Iyer, S. et al. (Eds.) (2022). Proceedings of the 30th International Conference on Computers in Education.
Asia-Pacific Society for Computers in Education

Relationship Analysis between Listener Face Direction and Utterance in Group Discussion

Nori MORISHIMA[a*], Izumi HORIKOSHI[b] & Yasuhisa TAMURA[c]
[a]*Graduate School of Science and Technology, Sophia University, Japan*
[b]*Academic Center for Computing and Media Studies, Kyoto University, Japan*
[c]*Faculty of Science and Technology, Sophia University, Japan*
*n-morishima-6r8@eagle.sophia.ac.jp

Abstract: The authors analyzed the relationship between face direction and utterance during group discussion. For groups of six students, face direction data were collected using video and OpenFace, and utterance data were collected using a microphone array and the Hylable system. The results of the analysis show that the average number of persons who turned their faces differed significantly when the speaker did or did not speak.

Keywords: Group discussion, Utterance, Face direction, Microphone array, OpenFace

1. Introduction

Active learning has gained research interest in recent years (Hartikainen, Rintala, Pylväs, & Nokelainen, 2019). One method of active learning is group discussion which involves frequent conversations. In general conversation, we communicate by combining verbal information, such as spoken language with nonverbal information, typically gestures and gaze (Jones & LeBaron, 2002). For example, in Hybrid meetings, it is difficult for remote participants to be involved in discussions (Yankelovich, Simpson, Kaplan, & Provino, 2007), which may be due to insufficient communication of verbal and nonverbal information. Research on conversation often focuses on verbal information. However, some research also focus on nonverbal information. This area is referred to as multimodal discourse analysis (MDA) (O'Halloran, 2011).

As an example of multimodal information, face direction, such as head pose, is important nonverbal information that indicates the target of speech and the alternation of roles in conversation (Murphy-Chutorian & Trivedi, 2008). As concrete research, Ito et al. (2022) focuses on nonverbal information in group discussions in which neural network models that estimate the persuasiveness of participants using head pose data were constructed. However, Ito et al. (2022) did not analyze the influence of nonverbal information on utterance during group discussion.

Therefore, in the present study, we focus on face direction and amount of utterance in a group discussion and hypothesized that a relationship exists between whether listeners turn their faces toward the speaker and amount of utterance. We used software to detect face direction and a microphone array to detect utterance to verify this hypothesis. We expect that the present study will provide insight for building a better learning environment on not only face-to-face group discussion, but also group discussion in HyFlex classes where communication using nonverbal information is difficult.

2. Method

2.1 Data Collection

We used two types of data: the amount of utterance for each person and face direction during the group discussion. In order to obtain the amount of utterance for each person, the Hylable microphone array was placed in the center of the group. This voice data was converted into amount of utterance. As for face direction, one iPad mini was placed in front of each subject to record the face of each subject. Face

directions were computed using OpenFace (Baltrušaitis, Zadeh, Lim, & Morency, 2018) on this video. In order to reduce the risk of COVID-19 exposure, the subjects wore masks with transparent mouthparts.

2.2 Experiment and Analysis

The subjects were 13 Japanese university students who consented to this research. The experiment was conducted on December 2 and 9, 2021. Three sets of 10-minute group discussions were held in order to obtain the data described in Section 2.1. For the experiment, the subjects were reorganized to make three groups with six participants in each (some students participated repeatedly). Before obtaining the data, we held a calibration for face direction. The subjects were instructed to direct their faces to specific angles, and we used the obtained baseline angle for the calibration. This calibration provides accuracy of face direction data.

In order to determine who made an utterance at a particular time, we defined the subject with the maximum amount of utterances in each time frame (5 s) as the speaker, and the other subjects as non-speakers.

3. Results and Discussion

3.1 Utterance Timing and Facing Timing

We visualized the timing when one subject made an utterance (utterance timing) and the timing when the other subjects turned their faces toward the subject (facing timing) (Figure 1). The blue dots indicate the utterance timing, and the red dots indicate the facing timing. Figure 1 shows the results for subject S011 in Group 1. Both A and B in Figure 1 are part of the timing in which Subject S011 spoke. Subject S011 spoke briefly in A and spoke longer in B. The number of subjects who turned their faces to Subject S011 was low in A, while the number of subjects who turned their faces to Subject S011 was high in B. Regarding this result, we assumed that the speaker made longer utterances because the face of the listener turned toward the speaker, and the speaker felt as if he/she was being asked for an opinion.

