International Journal of Information Technology & Decision Making
Vol. 6, No. 3 (2007) 427–442

THE WEARABLE SENSOR DEVICES FOR DETECTING CONVERSATIONAL EXPERIENCES

SADANORI ITO[*,†,‡], SHOICHIRO IWASAWA[†,‡], YASUYUKI SUMI[†,§] and KENJI MASE[†,‡,¶]

[*]*Graduate School of Engineering*
Tokyo University of Agriculture and Technology
Naka-cho, Koganeishi, Tokyo 184-8588, Japan

[†]*Department of Interaction Media*
ATR Media Information Science Laboratories
Sourakugun, Kyoto 619-0288, Japan

[‡]*Department of Network Robot*
ATR Intelligent Robotics and Communication Laboratories
Sourakugun, Kyoto 619-0288, Japan

[§]*Graduate School of Informatics, Kyoto University*
Sakyo, Kyoto 606-8501, Japan

[¶]*Information Technology Center, Nagoya University*
Nagoya, Aichi 464-8602, Japan
[*]*sito@cc.tuat.ac.jp*

We have developed a wearable device that records the activities of human–human and human–artifact interactions. Using microphones and cameras, the device imitates human perception, recording personal and social everyday-life experiences in multiple modalities, such as voice and visible scenes. These sensors record the perceived experiences continuously, and detect and index interactions from nonverbal behavior. The indexed stored experiences can serve as the first step toward a multimodal knowledge base created from daily life. An infrared LED ID tag system detects interactions, in terms of the ID and the relative positions of objects within the camera's visual field. In this study, we propose an "interaction scope" which is defined as the range of relative human–object positions that have a high probability of occurring in conversational interactions. Analysis of experimental conversational sessions confirms that this interaction scope exists and can represent these interactions naturally. We also demonstrate that our tag system effectively detects and measures the proposed interaction scope.

Keywords: Experience capturing; conversational experience; everyday-life computing; wearable device.

1. Introduction

The major advantage of ubiquitous and wearable computing lies in its potential for extending applications designed for personal use. We aim for a new social-computing

[*]Corresponding author. Tel./Fax: +81-42-388-7697.

paradigm that incorporates an interface based on social protocols.[1] This is not a simple extension of existing human–computer interfaces (HCI), but a migration of human–human interactions (HHI) to HCI. The conventional HCI design paradigm is largely dependent on human perceptual behaviors. HHI analysis, however, often takes an ethnographic approach, which requires tedious, manual labeling of behavior. Analyzing and modeling social protocols in the new paradigm, requires a novel system that can capture perceptual behaviors and social interactions automatically. A large amount of the interaction data collected by such a facility in a semistructured format, called the "interaction corpus", can serve as a behavioral knowledge for social-computing machines.

Interaction corpora that are recorded human experiences are usable in a multimodal knowledge base. We have prototyped applications for sharing the experiential knowledge that use the interaction corpus. For example, Fig. 1 shows a video summary system that records informal communication at an exhibition site, and summarizes valuable scenes providing multimodal knowledge contents for the community.

We propose a wearable interaction sensing and capturing system that can record natural conversational interactions and annotate segmented interactions of the obtained records in the form of the interaction corpus.[2] We introduce the concept of an "interaction scope" as the range of relative human–object positions with high probability of existing in these interactions. We discuss the viability of the

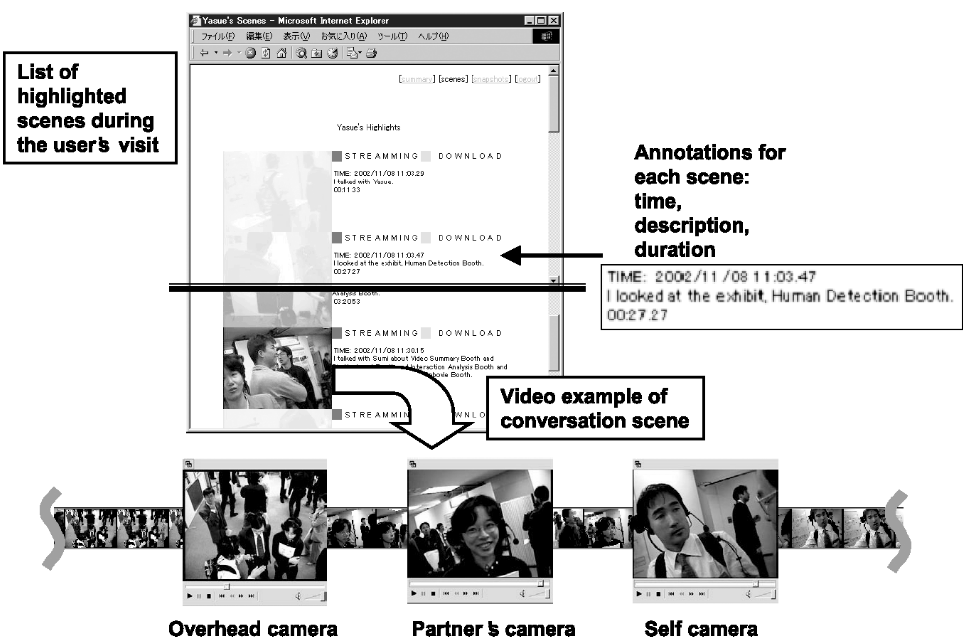


Fig. 1. Automated video summarization system.

