Human-entrained embodied interaction and communication technology for human-connected IoT design

Tomio WATANABE*
*Faculty of Computer Science and Systems Engineering, Okayama Prefectural University
111 Kuboki, Soja, Okayama, 719-1197 Japan
E-mail: watanabe@cse.oka-pu.ac.jp

Received: 18 June 2019; Revised: 11 October 2019; Accepted: 30 January 2020

Abstract

In human face-to-face conversation, embodied rhythms between speech and body motions such as nodding are mutually synchronized not only between talkers but also in a talker. This synchrony called entrainment in communication generates the sharing of embodiment in human interaction, which plays an important role in human interaction and communication. Focusing on the embodied entrainment, a human-entrained embodied interaction and communication technology has been developed by applying the entrainment mechanism of embodied rhythms of nodding and body movements to physical robots and CG characters in verbal communication. In particular, the technology for automatically generating communicative motions and actions from voice is put to practical use in communication robots and toys, media contents, e-learning and game software for a wide range of applications such as education, nursing and entertainment. The sense of unity and sharing of this technology supports the happy feelings and security, and it is the key to human interface that humans connect. From the viewpoint of human interface in the advanced media society and super aged society, the human-entrained embodied interaction and communication technology and the design theory of involvement are introduced as the basis of human connected Internet of Things (IoT) smart technology.

Keywords: Human interface, Human interaction, Embodied interaction, Embodied communication, Entrainment

1. Introduction

IoT is an original Internet of Things, from a system in which various things are connected to the Internet and mutually controlled by information communication, and it is expanded to a system that targets not only things but also people (Fig.1). Especially when targeting people, human interface is focused on how to design human-connected interface, such as the development of systems and technologies that promote and support human interaction and communication via information devices and media. The academic foundation of Information Communication Technology (ICT) related to IoT is the Claude Shannon's information theory that mathematically discusses information and communication. When we want to send a message to someone in a remote location, we will probably want to give it exactly and surely. By treating data probabilistically without entering into the importance and content of the message, the technical aspect of information communication on a noisy communication path was made clear by source coding, channel coding, etc., which is the basic concept of information theory we solved the problem. Of course, how to convey that information to people at the end is a big task however it is why we built the academic system by not entering the issue. How to communicate to people, and how to communicate in a form suitable for people based on the communication characteristics of people via the Internet is definitely important for supporting the sense of security and trust in human social systems with a sense of unity. In designing this human interface, first of all, it is important to know how people communicate with each other and the typical face-to-face embodied interaction and communication. From the viewpoint of human interface in the advanced media society and super aged society, on the basis of mainly the author’s publications, the human-entrained embodied interaction and communication technology and the design
2. Embodied communication

In face-to-face communication, not only verbal information, but also nonverbal information such as prosody information, nodding, motions and actions are interactively synchronized in rhythms not only between talkers but also in a talker. This phenomenon is observed in an infant’s movements in response to the mother’s speech as a primitive form of communication (Condon and Sander, 1974; Kobayashi et al., 1992). This synchrony of embodied rhythms in communication, referred to as entrainment, generates the sharing of embodiment in human interaction and communication. By sharing and entraining each other, we communicate smoothly (Cacioppo et al., 2014; Cornejo et al., 2017). Physiological aspects such as entrainment of heart rate variability and respiration closely related to emotional change also play an important role in interaction (Feldman et al., 2011). Interaction and communication through the entire body, including nonverbal information and physiological information, is called embodied communication, which establishes a relationship through each other's body. This sharing of embodiment creates a sense of unity and makes us feel a relationship with people. As we have acquired the language culture with the entrainment between the early childhood and the mother (parents), it may be extremely difficult to transmit and receive information in emotional empathy without the sharing of a sense of unity and embodiment by the entrainment of this embodied rhythm (Tramacere and Ferrari, 2016). Therefore, if this embodied communication mechanism is introduced to human interface, it is highly expected that an embodied communication system based on human characteristics can be realized. One major goal of the system development is to construct an embodied communication interface that has a sense of unity for mind connection through information devices and media for human connected IoT.

3. Embodied communication system for mind connection E-COSMIC

As a basis of human communication research, first focusing on the primary interaction and communication between mother and neonate, the author analyzed voice and image of the movement of the neonate's limbs in response to the mother's speech, and tried to apply the mechanism of the interaction to man-machine interface in 1978. Since then the author has been fascinated by the wonders and possibilities of this embodied interaction and communication for 40 years. During this time, we have analyzed and modeled the interaction between mother and neonate, the interaction and communication between adults, applied the entrainment of embodied rhythms such as nodding and gesture to the robot and CG character, developed the “Embodied Communication System for Mind Connection (E-COSMIC)” that allows us to feel a sense of unity, and we have developed the human-entrained embodied interaction and communication technology (Watanabe, 2011). The development is based on the project "E-COSMIC" as the research leader in the research area "Life information technology of advanced media society" by Core Research for Evolutional Science and Technology of Japan Science and Technology Agency (JST CREST).

