7ZB-02

Detecting and Analyzing Speaking Intention in Leader-led VR Group Discussions

Chenghao Gu[†], Jiadong Chen[†], Jiayi Zhang[†], Tianyuan Yang[†], Zhankun Liu[†], Shin'ichi Konomi[‡]

[†] *Graduate School and Faculty of Information Science and Electrical Engineering, Kyushu University*
[‡] *Faculty of Arts and Science, Kyushu University*

I. Introduction

Identifying speaking intentions of group members can assist the leader in smoothly guiding efficient discussions and increasing the participation of group members in virtual reality environments. In this paper, our main objective is to detect group members' speaking intentions within VR-based group discussions and investigate the relationship between leadership and the speaking intentions of group members. We utilize group members' sensor features available on off-the-shelf VR devices to construct personalized and general deep learning models for detection. The results show personalized models outperformed general models. We also found that for those low-engagement group members, the leader's sensor features improved the model's ability to detect speaking intentions. Based on the results we attained, we observed the behaviors of low-engagement group members when they held speaking intentions through video analysis and attempted to explore the reasons why leader sensor features take the impact on the detection results of low-engagement members' speaking intentions.

II. Related work

Existing research in supporting group discussion has revealed some issues regarding the learning process, group formation, and group dynamics. These issues can be traced back to impeded social interaction between group members. Therefore, many learning support tools have been developed to facilitate social interaction among group members, e.g., group awareness tools, robot-based group facilitators, virtual agents, and so forth. Yet, effective methods to facilitate social interaction in VR-based group discussions remain underexplored. Moreover, non-verbal features such as head movement, gaze, and hand gestures, as well as interaction features, have been commonly employed to analyze group dynamics in discussions. For instance, Chen et al. introduced a Kinect sensor-based approach to assess group discussion behavior, using algorithms to detect each group member's facial direction to infer their roles in the discussion [1]. Additionally, a novel model was proposed for recognizing Interaction Processing Analysis (IPA) categories, labeling 12 interaction categories based on multi-modal data [2]. In recent years, Chen et al. have investigated the utilization of sensor data from off-the-shelf VR devices and additional contextual information, such as user activity and engagement, for predicting opportune moments to send notifications using deep learning models [3]. To the best of our knowledge, there is limited prior research that aims to predict speaking intentions based on sensor data from off-the-shelf VR devices. Based on previous relevant studies [4], our emphasis is on investigating the impact of leadership on members' speech intentions.

III. Experiment Design

We conducted an experiment to understand speaking intentions in VR-based group discussions involving a leader and to explore deep learning models to detect speaking intentions. We recruited 24 participants (12 females, 12 males), who are university graduate students, aged 22-30 (M = 25.5, SD = 2.27). All participants were asked about their VR experiences before the experiment. Thirteen participants (54%) have prior VR experiences, and 11 participants (46%) have no prior VR experiences. We developed an experimental virtual environment (VE) by using the Unity engine's XR interaction tool to collect participants' sensor data. The VE resembles a meeting room with a large table and two whiteboards (see Fig. 1). Participants were divided into groups of 4 (2 males, 2 females) and each group entered the virtual environment for a survival game scenario involving a leader. During this scenario, a group of four participants is engaged in a survival game known as "Lost at Sea [5]". Besides, the group's task is to collectively rank 15 items based on their perceived importance for survival. Following a 20-minute discussion, the group is required to provide a final ranked list of the 15 items on a virtual environment's whiteboard. This scenario is characterized by a more problem-based communication, involving a clear task and objective.

Upon completion of the survival game, participants were instructed to annotate the time intervals during which they had the intention to speak. As for sensor data, we utilized VR devices to collect participants' positions, movements, and other physical actions during the discussions. The VR device employed in our experiment, Oculus Quest 2, is capable of providing sensor data from three components: the head-mounted display, the left-hand controller, and the right-hand controller. Each component enables the collection of data on position, rotation, velocity, acceleration, angular velocity, and angular acceleration.