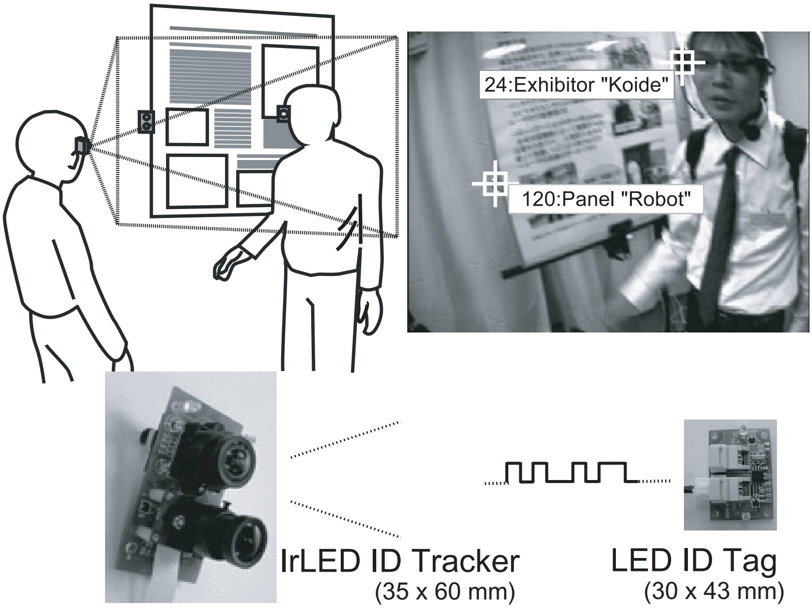


Fig. 2. Configuration of the wearable devices.

interaction recording system in terms of this interaction scope. The system detects and records interactions conducted in free space, such as exhibition sites, where the social protocols and personal behavior emerge. If the device properly detects interactions within the interaction scope, we can confirm the validity of the device design. We first measure the interaction scope with the help of motion capture and eye-tracker systems.

Our system includes an infrared ID tracker based on high-speed image sensor and infrared LED tag devices (see Fig. 2). The tracker detects the tag position within the view area. The system has two important functions:

(1) Recording video images of the visual experiences and detecting the target object in the scene. We investigated the optimum location of wearable cameras for recording the visual field.
(2) Interpreting the conversational interactions. The system determines not only the conversational groups but also detailed participatory relations among the conversational participants.[3]

These results are useful for indexing interaction records and analyzing conversational processes. To determine interactions, our system detects nonverbal behaviors in utterances, and the relative positions of participants. The detected range of relative positions with a high possibility of interaction is set by empirical knowledge. We investigate whether the device can detect actual conversational interactions during small-group interactions in free space.

2. Related Work

We can guess social relationships from the seating order of people in a meeting room. This suggests that the semantics of interaction can be intuitively understood from observable behavior. Ethnomethodology specialists have pursued detailed elaborations of this concept by studying such psychological and physical phenomena as interaction environments.[4–6]

One of the most important research topics in the wearable computing field is the analysis of characteristic elements, based on observations of recorded nonverbal behavior, obtained within the physical range of the recording equipment.[3,7,8]

The major purposes of collecting this data are to support interaction analysis by specialists and to assist in the user's collaborative and personal work, such as assisting in recall of event memories. Previous research used mechanisms such as infrared and radio-frequency transmission devices including radio-frequency identification (RFID) tags. Such small devices, attached to persons or things, can detect the behavior of people meeting and approaching things and places. The advantage of these devices is that, because they are comparatively inexpensive, and cause little psychological stress in users, they can be widely used not only in laboratories but also in everyday situations. The detected data are useful for creating indices that divide daily activities into segments.

However, all these research efforts share the limitations imposed by the efficiency of the devices used. Since most sensor devices detect only whether there are objective tags within the range area, it is difficult to detect more detailed interaction information such as the directions in which people are facing, or what they are looking at. Reliably detecting these interactions, despite the sensor limitations, depends on the compatibility of several factors: the nonverbal behavior of actual people, the radiation patterns of infrared and radio-frequency signals, and physical specifications, such as the sensitivity limits of the detector. If such compatibility is not investigated, the practical use of these devices might appear doubtful. This issue has rarely been investigated. In this paper, we investigate how to match the physical specifications of interaction-recording systems, including our proposed wearable device, with the conditions of actual interactions.

In contrast to approaches using small devices, some research projects have attempted to obtain more detailed information with high-efficiency fixed devices, such as motion trackers.[9,10] However, these are not appropriate for free spaces, such as the environments in our research. Another method uses wearable devices to record and detect behavior through image processing, using only a head-mounted camera (HMC).[11] The calculation costs of this approach are high, thus these methods are not appropriate for our research, which requires results in real time for many users. Another limit of previous work is that it did not match the device specifications to actual behavior, such as matching the camera's field angle to the person's field of view.

