WCR: A Wearable Communication Recorder Triggered by Voice for Impromptu Communication

Tomoo Inoue* and Yuriko Kourai*

*Graduate School of Library, Information and Media Studies, University of Tsukuba, Japan
inoue@slis.tsukuba.ac.jp, s0921731@u.tsukuba.ac.jp

Abstract - Everyday communication is not always planned in advance. There is a type of communication called "opportunistic communication" or "impromptu communication" that happens unintended in any place and in any moment. The conventional video recording equipment cannot capture this type of communication because it has to be set up in advance. This paper presents the development of a wearable communication recorder (WCR) for impromptu communication. To cope with a problematic issue of dealing with a large amount of video data in life log systems, the proposed wearable recorder records interpersonal communication selectively. It gets the recording cue from the first utterance of the user in a conversation and records the communication by video and audio as far back as some 30 seconds of the utterance. This can reduce useless recording and can record whole communication session at the same time. From the analysis of interpersonal communication, 10 seconds of backward recording is suggested to be acceptable. The performance of WCR was also examined in terms of the proper recognition of communication, which indicated over 90% of the recall rate.

Keywords: Wearable computing, Communication Research, Impromptu Communication, Nonverbal, Multimodal.

1 INTRODUCTION

Research on interpersonal communication has been actively conducted since 1950's, and its major progress was brought by the video recording technology. Observation was almost the only way for a researcher to study interpersonal communication which never happens again before the use of the video recording technology. The video recording technology allowed a researcher to investigate a piece of communication activity repetitively and to analyze it in fine time sequence by frame-by-frame observation. In other words, detailed communication analysis cannot be achieved without video-recorded communication.

It is well known that there are two types of interpersonal communication. One occurs at predetermined time and place while the other occurs opportunistically without prior appointment. Interpersonal communication that has been analyzed so far by video data falls into the former. Conventional video recording cannot capture the latter type of communication because the video recording equipment should be set up before the communication starts.

However, the video recording equipment has progressed. Wearable video recording equipment can be found recently partly owing to the wearable computing. It has been used in the life log research, for example.

We propose a wearable communication recorder to get the video data of the latter type of communication, or impromptu communication, due to the importance of the video recording of the communication when studying it. Impromptu communication can be recorded by using such a wearable video equipment. A problem is how to deal with a huge amount of recorded video when the video is always on. This is one of the major research issues in life log. Our approach is to record impromptu communication alone automatically.

The rest of the chapters are composed as follows. The related research is explained in Chapter 2. Our recorder is proposed in Chapter 3. The implementation of the recorder is described in Chapter 4. Daily informal communication is examined for the design of the recorder in Chapter 5. The initial investigation of impromptu communication is reported in Chapter 6, and the basic performance of the recorder is validated in Chapter 7. Conclusion is given in Chapter 8.

2 RELATED RESEARCH

2.1 Impromptu Communication

Impromptu communication is a type of informal communication. Informal communication is known to be important to keep and nurture human relationships and to facilitate group work. Because of this importance, there are many systems to support informal communication. VideoWindow is an early research system to support informal communication between distributed office rooms by the audio and video links [1]. Cruiser is an early research system to support informal communication between distributed desktops [2].

There are many other systems to support informal communication since then. However they are not for recording impromptu communication.

2.2 Communication Recording

Video recordings of communication have been made frequently for communication analysis.

Examples of such video use are the analysis of multi-party conversation by video recordings [3] and the analysis of body movement synchrony in psychotherapeutic counseling by video [4]. The video is recorded manually in these cases. The video recording of multi-party conversation in a meet-

ing with the automatic speaker identification has been realized recently [5]. However, all these are for communication which is scheduled in advance and which takes place at a fixed location.

There are systems to record communication in a limited area.

The Active Badge system is an early system to know the locations of the entire users. The user wears an infra-red transmitter named Active Badge. The networked sensor that is installed one or more in each room detects the Active Badge thus tells the location [6]. From the closeness of two or more badges, the system user is able to know the meeting of the people.

The Bat system is a subsequent of the Active Badge system. Ultrasound is emitted from the transmitter that the user wears and a number of receivers are installed on the ceiling, which makes high precision of locating possible [7].

Interactions between the user and the object can be recorded automatically using the infra-red markers on the objects and the infra-red receivers worn on the users in the limited area which the system is installed [8].

All these can record interactions more or less, but the markers are necessary. This means that the recording area is limited within the marker area.

