extremely) in order to rate the three modes on parameters such as "enjoyment," "ease of speaking" (a subject can speak easily and smoothly), "role play," and "operation" (a subject feels that he/she can manipulate the character) from the viewpoint of body interaction support during the generation of the character movement. Further, we recorded the extent of character operation by the data glove for modes A and C. The subjects randomly rested for a few minutes during the program production work.

In addition, after the completion of the recordings on the fifth (final) day, the subjects were asked to complete questionnaires for an overall evaluation of the work. The six items given below were rated on the seven-point bipolar scale for operation in modes A–C. The questionnaire addressed parameters such as "vivid movement," "movement to report content information," "rich movement," "reading efficiency," "work satisfaction," and "He/She wants to show his/her work to others."

Ten Japanese students (5 males and 5 females) participated in the experiments.

3.2 Results

Figure 8 shows the result of the evaluation carried out using NASA-TLX. The result of the analysis of variance is also shown in the figure. From the figure, we can

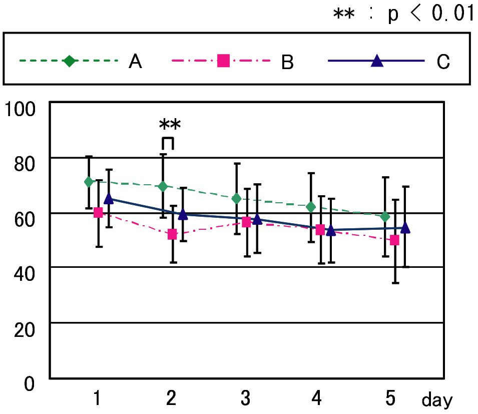

Fig. 8. Result of NASA-TLX

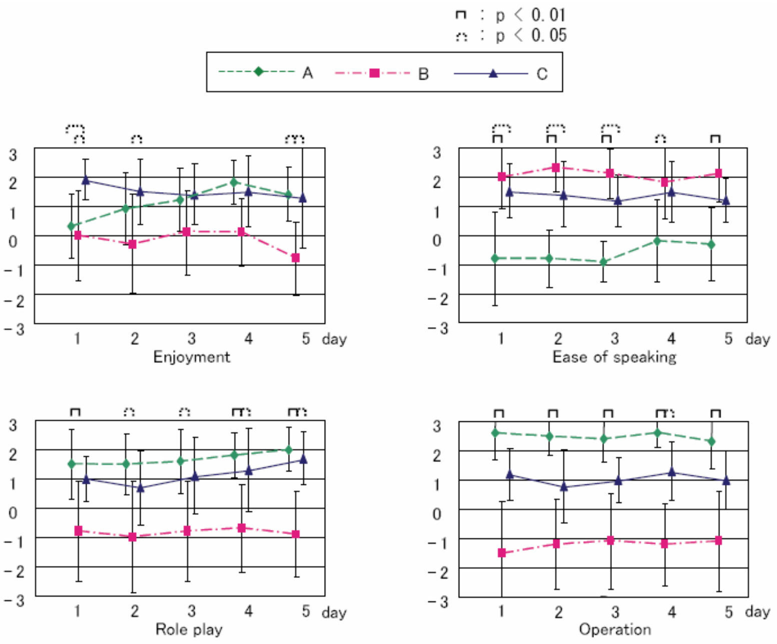

Fig. 9. Sensory evaluation

observe that the workload of each mode decreased with the passage of days. However, the workload in the case of mode A (hand motion input only) was higher than that in the case of the other modes.

Figure 9 shows the results of the seven-point bipolar rating for the five days on which the programs were recorded. The results of the Friedman test of each mode are also shown in this figure. For the parameter "enjoyment," modes A and C were rated higher than mode B. For "ease of speaking," mode B was rated the highest on all five days, and mode C was rated higher than mode A. For "role play" and "operation," mode A was rated higher than the other modes. However, mode C was also rated high for "role play."

