Chapter 9
Human-entrained Embodied Interaction and Communication Technology

Tomio Watanabe[1]

Abstract An embodied communication system for mind connection (E-COSMIC) has been developed by applying the entrainment mechanism of the embodied rhythms of nodding and body movements to physical robots and CG characters in verbal communication. E-COSMIC comprises an embodied virtual communication system for human interaction analysis by synthesis and a speech-driven embodied interaction system for supporting essential human interaction and communication based on the analysis that uses the embodied virtual communication system. A human-entrained embodied interaction and communication technology for an advanced media society is introduced through some applications of E-COSMIC. A generation and control technology of human-entrained embodied media is also introduced.

9.1 Introduction

In human face-to-face conversation, a listener's movements such as nodding and body motions are interactively synchronized with the speaker's speech. Embodied rhythms between voice and movement are mutually synchronized not only between talkers but also in a talker. The phenomenon is observed in an infant's movements in response to the mother's speech as a primitive form of communication [1, 2]. This synchrony of embodied rhythms in communication, referred to as entrainment, generates the sharing of embodiment in human interaction, which plays an important role in human interaction and communication. Entrainment in communication is also observed in physiological indices such as respiration and

[1] T. Watanabe
Faculty of Computer Science and Systems Engineering, Okayama Prefectural University
CREST of Japan Science and Technology Agency
111 Kuboki, Soja, Okayama 719-1197, Japan
e-mail: watanabe@cse.oka-pu.ac.jp

heart rate variability [3]. This embodied communication, which is closely related to behavioral and physiological entrainment, is an essential form of communication that forms the basis of interaction between talkers through mutual embodiment. Hence, the introduction of this mechanism to a human interface is indispensable to the realization of human-centered essential interaction and communication systems.

In this chapter, by focusing on the embodied entrainment, the human-entrained embodied interaction and communication technology through the development of the embodied communication system for mind connection (E-COSMIC) is introduced for supporting human interaction and communication [4]. E-COSMIC mainly comprises an embodied virtual face-to-face communication system and a speech-driven embodied interaction system, as shown in Figure 9.1. The former is developed for human interaction analysis by synthesis and the latter, for supporting human interaction and communication based on the analysis that uses the former. The effectiveness of the system is demonstrated by some actual applications on robot/CG and human interactive communications. With the aim of creating embodied media that unify performers and audiences for supporting the creation of digital media arts for entertainment and education, the generation and control technology of human-entrained embodied media is also introduced.