Because impromptu communication occurs at any place, we cannot apply these systems to record impromptu communication.

2.3 Life Log Research

A life log system is an always-on recorder of various events, user's behavior, and operated objects. Some of them take the video logs. They can be very large data because of the always-on feature of the life log systems. Dealing with very large data is not easy. Additionally, most of the data is useless when the user would like the video of a specific event. Thus information search or retrieval from this very large data is a major topic of research. Because of this, it may not be the best to record everything when the type of information to be needed is known. Our research is different from generic life log research and is focused on the recording of interpersonal communication.

Mixture of the video and sensor information has also been researched because image processing alone is not always enough to recognize objects in the video or to understand the context. Sensor information can be the annotation to the video recordings. Interpersonal communication has not been focused on in this type of research.

3 PROPOSAL OF A WEARABLE COMMUNICATION RECORDER

We propose a wearable communication recorder (WCR) to get the video data of impromptu communication. Because impromptu communication does not happen at a fixed place, conventional video recording equipment cannot fulfill the requirement of mobility. People also do not know when impromptu communication occurs because it happens without appointment. Typical impromptu communication is a short talk when two people come across in the hallway. Accordingly, the recording equipment should record interpersonal communication without the limitation of time and place. To fulfill this requirement, wearable equipment is applied.

Recording everything like a life log system may cause a problem of dealing with a large amount of video data although it is easy to do so. Recording communication alone is an option to avoid this problem. However it is disturbing for the user to turn on and off of recording manually every time communication happens. It is desirable that the system automatically turn on and off.

One of the important issues for the recording of interpersonal communication is the detection of initiation of communication and completion of communication. For the automatic recording of communication, utterance can be thought of the good candidate of the recording cue. However, it is not so simple to start recording when detecting the user's voice and to stop the recording when detecting the end of the voice. It is known that communication usually starts before the first utterance by eye contact or salutation [9]. In accordance with this, the recorder should record nonverbal cues which occur before the verbal utterance. Detecting these nonverbal cues robustly is not very easy in itself, and it needs measuring instruments that are not easy to wear.

To cope with this issue, our recorder applies the mechanism of a driving recorder. The driving recorder is a device commonly used in the business cars such as taxis and trucks. It records the video around the car as the evidence when it detects an impact by car accident or sudden braking. Because it is not good for the evidence video to start recording from the time of the impact, the video data is always buffered about a minute. The buffered data is then saved when the impact is detected by the accelerometer sensor.

The recording of interpersonal communication is thought to be possible by the voice detection and the mechanism similar to the recording mechanism of the driving recorder. The voice detection can be the trigger of saving buffered video data that include nonverbal cues which arise before verbal utterance. For completing the recording, it is also thought to be possible by extending the recording some more time after detecting the end of the voice.

With this explained mechanism, WCR is presumably able to record the whole communication session in any time and place while reducing the useless data.

4 IMPLEMENTATION OF THE WEARABLE COMMUNICATION RECORDER

4.1 Equipment

A USB camera for video communication (Microsoft LifeCam VX-6000) is used for the video recording. It has the 1.3 million pixel CMOS sensor that gives clear image. The view angle is 71 degree. It is with automatic adjustment of white balance that responds to the brightness. It is with directional noise-canceling microphone.

The built-in directional microphone of the USB camera is used for recording communication. Another throat microphone is used for assuring the detection of the user's utterance.

Figure 1: Appearance when wearing the system.

These are connected to the small laptop PC (Fujitsu LOOX P70-XN), which records the input information.

Figure 1 shows the appearance of the user when wearing the equipment.

4.2 Software

The operating system of the PC is Windows XP Professional. The program is written in C++. OpenCV 1.0 is used for video processing. Windows Multimedia API is used for audio processing.

WCR keeps detecting audio signal from the throat microphone. The video from the USB camera and the audio from the built-in microphone of the camera are buffered for some 30 seconds. When the audio from the throat microphone becomes bigger than a certain threshold, it is recognized as the utterance of the user of the system. Then the buffered video and audio is saved to the file, which is supposed to be the recording of the nonverbal communication before the utterance. On the other hand, when the audio from the throat microphone becomes and keeps less than a certain threshold after detecting the utterance, it is recognized as the completion of the utterance. The recording stops after a certain period of time from the completion.

The video is encoded in real time and saved as MotionJPEG with the resolution of QVGA. The each audio is saved as 8 bit monaural RIFF with 22kHz sampling rate.