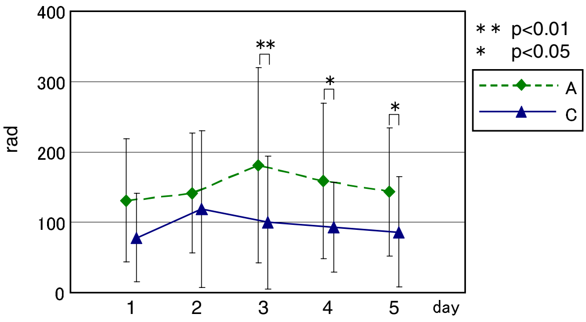

Fig. 10. Amount of data glove operation

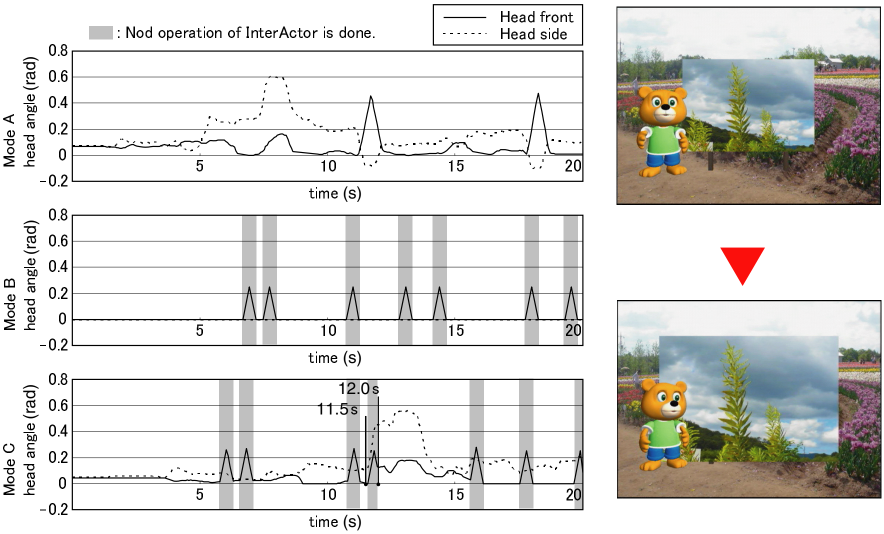

Fig. 11. Some head movements

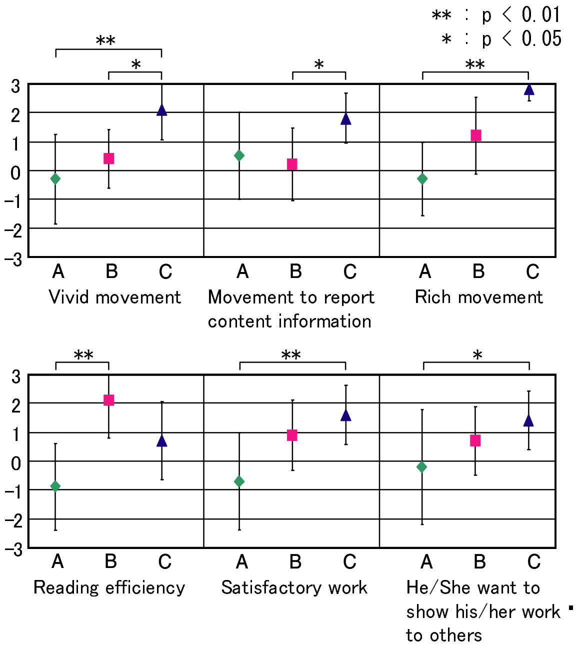

Fig. 12. Self evaluation of content

Figure 10 shows the extent of data glove operation in accordance with the hand motion input in modes A and C. The extent of operation shown in the figure is the total extent of hand and wrist joint operation used for character operation in the program recording of 1 min. The result of a t-test is also shown in the figure. Significant differences in the extents of operations are observed after the third day. Figure 11 shows some head movements and screenshots depicting these movements in mode C. The sections highlighted in gray are those where the nodding movements of InterActor were made. The character extensively turned to the right after approximately 8 s and 12 s in modes A and C. At this time, the character viewed the video. This figure shows a typical usage of the system (The content was recorded by a subject on the fifth day).

