Ubiquitous Experience Media

Kenji Mase
Nagoya University

Yasuyuki Sumi
Kyoto University

Tomoji Toriyama, Megumu Tsuchikawa, Sadanori Ito, Shoichiro Iwasawa, Kiyoshi Kogure, and Norihiro Hagita
Advanced Telecommunications Research Institute International

We exploit recorded activities as a good source of human–human and human–robot communication for sharing experience, memory, and knowledge. In particular, we're interested in audiovisual, ubiquitous, and wearable-experience-capturing technology as interaction-grounded lifelog tools. We've developed several devices—such as a wearable interaction tracker—that facilitate an indexed recording of human experiences.

Our ancestors developed various media to record their experiences: pen and paper, photography, video recording, and so on. As computers acquire a huge amount of memory capacity in ubiquitous computing environments, we can use computational media similarly, as novel and richer media to support our everyday activities of memorization and communication.[1]

We propose a wearable sensor client to provide people with first-person personal and mobile sensing as well as an environmental (stationary or ubiquitous) client for third-person public and stationary sensing.[2] First-person sensing involves a close body and self-perceptive sensing. The third-person sensing involves distant and overviewing perception. Any research on related topics should also cover the media interface design issues of such devices. Artificial partners such as a humanoid robot that can record people's activities in a second-person position with various embedded sensors could become ideal human–machine interfaces. The second-person position is advantageous for getting sympathy of a partner's emotions by closely sensing his/her situations. By integrating devices, they can function together to construct experiential media for people in a ubiquitous manner. Thus, we call these devices *ubiquitous experience media* (UEM).[3] This article presents an overview of our project, including the recent development of a new wearable device[4] as a lightweight UEM designed to attract wider use.

About the devices

We use video cameras, microphones, and other activity sensors as components of the UEM to monitor and record ourselves and our environment. We came up with an ID tag comprised of an infrared light-emitting diode (IRID tag) and infrared signal-tracking device (IRID tracker) to record the positional and situational contexts along with video/audio recording.[5]

If every person and every object in a space wore such an IRID tag, the IRID tracker wearer could tell what she or he is looking at or viewing. With such contextual indices, we structure the interactions based on the viewing aspect and positional proximity. Rather than waiting for usable image-based, object-recognition technology, we focused on how to record and describe human–human and human–object interactions. When robust recognition technology does become available, we can incorporate it in our system. It's quite feasible to realize such a sensor- and tag-intensive space in environments like hospitals, factories, energy plants, laboratories, and schools.

Our second step is to process large-scale interaction data captured with the UEM. We construct an *interaction corpus*, which is a large, semistructured set of interaction data. The captured data are automatically segmented into primitive behaviors and annotated semantically. We aim to use this corpus as a collection of experience elements to describe past experiences with other people. It's easy to collect highlighted actions by annotating tags—for example, to generate reconstructed diary contents[6] and visit/meeting summaries. The corpus can, of course, also serve as an infrastructure for researchers to analyze and model social protocols of human interactions. We exploit the extracted behavioral patterns for simulating and estimating future actions.

A computerized memory aid is also a challenging but promising application of our interaction corpus. Various high-density and multimedia digital recording devices have been developed and will continue to be improved. It's nearly possible to record and store the whole life of a person by video and audio as a multimedia autobiography onto a small magnetic disk or solid-state memory.[7] The interaction corpus' structure might lead to a new theory of human memory and an

1070-986X/06/$20.00 © 2006 IEEE
Published by the IEEE Computer Society

efficient way of using such an external memory for creative activities and exchanging information with other people. Even a system to support the simple briefing and reporting of events and experiences to colleagues at places such as business offices and hospitals would be helpful.

Many technological and social issues exist in realizing and deploying UEM: media processing, data mining, annotations, human interface design, aesthetic design, privacy concerns, and so on. The importance of privacy concerns varies depending on the balance of privacy cost and the system's benefit. If it's used in a closed environment such as a factory or hospital, users would more easily accept the system to achieve greater safety and efficiency. However, we need to add privacy protection for wider and more general use. One possible way is to automatically cover facial areas by mosaics in the sensing or early processing stages, an approach we're exploring seriously. For simplicity, we set aside privacy issues in this article and focused on the enabling technology.

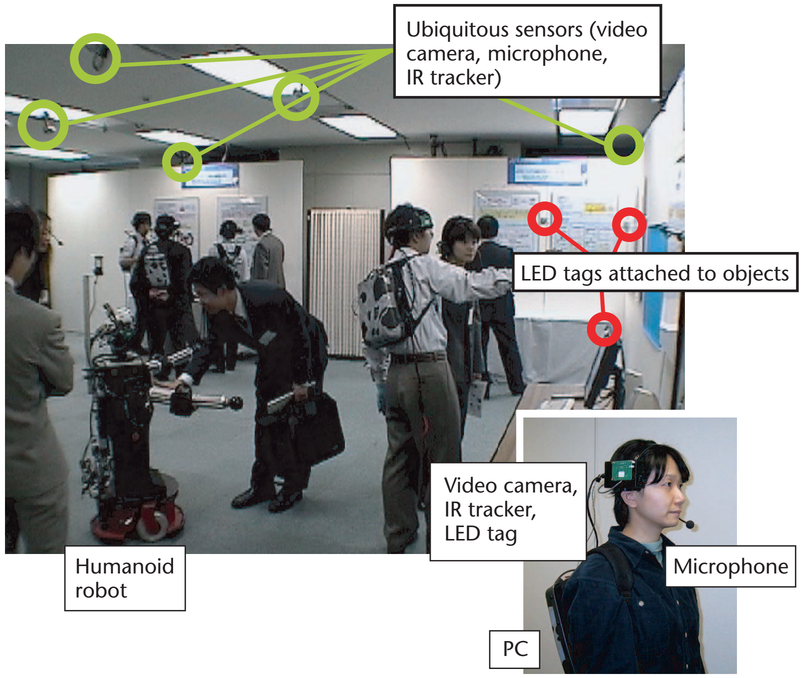


Figure 1. Ubiquitous experience media (UEM) sensor room setup (the first-generation wearable client).

