Analysis Environment of Conversational Structure with Nonverbal Multimodal Data

Yasuyuki Sumi  
Graduate School of Informatics, Kyoto University  
Kyoto 606-8501, Japan  
sumi@i.kyoto-u.ac.jp

Masaharu Yano  
Graduate School of Informatics, Kyoto University  
Kyoto 606-8501, Japan  
yano@ii.ist.i.kyoto-u.ac.jp

Toyoaki Nishida  
Graduate School of Informatics, Kyoto University  
Kyoto 606-8501, Japan  
nishida@i.kyoto-u.ac.jp

ABSTRACT
This paper shows the IMADE (Interaction Measurement, Analysis, and Design Environment) project to build a recording and analyzing environment of human conversational interactions. The IMADE room is designed to record audio/visual, human-motion, eye gazing data for building interaction corpus mainly focusing on understanding of human nonverbal behaviors. In this paper, we show the notion of interaction corpus and iCorpusStudio, software environment for browsing and analyzing the interaction corpus. We also present a preliminary experiment on multiparty conversations.

Categories and Subject Descriptors
H.1.2 [MODELS AND PRINCIPLES]: User/Machine Systems; H.4.3 [INFORMATION SYSTEMS APPLICATIONS]: Communications Applications

General Terms
Experimentation, Human Factors

Keywords
Multiparty conversation, Multimodal data analysis, Analysis environment

1. INTRODUCTION
In conversations, humans express our various intentions to others by nonverbal behaviors, such as the gaze direction, gestures, nodding, and back-channel feedback. Humans make such nonverbal behaviors with certain temporal/spatial patterns. We, humans, can unconsciously convey/understand mutual mental status and control the flow of our conversations with the nonverbal behaviors. Current computers, however, cannot understand the semantic structure of such human nonverbal behaviors.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyright 2010 ACM 978-1-4503-0414-6/10/11 ...$10.00.
interaction situations as visual data.

**Headworn microphones** By equipping all participants in the environment with wireless headworn microphones, we can record their individual utterances.

**Motion sensors** By equipping each participant and objects (potential focal points of conversations like posters) in the environment with markers, human motions and their relative positions can be recorded as 3D data. While there are some studies about 3D human motion detection with video information[5], motion capture system is useful for recording conversations with little constraint. We use the motion capture systems of the Motion Analysis Corporation.

**Eyemark recorder** We record the gaze directions of each participant in the environment as video and numerical data. We use two types of eyemark recorders: one is Mobile Eye from Applied Science Laboratories and another is EMR-9 from NAC Image Technology. Combining the coordinates from motion sensors and the numerical values of gaze directions, we can calculate the absolute coordinates of the gaze vector in the space.

In addition to these sensors, extra sensors such as polygraph can be used when necessary.

3. **ANALYSIS OF CONVERSATIONAL STRUCTURES BASED ON INTERACTION CORPUS**

We collect and annotate the captured data in the IMADE room based on hierarchical structure of interaction corpus as shown in Figure 2, to enable to use structured and coordinated, recorded sensor data when analyzing interactions.

The interaction corpus structure is separated into four layers by the interpretation steps and structured by accumulating such layers. The first layer is composed of the recorded raw data, and no interpretation is done in this layer. The second layer is the interaction primitive layer, where such primitive interaction activities as utterance and gazing are extracted from the sensor data. The third layer is formed by calculating the combinations of multiple primitive data concerning the relations between them. In the layer, we can observe several interesting social interactions such as joint attention, conversations sharing focal objects, etc. The fourth layer concerns the flow of interactions and comprises
The transitions of such situations as change of dominance of conversation group, mergers and separations of conversation field, and so on.

4. ICORPUSSTUDIO: ANALYSIS ENVIRONMENT

We developed iCorpusStudio, which is a software environment for browsing, annotating, and analyzing by prototyping interpretation. iCorpusStudio has two parts: one displays data and the other constructs interpretations. With iCorpusStudio, we can simultaneously view such recorded data as multiple video, audio, and motion while annotating the interpretations of interactions as labels.

Figure 3: Screenshot of iCorpusStudio

Figure 3 shows a screenshot of iCorpusStudio. The left side shows the videos and audio and the upper right displays the modeled view of the motion data. The lower right displays the labels and the numerical data retrieved by sensors for annotating and operating the data.

The entire interpretation process with iCorpusStudio is shown in Figure 4. We describe most of the processes in the follows.

4.1 Recording interaction interpretations with labels

iCorpusStudio stores the interaction interpretations as labels for the periods in which the interactions were done. While such annotating tools as ANVIL\(^1\) and ELAN\(^2\) exist, we can both manually and automatically apply labels by calculating the data with iCorpusStudio.