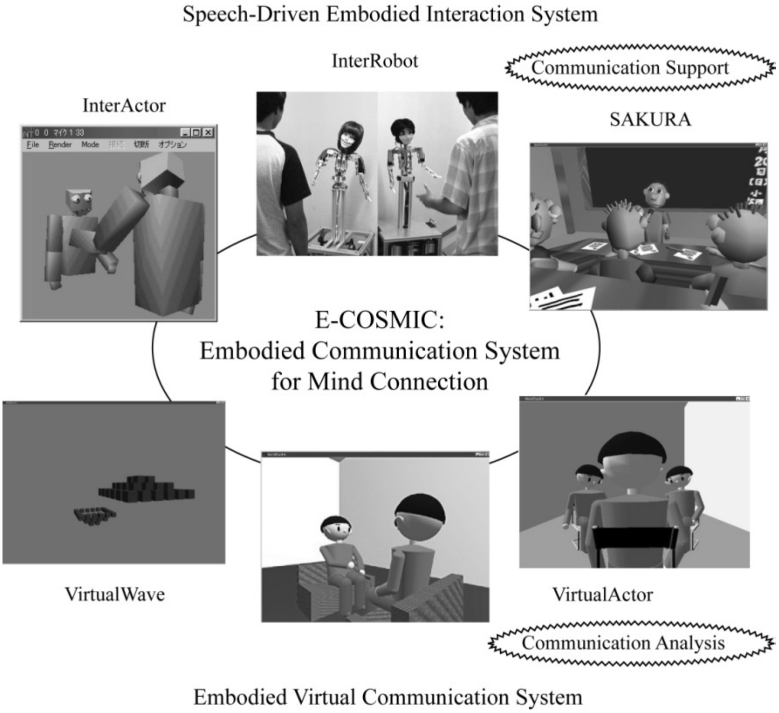


Figure 9.1 E-COSMIC

9.2 Embodied Virtual Communication System

The concept of an embodied virtual face-to-face communication system is illustrated in Figure 9.2. The figure presents a VirtualActor (VA), an interactive avatar, that represents the talker's interactive behavior such as gestures, nodding, blinking, facial color and expressions, paralanguage, respiration, *etc.*, based on one's verbal and nonverbal information as well as physiological information in a virtual face-to-face communication environment. Figure 9.3 provides an example of a virtual face-to-face scene with two VAs from the diagonal backward viewpoint of one's own VA. The motions of the head, arms, and body for each VA are represented based on the positions and angles measured by four magnetic sensors that are placed on the top of the talker's head, both wrists, and the back of the body [5]. Two remote talkers can communicate through their VAs and become aware of the interaction through the embodied interaction of VAs in the same virtual communication environment from any viewpoint. The analysis by synthesis for interaction in communication is performed by processing the behavior of VAs, such as cutting or delaying the motion and voice of VAs in various spatial relations and positions (Figure 9.4). For example, to examine the effects of only nodding on interaction, it is possible for a VA to represent just nodding without body motion even if the talker nods with body motion. Thus, the characteristics and relations of the embodied interaction between talkers are systematically clarified through the analysis by synthesis of interaction in communication by using the system in which talkers are the observers of interactions as well as the operators of interaction through their VAs. Further, physiological measurements such as respiration, heart rate variability, and facial skin temperature, as indices of emotional states in communication are utilized not only for quantitatively evaluating the interaction but also for transmitting the change in talkers' emotions through the VA affect display in which facial color and expressions are synthesized based on the measurement. An embodied virtual group communication system was also developed for three human interaction supports and analyses by synthesis, as indicated in Figure 9.1 [6].

Not only we have created a VA that represents human behavior precisely, we have also created an abstract avatar of a wave (VirtualWave, or VW) that is constructed from 6 × 6 cubes, as shown in Figure 9.5. Here, the communication function of the VA is simplified as a function of the motion of the VW in order to clarify an essential role of interaction. The rhythm of the VW is used to characterize the interactive rhythm, and this behavior is represented using only the motion of the head; the motion is measured by a magnetic sensor placed on top of the talker's head. This is because the head motion performs the essential function of regulating the flow of conversation, for example, nodding, by which each talker discriminates one's VW from the partner's VW and shares their interactions. A vertical shift such as nodding is expressed by an up-and-down displacement of cubes in which the wave approaches a quadrangular pyramid in shape, as shown in Figure 9.5. A horizontal shift such as back-and-forth is expressed by the parallel displacement of cubes in proportion to the shift. VW is represented by a frame rate of 30 f/s. The effects of the head motion on the interaction of VAs have already been demonstrated [7].

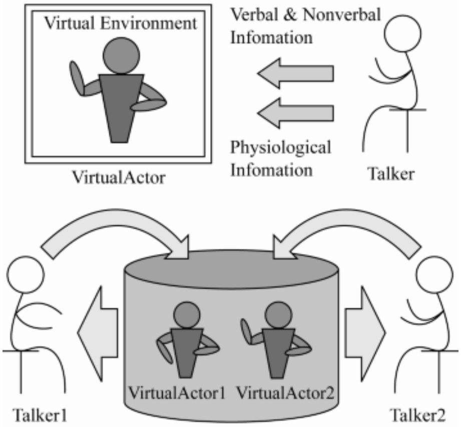


Figure 9.2 Concept of the embodied virtual communication system

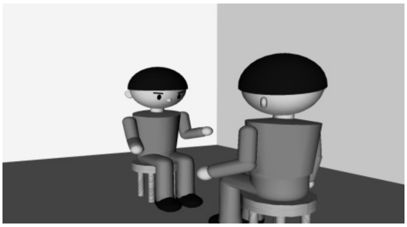


Figure 9.3 Example of a virtual face-to-face scene with two VirtualActors representing the talker and his/her partner

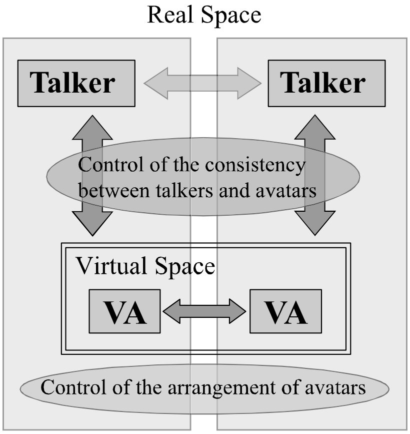

Real Space
Talker
Talker
Control of the consistency
between talkers and avatars
Virtual Space
VA
VA
Control of the arrangement of avatars