Ubiquitous experience media

We first prototyped a UEM system to record interactions among multiple presenters and visitors in an exhibition room. We installed and tested the prototype at the ATR Exposition public exhibition in November 2002 in Kyoto, Japan. Figure 1 is a snapshot of the exhibition room set up for the interaction corpus experiment. The room contained five booths. Each booth had two sets of ubiquitous clients that consisted of video cameras with IRID trackers and microphones. IRID tags were attached to possible focal points—such as on the posters and displays—to extract the social interactions in a presentation. Each booth presenter carried a wearable client consisting of a video camera with an IRID tracker, a microphone, and an IRID tag.

A second-generation wearable client employs a throat microphone and a head-mounted display (HMD). We used the throat microphone to detect speaking activity more clearly, and the HMD to provide augmenting information (such as people's names and the popularity rating of the poster in front of the person).

Third-generation IRID tracking system

Our system uses an infrared ID emission device (IRID tag) and an infrared-sensitive high-speed 128 × 128-pixel resolution complementary metal oxide semiconductor (CMOS) image sensor with a microprocessor. The IRID tag emits IDs coded by the Manchester coding scheme in 8 bits with an infrared LED on a 200-hertz (Hz) base frequency. Our image sensor runs at a 400-Hz refresh rate to find the position of a blinking spot and its identity. The spot is traced in a small search window placed on the previous detected position to reduce the computing cost on the microprocessor and to integrate moving blinking spots as one coded signal. The initial model used a sequential single-spot search, with tracking runs at 8 to 10 frames per second (fps). We used a parallel spot search for the third generation to achieve multiple spot searches at the same speed. The IRID tracker gives the *x*- and *y*- position and identity of any tag attached to an artifact or human in its field of view, as Figure 2 (next page) illustrates.

Because the IRID tracker senses infrared-spectrum images only, we use another sensor for the visible image capturing used in experience recording. The alignment of the axes of the IRID tracker and the visual sensor is important for visual reviewing and augmented reality applications. We recently developed the third-generation wearable client with a camera module, which features a combined coaxial IRID tracking sensor and image sensor (see Figure 3).

Incoming light through the tracker's front lens is split by a hot mirror into infrared and visible light and fed into CMOS and charge-coupled device (CCD) sensors. We achieved a lightweight fabrication of the module with a weight of 39 grams (without cable) and dimensions of 27 × 16 × 43 mm.

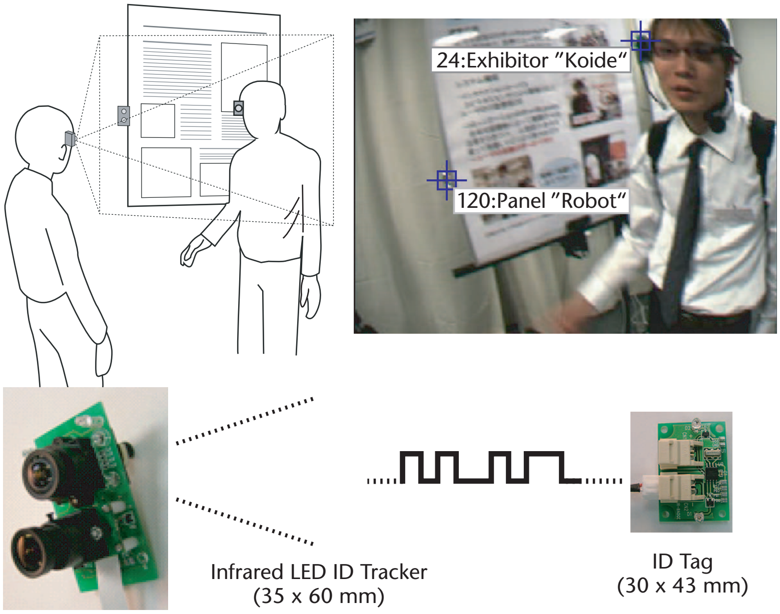


Figure 2. Second-generation model of an infrared identification (IRID) tracker system.

By wearing an IRID tracker on the ear or forehead, we can also estimate the user's approximate gazing. This approach assumes that gazing can be used as a good index for human interactions.[8] We investigated people's gazing behavior while designing the tracker module. Consequently, we chose a front lens with a 90-degree field of view. Figure 4 shows an experimental setup of casual meetings. We performed an analysis of relative head rotations to the target object with a motion capture system (mocap). We measured the head rotation with the mocap, while we measured the eye direction with an eye-tracking device. In a standing situation, the rotation distribution spreads to a 50-degree width with a unimodal peak, while in a sitting situation it spreads 70 degrees with bimodal peaks. We assumed that if the tracker sensor covered this width, we could replace the eye tracker with our IRID tracker. We adopted an additional 20 degrees in choosing the lens, taking into account the width of a visual target.[9]

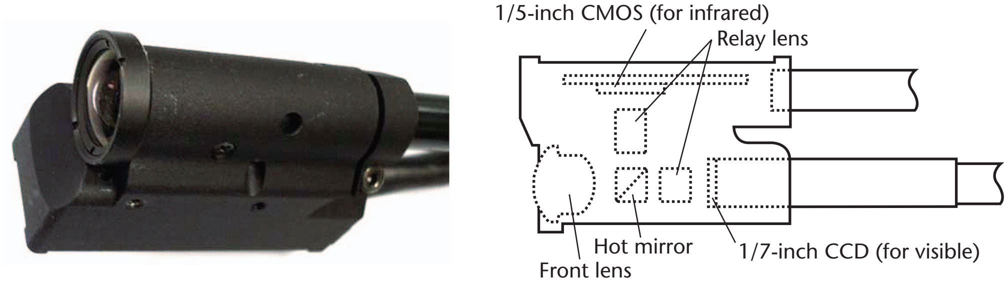
Wearable client PC

We used off-the-shelf portable PCs as the UEM controller initially. However, the users—including exhibition visitors—told us that because of the PC's weight, size, and the heat it created, it wasn't comfortable to wear for very long. We finally developed a custom-designed wearable client PC that uses an LC690132A LSI (Large-Scale Integration electronic integrated circuit) chip with an MPEG-4 encoder/decoder as its main controller. The PC has a serial port for the IRID tracker, a video-in port, a general analog-to-digital port, and a compact flash slot used for an IEEE 802.11b wireless local area network adapter. Figure 5 shows the PC and the IRID tracker module. The dimensions of the PC are $145 \times 113 \times 26$ mm, and the weight is 245 grams without the battery pack. With eight AA-type batteries, it can run for about 3 to 4 hours.

