Chang, M. et al. (Eds.) (2019). Proceedings of the 27th International Conference on Computers in Education. Taiwan: Asia-Pacific Society for Computers in Education

Multimodal Recording System for Collecting Facial and Postural Data in a Group Meeting

Yusuke SONEDA[a*], Yuki MATSUDA[a], Yutaka ARAKAWA[bc] & Keiichi YASUMOTO[a]
[a]*Nara Institute of Science and Technology, Japan*
[b]*Kyushu University, Japan*
[c]*JST PRESTO, Japan*
*soneda.yusuke.su2@is.naist.jp

Abstract: By the spread of active learning and group work, the ability to collaborate and discuss among the participants becomes more important than before. Although several studies have reported on that micro facial expressions and body movements give psychological effects to others during conversation, most of them are lacking in quantitative evaluation and there are few datasets about group discussion. In this research, we proposed a highly reproducible system that helps to make datasets of group discussions with multiple devices such as an omnidirectional camera (360-degree camera), an eye tracker and a motion sensor. Our system operates those devices in one-stop to realizing synchronized recording. To confirm the feasibility, we built the proposed system with an omnidirectional camera, 4 eye trackers, and 4 motion sensors. Finally, we succeeded to make a dataset by recording 8 times group meeting by using our developed system easily.

Keywords: Group discussion, active learning, multimodal communication dataset

1. Introduction

In the 21st century, the thinking and cooperating abilities gained through discussion have regarded significant skills (Arpan, 2017). A model called elaboration likelihood model (ELM) exists as a hypothesis as a communication model to persuade the listeners, and it is a content that has been actively studied for a long while (Richard, 1986). Small-group discussions are more effective on the retention of knowledge than classes in the large classroom. Hence, the classes utilizing small-group discussions should become mainstream in the future (Philip, Kerstin, Bruce, & Wilson, 2011). To improve the ability of discussion, it is important to assess the group discussion quantitatively. However, there are no related work that has realized easy and quantitative evaluation with multimodal devices.

We can divide communication with human beings into verbal communication and nonverbal one. Nonverbal communication has a large impression on the conversation partner and is a very important ability to build social relationships (Peter, 1987; Michael, 1988). Micro-expressions on the face has a large impression too, and it is important information for understanding the deep psychology of the human (Ekman & Friesen, 1975). On the other hand, analysis of the content of the conversation may include sensitive private information, so it is difficult to introduce a system that analyzes the content of the conversation in an actual educational scene. From the above, it is important to develop a system that evaluates the quality of communication from micro behavior and facial expression, and assists the students participating in the discussion.

Many comprehensive face image datasets for facial expression recognition are famous and some papers report the high accuracy results in emotion recognition (Kanade, 2000; Yan, 2014). But there are few datasets that have been labeled for multiple discussions. In this paper, we propose a system for making datasets about group discussion and describe the consideration of some datasets measured.

This article is organized as follows: Section 2 provides a description of related work, and Section 3 describes a proposed system and the experimental condition. Then, the results of the experiments are shown in Section 4, and Section 5 concludes this paper with future perspectives.

2. Related Work

Hori et al. (2012) developed the system that analyzes face direction and the contents of the statement real time with combining an omnidirectional camera and several directional microphones. But the audio analysis performed by this system is highly influenced by the form of the meeting room, so it works only in pre-optimized rooms. In addition, very expensive and high-spec equipment is required to analyze, it is difficult to introduce in general schools and companies.

Ohnishi et al. (2019) have developed the system that recognizes nodding, speaking and looking motions from acceleration and angular velocity data from a head motion with a 9-axis acceleration sensor. By amelioration of the algorithm, it is expected that the accuracy of recognition will be further improved, this paper reports head motion acceleration and angular velocity is very useful data in recognizing head behavior such as nodding. However, this system needs to synchronize the recorded video with the timestamp of the accelerometer and it is not automated. It is very laborious to match the labeling data of the recorded video with the timestamp of the acceleration sensor by human hands. If there is a time stamp gap between the video and the sensor data, there is a risk that the accuracy will decrease when performing activity recognition.