	Personalized Model					General Model				
Sensors features	Acc	F1.	Prec.	Recall	AUC	Acc	F1.	Prec.	Recall	AUC
Baseline	0.5098	0.4889	0.4729	0.5060	0.5071	0.5002	0.4702	0.4522	0.4897	0.4950
EEGNet	**0.6813**	**0.7132**	**0.6201**	0.8393	**0.6058**	**0.5823**	0.6057	**0.5041**	0.7585	**0.5633**
InceptionTime	0.6785	0.7069	0.5908	**0.8798**	0.6053	0.5338	**0.6293**	0.4803	**0.9122**	0.5312
MLSTM-FCN	0.6789	0.6968	0.6084	0.8153	0.5925	0.5383	0.5590	0.4840	0.6616	0.5390

TABLE I: Performance comparison of Personalized model and general model using sensor features(HMD, controllers, position and rotation in VE). Classification threshold is 0.5. Legend: “Acc”: Accuracy, “Prec.”: Precision., “F1.”: F1-score, “AUC”: AUROC

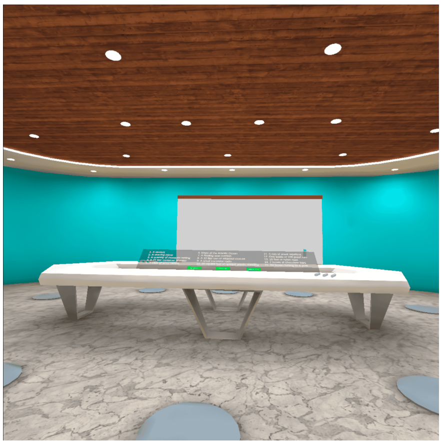

Fig. 1: The virtual environment of discussion

IV. Results

Following our main objective, we attempt to detect group members’ suppressed intentions to speak during VR-based group discussions using time-series sensor data. We explored three commonly used time-series classification models (EEGNet, InceptionTime and MLSTM-FCN). For the validation methods, we both trained personalized models for each group member by employing 10-fold cross-validation and employed a leave-one member-out cross-validation [3] to build a general model for all group members. Our baseline model is a random classifier, which assigns data as positive or negative samples with an equal 50% probability. Our results indicate that the personalized models outperform the general models across various evaluation metrics. Table I shows the performances of personalized and general models.

Furthermore, we concatenated the leader’s sensor features as contextual information with the member’s sensor features along the time axis. We found that for those members who spoke infrequently and had relatively low engagement during the discussion, the leader’s sensor features improved their personalized models’ ability to detect speaking intentions. To further support our previous findings and find some clues about why leader sensor features enhance the prediction of speaking intentions for those low-engagement group members, we attempted to conduct a preliminary analysis of the behaviors of low-engagement group members when they had speaking intentions, focusing on three aspects: hand gestures, head orientation, and body movement. We observed that when low-engagement group members held the intention to speak, leaders were almost looking around the survival item panel and whiteboard, engaged in discussions with other high-engagement group members, or organizing discussion topics. It seems to be difficult for leaders or other members to recognize being looked at by low-engagement members. Therefore, we speculate that these low-engagement group members, when holding speaking intentions, exhibit the behavior of looking toward the leader or other high-engagement group members, possibly seeking an opportunity to join others’ discussion or expecting assistance from the leader. In such cases, combining leader’s sensor features may assist the model in better learning the motion patterns of speaking intentions for these low-engagement members.

V. Conclusion and Future Work

In this paper, we examined the feasibility of detecting group members’ speaking intentions by using sensor data from commodity VR devices. We also found that leader’s sensor features improved the model’s ability to detect speaking intentions for low-engagement group members. After combining leader’s sensor features, the model’s F1-score, precision, and AUROC could be increased. To further support our results, We also analyzed first-perspective videos to observe the behaviors of low-engagement group members and tried to understand why leader sensor features affect the detection results of low-engagement members. We found that when these members held speaking intentions, they tended to gaze at the leader and other actively participating group members. Future research could explore the integration of biometric features or other relevant features to improve detection performances.

Acknowledgement

This work was supported by JSPS KAKENHI Grant Numbers JP20H00622, JP23H03507, JP23K02469.

References

[1] Z. Chen, W. Liu, W. Li, J. Xu, and W. Cheng, “Kinect-based behavior measurement in group discussion,” in *Proceedings of the 2019 The World Symposium on Software Engineering*, 2019, pp. 125–129.

[2] S. Li, S. Okada, and J. Dang, “Interaction process label recognition in group discussion,” in *2019 International Conference on Multimodal Interaction*, 2019, pp. 426–434.

[3] K.-W. Chen, Y.-J. Chang, and L. Chan, “Predicting opportune moments to deliver notifications in virtual reality,” in *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, 2022, pp. 1–18.

[4] J. Chen, C. Gu, J. Zhang, Z. Liu, and S. Konomi, “Sensing the intentions to speak in vr group discussions,” *Sensors*, vol. 24, no. 2, 2024. [Online]. Available: https://www.mdpi.com/1424-8220/24/2/362

[5] “lost at sea,” 2023, accessed 2023-06-09. [Online]. Available: https://insight.typepad.co.uk/lost_at_sea.pdf