The attachment of an IRID tracker module, IRID tag, and microphone on a human body is a difficult human–computer interaction (HCI) design issue. We decided to place the tracker module on the temple to obtain a closer position to the eye gaze and a fairly natural look (see Figure 5 on p. 24). The video, audio, and ID signals are sent by wire. The wire's weight can't be ignored, however. In the future, we might use a wireless link between the headset and the controller box if the power supply issue is effectively solved. We're investigating the optimal IRID tag position. We also have another headset model besides the one shown in Figure 5. It can hold several directional tags around the head so that the tracker can estimate the relative head rotation. The heat and weight problems are basically solved now, but the wiring remains a problem, as we learned in our user study.

Ubiquitous sensing

All of the data are stored in separated files of MPEG-4 and IRID logs synchronized by time



(a) **(b)**

Figure 3. Integrated IRID tracker module. The infrared and visible sensors are coaxially aligned. The input view is split by a hot mirror. CMOS = complementary metal oxide semiconductor; CCD = charge-coupled device.

stamps with the wearable client PC. Our system records audio as Pulse-Code Modulation (PCM) format files. To efficiently access and move individual sections of massive video data, we divided the data into smaller, 3 Mbyte-sized files and stored it onto disks. Here, the length for a 3-Mbyte MPEG-4 video segment is about a minute, where the data format is QVGA (320 × 240 pixels) at 15 fps. An alternative file format is to use Motion-JPEG. It's easier to edit its sequences into small segments, but the total record size becomes huge. We used the M-JPEG format with the initial system, using an off-the-shelf PC.

Our approach to using the UEM to record interactions integrates many sensors (a video camera, IRID tracker, and microphones) ubiquitously set up in rooms as well as wearable sensors (video cameras, IRID trackers, microphones, and other sensors and displays) to monitor humans as the subjects of interaction. In the current implementation, the system collects the data sent by multiple clients and a centralized server stores it. The clocks of all clients are synchronized within 10 ms by the Network Time Protocol (NTP). We also employ autonomous physical agents, such as humanoid robots, as social actors to proactively collect human interaction patterns by intentionally approaching humans. These devices compose the UEM as collaborative interaction media.

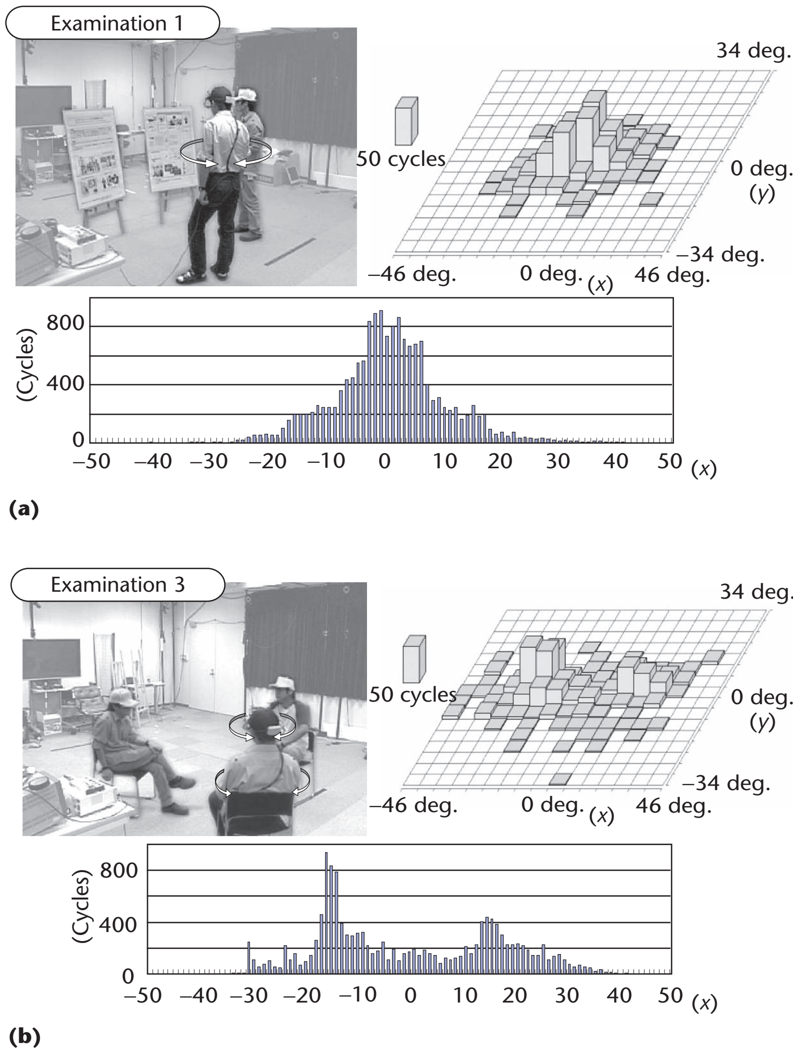


Figure 4. Distributions of head movements. (a) Standing and (b) sitting situations. The top left part of each figure shows the experimental setups. The top right part of each figure shows the frequency distributions of the difference angles between the head and the eye rotations along the pan (*x*) and tilt (*y*) axes. The bottom of each figure is its marginal distribution along the pan (*x*) axis. People generally move their heads more when standing, while two peaks appear for two targets in wide distribution in the sitting situation.