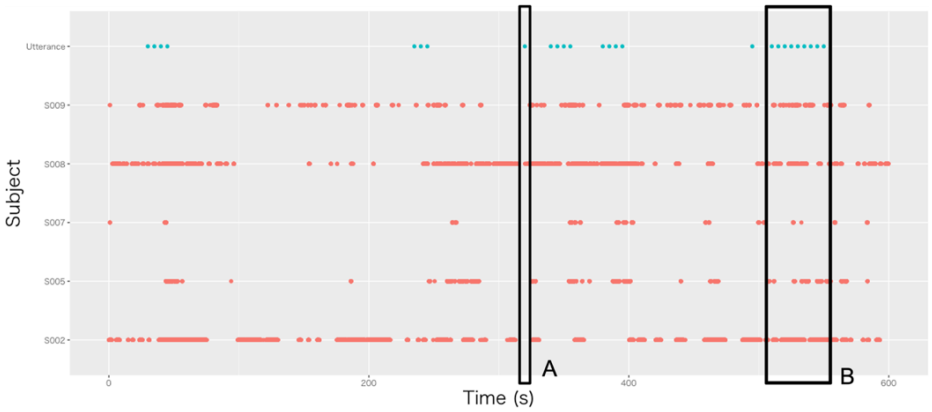

Figure 1. Utterance timing of speaker and facing timing of listener

3.2 Number of People Who Faced the Speaker and Whether the Speaker Made Utterances

Figure 2 shows the number of people who faced the speaker in two conditions: when the speaker was speaking (speaking timing) and when the speaker was not speaking (non-speaking timing). Orange indicates speaking timing and green indicates non-speaking timing. From Figure 2, we can observe the difference in the number of people who faced the speaker between speaking timing and non-speaking timing for all subjects except Subject S005 in Group 1. We interpreted this as it being easier for the speaker to speak when many listeners turned their faces in his/her direction.

Figure 2. Number of people who faced the speaker and whether the speaker made utterances

Next, we conducted an analysis of variance for the boxplot in Figure 2. The results show that, for all groups, the average number of people who faced the speaker was significantly different between speaking timing and non-speaking timing ($p < 0.001$).

Accordingly, since the average number of people who faced the speaker differs significantly between the two conditions, the possibility of a relationship between the number of people who faced the speaker, and the speaking timing/non-speaking timing is clear. Since the two types of data used in the present study were obtained at the same time, it is impossible to clarify the cause and the results from among face direction data and utterance data. However, we believe that, since the faces of the listeners turned toward the speaker, the speaker felt that being expected to speak and the willingness to speak increased. Thus, the number of people who faced the speaker was high during speaking.

4. Conclusion and Future Research

In the present study, we investigated the relationship between whether listeners turn their faces toward the speaker and amount of utterance in group discussions. As a result, having the face of the listener turned toward the speaker increases utterance by the speaker. In addition, a relationship may exist between the number of people who face the speaker and the amount of utterances. A future prospect is to research group discussions in HyFlex classes.

References

Baltrušaitis, T., Zadeh, A., Lim, Y. C., & Morency, L. P. (2018). OpenFace 2.0: Facial behavior analysis toolkit. In *2018 13th IEEE Int'l Conf. Automatic Face & Gesture Recognition* (*FG 2018*) (pp. 59-66). IEEE.

Hartikainen, S., Rintala, H., Pylväs, L., & Nokelainen, P. (2019). The concept of active learning and the measurement of learning outcomes: A review of research in engineering higher education. *Education Sciences*, *9*(4), 276.

Ito, A., Nakano, Y. I., Nihei, F., Sakato, T., Ishii, R., Fukayama, A., & Nakamura, T. (2022). Predicting persuasiveness of participants in multiparty conversations. In *IUI '22: 27th Int'l Conf. Intelligent User Interfaces* (pp. 85-88).

Jones, S. E., & LeBaron, C. D. (2002). Research on the relationship between verbal and nonverbal communication: Emerging integrations. *Journal of Communication*, *52*(3), 499-521.

Murphy-Chutorian, E., & Trivedi, M. M. (2008). Head pose estimation in computer vision: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *31*(4), 607-626.

O'Halloran, K. L. (2011). Multimodal discourse analysis. In K. Hyland (Ed.), *Continuum companion to discourse analysis* (pp. 120-137). Bloomsbury Publishing.

Yankelovich, N., Simpson, N., Kaplan, J., & Provino, J. (2007). Porta-person: Telepresence for the connected conference room. In *CHI '07 Extended Abstracts on Human Factors in Computing Systems* (pp. 2789-2794).