Figure 9.4 Analysis model for human interaction via embodied avatars

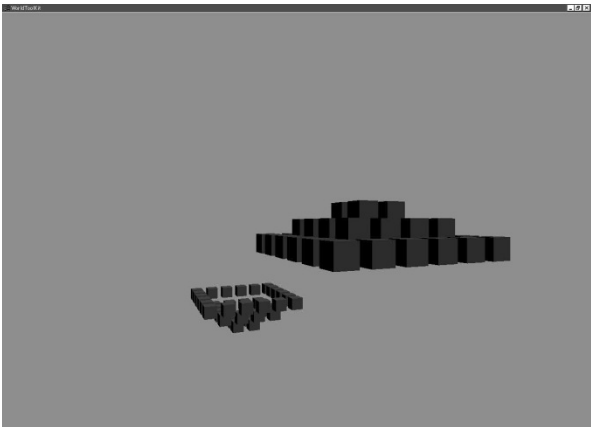


Figure 9.5 Example of VirtualWave's (VW) communication scene

9.3 Speech-driven Embodied Interaction System

Based on the human interaction analysis that uses the embodied virtual communication system, a speech-driven embodied interaction system is developed for supporting human interaction by generating the communicative motions of a physical robot referred to as InterRobot or a CG character known as InterActor; these communicative motions are coherently related to speech input [8]. The concept is presented in Figure 9.6. The system comprises two InterRobots (or InterActors) that function as both speaker and listener based on speech input. When Talker 1 speaks to InterRobot 2, InterRobot 2 responds to Talker 1's utterance with an appropriate timing through its entire body motions, including nodding, blinking, and actions, in a manner similar to the body motions of a listener. Thus, Talker 1 can talk smoothly and naturally. Subsequently, the speech is transmitted via a network to the remote InterRobot 1. InterRobot 1 can effectively transmit Talker 1's message to Talker 2 by generating the body motions similar to those of the speaker based on the time series of the speech and by simultaneously presenting both the speech and the entrained body motions. This time, Talker 2 in the role of a speaker achieves communication in the same way by transmitting his/her speech via InterRobot 1 as a listener and InterRobot 2 as the one talking to Talker 1. Thus, in this manner, two remote talkers can enjoy a conversation via InterRobot. The information transmitted and received by this system is only through speech. Of significance is the fact that it is a human who transmits and receives the information; the InterRobot merely generates the entrained communicative movements and actions based on speech input and supports the sharing of mutual embodiment in communication.

With regard to a listener's interaction model, the nodding reaction model from a speech ON–OFF pattern and the body reaction model from the nodding reaction model are introduced (Figure 9.7). When *Mu*(*i*) exceeds a threshold value, nodding

$M(i)$ is estimated as the weighted sum of the binary speech signal $V(i)$. The body movements are related to the speech input by operating both the neck and one of the wrists, elbows, arms, or waists at the timing over the body threshold. The threshold is set lower than that of the nodding prediction of the MA (moving average) model, which is expressed as the weighted sum of the binary speech signal to nodding. In other words, for the generation of body movements when InterActor functions as a listener, the time relationships between nodding and other movements are realized by varying the threshold values of the nodding estimation.