Interaction corpus

Earlier we briefly defined an interaction corpus. To define it more thoroughly, an interaction corpus is a captured collection of human behaviors and interactions among humans and artifacts. Figure 6 (next page) illustrates how we process the data from our UEM to develop the interaction corpus and how applications use this data.

Here, we extend the definition of the UEM to include not only sensors but also actuators and spaces, where people have experiences, share experiences, and recreate experiences. This is because our experience is based on interacting with other humans and artifacts in a space. The trigger by the actuators of the artifacts, for example, can be useful context information for describing the interaction in rich detail. Intelligent and autonomous artifacts (robots) will be placed everywhere and will make the experience enjoyable and recordable.

From sensing to indexing

The UEM records the experience events. The recorded events and interactions are annotated automatically, semiautomatically, or manually depending on the complexity of tasks and objects. If we use ID-emitting devices such as our UEM, we can capture relative positional information automatically and use it to describe human interactions with objects. An intelligent vision system might provide the names of objects and people in a captured scene in the future. Discussion recording could also be automatic. However, the automatic estimation of the importance of topics is not an easy task. Our current

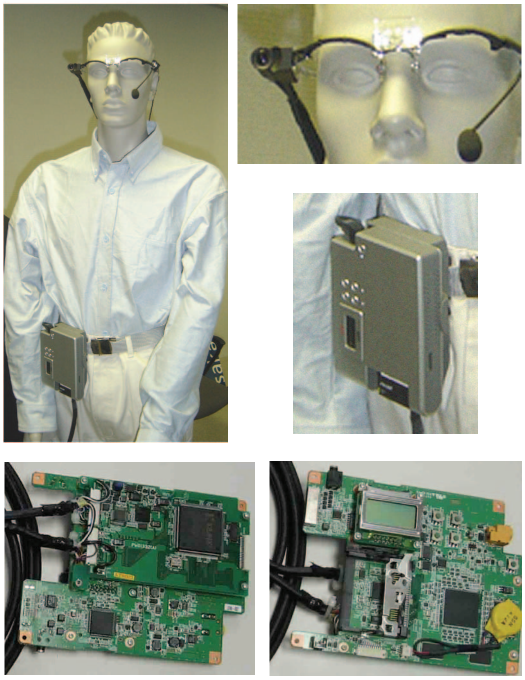


Figure 5. Lightweight wearable PC and third-generation UEM. The IRID tag is placed in the middle of the user's forehead.

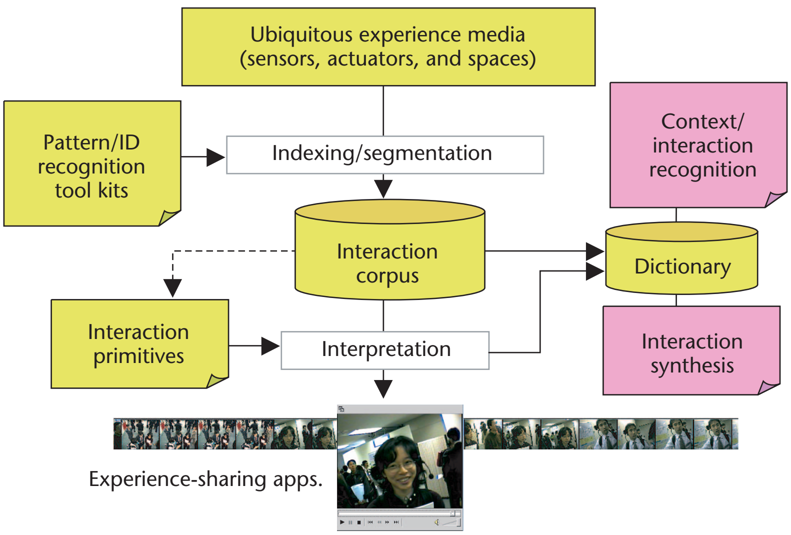


Figure 6. Interaction corpus developed from UEM inputs. The UEM data is segmented and indexed by multimedia processing. The corpus is used in various experience-sharing applications by interpreting the content in a service domain. The corpus will be refined in the future to create a general-purpose interaction dictionary.

semantic interpretation can label the event as group discussion, joint attention, and so on. We focus on context analysis rather than content analysis as an initial step. Once we clearly understand interactions with a syntactic structure, we'll be able to handle content with semantic analysis. Also, annotation could be created personally or socially by a group of people. Both types are used to segment and annotate the interactions and to index each interaction.

Interaction corpus structural design

We propose a four-layered data structure for the interaction corpus to manage the interactions systematically. Its lower layers are designed to be universal, while the top (fourth) layer is application oriented. The combinations can be made in a top-down manner by interaction designers or in a bottom-up manner where combination patterns are extracted from the real interaction data with some data-mining methods. We describe the top-down approach in this article. For a bottom-up approach to extract the composites, see Morita et al.[10]

The first (lowest) layer is composed of all of the captured data stored in the data server. These data are recorded with the time information in a hierarchical structure. The second layer consists of segmented interaction scenes from the IRID tracker data and speech activity. We define interaction primitives in the third layer as elemental intervals or moments of activities. For example, a video clip that contains a particular object (such as a poster or a user) in the view constitutes a "Look" event. Since the identifications and the viewing coordinates of all objects are known from the IRID tracker and tags, it's easy to determine these events. We then interpret the meaning of events by compositing the primitives in the fourth layer across different users and sensors to obtain a more socially oriented description.

Let's look into the details of each layer.

Raw-data layer

The first (lowest) layer, called the *raw-data layer*, stores pairs consisting of raw sensor data and their time stamps. These are instantly obtained and used in immediate applications. In the current implementation, the information of the IRID tracker (ID and position), video, and audio are used, and their data are stored separately with time stamps.