An outline of E-COSMIC is shown in Fig.2. This system comprises an embodied virtual communication system for analyzing and understanding interaction and communication via each talker's avatar in a virtual environment where various types of communication information can be controlled, and a speech-driven embodied interaction system that

Fig.1 Human connected IoT.
automatically generates communicative motions and actions and supports communication based on the analysis that uses the embodied virtual communication system.

**Speech-driven embodied interaction system**

![Embodied Communication System](image)

3.1 Embodied virtual communication system

An embodied virtual communication system is a system that synthetically analyzes embodied communication by processing nonverbal information and physiological information of talkers in a virtual environment that can control various types of communication information (Watanabe et al., 2004a). The spatial arrangement, position, and background of the virtual actor (VA), an embodied avatar of the talker, can be freely changed to express information such as body movements, prosody, pupil responses, gaze, and laughing expressions (Fig. 3). Figure 4 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. 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. It is possible to analyze embodied interaction and communication by removing, adding, processing various information such as the influence of time lags. In particular, by using the contradictory induction method that contradicts the actions of a talker and the VA in the unconscious perceptual-motor system to cause the contradiction, the process of the contradiction being resolved can be investigated (Fig.5). For example, if the listener's VA head movement stops inconsistently, or if a nod response is added inconsistently, the effects of the head movement or nod response in embodied communication can be clarified for the first time by the experience/verification through the system. By associating and generating both of the conversational partner’s VA and one's own VA in the same space on the Internet, the interaction is promoted with the sharing of embodiment, so that communication can be smoothly performed.
3.2 Speech-driven embodied interaction system

Based on the analysis results of the embodied virtual communication system, focusing on embodied rhythms such as nodding and gestures, the speech-driven embodied system with embodied CG character InterActor or the physical interaction robot InterRobot which automatically generates the communicative motions and actions as the speaker and the listener has been developed (Watanabe et al., 2004b).

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 (Fig. 6). The timing of nodding is predicted using a hierarchy model consisting of two stages: macro and micro. The macro stage estimates whether a nodding response exists or not in a duration unit, which consists of a talkspurt episode $T(i)$ and the following silence episode $S(i)$ with a hangover value of $4/30$ s. The estimator $M_u(i)$ is a moving-average (MA) model, expressed as the weighted sum of unit speech activity $R(i)$ in (1) and (2). When $M_u(i)$ exceeds a threshold value, nodding $M(i)$ also becomes an MA model, estimated as the weighted sum of the binary speech signal $V(i)$ in (3). The body movements are related to the speech input by operating both the neck and one of the wrists, elbows, arms, and waists at the timing over the body threshold. The threshold is set lower than that of the nodding prediction of the MA 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.

$$M_u(i) = \sum_{j=1}^{j} a(j)R(i-j) + u(i)$$  \hspace{1cm} (1)

$$R(i) = \frac{T(i)}{T(i) + S(i)}$$  \hspace{1cm} (2)

$a(j)$: linear prediction coefficient  
$T(i)$: talkspurt duration in the $i$-th duration unit  
$S(i)$: silence duration in the $i$-th duration unit  
$u(i)$: noise

$$M(i) = \sum_{k=1}^{K} b(j)V(i-j) + w(i)$$  \hspace{1cm} (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 the operation when the power of speech is over the threshold.

We also developed the speech-driven embodied group-entrained communication system SAKURA that uses
multiple InterActors, as well as its use as a human-to-human or human-to-robot interactive interface (Watanabe and Okubo, 2003). In the educational effect experiment using SAKURA, the transmission effect is significantly different between the group who watched the system performing entrained reaction and the group who watched the system performing dull reaction from the same recorded speech. The very characteristic result which shows the importance of interaction-activated communication was confirmed. While InterActor as an electronic character is expected to be widely applied as a GUI, the physical version of SAKURA with InterRobots as a physical medium dramatically enhances the sharing of embodiment by entrainment as well as presence. The system was also demonstrated repeatedly at various events in Japan and abroad, and was exhibited regularly in the National Museum of Emerging Science and Innovation, which was highly valued by many people, including not only children but also elderly people, not only as the theory of physical communication effects but also as the permanent demonstration system (Fig. 7). With the nodding reaction of InterRobot listening to a child's proud story with enthusiasm, we can experience the possibility of the system technology that makes us feel happy.