Figure 2: Devices used for recording daily informal communication: Omnidirectional camera (left) and a conference microphone (right).

5 INVESTIGATION OF DAILY INFORMAL COMMUNICATION

5.1 Aim

To know the appropriate buffering time in WCR, communication was investigated. Although the actual target of WCR is impromptu communication, daily informal communication was investigated because the targeted communication cannot be captured without a wearable continuous recording system such as our system and because daily informal communication is thought to be more similar to impromptu communication than other types of communication.

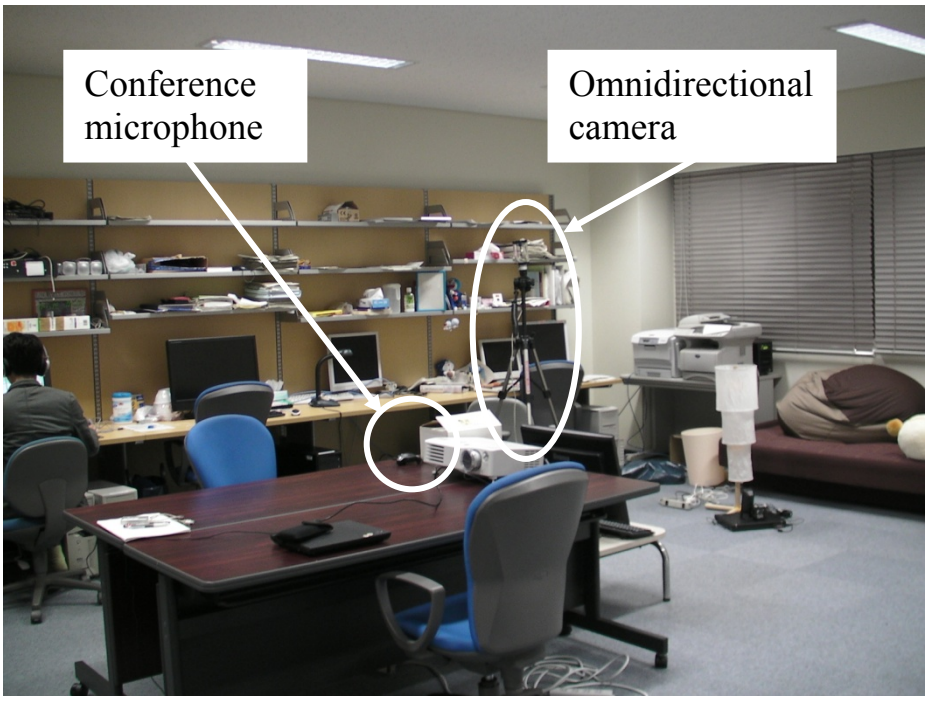

Figure 3: Recording environment of daily informal communication.

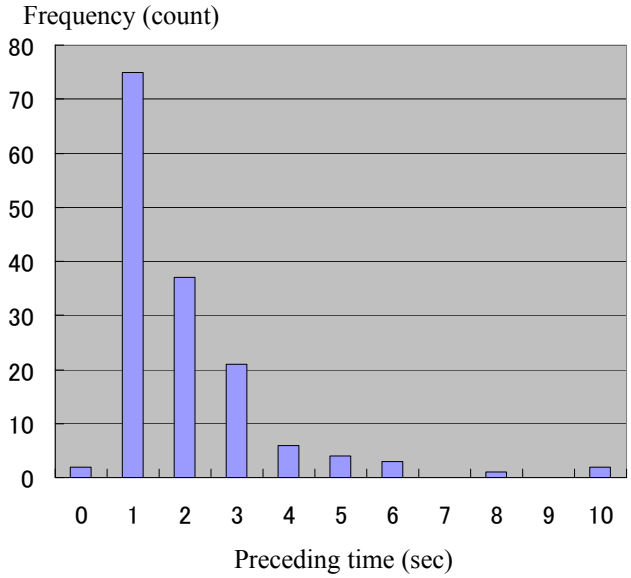

Figure 4: Preceding time of communication behavior before the first utterance.

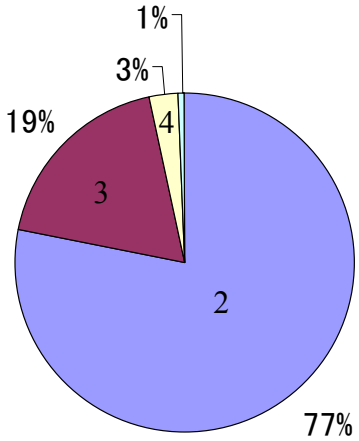

Figure 5: Number of participants in the communication.