Figure 12 shows the result of the overall evaluation after the end of the experiment. The result of the Friedman test is also shown in this figure. The ratings for most of the parameters among "vivid movement," "rich movement," "work satisfaction," and "He/She wants to show his/her work to others" were the highest in the case of mode C; mode A had the lowest ratings. In particular, mode C was rated higher than the other modes on the parameters "vivid movement" and "rich movement." For the parameter "reading efficiency," mode B was rated higher than the other modes. The ratings for the parameter "movement to report content information" were higher in the case of mode C than in the case of mode B.

4 Discussion

The subjects' evaluation changed after they participated in the experiment. The results of the evaluation using NASA-TLX show that the workload of content production decreased. In the case of modes A and C, the ratings for the parameter "role play" increased gradually, suggesting that the subjects gained experience in character operation by hand motion input. On the first day, the ratings of the parameter "enjoyment" in mode C were higher than those in mode A; on the fifth day, these ratings were almost the same. However, after the third day, a significant difference was observed between the extents of character operation in modes A and C. On the fifth day, the ratings for this parameter were high in the case of both mode A and mode C. In the case of mode B, the parameter "operation" was rated low; however, the parameter "ease of speaking" was rated high during the five days. These results demonstrate the effectiveness of InterActor in providing communications support. The rating for the parameters "ease of speaking" and "operation" in the case of mode C was intermediate between the ratings in the case of modes A and C. In addition, the results of the evaluation using NASA-TLX and the rating for the parameter "role play" in the case of mode C were intermediate between those in the case of the other two modes. Therefore, we can conclude that mode C is well balanced between mode A and mode B.

5 Conclusion

An embodied interactive character termed InterPuppet was used for the production of video content, and its performance was evaluated from the viewpoint of content creators. First, we outlined the method for video content production for making the best possible use of InterPuppet; we also described the evaluation system. A sensory evaluation and the behavioral analysis of the production in three modes—hand motion input only, InterActor, and InterPuppet—for five days demonstrated the effectiveness of the proposed InterPuppet system. It was also found that InterPuppet received high ratings for the parameters “enjoyment,” “ease of speaking,” and “role play.” Moreover, character motion received high ratings for the parameters “vivid movement,” “movement to report content information,” and “rich movement”.

We have developed a learning support system using InterActor, in which InterActors are superimposed on video images such as educational programs [6]. By providing support for content production, we can produce attractive content by superimposing two or more InterActors.

Acknowledgements. This work under our project "Generation and Control Technology of Human-entrained Embodied Media" has been supported by CREST (Core Research for Evolution Science and Technology) of JST (Japan Science and Tech-nology Agency).

References

1. SIGGPRAPH (2008), `http://www.siggraph.org/s2008/`
2. Morishima, S.: Face and Gesture Cloning for Life-like Agent. In: Proceedings of the 11th International Conference on Human-Computer Interaction, pp. 2044 (2005)
3. Kobayashi, N., Ishii, T., Watanabe, T.: Quantitative Evaluation of Infant Behavior and Mother Infant Interaction. Early Development and Parenting, 23–31 (1992)
4. Watanabe, T., Okubo, M., Nakashige, M., Danbara, R.: InterActor: Speech-driven embodied interactive actor. International Journal of Human-Computer Interaction 17(1), 43–60 (2004)
5. Yamamoto, M., Watanabe, T.: Development of an Embodied Interaction System with InterActor by Speech and Hand Motion Input. In: CD-ROM of the 2005 IEEE International Workshop on Robots and Human Interactive Communication, pp. 323–328 (2005)
6. Watanabe, T., Yamamoto, M.: An Embodied Entrainment System with InterActors Superimposed on Images. In: Proceedings of the 11th International Conference on Human-Computer Interaction, p. 2045 (2005)