Segmentation layer

The second layer is called the *segmentation layer*. The data is segmented automatically by preprocessing for each modality of data, such as an IRID data stream and microphone volume. For the IRID data stream, the segmentation layer stores fragments of short gazing information in the raw-data layer. These are combined into a more meaningful cluster by merging neighborhood gazing fragments if they're temporally close to each other. The remaining isolated and short IRID data are considered noise and deleted. For the microphone volume, the segmentation layer uses a conventional threshold technique to first define a speech event. Then the merging and noise deletion are done to define an utterance segment.

The clustering algorithm is a sequential one. It first merges consecutive events into a current cluster when their interval is less than a maximum (max) interval. If a new event appears over the max interval, it becomes a seed of another cluster. After all signals are merged into segments, shorter segments of less than a minimum (min) duration are deleted as noise, as Figure 7 shows. We used parameters of 8 or 4 seconds for tag data and 4 or 2 seconds for speech data as the max interval and min duration, respectively. These parameters should be set empirically based on the sensor sensitivity and the speaking and gazing behavior of a situation.

We constructed the segmentation layer based on the activity-detection modules of speech, vision, actions, and so on. Thus, the second and upper layers of the proposed corpus depend on the detection algorithm's quality and robustness. It evolves as the media technology progresses. The algorithms used should be coded in the corpus.

Primitive layer

The third layer, called the *primitive layer*, stores interaction primitives, each of which we define as a unit of interaction. In this layer, an interpretation of interaction is attached to the segmentation layer data. This process is similar to the morphological analysis of human language. We extract the interaction primitives based on gazing information from the segmented ID (tag) data and the utterance information from the segmented utterance. The former is a binomial interaction represented as "A look at B," while the latter is a monomial interaction, "A speak."

Using the multimodal segments in combination or alone, we define various interaction primitives. We believe that both gazing and utterance provide important data for finding human–human or human–object interactions. For example, when the UEM of user A detects the gazing of a tag of person B, we call the interaction primitive a "Look_at" situation. When the UEM detects speaking, it calls the interaction primitive "Speak." Moreover, when the system detects "Look" and "Speak" simultaneously, the interaction is called "Talk_to."

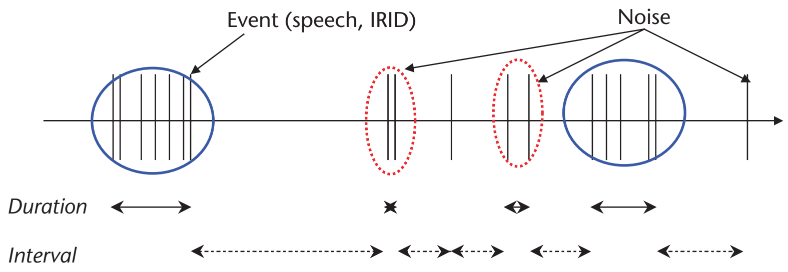


Figure 7. Segmentation of events to interactions. Consecutive events are merged to a cluster. Shorter segments are deleted as noise.

Figure 8a shows some examples of the interaction primitives. An object such as a presentation poster, using the stationary camera attached to it, looks at the human in front of it. In this manner, we can introduce other modality primitives when an appropriate sensor is available, such as "Touch," "Walk," and "Sit."

Composite layer

The fourth (top) layer, called the *composite layer,* stores the composite interactions that are more socially oriented and application dependent. By combining the interaction corpus' composites, we can better represent the scenes of interaction. In the exhibition situation, for example, we store correlation interaction data between two or more clients (humans' or objects' data). If two people

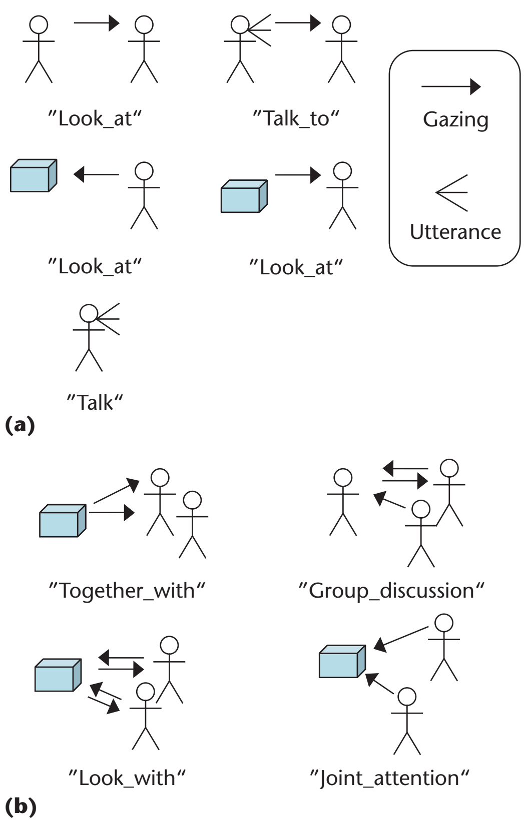


Figure 8. Structure of the interaction corpus. (a) Samples of primitive interactions. (b) Samples of composite interactions.

