Multimodal collaborative workgroup dataset and challenges

Vanessa Echeverria[1], Gabriel Falcones[2], Jaime Castells[2], Roger Granda[2], and Katherine Chiluiza[3]

[1] Connected Intelligence Centre, University of Technology Sydney, Broadway, Ultimo, NSW 2007, AUS
`vanessa.i.echeverriabarzola@student.uts.edu.au,`
[2] ESPOL Polytechnic University, Escuela Superior Politécnica del Litoral, ESPOL, Centro de Tecnologías de Información, Campus Gustavo Galindo Km 30.5 Vía Perimetral, P.O Box 09-015863, Guayaquil, Ecuador
`{gabriel.falcones,jaime.castells,roger.granda}@cti.espol.edu.ec`
[3] ESPOL Polytechnic University, Escuela Superior Politécnica del Litoral, ESPOL, Campus Gustavo Galindo Km 30.5 Vía Perimetral, P.O Box 09-015863, Guayaquil, Ecuador
`kchilui@espol.edu.ec`

Abstract. This work presents a multimodal dataset of 17 workgroup sessions in a collaborative learning activity. Workgroups were conformed of two or three students using a tabletop application in a co-located space. The dataset includes time-synchronized audio, video and tabletop system's logs. Some challenges were identified during the collection of the data, such as audio participation identification, and user traces identification. Future work should explore how to overcome the aforementioned difficulties.

Keywords: collaboration; group work; collocated spaces; multimodal learning analytics

1 Introduction

Computer Science (CS) students are required to develop teamwork abilities to be successful in their professional life. Nevertheless, the state of the art presented in [9] pointed out that CS students are lacking skills in many different areas, including technical and interpersonal skills (communication, teamwork, critical thinking). Therefore, educators face a big challenge in supporting their students on the development of these skills. One way to improve these skill is through the implementation of Collaborative learning activities in co-located spaces [5].

Multimodal Learning Analytics (MMLA), which has been explored in recent years [7][12] , allows to extract and analyze useful information from different sources collected during learning processes, to understand how students learn [8]. In this context, capturing most of the data while performing collaborative learning activities is one of the big challenges faced in MMLA applications, due to the complex task and the availability of technology and sensors.

One particular approach that has been used to capture traces of participants in co-located workgroup activities is the use of multitouch tabletops through tabletop systems (TS). Several studies have included these artifacts. Nonetheless, most studies only involved the use of one modality (user's touch) for analysis [6] [4] [3] [10] [1], lacking of a deeper granularity in the analysis of the collaborative task.

The main contribution of this paper is to provide a multimodal dataset of collaborative learning activities in co-located spaces using a TS proposed by [11] with the goal of augmenting collaboration and discussion between peers. Information of the collaborative activity is obtained from three sources: tabletop system's log, recorded audio, and recorded video.

This paper is structured as follows: section 2 shows how the collaborative sessions were organized. Data collection setting is described in section 3. Next, our data set is explained in section 4. In section 5, challenges and future of the data collection process are presented.

2 Participants and Task

Participants were 47 undergraduate students from CS, they were enrolled in an introductory Database Systems course. Students conformed groups of two or three members, based on their own affinity. During each session, groups were asked to collaboratively solve a database design problem and generate an entity-relationship (ER) model. They used the tabletop system to draw the design solution. Each student could interact with the system using their hands and tablets. The tablets were used to read the task description and create objects (entities and attributes). Students used their hands to move or delete objects over the tabletop and create relations between them. Each collaborative session had a duration of 20-30 minutes approximately.

3 Data collection setting

To capture student's collaboration process, several input devices recorded student activities on the tabletop. Those devices includes: Video-camera and a Kinect v2. Figure 1 depicts the setting of the system. Our dataset includes: Audio, Video and tabletop system's logs. In the following paragraphs it is explained how the capture of each information source was carried out.

3.1 Audio

To record the audio, the microphone array of a Kinect v2 was located in front of the students. An application was developed using the Kinect SDK in order to estimate which student is talking at any time, based on the angle of the

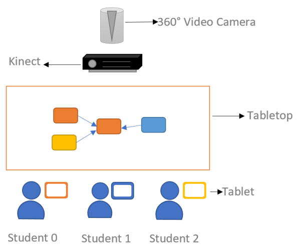

Fig. 1. System configuration

audio source, assuming that students never change their position around the tabletop. It is worth to say that this application does not recognize multiple participants talking at the same time. As a result, a CSV file with student's speaking intervention information was generated, along with a 4-channel wav audio file.