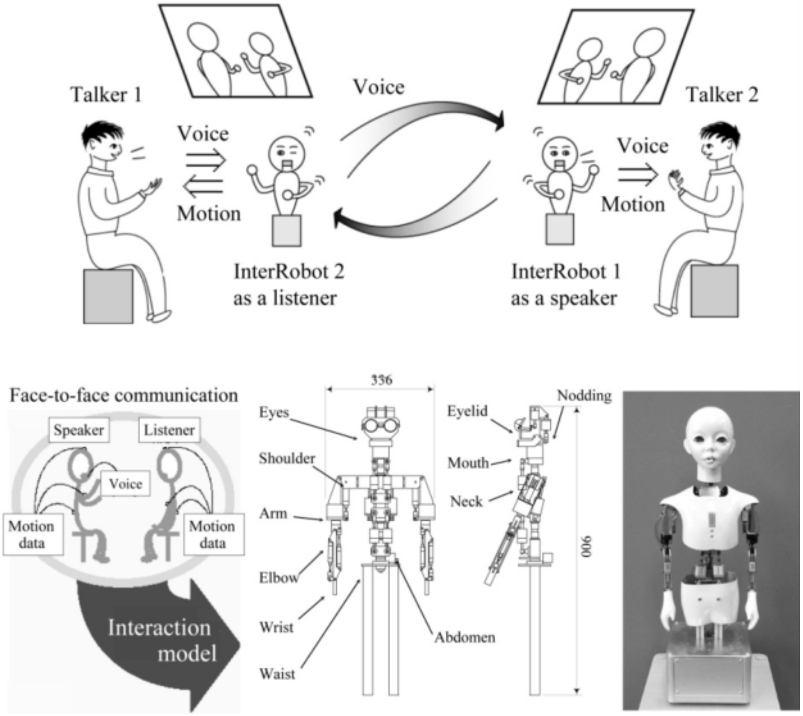


Figure 9.6 Concept of the speech-driven embodied interaction system

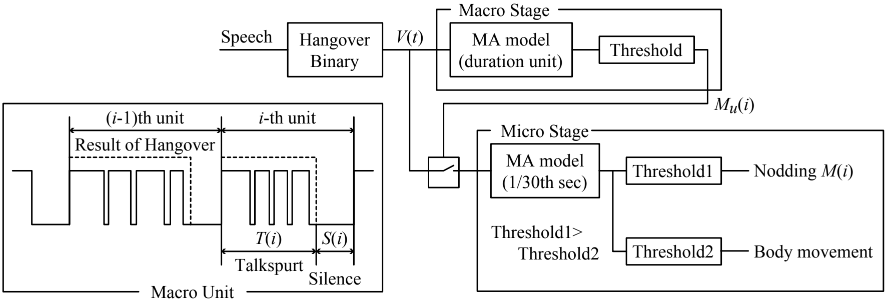


Figure 9.7 Interaction model

$$M_u(i) = \sum_{j=1}^{J} a(j)R(i-j) + u(i) \tag{9.1}$$

$$R(i) = \frac{T(i)}{T(i)+S(i)} \tag{9.2}$$

$a(j)$: linear prediction coefficient

$T(i)$: talksqurt duration in the *i*th duration unit

$S(i)$: silence duration in the *i*th duration unit

$u(i)$: noise

$$M(i) = \sum_{j=1}^{K} b(j)V(i-j) + w(i) \tag{9.3}$$

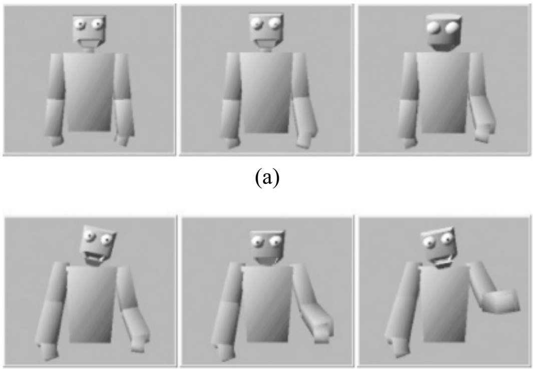
$b(j)$: linear prediction coefficient

$V(i)$: voice

$w(i)$: noise

The body movements as a speaker are also related to the speech input by operating both the neck and one of the other body actions at the timing over the threshold, which is estimated by the speaker's interaction model as its own MA model of the burst-pause of speech to the entire body motion. Because speech and arm movements are related at a relatively high threshold value, one of the arm actions in the preset multiple patterns is selected for operation when the power of speech is over the threshold. The expressive actions of InterActor are shown Figure 9.8.

We developed a system superimposed on a nodding response model for the analysis by synthesis of embodied communication under the expected conditions of promoted interaction. In addition, for this system, we performed experiments



(a)

(b)

Figure 9.8 Expressive actions of InterActor: (a) listener's action, and (b) speaker's action

for the analysis by synthesis of embodied communication by examining the sensory evaluation and the voice–movement analysis while inconsistently adding nodding responses in VA. We found that the cross-correlation between the talker’s voice and the listener’s head movement in the inconsistently activated condition increases at a significance level of 1% compared to that observed under normal conditions. The result also demonstrates that the system superimposed over the nodding response promoted interaction in embodied communication [9]. Furthermore, we have also developed a speech-driven embodied entrainment system called “InterWall” in which interactive CG objects behave as listeners on the basis of the speech input of a talker (Figure 9.9). This system can support human interaction and communication by producing embodied entrainment movements such as nodding on the basis of the speech input of a talker. We confirmed the importance of providing a communication environment in which not only avatars but also CG objects placed around the avatars are related to virtual communication.

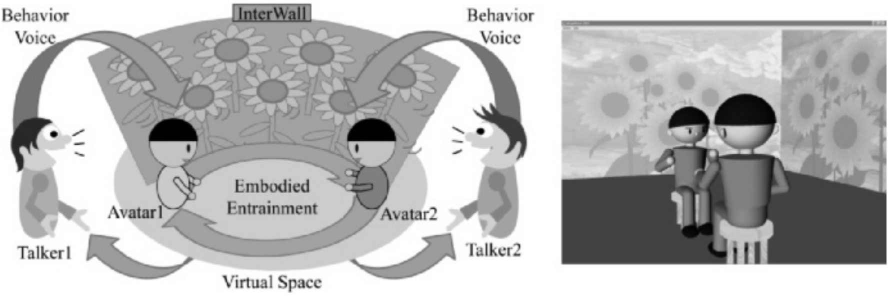


Figure 9.9 InterWall

Figure 9.10 illustrates a speech-driven embodied group-entrained communication system referred to as SAKURA [10]. SAKURA activates group communication in which InterActors are entrained to one another as a teacher and some students in the same virtual classroom. By using SAKURA, talkers can communicate with a sense of unity through the entrained InterActors by using only speech input via the network. Figure 9.11 depicts a physical version of SAKURA with InterRobots. Their entrained movements and actions based on speech can activate and assist human embodied interaction and communication. Figure 9.12 indicates another physical version of SAKURA with four InterRobots and one InterActor, which is exhibited in the National Museum of Emerging Science and Innovation where visitors can enjoy a dynamic experience of embodied communication. They perceive the effects of group-entrained communication environment intuitively and recognize the importance of embodied communication.

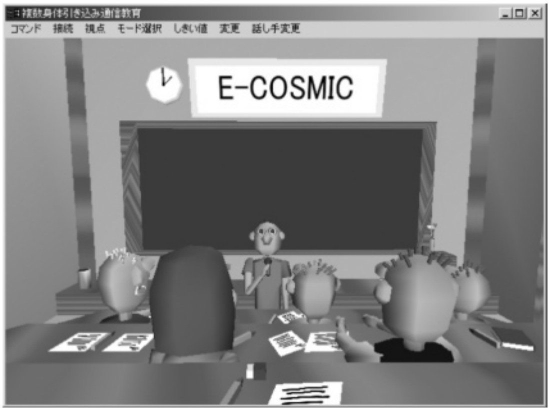


Figure 9.10 SAKURA: The speech-driven embodied group-entrained communication system