Thus, to make datasets for group discussions in various educational scenes, the reproducibility of the recording system and the mechanism synchronize each device data are important. We propose a system that has two major features. The first feature is that it is a highly reproducible system which can easily make group discussion datasets and can be used in any room. The second feature is that the proposed system uses multiple types of devices to acquire more data and each timestamp does not slide from the exact time by operating in one-stop.

3. Proposed Methods

3.1 Assumed environment

Four people hold a five minutes discussion two times per experiment. The topic in this discussion does not depend on the knowledge possessed by the participants, each participant can talk equally. We prepared the themes which the answer is narrowed to two ways. For example, “If you have a time machine, which do you want to travel to the past or the future ?”, “Which do you like cat or dog ?” and so on. From the presented agendas, we asked the participants to select two themes which all of them don’t have the same opinion, and each participant discusses their views. After each discussion is over, we make surveys of discussion for them.

3.2 Devices

We developed a measurement system using three types of devices to make data sets for multi-person communication. The devices are THETA V, Pupil Mobile Eye Tracking Headset, (hereinafter called “Pupil”) and LPMS-B2.

THETA V (https://theta360.com/en/about/theta/v.html Last access 15 Aug 2019) is the omnidirectional camera developed by RICOH, it can record video with 29.97 fps frame rate and 4K pixels. In addition, since it can behave like a Wi-Fi base station, it is possible to get shooting commands from a PC or mobile terminal. It records the faces and bodies of every participant, we prepared only one.

Pupil (https://pupil-labs.com/pupil/ Last access 20 May 2019) is a glasses-type eye tracking system developed by Pupil Labs. From the eye direction and movement, it is possible to measure which direction the participant was looking at. In the case of this study, it is used to detect the person who is watching during the discussion. Pupil connects to a laptop to perform analysis and recording. In this experiment, we used Surface provided by Microsoft. When Pupil measurement application is activated, it can receive commands for recording from the background, so it is possible to send commands for shooting the video remotely using wired LAN or Wi-Fi. Each participant attaches Pupil, we prepared four Pupils and four Surfaces.

LPMS-B2 (https://lp-research.com/lpms-b2/ Last access 15 Aug 2019) is a lightweight 9-axis wireless motion sensor developed by LP-Research. The measure of this device is 39 × 39 × 8mm and

weight is 12g. It can transmit acceleration information to a PC or mobile terminal via Bluetooth. OpenMAT, GUI open source application, control and collect data sent from LPMS-B2. LPMS-B2's timestamp recording starts when a measurement is started by OpenMAT. Moreover, since it is possible to simultaneously acquire data from several LPMS-B2s, acceleration data obtained from them don't have a difference of timestamp. In this experiment, the sampling frequency was set to 100Hz, and data on the 9-axis acceleration of the head was acquired by attaching to a Pupil. Each Pupil has LPMS-B2, so we prepared four LPMS-B2s.

3.3 Proposed System

In this research, we define data measured by these three types of devices and data annotated to omnidirectional video as a dataset in the group discussion. As we described in Section 2 (Related Works), for accurate analysis, it is very important that each timestamp of the data acquired by several devices is not out of alignment. We developed a system which sends commands for the THETA V to record the video via Wi-Fi, for Surfaces to record the pupil video via wire connection and for OpenMAT to start acquiring data from LPMS-B2. A diagram of the proposed system is shown in Figure 1.

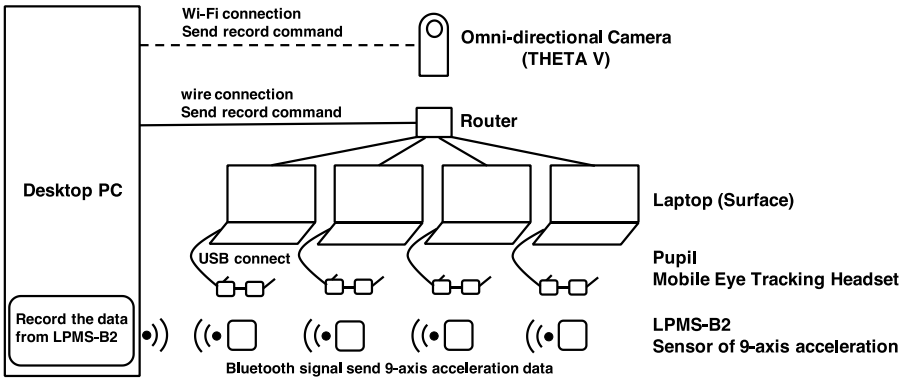


Figure 1. Proposed System Diagram.