5.2 Method

The omnidirectional camera (VSTONE VS-C42U-300-TK) and the conference microphone (BUFFALO BSKP-CU201) (Figure 2) were set up at the room which is approximately 9 m x 6 m (Figure 3). Communication in the room was recorded 27.5 hours in 5 days. The video was recorded in 2048 x 1536 (QXGA) with 6 fps. The elevation angle of the camera view is approximately 15 degree and the depression angle is approximately 60 degree. The microphone is omnidirectional with the noise reduction.

5.3 Result

For the every communication, the time of communication behavior before the first utterance was clocked. The result is shown in Figure 4. The total number of communication was 151. The average time of communication behavior before the first utterance was 2 seconds. The first utterance came less than 3 seconds from the beginning in nearly 90% of communication. The longest interval between the beginning of communication and the first utterance was 10 seconds. It

Table 1: Time between the initiation of communication and the first utterance of the user

communication ID	1	2	3	4	5	6	7	8	9
time (sec.)	1.5	0.0	4.6	3.7	3.0	4.1	2.8	0.7	2.5
communication ID	10	11	12	13	14	15	16	17	
time (sec.)	4.0	0.0	3.2	1.2	4.4	2.0	1.8	1.4	

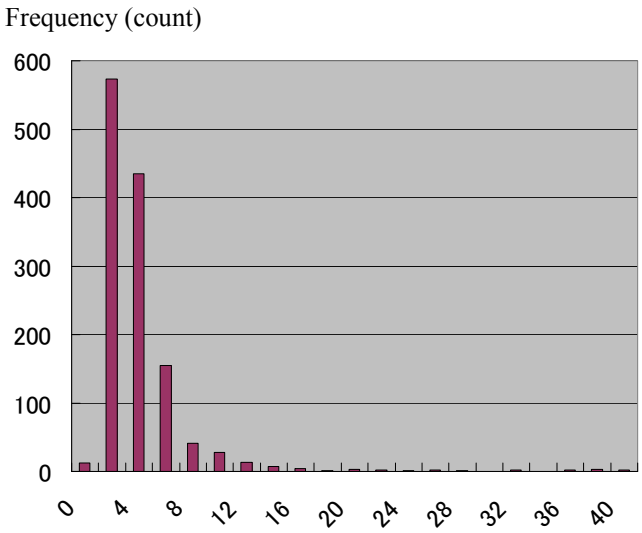

Figure 6: Intervals between the utterances.

seems to be reasonable that the buffering time of the WCR is 10 seconds.

Additionally, the number of participants in the communication is shown in Figure 5. Dyadic communication occupies over 3 out of 4 communications, and most of the rest was triadic communication.

6 INITIAL USE OF THE WEARABLE COMMUNICATION RECORDER

6.1 Aim and Method

To examine the feasibility of the system, we used it on trial. The buffering time was set 30 seconds. This means that the video and audio were recorded from 30 seconds back before the trigger of the first utterance. From the trial data, it was examined how long initiation of communication come before the first utterance of the user. The duration between the utterances was also examined to get the clue of the completion of communication.

The trial user was one male university student. He wore WCR a day from 10am to 10pm. He communicated with others as same as everyday.

6.2 Result

The total recording time was 73 minutes. The speaker and the time of utterance, initiated time of communication by observation were annotated to the video and audio data. This was done by the video annotation software named CIAO.

The communication occurred 17 times in the data. The moment the user became aware of the communication target was regarded as the initiation of communication. The time between the initiation of communication and the first utterance of the user is shown in Table 1. Because the sample number is small, its distribution should not be concluded. But it is suggested that most utterance began within 5 seconds from the initiation of communication from Table 1. All the communications were dyad. The time may become longer in the communication by more than 3 people. It is suggested that preparation of 30 seconds of buffering time is long enough and probably could be shorter, which is easy to implement.

Then intervals between utterances were examined. It is shown in Figure 6 with the scale of 2 seconds for X axis. The most frequent interval was between 2 to 4 seconds. Over 90% was within 8 seconds and over 98% was within 16 seconds. There were data over 30 seconds, but they were about different topics or talking himself. The longest interval in a sequence of conversation was 20.2 seconds. It is suggested that the recording can stop about 20 seconds after the completion of the last utterance.

