

Development of a wearable communication recorder triggered by voice for opportunistic communication

Tomoo Inoue* and Yuriko Kourai*

*Graduate School of Library, Information, and Media Studies, University of Tsukuba, Japan
inoue@slis.tsukuba.ac.jp, s0512184@u.tsukuba.ac.jp

Abstract – Everyday communication is not always planned in advance. There is type of communication called “opportunistic communication” that happens unintended in any place and in any moment. Conventional video recording equipment cannot capture this type of communication because it has to be set up in advance. This paper presents development of a wearable communication recorder (WCR) for opportunistic communication. To a problematic issue of dealing with large amount of video data in life log systems, the proposed wearable recorder only records interpersonal communication. It gets recording cue by utterance of a user and records communication 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.

Keywords: Wearable computing, Communication Research, Opportunistic Communication, Nonverbal, Multimedia.

1 INTRODUCTION

Research on interpersonal communication has been actively conducted since 1950's, and its major progress was brought by video recording technology. Observation was almost the only way for a researcher to study interpersonal communication that never happens again before the use of video recording technology. Video recording technology enabled 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 opportunistic communication because video recording equipment should be set up before communication starts.

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

We propose a wearable communication recorder to get video data of opportunistic communication from the background mentioned above. Opportunistic communication can be recorded by using such wearable video equipment. Problem is how to deal with huge amount of video records when video is always recorded. This is a major research issue in life log. Our approach is to record opportunistic communication alone automatically.

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

2 RELATED RESEARCH

2.1 Opportunistic Communication

Opportunistic 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 audio and video link [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 and analysis of opportunistic communication.

2.2 Communication Recording

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

Examples of such video use are analysis of multi-party conversation by video recordings [3] and analysis of body movement synchrony in psychotherapeutic counseling by video [4]. Video is recorded manually in these cases. Video recording of multi-party conversation in a meeting with automatic speaker identification has been realized recently

[5]. However, all these are for communication that is scheduled in advance and takes place at fixed location.

There are recording of communication in a limited area. The Active Badge system is an early system of personal location. A user wears an infra-red transmitter named Active Badge. 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, a system user is able to know the meeting of people. The Bat system is subsequent of Active Badge system. Ultrasound is emitted from a transmitter that a user wears and a number of receivers are installed on the ceiling, which makes high precision of locating possible [7]. Automatic recording of interactions is done using infra-red markers on objects in an area and infra-red receivers worn on users [8]. All these can record interactions more or less, but markers are necessary. This means that recording area is limited.

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

2.3 Life Log Research

Life log system is an always-on recorder of various events, user's behavior, and operated objects. It gains very large data because of its always-on feature. Dealing with very large data is not easy. Additionally, most of the data is useless. Information retrieval from this very large data is a major topic of research because of this feature. 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 recording of interpersonal communication.

Mixture of video and sensor information has been also researched because image processing alone is not assured to recognize objects in the video or to understand context. Sensor information can be annotation to the video recordings. It has not focused on interpersonal communication.

3 PROPOSAL OF WEARABLE COMMUNICATION RECORDER

We propose a wearable communication recorder (WCR) to get video data of opportunistic communication. Because opportunistic communication does not happen at a fixed place, conventional video recording equipment cannot fulfill the requirement of mobility. People also do not know when opportunistic communication happens because it happens without appointment. Typical opportunistic communication is short talk when 2 people come across in the hallway. Accordingly, recording equipment should record interpersonal communication without limitation of time and place. To fulfill this requirement, wearable equipment is applied.

Recording everything like a life log system may cause problem of dealing with 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 a 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 detection of initiation of communication and completion of communication. For automatic recording of communication, utterance can be a cue. It is not so simple to start recording when detecting user's voice and stop the recording when not detecting the voice. It is known that communication usually starts before first utterance by eye contact or salutation [9]. In accordance with this, the recorder should record nonverbal cues that occur before verbal utterance. Detecting these nonverbal cues robustly is not very easy in itself, and it needs measuring instruments that are not easy to be wearable.

To this issue, our recorder applies mechanism of a driving recorder. A driving recorder is a device commonly used in business cars such as a taxi. It records video around the car as evidence when it detects impact by car accident or sudden braking. Because it is not good for evidence video to start recording from the impact, video data is buffered about a minute. The buffered data is saved on impact.

Recording of interpersonal communication is thought to be possible by voice detection and mechanism similar to the recording mechanism of a driving recorder. Voice detection can be the trigger of saving buffered video data that include nonverbal cues that come before verbal utterance. For stop recording, it is also thought to be possible by extending the recording some more time after not detecting voice.