Originally it is necessary to manually activate all devices and record a discussion, our proposed system possible to automatically acquire all data simultaneously. Hence, the timestamps of acquired data match with each other, we can make easily high-quality multi-sensor datasets about group discussion.

Figure 2 shows the image which a participant wearing Pupil with LPMS-B2. Figure 3 shows the image of an actual experiment.

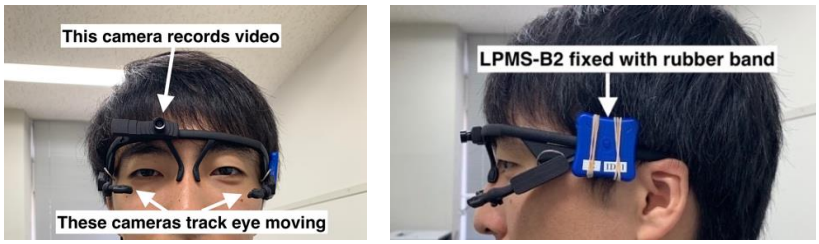


Figure 2. Pupil with LPMS-B2.

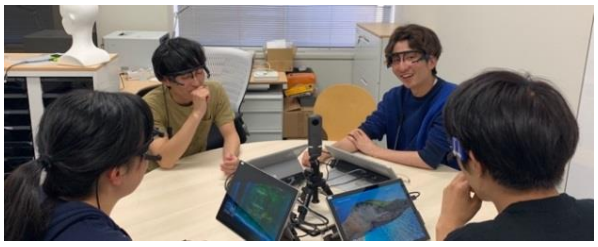


Figure 3. Experiment image of group discussion.

3.4 Annotation

Each participant annotates videos of THETA V about their behavior by using ELAN, a software that can perform time-series annotation on video. Figure 4 shows the image of annotating with ELAN. THETA's video is expanded in a rectangle. When annotating, we ask them to annotate with paying attention to the reason for behavior. Table 1 shows a list of labels of annotation. These labels are decided with reference to the book of Peter (1987).

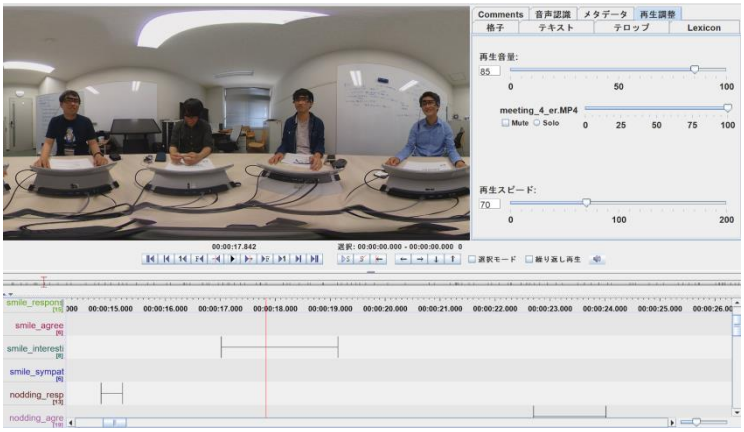


Figure 4. Annotation with ELAN.