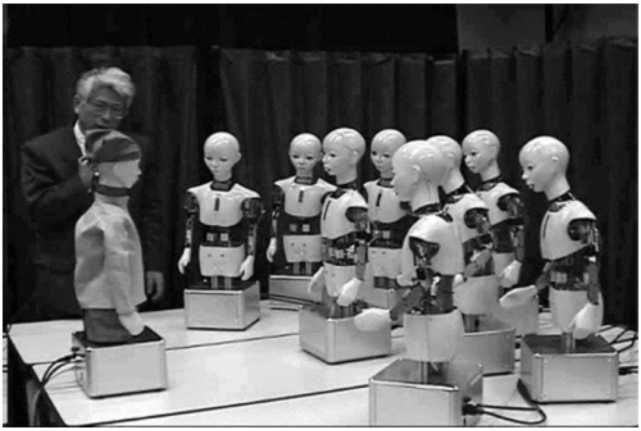


Figure 9.11 Physical version of SAKURA with InterRobots

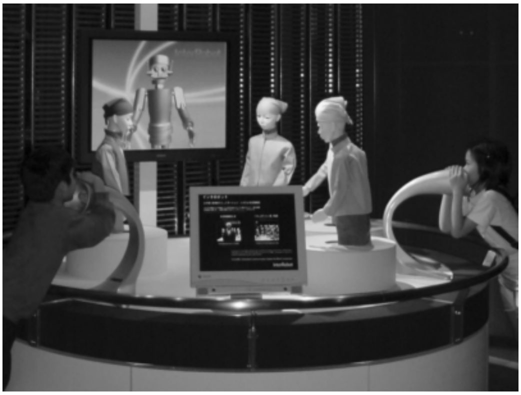


Figure 9.12 Speech-driven embodied interaction system with InterRobots and an InterActor in the National Museum of Emerging Science and Innovation

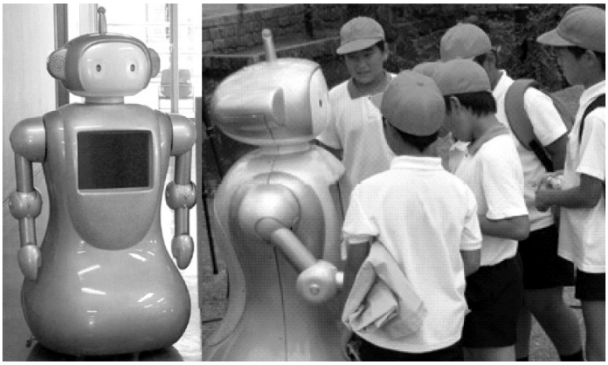


Figure 9.13 Interaction scene between an InterRobot and children

9.4 Embodied Interaction and Communication Technology

In this section, some actual applications of InterRobot/InterActor to human interface are introduced. Figure 9.13 depicts an interaction scene between an InterRobot and children. This InterRobot is commercially available and marketed for kindergarten use. Children enjoy and are excited about having conversations with the InterRobot, while the teacher standing behind the InterRobot enjoys talking and encouraging children in a new communication mode from a completely different standpoint, just changing to a friend and so forth. By focusing on an animal character, which is the type most preferred by children, an animal-type InterRobot/ InterActor known as InterAnimal has been developed in order to encourage and cheer up children, as depicted in Figure 9.14. The bear-type InterAnimal shown in Figure 9.15 was a popular exhibit in the 2005 EXPO. Figure 9.16 illustrates a toy version of InterRobot with the function of a listener, which generates the listener's actions of nodding, tilting his/her head, and moving his/her arms up and down, based on speech input. The stuffed toy bear is eager to listen without ever uttering a word. It is commercially available and marketed under the name of Unazukikun.

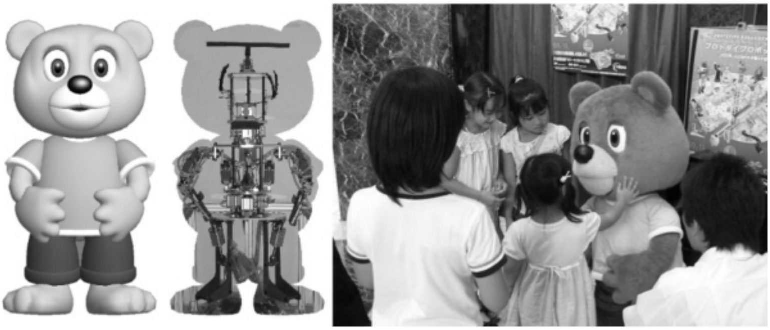


Figure 9.14 InterAnimal: Animal-type InterActor

The InterActor, as indicated in Figure 9.17, is also commercially available under the name of InterCaster through which news and media contents are effectively and cordially transmitted in a commercial program. By superimposing In-

