

Analysis and Visualization of Group Discussion Based on Sound Source Angle Obtained Using Kinect

Junichi TAGUCHI^{a*}, Izumi HORIKOSHI^a & Yasuhisa TAMURA^b

^a *Graduate School of Science and Technology,
Sophia University, JAPAN*

^b *Dept. Information and Communication Sciences,
Sophia University, JAPAN*

*j-taguchi-7ew@eagle.sophia.ac.jp

Abstract: This study proposes a method to analyze and visualize the utterance sequence and transition of a face-to-face group discussion. This method uses sound source angles obtained using the Microsoft Kinect audiovisual sensor. An instructor organizes several learner groups and conducts group discussions. When these group discussions are underway concurrently, it is difficult for the instructor to effectively monitor all groups. To solve this problem, we propose a method to analyze and visualize the utterance sequence and transition of each group automatically by using only sound source angle data that is estimated automatically using the four-microphone array in the Kinect sensor. We developed a program to extract sound source angle data from the Kinect sensor, transform it to designated speakers, and visualize the utterance sequence and transition to enable the instructor to recognize the state of a group. Equipment setup and data analysis are simple and easy in this method. We conducted an experiment and the results showed that almost all utterance sequences and transitions could be visualized successfully using the proposed method.

Keywords: Group discussion, Kinect, Learning analytics

1. Introduction

Learner-centered activities like group discussion, debate, and group work are called “active learning,” and they have become increasingly popular in real classrooms (Chickering & Gamson, 1987). Bonwell and Eison (1991) mentioned that “discussion in class is one of the most common strategies promoting active learning” (p.6).

In a group discussion, an instructor organizes students into groups to discuss topics concurrently. In real classes, it is difficult for an instructor to effectively monitor all groups simultaneously. To solve this problem, studies have investigated the use of different devices for monitoring students. Among examples of using audio information, Zancanaro, Lepri, and Pianesi (2006) classified functional roles in a group by using close-talk microphones and a voice activity detector. Tomiyama, Nihei, Nakano, and Takase (2018) segmented group discussions by using a headset microphone and an inertial motion unit attached to the participants.

Many studies have also focused on computer-supported collaborative learning (CSCL). Among examples of visualizing or emphasizing the discussion context, Lobel, Neubauer, and Swedburg (2005) visualized the degree of contribution to a discussion through a diagram by using Pajek, a free software for large network analysis. Hayashi, Kojiri, and Watanabe (2010) detected and emphasized noteworthy utterances in a text-chat discussion by using a discussion support interface.

This study proposes a method to analyze and visualize the utterance sequence and transition of each group automatically using sound source angle data. Equipment setup and data analysis are simple and easy in this method, making it suitable for use in actual classes. This study aims to verify whether the utterance sequence and transition of a group discussion can be analyzed and visualized based on sound source angle data obtained using the Microsoft Kinect audiovisual sensor.

2. Method

2.1 Developed Functions

To achieve our objective, we developed a set of C# and R programs to provide the functions below.

- (1) Extract a sound source angle data sequence using the Kinect sensor (Program-1, C#).
- (2) Transform the sound source angle data sequence to an utterance sequence (Program-2, R).
- (3) Generate an utterance transition diagram based on the utterance sequence (Program-3, R).

The utterance transition diagram enables the instructor to recognize the state of the group. The four-microphone array in the Kinect sensor estimates the sound source angle when it recognizes sound around itself (except for environmental sound). Program-1 was based on a sample one provided by Microsoft; it outputs the sound source angle (from -50° to 50° , shown in Figure 1), time (in millisecond), and reliability (from 0 to 1) every time a sound is detected.

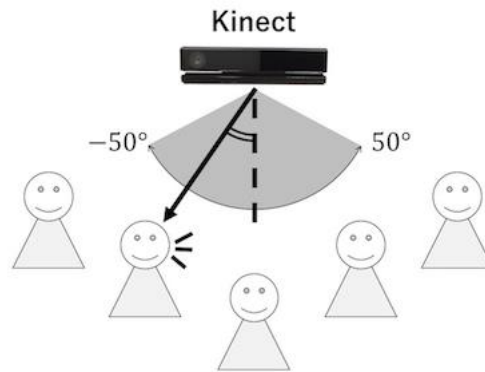


Figure 1. Sound Source Angle

2.2 Experiment

We conducted an experiment to verify our method. The target class was an Information Literacy course at Sophia University. The subjects were five students who agreed to participate in our study. In the target class, each student made a presentation in his or her group. After the presentation, the students had a discussion to improve the presentation. We focused on and analyzed one of the discussions students had in the class.

Five subjects were placed in a fan shape within the sound source angle acquisition range of the Kinect sensor. We also recorded a video of the discussion to verify the accuracy of data acquisition.

3. Result and Discussion

Figure 2 shows the time series data of the sound source angle. The horizontal axis shows the elapsed time since the start of the discussion and the vertical axis, the sound source angle. These data were acquired using Program-1, as mentioned in Section 2.1. We analyzed these data using Program-2. Figure 3 shows the analysis result, namely, the utterance sequence. The horizontal axis shows the elapsed time since the start of the discussion and the vertical axis, the subject ID. The results indicated that the utterance transition and sequence could be visualized using sound source angle data. Note that subject S5 did not speak during the discussion, so there were no corresponding data in this figure.