Table 1

The contents of labeling

Label name	Category	Detail
smile	response	does not have special meaning generated by the listener
	agree	smile means consent
	interesting	smile that occurs when the discussion is interesting
	sympathy	seeking empathy
nodding	response	does not have special meaning generated by the listener
	agree	nodding means consent
talk	description	statement when explaining your opinion
	objection	statement when denying the opinion of the speaker
	agree	statement when you agree with the speaker's opinion
	say	giving the right to speak to a third person
other	-	other behavior

4. Result

4.1 Data Collection

16 students from our laboratory cooperated, we made a dataset of eight discussions. These participants have their native language Japanese, 14 males, 2 females, and ages 22 to 24. The total number of hours is about 70 hours in total to make this dataset. Most of the experiment time was spent mainly on annotation.

After each discussion, we conducted a survey in Table 2. A1 and A2, if the participant's opinion is the former, the answer is point 1. On the other hand, if it is the latter, the answer is point 5. Therefore, point 3 means a neutral opinion. B1 to B8, the participants answer each item between 1-5 point, closer 1 means "No" and closer 5 means "Yes".

Table 2

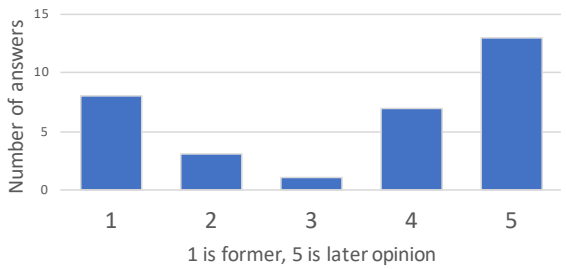
Discussion survey questions

Survey ID	The content of survey
A1	Before the discussion, is your opinion the former or the latter?
A2	After the discussion, is your opinion the former or the latter?
B1	I was satisfied with the discussion.
B2	I could talk my own opinion.
B3	I heard what people with the same opinion say.
B4	I heard what people with the opposite opinion say.
B5	The discussion was enjoyable.
B6	The group often cut off my speech.
B7	I would like to have discussion with this group in future.
B8	I have an attention for the omnidirectional camera.

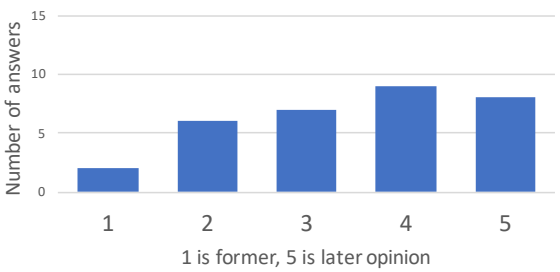
4.2 Discussion of findings

The results of the survey are shown in Figure 5. Comparing (a) and (b) shows that the participant's opinion approaches neutral opinion through the discussion. This is because the survey ID B4: "I heard what people with the opposite opinion say." is 4.125 point on average, so it seems that there was a change in opinion through hearing the opposite opinions.

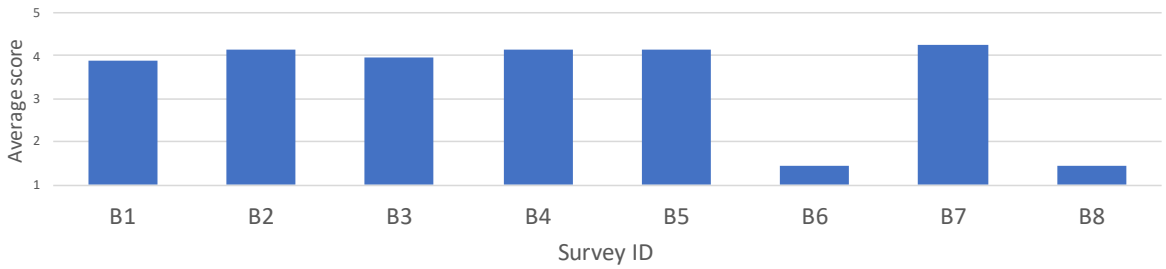
Survey ID B8: "I have an attention for the omnidirectional camera." has a score of 1.43 point on average, most people did not care about the omnidirectional camera. Only one person answered 4 point and we interviewed directly after the end of the experiment. The participant said, "I was concerned about omnidirectional camera recording the video". However, most participants stated, "I was not particularly my mind on the omnidirectional camera". Placing the omnidirectional camera in the center of the table does not have a psychological impact.