E-COSMIC, which integrates these two systems, i.e. the embodied virtual communication system and the speech-driven embodied interaction system, is the embodied communication system for mind connection that can share embodiment by entrainment through the body in the super-aging society and advanced media society. This system technology was one of the representative results of research on the follow-up of the 3rd Science and Technology Basic Plan by the Ministry of Education, Culture, Sports, Science and Technology Science and Technology Policy Research Institute, and was selected as the human-entrained communication technology in "Innovator Japan: the realization of a ubiquitous society that attracts the world". This body-to-body communication system is also a semantic generation system that generates and creates meaning in interaction, and is the fundamental technology/system of human interface that people connect with. In particular, the system/technology that automatically generates rich communication operations from voice is called iRT (InterRobot technology), is put to practical use in communication robots, toys, media content, e-learning, game software, etc., for a wide range of applications such as education, nursing and entertainment as shown in Fig. 8.
4. Generation and control technology of human-entrained embodied media

The placement of multiple InterActors and InterRobots creates an environment for interaction with group entrained effects. The speech-driven embodied group-entrained communication system SAKURA is named after "Sakura", which is important for interaction-activated communication. In the recording of various performances and talk shows, if the audiences fully exerts the function of so-called Sakura, the exciting audiences support creating a media work that attracts themselves by maximizing the performance of the actor by the sense of unity between the actor and the audiences. The actor can perform naturally with the best performance by sharing the context and feeling the sense of unity. The exciting method of using the audience as Sakura has a great influence on the quality of the media work, as the success or failure of the audiences as Sakura is directly reflected in the performance of the actor. Therefore, a system for synthesizing, analyzing, processing, and evaluating in virtual and real environments by group embodied entrainment control based on the experience of this field would be expected to stably support the production of the highest quality 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 developed 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 (Watanabe, 2018).

The system generates group embodied entrainment by virtual audiences such as CGs and robots, and also by adding virtual audiences to the real audience, not only the actor but also audiences can be entrained to maximize the performance of the actor. The generation and control technology of human-entrained embodied media can be developed as the fundamental technology of human interface that people connect as content generators on the Internet.

People communicate to connect. To do that, communication itself must be a mechanism for the brain and body to rejoice. Certainly, conversations that are full of sense of unity and sharing are really fun. It is a pleasure that the other person is involved with us and synchronized with our rhythm. This embodied entrainment is the essence of communication. The embodied interaction toys "Pecoppa" and “Hanappa”, which respond with an appropriate timing of nodding to speech sounds, were commercialized by applying the embodied entrainment technology (Fig.9). They make a nodding response based on speech input and support the sharing of mutual embodiment in communication. They use a material called BioMetal made of a shape memory alloy as their driving force. We can experience the importance and possibility of embodied communication, and the essence of communication (Giannopulu et al., 2016; Giannopulu et al., 2018; Asada, 2015)
5. Design of interaction-activated communication

The embodied communication system for mind connection E-COSMIC and the generation and control technology of human-entrained embodied media have been proposed as the human-entrained embodied interaction and communication technology. It focuses on the embodiment and entrainment in human interaction and communication, and smoothly promotes embodied interaction and communication through information devices and media for human interface connecting people. Through information devices, human interaction and communication that are difficult to analyze directly can be synthetically modeled, enabling the development of a high-affinity interface based on human essential characteristics. An example is the embodied virtual communication system for synthetically analyzing embodied communication by measuring and processing nonverbal information and physiological information of talkers in a virtual environment capable of controlling various types of communication information. Embodied communication technologies that enhance realism have been also developed such as five-sense communication technology, robot media technology, and mutual communication technology that uses the presence of each other, focusing on the shadow of the body that is inseparably and unconsciously awakened (Jung et al., 2018; Leite et al., 2013; Sejima et al., 2016). These research and development lead to a sense of unity that is the premise of communication, and also a sense of security and credibility, and is expected as an emerging trial of the design theory of interaction-activated communication.

Since Descartes, leaps and bounds in science and technology have been brought about by separating and objectively looking at self and objects. However, these achievements are essentially the achievements in the area where human beings are not involved, and there is a big contradiction in thinking that this methodology can be applied to humans or society as it is. In the field of human interface for truly human beings, communication and interaction in particular are clearly inadequate with the self-other separation method, and a design theory related to creating an interaction environment is desired. The human-entrained embodied interaction and communication technology has been proposed as a self-other non-separating research method and system. In order to systematically analyze and understand the embodied communication characteristics underlying the interaction, as suggested by the author, it is essential to develop systems and technologies from the perspective of self-orientation and non-separable communication that captures the interaction environment and positions oneself from the environment where the interlocutor is both an observer of the dialogue and an operator of the dialogue information.

The human-entrained embodied interaction and communication technology is a powerful tool for communication support based on the acquisition of embodied rhythms that are essential for language acquisition from infants, and structurally approaches to elucidating embodied communication for connecting people and communicating thoughts via information devices. There is nothing superior to meeting people with each other, however, other people to actually meet each other. We have developed with the thought, however like communication via InterActors and InterRobots, it can contribute to the embodied interaction promotion and embodied communication support beyond the actual meeting in some cases.