The next section describes a method of recording and detecting interactions using our system, and presents a method of investigating the compatibility of the device specifications with actual behavior.

3. A Wearable Device for Measuring the Interaction Scope

The system we developed consists of both a wearable client device that records user interactions and a database server that collects the recorded information via a wireless LAN. The wearable device incorporates a microphone to record acoustical experiences, an HMC that captures the user's view field to record the visual experience, and an infrared ID tracker and tag for detecting objects within the viewing field angle. The tag has one or more infrared LEDs that transmit unique Manchester-coded 8-bit ID numbers by blinking at 200 Hz. The tracker consists of an M64283 image sensor with a field angle of 90°, 128×128 pixel resolution, and a maximum frame rate of 400 Hz. The c8051f125 microprocessor decodes the captured IDs of the tag LEDs, and outputs the ID numbers and X–Y coordinates. The tracker detects all n tags captured at $150 + (100 \times n)$ ms and sequentially outputs the results. The tracker can follow the tag movement at a rate of $56.25°\mathrm{s}^{-1}$ in the horizontal and vertical directions. The camera, tracker, and tag attach to the user's head, allowing simultaneous recording of the user's view field and the detection of captured objects in real time. Figure 3 shows the image and the interaction scope specification physically assembled by the tag system. For instance, it is possible to detect behaviors observed by users, as well as the people located at positions where they can look at other users. It is also possible to determine interactions,

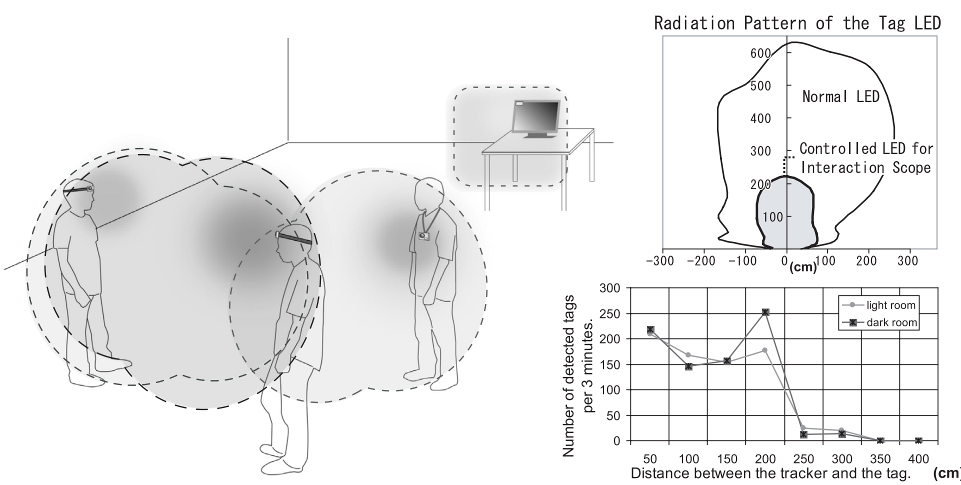


Fig. 3. Interaction scope is physically assembled by the tag system.

such as conversations and group discussions, if alternating utterances and users are detected simultaneously.

Tags can be attached to many things and locations. For example, at an exhibition, they are attached to locations such as booths and to important objects such as posters and exhibited objects. The results of tag detection determine interactions such as what object or people the users are looking at or with whom they are talking.

To ensure that these determinations are correct, nonverbal behavior must be recorded and detected as the interactions occur. The following two requirements ensure compatibility between the system and the observed interaction. First, the recording range of the tracker and camera must include the user's gaze target and approximate the user's view fields. Second, the targets of detected tags must correctly reflect the targets of the user's actual interactions. For example, visual behavior is different when there is a poster 3 m in front of the user as opposed to when there is a PC display 3 m in front of the user. In latter case visual behavior is unlikely, since a high-resolution display is generally used at short range. Thus, we assume that the effective range of visual behavior depends partly on the resolution of the target, as well as its distance and angle from the target.[10,12]

We suggest that a unified relative position, constructed by the angle of users view field and the distance and angle from the target, which are common attributes in visual interactions, provides effective nonverbal behavior as a standard for determining interaction. In our research, the relative position range, called the interaction scope, is achieved through the specifications of the tag radiation pattern and the field angle of the HMC and tracker. The latter is controlled by an optical lens. The tag's radiation pattern and the light divergence are controlled by tag's output power and a light-shaping diffuser with a holographically recorded diffusion layer. These detailed device specifications are determined empirically, based on the results of previous trials.[1] In the following sections, we investigate the viability of determining and recording interaction with these devices in actual situations in which small-group interaction proceeds in free space.

4. User Gaze Direction and HMC Field Angle

This section gives the results of experiments with conversations by subjects wearing the proposed devices. Our purpose was to measure the user's gaze direction, substituting the direction of the head for the gaze direction, to determine its compatibility with the device features.