By viewing the videos, we can label the period in which a subject turns from one thing to another and annotate a certain period of a conversation from the audio data as ‘subject A is speaking,’ ‘saying “I know it,”’ or ‘talking about traveling,’ and so on.

Additionally, we can get a more abstract interpretation of conversations by computing from labels. For example, we can calculate the periods of joint attentions from the gazing labels of each subject and compute the periods in which one subject turns toward the location at which another subject is pointing one second before from the temporal relations between pointing and gazing. Furthermore, iCorpusStudio can reduce label operation procedures so that we can apply one interpretation procedure for data to another.

4.2 Visualization of human behaviors from motion data

Motion data visualized based on 3D CG (Computer Graphics) human models is useful to view the essence of the participants’ behaviors from any angles. However, collected raw data from the motion capture system is only collections of points, and has difficulty to find conversation structures.

To solve this problem, iCorpusStudio supports to create 3D models from the coordinates of the points that visualize human behaviors in conversations. For example, a head model is formed as a sphere that encloses the points of the top of the head and the neck, a body’s trunk is formed as a cylinder of the shoulder to the waist, and an arm is formed as a cylinder of the shoulder to the wrist through the elbow.

In addition, using calculated vectors from data of eyemark recorders, we can view 3D vectors of the gaze directions. With models, we can easily determine the directions of the arms and heads, and calculate the coordinate relations of participants.

4.3 Pre-symbolic interpretation of interactions with numerical data

iCorpusStudio can show plotted graph of such numerical values as coordinates and angles from the 3D data recorded by motion sensors and eyemark recorders. For example, Figure 5 shows the plotted data of the distance between a human’s body and his/her hand recorded by motion sensors. When the hand is away from the body at the high values, we can interpret that he/she was gesturing in those periods.

iCorpusStudio can also show videos of those periods with which we can verify that he/she was gesturing in all of these periods. iCorpusStudio can not only display values from the 3D data but also calculate values with those data and display them. Figure 6 is an example of calculation with

\(^1\)http://www.anvil-software.de/
\(^2\)http://www.lat-mpi.eu/tools/tools/elan
iCorpusStudio, in which we calculated the series of angles between the directions of the head and the relative location to another subject, differentiated the angles, and obtained the labels as ‘quick turning to another participant’ at the periods of higher values. By using such operations, we can quickly and roughly pinpoint interesting scenes from exhaustive huge data.

Other operations for numerical values include arithmetic operations, smoothing, labeling series of values with maximum values among a set of series at each period, and calculating the statistics of the series of values in periods with particular labels, etc.

5. PRELIMINARY EXPERIMENT ON MULTIPARTY CONVERSATION

We have recorded various types of conversational interactions using the IMADE room and built a recording corpus using iCorpusStudio. This section shows a preliminary report of analysis of multiparty conversation.

Figure 7 shows the recording settings of the experiment. In the experiment, we expected to observe several macroviewpoint phenomena such as moves of conversational fields, topic transition; and micro-viewpoint phenomena such as cooccurrence pattern of speech, gaze, and gesture.

Three participants attended the conversation and were instructed to move among the panels while they discussed the poster contents and viewed the panels. We located six panels of posters shown in Figure 7. The panels are set of old paintings illustrating city lives, landscape, and buildings in mid-age Kyoto. The participants were students of Kyoto University, so we intend them freely chat about their own daily lives in Kyoto inspired with the old paintings.

They were equipped with eyemarker recorders, headworn microphones, and markers for motion capture system to record their behaviors and utterances. Cameras were also mounted around the environment. The markers were attached at four points on the head, one at the back of neck, one at the center of the back, two at the shoulders, and one at both elbows and both wrists. We also attached markers at the four corners of each panel to record the panel coordinates.

We instructed the participants to freely chat and finish the chat at anytime they satisfy. Resultly, we could have two sets of conversational data of around 30 minutes. These settings enabled us to record such gestures as pointing, the reactions to those gestures, the transitions of the conversation situations with changes of topics, and the transitions of gazes with changes in the targets of interest.

We have been analyzing the data using iCorpusStudio and found many interesting patterns of nonverbal behaviors. For example, this data contains many types of conversational scenes such as discussion scene along with the individual panel, chat with F-formation, separately exploring, etc. We could confirm that those scenes were easily and automatically classified based on distribution of head directions and distance among the three participants.

6. ACKNOWLEDGMENTS

Earlier version of iCorpusStudio was developed by Hiroyuki Kijima and Atsushi Nakata. Building the IMADE room was done with many colleagues, especially, Ryohei Fukuma, Takuma Nakazawa, Hisao Setoguchi, Hiroshi Katsumi, and Hiroyasu Saiga. The authors are deeply grateful to Tetsuya Kawahara, Katsuya Takanashi, and Mayumi Bono for valuable discussion.

7. REFERENCES