The present Internet has various problems due to its embodiment being suppressed, although information exchange and message exchange are carried out. The space-time of communication on the Internet is obviously generated by each communicator individually, and the mechanism for sharing the space-time is embodied entrainment. Thus the introduction of embodied interaction and communication technology will be a key to the sound development of the IoT in which things and people will be closely connected. For embodied communication for mind connection, it
is essential to share the sense of unity and embodiment through embodied entrainment of the whole body. With the rapid development of information and communication infrastructure towards a full-fledged IoT era (Bandyopadhyay et al., 2011; Wang et al., 2016), the human-entrained embodied interaction and communication technology is a key technology for supporting rich and high quality living enjoying interaction and communication, and would contribute to the development of systems that highly integrate cyberspace and physical space as the basis of human connected IoT smart technology.

References

Asada, M.: Towards Artificial Empathy, International Journal of Social Robotics, Vol.7, No.1 (2015), pp.19-33.

Bandyopadhyay, D. and Sen, J.: Internet of things: Applications and challenges in technology and standardization, Wireless Personal Communications, Vol.58, No.1 (2011), pp.49-69.

Cacioppo, S., Zhou, H., Monteleone, H. Majka, E.A. et al.: You are in sync with me: neural correlates of interpersonal synchrony with a partner, Neuroscience 277 (2014), DOI: 10.1016/j.neuroscience.2014.07.051.

Condon, W.S. and Sander, L.W.: Neonate movement is synchronised with adult speech: interactional participation and language acquisition, Science, No.183 (1974), pp.99-101.

Cornejo, C., Cuadros, Z., Morales, R., and Paredes, J.: Interpersonal coordination: methods, achievements, and challenges, Frontiers in Psychology, 8:1685 (2017), DOI: 10.3389/fpsyg.2017.01685.

Feldman, R., Galili, G., Singer, M., and Louzoun, Y.: Mother and infant coordinate heart rhythms through episodes interaction synchrony, Infant Behavior and Development. Vol.34 (2011), pp.569–577.

Giannopulu, I., Montrrynaud, V. and Watanabe, T.: Minimalistic toy robot to analyze a scenery of speaker-listener condition in autism, Cognitive Processing, Springer (2016), DOI: 10.1007/s10339-016-0752-y.

Giannopulu, I., Terada, K. and Watanabe, T.: Communication using robots: a Perception-action scenario in moderate ASD, Journal of Experimental & Theoretical Artificial Intelligence (2018), DOI: 10.1080/0952813X.2018.1430865.

Jung, H. and Song, Y.E.: Robotic remote control based on human motion via virtual collaboration system: A survey, Journal of Advanced Mechanical Design, Systems, and Manufacturing, Vol.12, No.7 (2018), DOI: 10.1299/jamdsm.2018jamdsm0126.

Kobayashi, N., Ishii, T. and Watanabe, T.: Quantitative Evaluation of Infant Behavior and Mother-Infant Interaction, Early Development and Parenting, Vol.1, No.1 (1992), pp.23-31.

Leite, I., Martinho, C. and Paiva, A.: Social Robots for Long-Term Interaction: A Survey, Journal of Social Robotics, Vol.5, No.2 (2013), pp.291-308.

Sejima, Y., Watanabe, T. and Jindai, M.: Estimation Model of Interaction-activated Communication based on the Heat Conduction Equation, Journal of Advanced Mechanical Design, Systems, and Manufacturing, Vol.10, No.9 (2016), Paper No.15-00548, pp.1-11.

Tramacere, A., and Ferrari, P. F.: Faces in the mirror, from the neuroscience of mimicry to the emergence of mentalizing, Journal of Anthropological Sciences, Vol.94 (2016), pp.113–126.

Wang, L. and Wang, G.: Big data in cyber-physical systems, digital manufacturing and industry 4.0’, International Journal of Engineering and Manufacturing (IJEM), Vol.6, No.4 (2016), pp.1-8.

Watanabe, T.: Human-entrained Embodied Interaction and Communication Technology, Emotional Engineering, Springer (2011), pp.161-177.

Watanabe, T., Ogikubo, M. and Ishii, Y.: Visualization of Respiration in the Embodied Virtual Communication System and Its Evaluation, International Journal of Human-Computer Interaction, Vol.17, No.1 (2004a), pp.89-102.

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 (2004b), pp.43-60.

Watanabe, T.: Embodied Communication for Mind Connection, Journal of the Japan Society of Mechanical Engineers, Vol.121, No.1195 (2018), pp.14-17 (in Japanese).

Watanabe, T. and Okubo, M.: SUKURA: Voice-Driven Embodied Group-Entrained Communication System, Proc. of International Conference of Human-Computer Interaction 2003, Vol.2 (2003), pp.558-562.