4.1. *Direction of head and gaze direction*

Visual acuity in the foveal vision area (2° from gaze direction) is high, but it is drastically lower in the peripheral vision area (160° from gaze direction). Because the effective view field for the actual cognitive processes involving texts and shapes is limited to approximately 20° from the gaze direction, most visual experiences "exist" within this range. That is, observation from the gaze direction is the most significant factor in determining the nature of interaction related to looking at objects, interaction among people, and interaction between people and things. To record a visual experience as a video image, the range of the effective view field should be considered. It is also necessary to have an appropriate target of the gaze direction as the main composition element as well as a field angle that can maintain sufficient image quality.

It is difficult to measure the gaze direction. There are two methods for making this measurement precisely: physical contact type and physical non-contact type. The former devices are large and cause users much physical stress. Because the measurable distance of the latter devices is limited, they have difficulty of measuring the line of sight of people moving around. Therefore, to reduce physical stress, and adapt to human movement, small sensor devices are worn on the head and shoulders, and the system uses measurements of the head and shoulder directions to calculate substituted for the gaze direction. However, little research has been done to verify the effectiveness of devices used to approximate the gaze direction.

In terms of the physiologic factors, the gaze direction is changed by contraction of the extrinsic eye muscles within the limits of ocular motility which are 50° in every direction. The range of movement of foveal vision limited by ocular motility is called the field of fixation. In this range, we see objects in full detail without moving our heads. In using the HMC, a field angle of about 120° is necessary to record the entire range that can be seen physiologically, that is, within the field of fixation and the effective view field. However, such a wide angle deteriorates image quality, and it is not appropriate for recording.

For cameras worn on the shoulders or chest, a range of 280°, which including 160° for the rotation angle of the neck, is needed to see the entire physiologic range. Accordingly, the recorded range by using such a method cannot be regarded as the approximate visual field. Another problem is that such methods cannot detect changes in the gaze direction's target without changing the posture of the shoulders or chest. Visual fixation over a certain amount of time stresses on the extrinsic eye muscle. To avoid this stress during prolonged visual fixation, people make orienting movements of the head and body in the direction of the gaze direction. Due to these orienting movements, the recordable field angle of a shoulder- or chest-mounted camera does not necessarily capture the entire physiologic range. However, how often movement of the gaze direction causes corresponding orienting movements differs in each environment. For example, a wide-angle HMC is generally used for short-distance recording, such as work on a table, while a narrow-angle HMC is generally used for long-distance recording. Orienting movements are seldom involved in table work because the targets direction are distributed within short distances over wide ranges, and the targets change frequently. We considered this a major reason for using a wide-angle camera.

The first model of our HMC, used for recording the visual field of visitors at exhibition sites, had a 44° horizontal field angle. In these recordings, users were often observed having conversations with partners out of the field angle. Based on this experience, the new system has a 90° horizontal field angle.

To investigate the compatibility of HMC field angle with recorded view fields, we measured the variance between the directions of the head and the gaze direction in small-group interactions, such as those at exhibition sites.