(a) A1: Opinion bias before discussion



(b) A1: Opinion bias before discussion



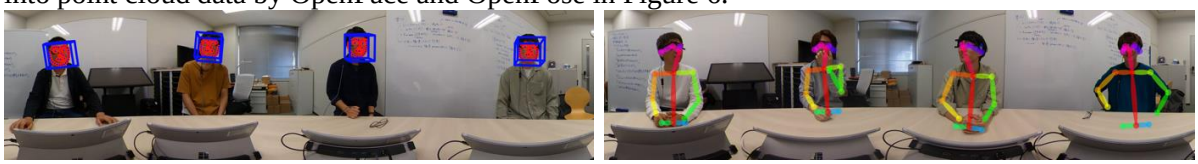
(c) B1-B8: Average scores for each question

Figure 5. Results for the survey listed in Table 2.

5. Conclusions and Future Work

The ability to discuss in a group discussion is expected to become even more important in the future. We proposed a system for making datasets on group discussions in one-stop, and we have made up a dataset of eight discussions. We confirmed that high-quality datasets could be made by using three types of devices and constructing a system which records consistent datasets. Also, the participants do not attract attention to the omnidirectional camera placed at the center, so our proposed method is considered to be a system that is useful to introduce to the various education scenes. We plan to increase the number of experiments and acquire a dataset.

At the current stage, videos recorded by THETA V directly deal with face images of the participants in the discussion, it is difficult to release them as open data from the viewpoint of privacy. Applying OpenFace: it converts face image to point cloud data, and OpenPose: it converts body image to point cloud data, to the THETA V's video, we will release as open data in the future in a state where it is easy to analysis and protecting private information. There are converted images of the discussion into point cloud data by OpenFace and OpenPose in Figure 6.



(a) A1: Opinion bias before discussion (b) A1: Opinion bias before discussion

Figure 6. The point cloud data of face and body.

Acknowledgements

This research is partially supported by JST PRESTO and Innovation Platform for Society 5.0.

References

A. Ohnishi, K. Murao, & T. Terada (2019). A method for structuring meeting logs using wearable sensors. Internet of Things, 5, 140-152. doi:10.1016/j.iot.2019.01.005.

Argyle M (1988). Bodily Communication. *London: Routledge.*

Arpan. C (2017). Active Learning through Discussion. Bridge 1-4.

Ekman. P, & Friesen. W, V (1975). Unmasking the face: a guide to recognising emotions from facial clues. Englewood Cliffs, New Jersey: Prentice-Hall.

Philip, H. Pollock, Kerstin. H., & Bruce. M. Wilson (2011). Comparing the Benefits of Small-Group and Large-Class Discussions. *Journal of Political Science Education*. 48-64. doi:10.1080/15512169.2011.539913

Peter. E, Bull (1987). Posture and Gesture. *Pergamon Press.*

Richard. E, Petty,& John T Cacioppo (1986). The Elaboration Likelihood Model of Persuasion. *Springer Series in Social Psychology. Communication and Persuasion*, 1-24.

T. Hori, S. Araki, T. Yoshioka, M. Fujimoto, S. Watanabe, T. Oba, A. Ogawa, K. Otsuka, D. Mikami, K. Kinoshita, T. Nakatani, A. Nakamura & J. Yamato (2012). Low-latency Real-Time Meeting Recognition and Understanding Using Distant Microphones and Omni-directional Camera. *IEEE Transaction on Audio, Speech, and Language Processing 20(2)*, 499-513. doi:10.1109/TASL.2011.2164527.

T. Kanade, Jeffrey, F. Cohn., & Y. Tian (2000). Comprehensive Database for Facial Expression Analysis. *Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition*.

W, J. Yan., X, Li., S, J Wang., G Zhao, Y, J. Liu., Y, H. Chen., & X Fu(2014). CASME Ⅱ:An Improved Spontaneous Micro-Expression Database and the Baseline Evaluation. *PLoS ONE*. doi:10.1371/journal.pone.0086041