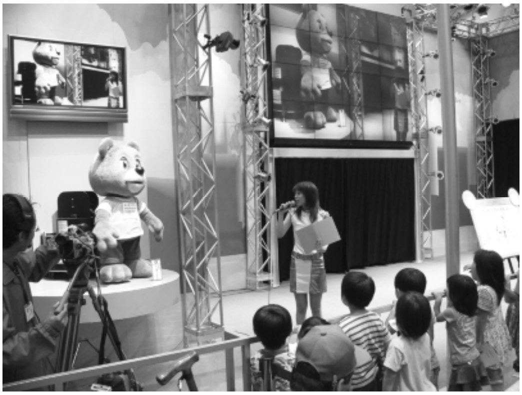


Figure 9.15 InterAnimal in EXPO 2005

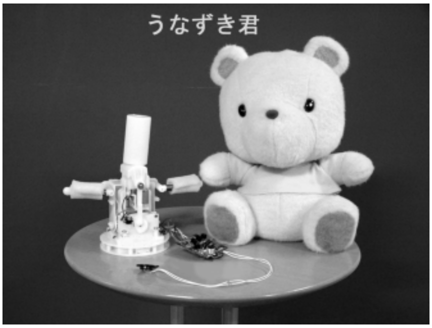


Figure 9.16 Toy version of InterRobot

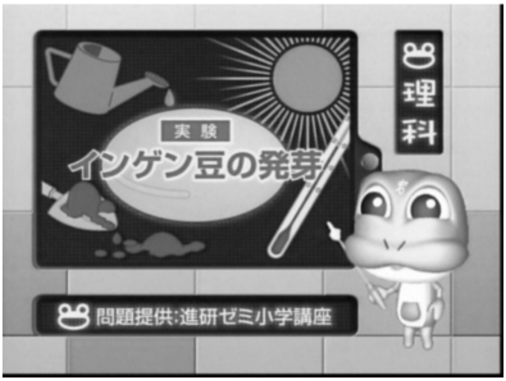


Figure 9.17 InterCaster in an educational program

terActors as listeners on the video images of a lecture such as an education program, the InterActor-superimposed learning support system has been developed, as illustrated in Figure 9.18 [11]. The system provides group-entrained interaction effects for audiences who watch the video, in which two InterActors at the bottom of reduced-size images are entrained with the lecturer's speech.

The InterActor is a speech-driven CG-embodied interaction character that can generate communicative movements and actions for an entrained interaction. An InterPuppet, on the other hand, is an embodied interaction character that is driven by both speech input, similar to the InterActor, and hand motion input, like a puppet. Therefore, humans can use the InterPuppet to communicate effectively by using deliberate body movements as well as natural communicative movements and actions [12]. An advanced InterPuppet with a cellular phone-type device was developed as indicated in Figure 9.19. On the display, two characters – representing the talker and his/her partner – were arranged at an angle of 90° to each other, based on the finding of the conversation arrangement of the InterActors. The

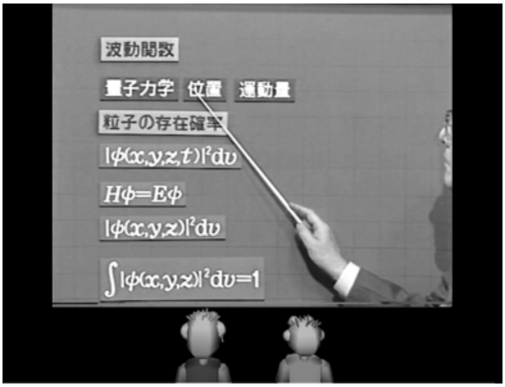
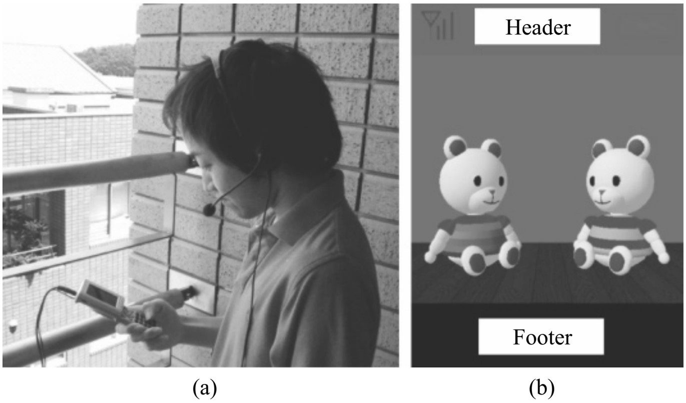


Figure 9.18 InterActor-superimposed learning support system



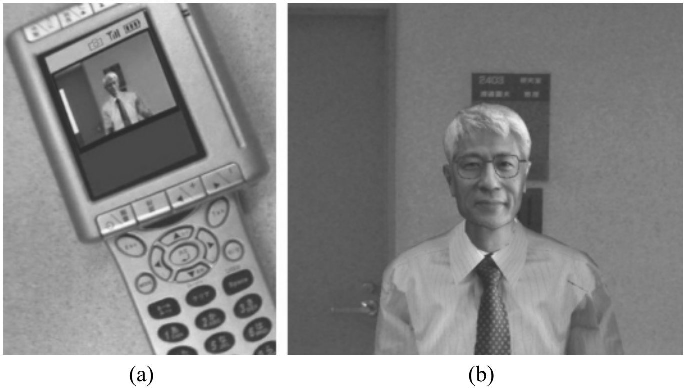
(a) (b)

Figure 9.19 InterPuppet with a cellular phone-type: (a) an image of someone using InterPuppet and (b) screen shot

screen comprises installed headers and footers similar to those in cellular phones. The character resembled an animal type, which could be reminiscent of a doll. When the talker speaks to another user, the talker’s InterPuppet behaves as the speaker and the partner’s InterPuppet responds as a listener. Then, voice is transmitted via the network to the remote partner and the InterPuppet. If the user inputs his/her hand motion, the hand motion is converted to InterPuppet movement and is included in the movements of the InterActor [13].