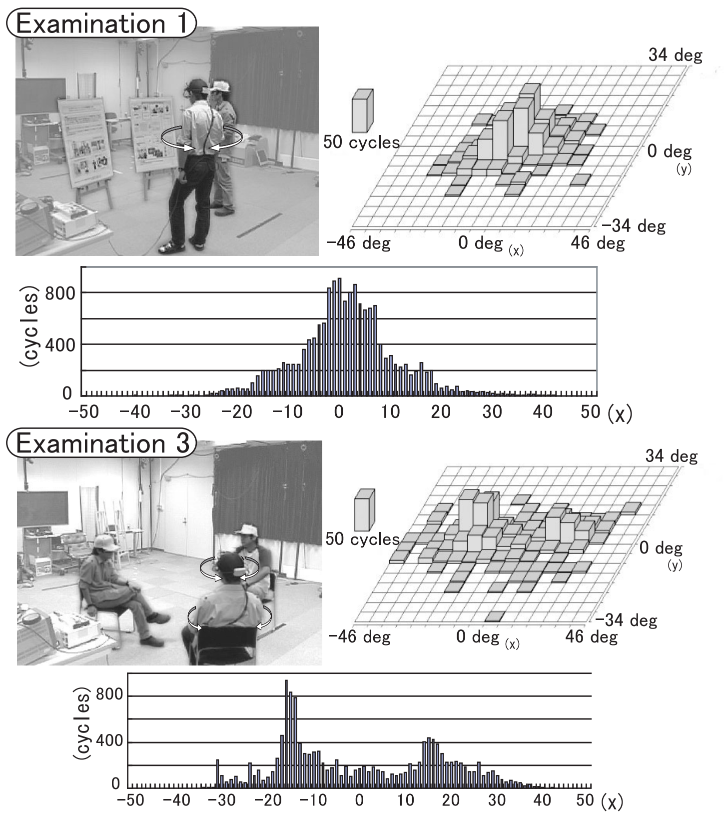
4.2. *Experiments and results*

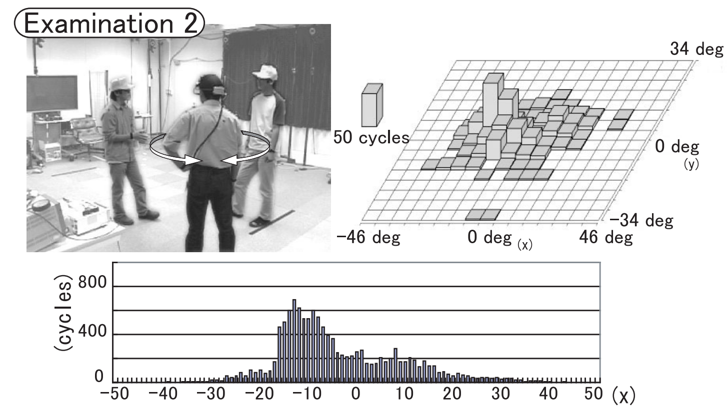
Experiment 1 recorded an exhibitor who positioned himself near two posters to explain and discuss their contents to visitors. This simulates typical interaction at an exhibition booth. In Experiments 2 and 3, discussions between subjects were recorded to compare the effect on orienting movements from environmentally restrictive postures, such as sitting in chairs. In Experiments 1 and 2, users were standing and thus free to change position and posture. In Experiment 3, users remained sitting posture, in chairs equally spaced around a 2 m diameter circle. Each experiment lasted 5 min. One subject wore the gaze direction measuring device, and in Experiment 1, the visitor wore it as well. In each experiment, participants discussed how to solve the problems of wearable devices for recording interactions. The subjects in each experiment had knowledge of this topic. We used an EMR-8B EyeMark Recorder made by Nac Inc. to measure the gaze direction. A pupil and corneal reflex technique could detect a maximum ocular motility of about 46° with 0.1° of precision. The wired measuring device worn by subjects weighed about 250 g.

Figure 4 illustrates the distribution of the number of visual fixations in each experiment. Visual fixation is defined as the condition in which the gaze direction changes less than a range of 2° within 100 ms. There were many changes of the gaze direction among different targets, such as other subject and the poster, depending on the interaction situation. The gaze direction moved among targets 71 times in Experiment 1, 80 times in Experiment 2, and 81 times in Experiment 3.

Visual fixations in Experiments 1 and 2 were centered at ($x = 1.2°$, $y = -1.5°$) and ($x = -6.2°$, $y = 1.3°$), respectively, relative to the frontal head direction. These fixations are concentrated in the range of 10° to 13° horizontally and 5° to 6° vertically. Most visual fixations fall in the range of about 50° horizontally, forming the unimodal pattern shown in Fig. 4. The relative positions of the posters, other subjects, and the targets lines of sight do not appear in the distribution of the direction of the gaze direction because the head's positioned was fixed to the target by movement involving the rotation of the head. This confirmed that, in a standing posture with free orienting movements of the head and body, the range of the fixation field is about 50°. Therefore, a camera worn on the head with a typical field angle of 40° is not appropriate for recording view field targets while being worn on the head.

An even wider field angle is needed for sitting postures. In the sitting posture of Experiment 3, visual fixation points were distributed within a range of 70°





Exam	Center		Standard Deviation		Distribution pattern	Peak
	X	Y	X	Y		
1	1.2	−1.5	10.1	5.4	single-peaked	0
2	−6.2	1.3	13.1	6.1	single-peaked	−13
3	−3.2	2.2	17.7	5.8	double-peaked	15, −16

Fig. 4. Experimental setup and results of eye fixation.

horizontally, forming the bimodal pattern shown in Fig. 4. Peaks appeared at 15° and −16° horizontally. Conversation partners sat 30° to the left and right of the subject. Subtracting the magnitude of the peaks of gaze direction movement from the angular distance between conversation partners shows that subjects made both a visual line movement of 15° and a posture rotation of 15° to see targets, due to the posture limitations imposed by the chair. Therefore, it is difficult to record a video image of this experience with a narrow-angle HMC. In such environments, a sensor worn on the shoulders or chest is not appropriate, because a wide variance exists between the gaze direction and the sensors.

Our research recorded and determined interactions using HMC and trackers in small groups and at short distances. In such environments, interactions between gaze targets about 50–100 cm wide, such as a poster or a person, when users were about 200–300 cm from the target, require a 10–30° horizontal angle to fit the target within the field angle. To record the target objects as video images and detect tags, the total angle of the visual fixation range was obtained from Experiments 1 and 2; consequently, a field angle of about 80°, corresponding to the target objects, was considered appropriate.

The conclusions obtained from the experimental results indicate that empirically fixed field angles of 90° for the HMC and the tracker are suitable.

5. Compatibility of Relative Position and the Interaction Scope

This section addresses the effectiveness of using the detection results of relative positions obtained by trackers and tags to detect interactions. The interaction scope is defined as the set of relative positions that are highly likely to exist within the visual interactions of a person. This concept uses the tracker's substituted view field angle and the viewable tag range. Conversational interactions are detected using the detected results of targets inside the interaction scope, as well as other modalities.