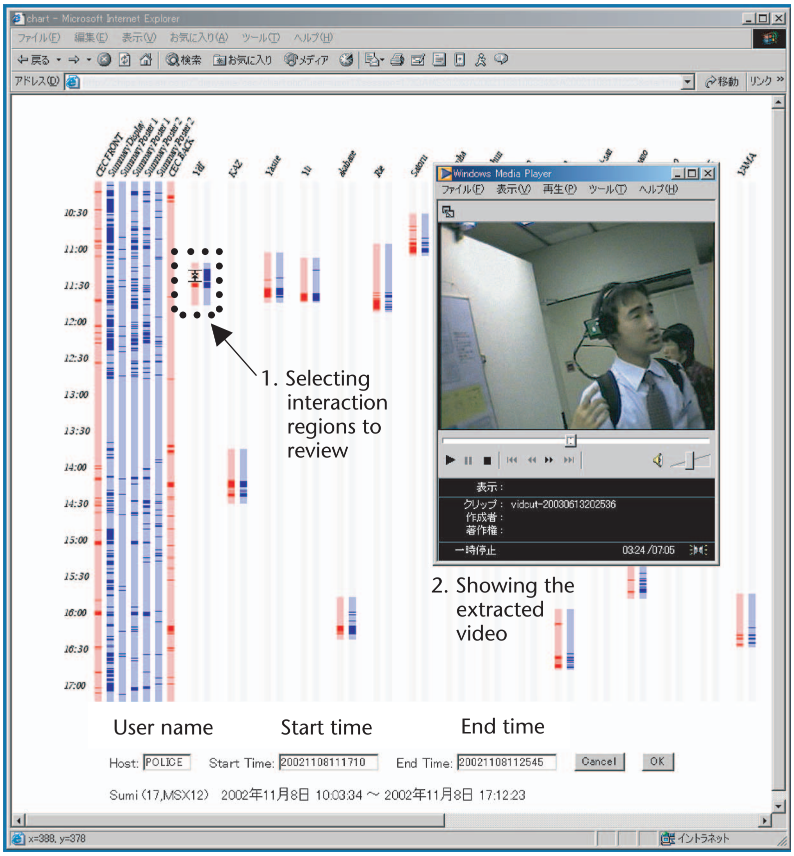


Figure 9. Interaction corpus viewer. The interaction corpus is searchable by posing a query with the user's name and time of an event, then selecting an interaction region. Corresponding video clips are replayed in a pop-up window.

"Talk_to" each other, the interaction is recorded as "Talk_with." When three or more people are interpreted as "Talk_to" or "Look_at," the composite is interpreted as "Group_discussion." Figure 8b shows examples of the composites.

Interaction corpus applications

We can use the interaction corpus as a well-structured experience data collection that's shared with other people for communicating and creating further experiences.

We aim to use this corpus—as a context-elements collection—to share previous experiences with others. The interaction corpus is also used in other applications, such as interaction facilitation by humanoid robots or HMDs for experience sharing by summarizing experiences. We demonstrated these in an exhibition setting with research poster presentations at the 2003 ATR exposition.

Database query

As a basic analysis tool, we developed a corpus viewer that we can use to browse any event involving multiple users by posing a query for a specific event search. Figure 9 shows a snapshot of the corpus viewer. It shows the color-coded interactions of the primitive and composite layers in chronological order for each user and client. Since the database is stored using a MySQL database system, you can pose a query such as "user A from Start_time to End_time" for all interactions involving user A. A GUI-style interface lets you choose a region of interest to review the experience. Social scientists have used the viewer for interaction analysis in our lab.

Experience summary

We were able to extract appropriate scenes from the viewpoints of individual users by clustering events having spatial and temporal relationships. We created a summary page by chronologically listing scene videos, which we automatically extracted based on interactions. The system assigned each scene with annotations, such as time, description, and duration. The descriptions were automatically determined according to the interpretation of extracted interactions by using templates—for example, I talked to someone; I stayed with someone; and I looked at something.

We also provided a summary video for a quick overview of the events that the users experienced. To generate the summary video, we used a simple format in which relevant scenes of at most 15 seconds were assembled chronologically with fading effects between the scenes. The system selected scenes from the interactions of the composite layer, focusing on a particular user whose name was an SQL query. The summary search engine looks for interactions that have the user's name (ID) in the corpus database, and if the interaction length is longer than a preset duration, the interaction is chosen as part of the summary. We designed these criteria for this open laboratory exhibition by supposing that the visitor might want to get reports on long visits or meetings.

We didn't limit the event clips used to make up a scene to those captured by a single client's UEM. For example, as Figure 10 shows, for a summary of a conversational "Talk_with" scene, the video clips used were recorded simultaneously by a fixed camera (see Figure 10a) on the ceiling that captured both users, the camera (see Figure 10b) of the conversation partner, and the camera (see Figure 10c) worn by the user herself. Our system selected appropriate video clips to make a summary by consulting the speech events (volume levels) of the users' individual voices. The system assumes that the IRID tag worn by a user indicates that his or her face is in the video clip if the associated IRID tracker detects it. Therefore, the integration of video and audio from different worn sensors could generate a scene of a speaking face from one user's camera with a clearer voice from another user's microphone. This exchange feature is invoked when the situation is recognized as "Talk_with" and is annotated accordingly in the corpus.

In the current system, the generated summary is a simple sequence of events only bounded by the vocabulary of primitive scenes. The summary itself effectively expresses the experience. However, we think that a directed summarization would help a person enthusiastically tell others about the experience. We have already developed the ComicDiary,[11] a directed diary-generation system using comic presentation techniques. It incorporates predefined stories by directors to generate storytelling based on a fraction of experience records stored in a personal digital assistant. We plan to integrate the system with the video summary system to generate a storytelling summary based on the interaction corpus.

Interaction facilitation

Interaction facilitation has two aspects. One aspect is to facilitate interactions for the benefit of a capturing system. In our sensor-room setting, we provided wearable and ubiquitous clients. If a user's position isn't appropriate for a stationary sensor (such as a camera or microphone) because of its distance and orientation, we might want to bring the user to a suitable location to record his or her voice and picture clearly for later use. To help with this, we introduced facilitating entities (facilitators), such as robots and visual guides, to aid in capturing data.

The other aspect is to facilitate people's interaction with other people or artifacts for people to enjoy talking with others and investigating the events. In our exhibition scenario, visitor assistance was helpful to many visitors.