Figure 9.20 shows the system appearance and the CG image of the system. In this system, communications using a videoconferencing telephone become possible by presenting a real-type CG character with the user's characteristic as well as an anonymous CG character. It is difficult for a videoconferencing cellular telephone to continuously capture the user with the camera when the user speaks while carrying it, because the camera is set in the body of the cellular phone. In addition, unexpected backgrounds and real-time telecasted images cannot be recognized by a videoconferencing telephone. If the InterPuppet characteristic is ideally employed, a real-type CG character can express intentional body movements that include intentions and emotions in addition to entrainment body motion. Therefore, remote face-to-face communications such as videoconferencing telephones will be enabled and involve only speaking and easy key operations, instead of needing to take care of the camera as well as the environment. In addition, the user can prepare his/her own face background and can freely set the conversation arrangement and the aspect.

Figure 9.21 shows the speech-driven embodied entrainment systems InterPointer and InterVibrator which support embodied interaction and communication during presentations [14]. InterPointer is a pointing device with a visualized response equivalent to the nodding response and its response time to the speech input is similar to that of a listener. InterVibrator is a vibration device with a vibratory response by nodding response in the same way. An integrated system of InterPointer and InterVibrator for supporting interactive presentation is developed.



(a) (b)

Figure 9.20 Application to video conferencing telephone: (a) system appearance, and (b) CG image

Figure 9.22 shows an example of a presentation using the system. It is expected to support interactive communication by synchronizing the embodied rhythms using a visualized response and a vibratory response.

A speech-driven embodied entrainment chair system called “InterChair” is being developed for supporting embodied interaction (Figure 9.23). The system generates bodily responses equivalent to nodding in the same timing as listener to speech input. The system naturally forces a user sitting on it to nod with backward–forward motions of 0.01 G. The nodding responses activate embodied interactions among a speaker and listeners around the user.

Furthermore, we have also developed an embodied entrainment character chat system called “InterChat” by introducing an enhanced interaction model that can create motions where the communicative motions and natural actions can be generated easily from both the typing rhythms and the meaning of the words of sending and receiving messages because the typing rhythms resemble the speech rhythms (Figure 9.24). The system is expected to be the basis for a new type of interactive chat communication with embodied entrained characters. The development expands the applications of the embodied interaction and communication technology from a voice input to a typing input.