Listeners are generally seated in an auditorium and do not engage in explicit behavior, but listener detection is difficult in a space through which users can move freely. This requires multiple modalities and a new detection method. It is easy, however, to detect speakers when observing discussions. Based on the hypothesis that a speaker faces a listener when talking, we developed a system to detect conversational interactions that identifies objects detected within the interaction scope of the speaker as listeners. The following section describes the experiments made to investigate the accuracy of the detection method.

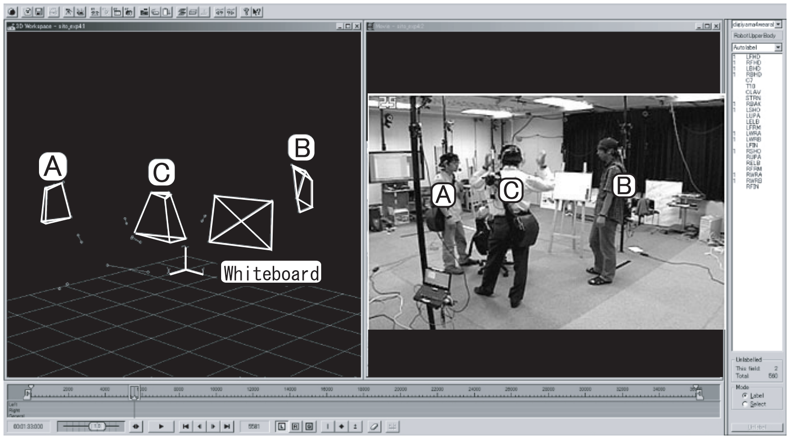
5.1. *Experiment*

In this experiment, we recorded 10 min of free discussion among three subjects wearing devices developed to record experience in a space where free movement is allowed. This space simulates the conditions of an exhibition site. The participants discussed wearable recording device of experience. As related exhibits for the discussion, we added a mannequin wearing a device, and a whiteboard for writing notes. Subjects were allowed to move around freely and to use and move the exhibits.

A Vicon motion capture system measured the positions of the subjects and exhibits. Vicon is consists of 12 infrared cameras with an infrared irradiation function, set in a 7.5×10 m room, and spherical passive markers (1 cm diameter) made of a retroreflective material. Vicon reconstructs the three-dimensional position of each marker from the two-dimensional position captured by the cameras. Vicon's measurable area is 2.5×3.5 m, and it has a temporal resolution of 60 Hz and a spatial resolution of 1 mm. As shown in Fig. 5, markers were attached to subjects and exhibits. Subjects had four markers on the cap, two on the shoulders, and one on the right shoulder blade. We named the marker on the left front of subject A's head *LFHDa*, the one on the right front of the head *RFHDa*, and the one at the center of the back of the head *BAHDa*. By establishing the median of *LFHDa* and *RFHDa* as *CFHDa*, we defined the direction of the head as $\vec{va} = \overrightarrow{BAHDaCFHDa}$. *CFHDa* was also defined as the position of the body. This three-dimensional coordinate information is stored in the server. The video images and sounds recorded by each recording device are stored in the database server after synchronizing the time information of the client with that of the server by Network Time Protocol (NTP) via wireless LAN. The input from a throat microphone is gate processed at the power level. Its results are stored as the speaking section. The above devices obtained ten minutes of experience data, including sound, information about the speaker, and the absolute positions of the subject and exhibit.

5.2. *Results*

After the experiments, the speaker (the subject himself) watched the recorded video and manually labeled the listener, to identify to whom he was talking. He also recorded the correct relationship information between the speaker and listener. Conversations when the speaker was unconscious of a listener or was speaking to several listeners were considered outside of the labeling target. The percentage of unlabeled speaking was 5.29% as measured by time duration and 11.76% as measured by number of occurrences. Unlabeled speaking of less than 10 s denotes actions such as nods. The speaker is subjectively conscious of listeners during most of the time spent speaking outside of the labeling target. Using the manual labeling, the relative positions of subjects A and B, and those of A and C, were sampled at 1/3 s, intervals for distance and angle. Figure 5 shows the mapped results. During the 10 minutes of the experiment, the standard deviation was dispersed over distance (34.91 cm) and angle (61.08°); the speaking time, in contrast, was centered at a distance (25.62 cm) and angle (31.75°). An HMC and tracker with a 90° field angle range accounted for 88% of the time that conversations were observed. These results indicate that the speaker is apt to face the listener when the listener is identified during speaking.