Robot facilitation. In an exhibition situation, an assistant often guides a visitor to pay attention to a specific event. In our prototype system, a humanoid communication robot, Robovie-II, served as such an assistant. Robovie wears an IRID tag and tracker and engages in communication with the visitor in front of it by referring to the already accumulated interaction corpus to create new interactions. We demonstrated several methods of interaction facilitation with Robovie, such as

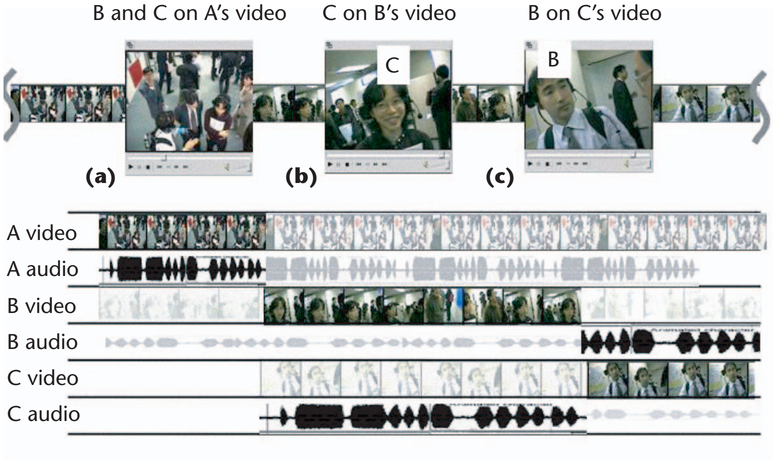


Figure 10. Experience summary by three clients. (a) Footage from a fixed camera on the ceiling captures both users. (b) Footage from camera of a conversation partner. (c) Footage from the user's camera.

- calling the visitor's name as a greeting and getting his or her attention by posing a query;
- guiding the visitor by gesture and voice to other places (booths) of his or her interest based on room conditions, such as the space becoming crowded (by checking the current number of visitors at each booth);
- talking about other people who have visited before or who have met Robovie; and
- giving informative announcements about the place the person is visiting, such as local weather information and the total number of visitors.

Robovie is good at attracting the attention of people and guiding them to places because it has a humanoid upper body, eyes, head, and hands. As these body parts move quickly to make gestures naturally, people are drawn to the robot's facilitation[12] (see Figure 11). In our experiment

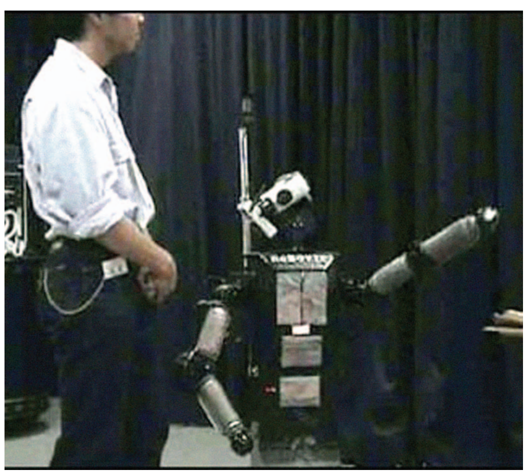


Figure 11. Robovie's interaction facilitation. Robovie is pointing at a poster with hand and head gestures, using eye camera gazing to obtain the user's attention.

on robot interaction, a human subject tended to respond better unconsciously to the robot's pointing gesture when it used all of its parts to make it (for example, aligned to the subject, gazing and pointing to the target while making occasional eye contact) than while using fewer parts' movements. These services are possible because the robot can discern the user's situation and the situation of other booths from the concurrent interaction corpus.

Visual and auditory facilitation. Robovie's interaction functions provided visitors with public and personal assistance. However, it was only given when the visitor encountered the robot. To provide a continuous facilitating service, we loaned visitors an HMD with a wearable computer to provide them with augmented visual information. The display recommended persons and booths that the visitor might like to meet and visit. The system selects a person of interest from the previous interaction corpus data based on the similarity of an interest vector. The vector is formed from the stay time at each booth as an element. The HMD computes similarity by taking an inner product of the vectors. It performs booth recommendation in a similar fashion. The HMD also shows augmenting information about the person and the booth in front of the user. The similarity rate of interest between people is calculated and displayed on the HMD's screen. It also displays the popularity of visited booths on screen.

Conclusion

We first constructed the UEM to capture experience in the first-person view and in the third-person view and then extended it to the second-person view by incorporating a humanoid robot as a partner. Multiple multimodal sensors were either worn or placed ubiquitously in the environment. Through a demonstration of our system at exhibition events, we were able to provide users with a video summary at the end of their experience, made on the fly. The videos were well accepted by the participants in the exhibition event, but the quality of summarization should be investigated in a controlled environment.

In the future, we hope to develop a system that will let researchers quickly query specific interactions by using simple commands and provide enough flexibility to suit various needs. We prototyped three generations of wearable clients for the UEM and achieved the development of a fairly small module with an IRID tracker. We'll continue its evaluation in real situations. For instance, we want to bring the wearable clients outdoors for nature observations, recording the experience and discussing the findings in a collaborative learning setting.

We often discuss developing an intelligent human interface for a computer system that performs a certain task for humans. Rather, future intelligent systems will become our symbiotic partners in the form of tangible or embedded companions in daily life. An important requirement to realize such a new paradigm is for future intelligent systems to have and use common sense and knowledge of human sociability. We can construct such a system by considering the corpus of interactions that we've gathered. **MM**

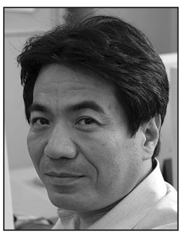
Acknowledgments

Tetsuya Matsuguchi, Tetsushi Yamamoto, Sidney Fels, Atsushi Nakahara, Yasuhiro Katagiri, Masashi Takahashi, and members of Advanced Telecommunications Research Institute International's (ATR's) Media Information Science and Intelligent Robotics and Communications Laboratories made valuable contributions to this work. The National Institute of Information and Communications Technology supported this research.