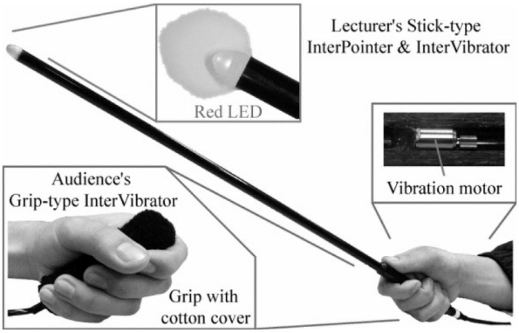


Figure 9.21 InterPointer and InterVibrator

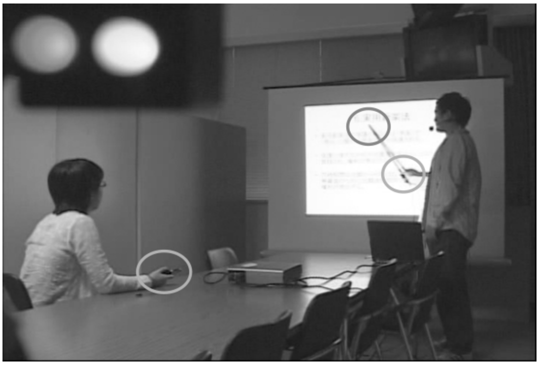


Figure 9.22 Example scene of presentation using InterPointer and InterVibrator

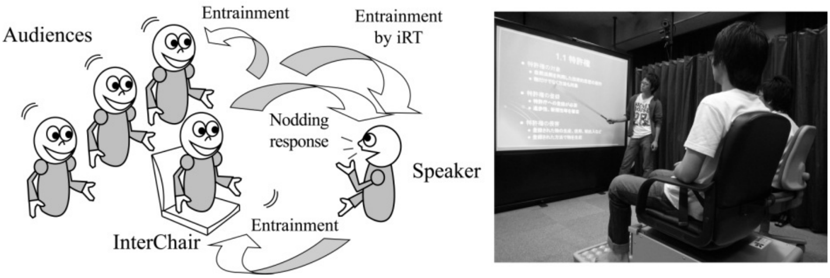


Figure 9.23 Concept of InterChair

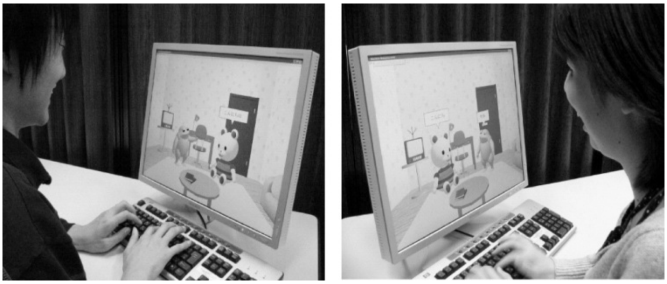


Figure 9.24 InterChat

Figure 9.25 Human-entrained embodied media

With the aim of creating embodied media that unify performers and audiences for supporting the creation of digital media arts for entertainment and education, we will develop a generation and control technology of human-entrained embodied media by developing and integrating the following three technologies: (1) “embodied entrainment media technology” to set embodied media alight with virtual audiences’ entrained responses; (2) “embodied space and image media technology” to integrate and display special media with embodied audiences; and (3) “embodied acoustic media technology” to produce music and embodied acoustics from body motions. Figure 9.25 depicts our embodied media exhibited in the National Museum of Emerging Science and Innovation. What I want to convey with the media is the mystery and importance of embodied interaction and communication.

9.5 Conclusions

The human-entrained embodied interaction and communication technology for an advanced media society was proposed through the development of E-COSMIC for supporting essential human interactive communication based on the entrainment mechanism of the embodied rhythms between speech and body movements such as nodding. Some actual applications pertaining to robot/CG and human interactive communication were also demonstrated. In particular, the speech-driven embodied interaction system, such as InterRobot and InterActor, is a robust and practical communication support system for everyday living, which activates embodied interaction and communication in a new communication mode by using only speech input. The speech-driven entrainment technology for enhancing interaction and communication would be expected to form the foundation of mediated communication technologies as well as the methodology for the analysis and understanding of human interaction and communication, and allow the development of a new embodied communication industry for supporting essential human interactive communication.

References

1. Condon W.S., Sander L.W.: Neonate movement is synchronized with adult speech: interactional participation and language acquisition. Science **183**:99–101 (1974)
2. Kobayashi N., Ishii T., Watanabe T.: Quantitative evaluation of infant behavior and mother–infant interaction. Early Development and Parenting **1**:23–31 (1992)
3. Watanabe T., Okubo M.: Evaluation of the entrainment between a speaker’s burst-pause of speech and respiration and a listener’s respiration in face-to-face communication. In: Proceedings of the 6th IEEE International Workshop on Robot–Human Interactive Communication (RO-MAN’97), pp. 392–397 (1997)
4. Watanabe T.: E-COSMIC Embodied communication system for mind connection. In: Proceedings of the 13th IEEE International Workshop on Robot–Human Interactive Communication (RO-MAN 2004), pp. 1–6 (2004)

5. Watanabe T., Ogikubo M., Ishii Y.: Visualization of respiration in the embodied virtual communication system and its evaluation. Int. J. Human Comp. Interaction **17**(1):89–102 (2004)
6. Shintoku T., Watanabe T.: An embodied virtual communication system for three human interaction support and analysis by synthesis. In: Proceedings of the 5th IEEE International Symposium on Computational Intelligence in Robotics and Automation (CIRA2003), pp. 211–216 (2003)
7. Watanabe T., Okubo M.: An embodied virtual communication system for human interaction sharing. In: Proceedings of the 1999 IEEE International Conference on Systems, Man, & Cybernetics, pp. 1060–1065 (1999)
8. Watanabe T., Okubo M., Nakashige M., Danbara R.: InterActor: speech-driven embodied interactive actor. Int. J. Human Comp. Interaction **17**(1):43–60 (2004)
9. Sejima Y., Watanabe T., Yamamoto M.: Analysis by synthesis of embodied communication via virtual actor with a nodding response model. In: Proceedings of Second International Symposium on Universal Communication (ISUC2008), pp. 225–230 (2008)
10. Watanabe T., Okubo M: SUKURA voice-driven embodied group-entrained communication system. In: Proceedings of HCI International 2003, vol. 2, pp. 558–562 (2003)
11. Watanabe T., Yamamoto M.: An embodied entrainment system with interactors superimposed on images. In: Proceedings of HCI International 2005, vol. 4, pp. 1–6 (2005)
12. Osaki K., Watanabe T., Yamamoto M.: Development of an embodied interaction system with interactor by speech and hand motion input. Trans. Human Interface Soc. **7**(3):89–98 (2005)
13. Osaki K., Watanabe T., Yamamoto M.: A cellular phone type of mobile system with speech-driven embodied entrainment characters by speech and key inputs. Trans. Human Interface Soc. **10**(4):73–83 (2008)
14. Nagai H., Watanabe T., Yamamoto M.: A speech-driven embodied entrainment system with visualized and vibratory nodding responses as listener. Trans. Jpn. Soc. Mech. Eng. **75**(755):163–171 (2009)