	Entire experiment time		Speaking time	
	Distance	Angle	Distance	Angle
Mean	149.26	-10.10	127.02	5.00
Variance	12186.96	3730.57	6557.21	1006.72
Standard Deviation	34.91	61.08	25.62	31.75

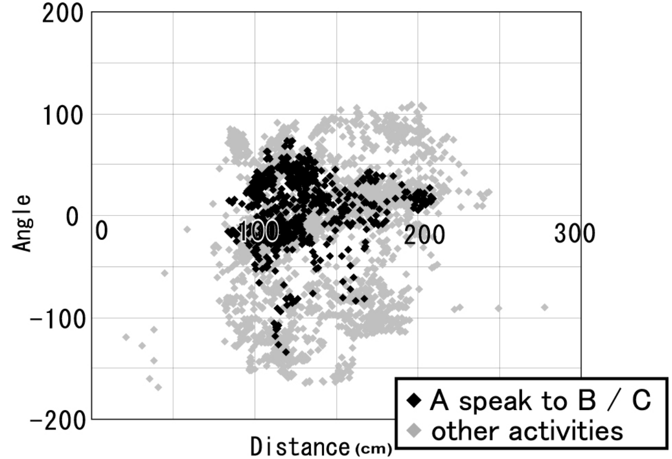


Fig. 5. Experimental setup of position measurement and relative positional map.

We also confirmed that the tracker and tags detected within a user's interaction scope have perform adequately to detect the interaction process between speakers and listeners. Outliers were also observed, however, and their characteristics are discussed in the following section.

6. Discussion

In this section, we analyze the scenes containing outliers to confirm the applicability and limitations of detecting interactions with our method. In these scenes, the actual interaction conditions that existed were different from the detected results because of our system's deviation from the detection criteria for interactions. The outliers observed during the experiment in Sec. 5 were measured as a short 5 s interaction between subjects A and B and one 4 s interaction between A and C. The relative angle of B as seen from A is concentrated in the range of ±50°. In contrast, the relative distance is distributed widely over a range of 70–200 cm. The movements in such outliers are shown in Fig. 6, which illustrates the relative distance of subject B extended to around 200 cm and a change in the image of B's view field. Figure 6 also shows the mapped relative positions of A and B and illustrates the observed time of B's movement with light-colored dots. Subject B significantly changed his position to face toward and to point his finger at the mannequin when the conversation turned to the experience-recording device worn by the mannequin. This scene was

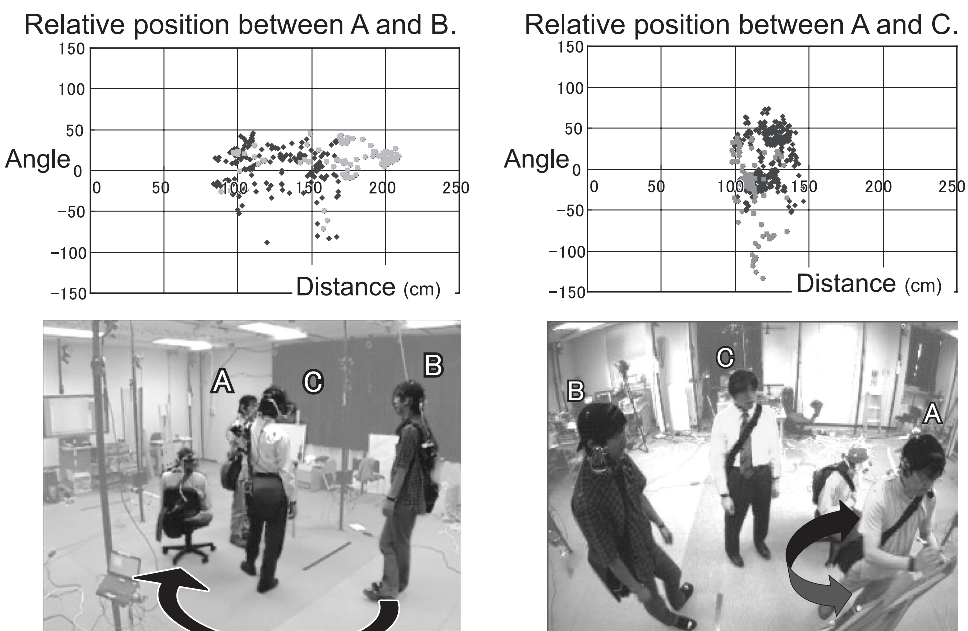


Fig. 6. Relative positional map of interaction with outlier.

also recorded by B's view field camera. The relative position of C as seen from A is comparatively stably distributed within 100–150 cm; in contrast, the relative angle is distributed in a range over 100°, although it is generally concentrated within ±50°. Since this is beyond the ocular motility limit, there was distinctly continuous time when the subject conducted conversation without the listeners in sight. The movement of subjects in such outlier scenes was observed. It is apparent that subject A wrote a related memo on the whiteboard while speaking to C. The recorded view field scenes from the cameras of subjects A and C are shown in Fig. 6. C continuously looked at A, whereas A looked at C and the whiteboard alternately.

The above results indicate that a conflict of modality and a selection of modality occurred during the conversations. In the case of the outliers in the conversation between A and C, these participants both viewed and wrote on the whiteboard simultaneously while holding their conversation. As a result, differences were found between the speaker's subjective consciousness of the listener and the detected results of the interaction target by head direction. From recording experience viewpoint, observation of the outliers does not have to be considered. The experience-recording device records an image of the process of writing on the whiteboard but records the speakers conversation as sound. This closely approximates the user experience. The detected whiteboard as the line of sights target provides an effective operational target; however, using the detected results for the view field direction and speaking behavior, the system inappropriately interprets the whiteboard as the listener.