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

4 IMPLEMENTATION OF WEARABLE COMMUNICATION RECORDER

4.1 Equipment

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

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

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

Figure 1 shows appearance of a user when wearing these equipments.

4.2 Software

OS of the PC is Windows XP Professional. Program is written in C++. OpenCV 1.0 is used for video processing. Windows Multimedia API is used for audio processing.



Figure 1: Appearance when wearing the system.

WCR keeps detecting audio from the throat microphone. Video from the USB camera and 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 utterance of the user of the system. Then buffered video and audio is saved to a file, which is supposed to be recording of 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 recognizing utterance, it is recognized as completion of the utterance. Recording stops after a certain period of time then.

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

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 opportunistic communication, daily informal communication was investigated because the targeted communication cannot be captured without our system and because daily informal communication is thought to be more similar to opportunistic communication than other type of communication.



Figure 2: Devices used for recording daily informal communication: Omnidirectional camera (left) and conference microphone (right) Appearance when wearing the system.

5.2 Method

An omnidirectional camera (VS-C42U-300-TK) and a conference microphone (BUFFALO BSKP-CU201) (Figure 2) were set up at a room which is approximately 9 m x 6 m (Figure 3). Communication in the room was recorded 27.5 hours in 5 days. Video was recorded in 2048 x 1536 (QXGA) with 6 fps. Horizontal view angle of the camera is 10 – 15 degree on the upper side and 55 degree on the lower side. The microphone is omnidirectional with noise reduction.

5.3 Result

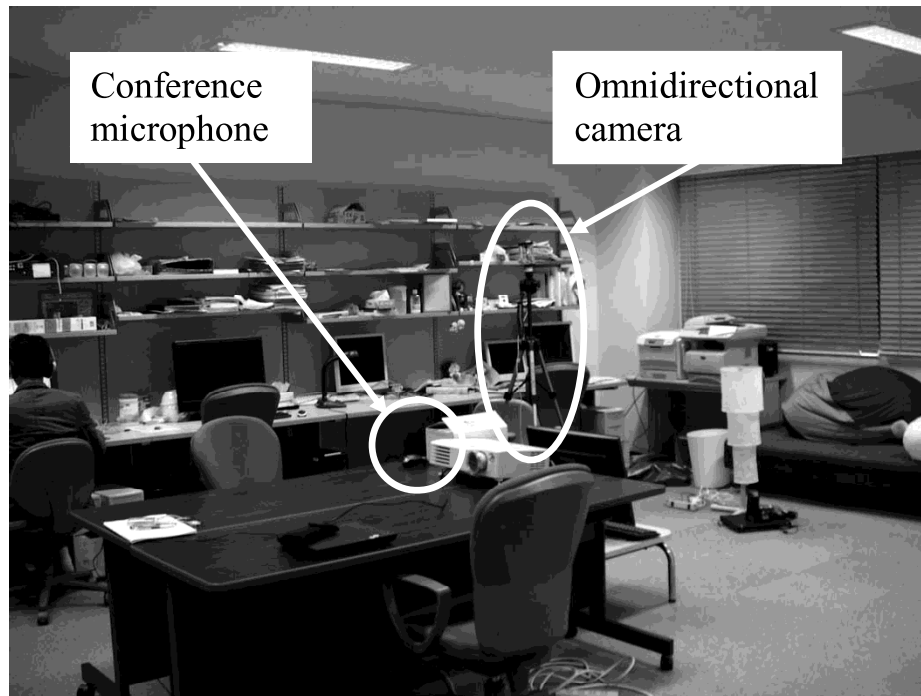


Figure 3: Environment of recording daily informal communication.