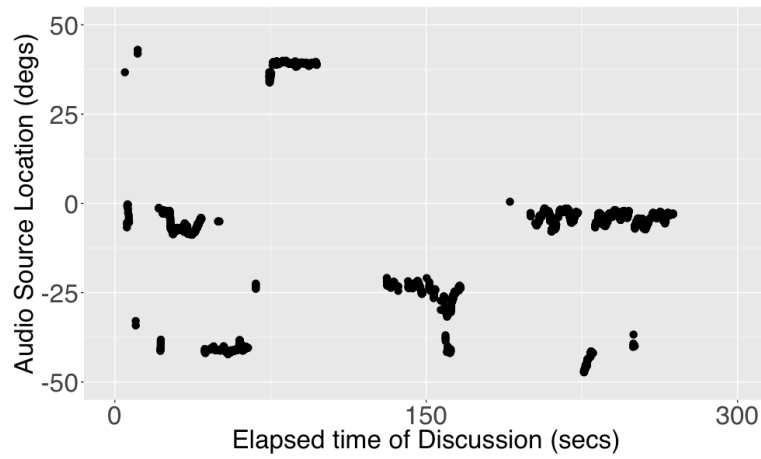


Figure 2. Time series data of sound source angle

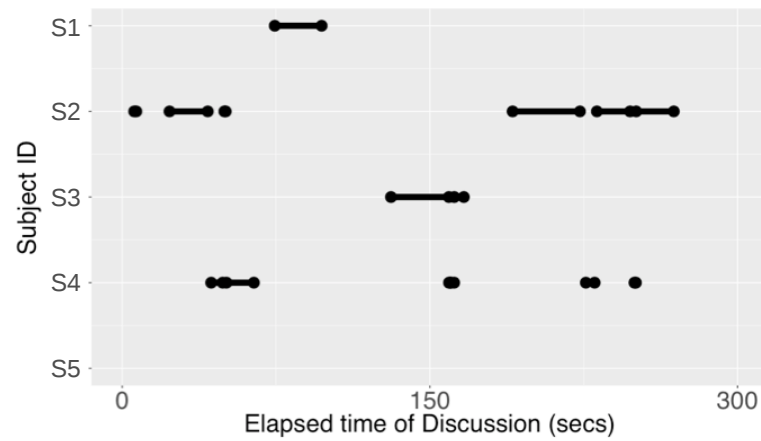


Figure 3. Utterance sequence

Program-3 was used to generate the utterance transition from the utterance sequence, as shown in Figure 4. In this figure, each node indicates a subject. An edge between S3 and S4 means that subject S4 spoke before or after S3. Note that node X indicates a silent state. The size of each node represents the total utterance time of each subject (in the case of X, silent time). In this discussion, subject S4 was the presenter who made a presentation right before the discussion.

This figure shows many edges between S2 and S4, indicating a relatively active discussion between S2 and S4. The edges into and out of node X (silence state) indicate that three subjects spoke after silence (S1, S2, and S3). Among them, S2 spoke two times after silence. Therefore, the proposed system can also detect and identify students who broke a silence.

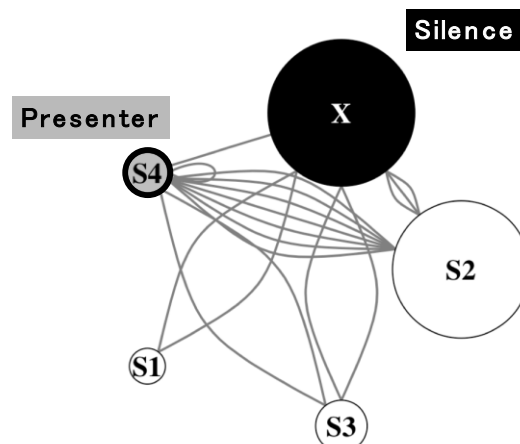


Figure 4. Utterance transition

4. Conclusion and Future Works

This study aimed to verify whether the utterance sequence and transition of a group discussion could be analyzed and visualized based on sound source angle data obtained using the Kinect sensor. The results indicated that the utterance sequence (Figure 3) and utterance transition (Figure 4) could be visualized successfully by using sound source angle analysis.

In this study, we analyzed only one discussion in one group of the whole class. Therefore, in the future, we will analyze and compare results with other class discussions, collect data for the whole class, and improve the accuracy of correcting data. We also noted the problem that low-volume voice utterances could not be detected because the Kinect sensor regarded them as environmental sound. To solve this problem, we will try to find appropriate alignments of the subject and the Kinect sensor.

References

- Bonwell, C. C., & Eison, J. A. (1991). Active Learning: Creating Excitement in the Classroom. *ASHE-ERIC Higher Education Reports*. ERIC Clearinghouse on Higher Education, The George Washington University.
- Chickering, A. W., & Gamson, Z. F. (1987). Seven Principles for Good Practice in Undergraduate Education. *AAHE Bulletin*, 3, 7.
- Hayashi, Y., Kojiri, T., & Watanabe, T. (2010). A Method for Detecting Focusing Utterance through Discussion in Collaborative Learning. *Proc. of 18th International Conference on Computers in Education* (pp. 186-190).
- Lobel, M., Neubauer, M., & Swedburg, R. (2005). Comparing How Students Collaborate to Learn About the Self and Relationships in a Real-Time Non-Turn-Taking Online and Turn-Taking Face-to-Face Environment. *Journal of Computer-Mediated Communication*, 10(4), JCMC10416.
- Tomiyama, K., Nihei, F., Nakano, Y. I., & Takase, Y. (2018). Identifying Discourse Boundaries in Group Discussions using a Multimodal Embedding Space, *Joint Proc. ACM IUI 2018 Workshops*.
- Zancanaro, M., Lepri, B., & Pianesi, F. (2006). Automatic Detection of Group Functional Roles in Face to Face Interactions. *Proc. 8th International Conference on Multimodal Interfaces* (pp. 28-34). ACM.