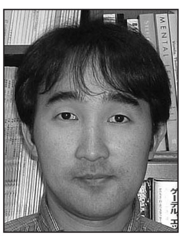
References

1. V. Bush, "As We May Think," *The Atlantic Monthly*, vol. 176, no. 1, July 1945, pp. 101-108.
2. R. Jain, "Digital Experience," *Comm. ACM*, vol. 44, no. 3, 2001, pp. 38-40.
3. K. Mase et al., "Interaction Corpus for Experience Sharing Using Ubiquitous Experience Media," *Proc. 14th Int'l Conf. Artificial Reality and Telexistence*, 2004, pp. 55-61.
4. M. Tsuchikawa et al., "Low-Stress Wearable Computer System for Capturing Human Experience," *Proc. 9th IEEE Int'l Symp. Wearable Computers*, IEEE CS Press, 2005, pp. 27-33.
5. Y. Sumi et al., "Collaborative Capturing an Interpretation of Experiences," *Pervasive Computing: A Collection of Contributions Presented at PERVASIVE 2004*, Oesterreichische Computer Gesellschaft, 2004, pp. 253-258.
6. B. Clarkson, K. Mase, and A. Pentland, "The Familiar, a Living Diary and Companion," *Proc. Computer–Human Interaction 2001 Extended Abstract*, ACM Press, 2001, pp. 271-272.
7. J. Gemmel et al., "Passive Capture and Ensuing Issues for a Personal Lifetime Store," *Proc. 1st ACM Workshop on Continuous Archival and Retrieval of Personal Experiences*, ACM Press, 2004, pp. 48-55.

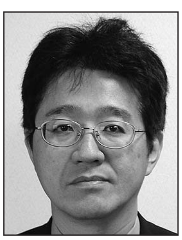
8. R. Stiefelhagen, J. Yang, and A. Waibel, “Towards Tracking Interaction between People,” *Proc. Intelligent Environments Am. Assoc. for Artificial Intelligence Spring Symp.*, AAAI Press, 1998, pp. 123-127.
9. S. Ito et al., “Implementation and Evaluation of Wearable Sensor Devices for Recording and Detecting Conversational Interaction,” *Proc. 9th IEEE Int’l Symp. Wearable Computers*, IEEE CS Press, 2005, pp. 130-137.
10. T. Morita et al., “A Pattern Mining Method for Interpretation of Interaction,” *Proc. 7th Int’l Conf. Multimodal Interfaces*, ACM Press, 2005, pp. 267-273.
11. Y. Sumi et al., “ComicDiary, Representing Individual Experiences in a Comics Style,” *Proc. Conf. Ubiquitous Computing*, LNCS 2498, G. Borriello and L.E. Holmquist, eds., Springer, 2002, pp. 16-32.
12. Y. Koide et al., “An Approach to Integrating an Interactive Guide Robot with Ubiquitous Sensors,” *Proc. 2004 IEEE/Robotics Soc. of Japan Int’l Conf. Intelligent Robots and Systems*, IEEE CS Press, 2004, pp. 2500-2505.



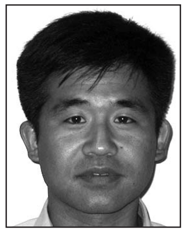
Kenji Mase is a professor at the Information Technology Center of Nagoya University. His research interests include image processing, ubiquitous computing, and their applications for computer-mediated communications. Mase received his MS and PhD in information engineering from Nagoya University.



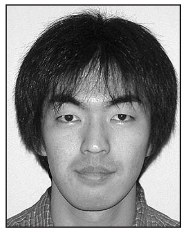
Yasuyuki Sumi is an associate professor in the Graduate School of Informatics at Kyoto University. His research interests include knowledge-based systems, creativity-supporting systems, interface and social agents, ubiquitous and wearable computing, Web intelligence, multimedia processing, and their applications for facilitating human interactions and collaborations. Sumi received his MEng and DEng in information engineering from the University of Tokyo.



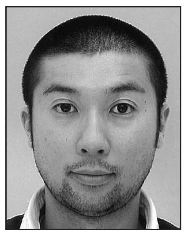
Tomoji Toriyama is a group leader of the Knowledge Science Laboratories at ATR, Japan. His research interests include Large-Scale Integration electrical integrated circuit (LSI) architecture design methodology, human interface, and image processing. Toriyama received his PhD from Toyama Prefectural University, Japan.



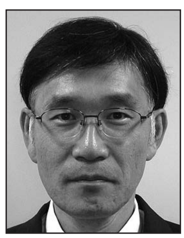
Megumu Tsuchikawa is a manager of the Intellectual Property Center at NTT, Japan. His research interests include image processing and image understanding. Tsuchikawa received his MEng in mechanical engineering from Waseda University, Japan.



Sadanori Ito is a lecturer at the Graduate School of Engineering, Tokyo University of Agriculture and Technology. His research interests include knowledge-based systems, creativity-supporting systems, and ubiquitous and wearable computing. Ito received his MS and PhD in knowledge science from the Japan Advanced Institute of Science and Technology (JAIST).



Shoichiro Iwasawa is a research expert at the Universal Media Research Center at the National Institute of Information and Communications Technology (NICT), Japan. His interests include digital face cloning and synthesis and motion-capture data processing. Iwasawa received his MEng and PhD in electrical engineering and electronics from Seikei University, Japan.



Kiyoshi Kogure is the director of the ATR Knowledge Science Laboratories. His research interests include intelligent environments, intelligent agents, and natural-language processing. Kogure received his PhD in electrical engineering from Keio University.

Norihiro Hagita established the ATR Media Information Science Labs and ATR Intelligent Robotics and Communication Labs. His major interests are communication robots, network robot systems, interaction media, and pattern recognition. Hagita received his MS and PhD in electrical engineering from Keio University.

Readers may contact Kenji Mase at mase@nagoya-u.jp.

For further information on this or any other computing topic, please visit our Digital Library at http://computer.org/publications/dlib.