This is a limitation of a method that detects interactions from nonverbal behavior using unified information from sensors that capture different modalities, such as utterances and visibility. When different modalities are used for different targets, accurate detection results could not be obtained. However, the human behavioral context at an abstract level can be inferred from the use of temporally continuous aggregation of individually detected interactions. This method can be an effective compensation for this misdetection.[2] For instance, Fig. 7 shows the results of interaction detection in Experiment 1 in Sec. 4. From the detected results, it could make episodic records such as "the exhibitor had a conversation with a visitor while looking at posters 1 and 2".

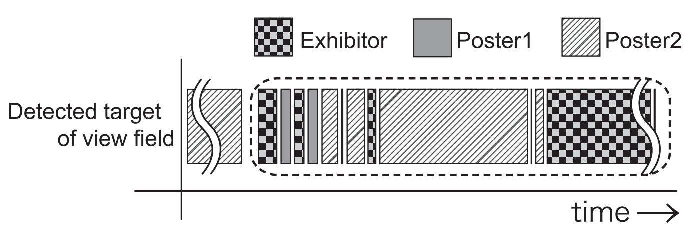


Fig. 7. Transition of gaze object in Experiment 1 in Sec. 4.

7. Conclusion

This study evaluated the performance of wearable devices in recording and detecting interaction experiences. When users were able to move around freely, in a standing posture, we experimentally confirmed the hypothesis that 90° view field angle was sufficient. With a camera and tracker substituting for the view field, we examined recording and detecting the images of experience with multiple devices, each having a different field angle. The measured results of the movement of the actual gaze direction confirmed that a horizontal field angle of 90° is appropriate in a standing posture. Furthermore, in such restrictive postures as sitting, it is clear that the appropriate field angle becomes wider.

To detect the relations of participants in conversation, we used the results from the interaction scope, which detected the relative positions of high probability in existing interactions by substituting the range of the tracker's visual field and the tag's viewable range for actual interaction phenomena. We verified the functionality of detecting conversational relations, by comparing automatic detection results to what the speaker's verbal report. On the other hand, it was found that the circumstances of the modalities used for different interactions, such as writing on a whiteboard along with speaking, are difficult to detect in real time.

In future work, we will investigate a detection method that uses heuristics and temporal aggregates, based on the device capabilities studied in this work.

And we will consider the methods to make effective use of interaction corpus as knowledge base.

References

1. Y. Sumi, S. Ito, T. Matsuguchi, S. Fels and K. Mase, Collaborative capturing and interpretation of interactions, *Pervasive 2004 Workshop on Memory and Sharing of Experiences* (2004), pp. 1–7.
2. M. Takahashi, S. Ito, Y. Sumi, M. Tsuchikawa, K. Kogure, K. Mase and T. Nishida, A layered interpretation of human interactions captured by ubiquitous sensors, *The 1st ACM Workshop on CARPE* (2004), pp. 32–38.
3. T. Choudhury and A. Pentland, Modeling face-to-face communication using the sociometer, *International Conference on Ubiquitous Computing 2003* (WA, 2003).
4. E. T. Hall, *The Hidden Dimension* (Doubleday, New York, 1966).
5. E. Sommer, *Personal Space: The Behavioral Basis of Design* (Prentice-Hall, New Jersey, 1969).
6. R. Baker, *Ecological Psychology: Concepts and Methods for Studying Human Behavior* (Stanford University Press, Stanford, 1968).
7. A. K. Dey, D. Salber, G. D. Abowd and M. Futakawa, The conference assistant: combining context-awareness with wearable computing, *The 3rd International Symposium on Wearable Computers* (IEEE, 1999), pp. 21–28.
8. R. Borovoy, F. Martin, S. Vemuri, M. Resnick, B. Silverman and C. Hancock, Meme tags and community mirrors: moving from conferences to collaboration, in *Proceedings of the 1998 ACM Conference on Computer Supported Cooperative Work (CSCW'98)* (1998), pp. 159–168.

9. P. Chiu, A. Kapuskar, S. Reitmeier and L. Wilcox, Meeting capture in a media enriched conference room, in *Proceedings of the Second International Workshop on Cooperative Buildings (CoBuild'99)* (Springer, London, 1999), pp. 79–88.
10. E. Stiefelhagen, J. Yang and A. Waibel, Modeling focus of attention for meeting indexing based on multiple cues, *IEEE Transactions on Neural Networks* **13**(4) (2002) 928–938.
11. S. Mann, Humanistic computing: WearComp as a new framework for intelligence signal processing, in *Proceedings of the IEEE* **86**(11) (1998) 2123–2125.
12. J. Gibson and A. Pick, Perception of another person's looking behavior, *American Journal of Psychology* **76** (1963) 386–394.