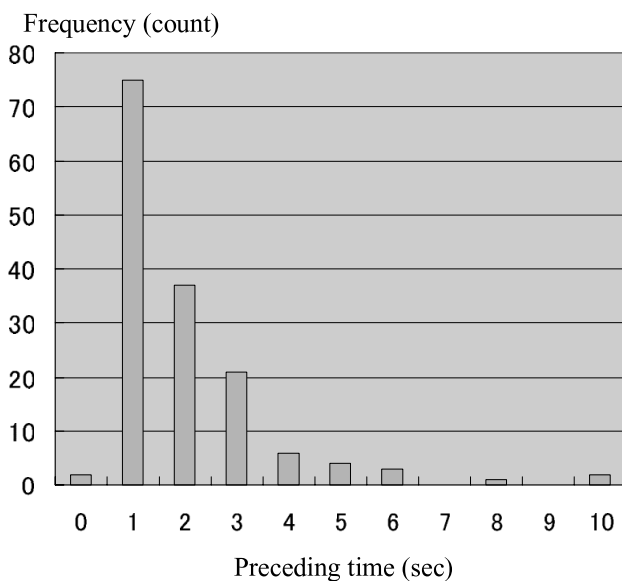


Figure 4: Preceding time of communication behavior before first utterance

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

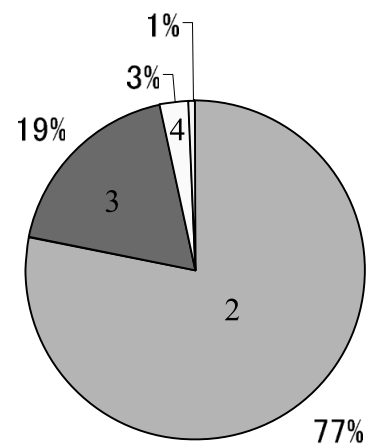


Figure 5: Number of participants in a communication

Additionally, the number of participants in a communication is shown in Figure 5.

6 INITIAL USE OF WEARABLE COMMUNICATION RECORDER

6.1 Aim and Method

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

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	

Frequency (count)

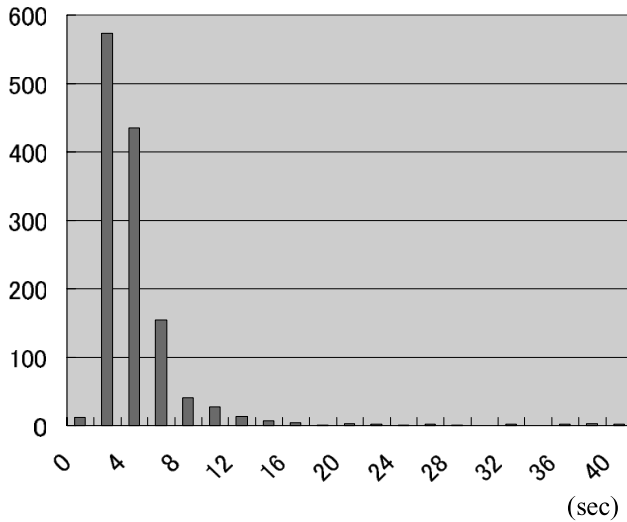


Figure 6: Interval between utterances

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

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

Total recording time was 73 minutes. Speaker and the time of utterance, initiation time of communication by observation were annotated to the video and audio data. This was done by the video annotation software named CIAO.

Communication happened 17 times in the data. Initiation of communication was determined as the point that the subject became aware of the communication target. The time between initiation of communication and the first utterance of the subject is shown in Table 1. Because the number of sample is small, its distribution should not be concluded. But it is suggested that most utterance began within 5 seconds from initiation from the Table 1. It should be noted that all communication were between 2 people. The time is supposed to become longer in the case of 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. Figure 6 is the graph of interval 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. Some data were 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 recording can stop about 20 seconds after completion of the last utterance.

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

7 PERFORMANCE EVALUATION OF WEARABLE COMMUNICATION RECORDER

7.1 Aim and Method

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

A female university student wore the WCR in a room 13 hours in 2days. The room and the setup were the same as in the section 5. All communication in the room was recorded by the omnidirectional camera and the conference microphone, which became the correct answer of communication in the room. The recording taken by the WCR was matched with the answer, thus evaluated the reliability of the WCR. The buffering time was set to 10 seconds.

7.2 Result

The number of actual communication was 32. The number of recording of WCR was 39, which means the WCR recognized these as communication. Among these 39, the actual communication was 30.

The WCR could pick up 30 out of 32, i.e. 94 % of all communication, which is known as recall rate. The WCR also sometimes mistakenly recognized the noise event as communication. The rate of actual communication among all recordings was 30 out of 39, i.e. 77 %, which is known as precision rate.

8 CONCLUSION

We proposed and implemented a wearable communication recorder (WCR) to get video data of opportunistic communication. Different from other life log recorder or conventional video recording equipment, WCR focuses on recording opportunistic communication efficiently. Video

recording has played an important role in communication analysis but has been applied to limited communication. Significance of WCR is to expand its target. Although current prototype should be improved in some points such as appearance and functions as a wearable computer, it is expected that analysis of opportunistic communication progresses with this type of system.

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).