Although this initial use is not enough to give the detailed data about communication, feasibility of the system was confirmed.

7 PERFORMANCE EVALUATION OF THE WEARABLE COMMUNICATION RECORDER

7.1 Aim and Method

The reliability of the automatic recording of WCR was examined in terms of the recognition of communication.

A female university student wore WCR in a room 13 hours in 2days. The room and the setup were the same with what was described in Section 5. All communication in the room was recorded by the fixed camera and the conference microphone. This recording became the reference to know the communication in the room. The recording by WCR was matched with the reference recording. An example of the fixed camera recording and WCR recording is shown in Figure 7. The reliability of WCR was evaluated this way. The buffering time was set to 10 seconds. The recording stopped automatically after 30 seconds from the last utterance.

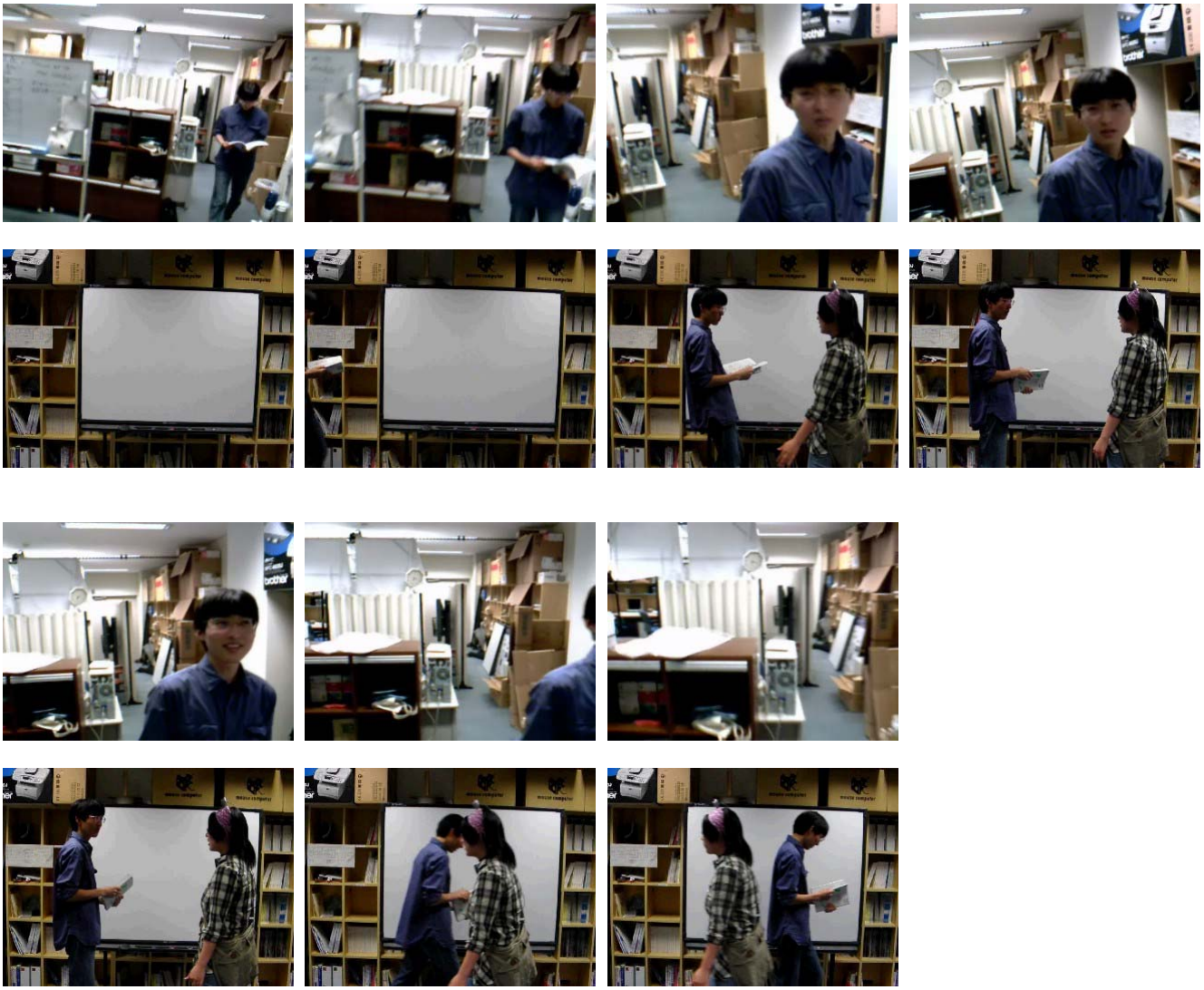

Figure 7: Example of WCR recording and the fixed camera recording (Top: WCR recording, Bottom: fixed camera recording. Time-line from left to right).

7.2 Result

The number of the communication with the user was 32. The number of the recordings by WCR was 39, which means WCR recognized these recordings as communication occurrences. Among these 39, the number of communications that were recorded properly was 30.

This shows that WCR could record 30 out of 32, i.e. 94 % of the communication events properly. The rate is known as the recall rate. One of the two communications that WCR failed to pick up was because of the too small voice of the user. The utterance volume was under the detection threshold. The other communication that WCR failed to record properly included the utterance interval over 30 seconds. WCR recognized this communication as two different communications.

WCR also sometimes mistakenly recognized the noise event as communication. The rate of the properly recorded communication among all the recordings was 30 out of 39, i.e. 77 %. The rate is known as the precision rate. WCR applies the throat microphone for the user's voice detection. The recording is controlled by this voice detection. Because the throat microphone does not pick up the environmental noise, the environmental noise hardly effect to the result. The number of people in the room was seven at a maximum. The recognition errors were due to the sound the user made such as sneezing and cough.

8 CONCLUSION

We proposed and implemented the wearable communication recorder (WCR) to get the video data of impromptu communication. Different from other life log recorder or conventional video recording equipment, WCR focuses on recording impromptu communication efficiently. The video recording has played an important role in communication analysis but has been applied to the limited communication setup. The significance of WCR is to expand its target. Al-

though current prototype has points to be improved such as its appearance or the functions as a wearable computer, it is expected to advance the analysis of impromptu communication.

REFERENCES

[1] R. S. Fish, R. E. Kraut, R. W Root, and R. E. Rice, Video as a technology for informal communication, Communications of the ACM, Vol.36, No.1, pp.48-61 (1993).
[2] R. W. Root, Design of a multi-media vehicle for social browsing, Proceedings of CSCW'88, pp.25-38 (1988).
[3] K. Ueda, S. Yoshikawa, Y. Den, C. Nagaoka, Y. Ohmoto, and M. Enomoto, Analysis and modeling of conversation, Journal of Japanese Society for Artificial Intelligence, Vol.21, No.2, pp.169-175 (2006). (in Japanese)
[4] C. Nagaoka, and M. Komori, Body movement synchrony in psychotherapeutic counseling: a study using the video-based quantification method, IEICE Transactions, Vol.E91-D, No.6, pp.1634-1640 (2008). (in Japanese)
[5] AIST Press Release, http://www.aist.go.jp/aist_j/press_release/pr2008/pr20081014_2/pr20081014_2.html (2008) (in Japanese)
[6] R. Want, A. Hopper, V. Falcao, and J. Gibbons, The active badge location system, ACM Trans. on Information Systems, Vol.10, No.1, pp.91-102 (1992).
[7] M. Addlesee, R. Curwen, S. Hodges, J. Newman, P. Steggles, A. Ward, and A. Hopper, Implementing a sentient computing system, IEEE Computer, Vol.34, No.8, pp.50-56 (2001).
[8] Y. Sumi, S. Ito, T. Matsuguchi, S. Fels, and K. Mase, Collaborative capturing and interpretation of interactions, Trans. IPSJ, Vol.44, No.11, pp.2628-2637 (2003). (in Japanese)
[9] Tang, J. C., Approaching and Leave-Taking: Negotiating Contact in Computer-Mediated Communication, ACM Trans. on CHI, Vol.14, No.1, Article5 (2007).

(Received August 29, 2009)
(Revised August 10, 2010)

Tomoo Inoue is Associate Professor of the Graduate School of Library, Information and Media Studies at University of Tsukuba. His research interests include understanding and supporting human interaction, CSCW, and advanced learning support systems. He received Ph.D. from Keio University in 1998. He is a recipient of the Best Paper Award and the SIG Research Award from the Information Processing Society of Japan (IPSJ).

Yuriko Kourai is a graduate student of the Graduate School of Library, Information and Media Studies at University of Tsukuba. Her research interests include communication support systems.