3.2 Video

Each group session was recorded with a Lucy 360 camera, and later processed and transformed into a wide screen video. The camera was situated in front of the students.

3.3 Application's Log file

The tabletop application recorded every action the students made in a log file, which was a plain text file. Every action was saved on a line of text, which indicated the student and the name of the action performed, along with the date and time of when the action was performed.

3.4 Capture and synchronization

Every different source of capture was manually initiated by a member of the research team. However, since there was a difference in seconds between every input device, all of the files had to be manually synchronized later. The CSV audio files contain a time-stamp in case it is needed for synchronization in the future.

An additional Excel file was created to join the files from the audio capture and the log file. A new column was added to display the relative time, which started on 0 when the audio recording began.

4 Dataset Records

All the collected data is public available from `http://www.cti.espol.edu.ec/tabletopDataset.html` prior to sign a collaborator agreement. The dataset consists of 17 collaborative sessions. For each group the following files were saved:

- A 360° video file in mp4 format, recorded at 15 FPS, with a resolution of 1152x320.
- A 4-channel wav format audio file.
- An audio log file in a CSV format with the following columns: *angle; confidence; timestamp; student id*, at a rate of 16FPS (average). The *angle* value is within the range of -50° and 50°. The *confidence* value is a number between 0 (lower confidence) and 1 (higher confidence). The *timestamp* is recorded for further synchronization if needed. The *student id* is assigned as 0, 1 or 2 (see Fig.1).
- A tabletop system's log file in a CSV format. Each line contains an action performed by a student. The actions recorded can be one of the following: create, delete or move entity; create, delete or move relation; join notes (to create entities); and split entity into notes. The timestamp of the action was also saved, as well as the student who performed the action. The identification of which student created an entity comes from the tablet. The identity of actions performed over tabletop objects were estimated using an implementation of the approach proposed by [2].

5 Challenges and future work

Collecting and analyzing data from multiple sources comes along with some challenges described below:

- The identification of student's speaking participation. Since we are using a single audio stream to identify student's oral interventions it's difficult to determine when two or more students speak at the same time.
- Identification of actions' ownership on the tabletop. The application that was used for user's identification has an accuracy of about 90%.
- Synchronization of the multiple inputs. Given that the audio and video recording started at different time, they had to be manually synchronized after the recording session.

Further research implies more reliable methods for tracking both the student's speaking participation and their interactions on the tabletop. Also, a centralized sever could be the solution for automatically synchronize all captured data. Learning analytics applications on this dataset may explore awareness and reflection of collaborative learning process through on time feedback; modeling groupwork behavior for predicting failure or success of collaborative tasks; automatic evaluation of group performance and monitoring student's collaborative learning skills over time.

References

1. A. Clayphan, A. Collins, C. Ackad, B. Kummerfeld, and J. Kay. Firestorm: A brainstorming application for collaborative group work at tabletops. In *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces*, ITS '11, pages 162–171, New York, NY, USA, 2011. ACM.
2. A. Clayphan, R. Martinez-Maldonado, C. Ackad, and J. Kay. An approach for designing and evaluating a plug-in vision-based tabletop touch identification system. In *Proceedings of the 25th Australian Computer-Human Interaction Conference: Augmentation, Application, Innovation, Collaboration*, OzCHI '13, pages 373–382, New York, NY, USA, 2013. ACM.
3. G. Falcones, M. Wong-Villacrés, V. E. Barzola, and K. C. Garcia. Enhancing quality of argumentation in a co-located collaborative environment through a tabletop system. In *2016 IEEE Ecuador Technical Chapters Meeting (ETCM)*, volume 01, pages 1–6, Oct 2016.
4. R. X. Granda, V. Echeverría, K. Chiluiza, and M. Wong-Villacrés. Supporting the assessment of collaborative design activities in multi-tabletop classrooms. In *Computer Aided System Engineering (APCASE), 2015 Asia-Pacific Conference on*, pages 270–275. IEEE, 2015.
5. D. W. Johnson and R. T. Johnson. Social skills for successful group work. *Educational leadership*, 47(4):29–33, 1990.
6. R. Martínez, A. Collins, J. Kay, and K. Yacef. Who did what? who said that?: Collaid: An environment for capturing traces of collaborative learning at the tabletop. In *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces*, ITS '11, pages 172–181, New York, NY, USA, 2011. ACM.
7. X. Ochoa, K. Chiluiza, G. Méndez, G. Luzardo, B. Guamán, and J. Castells. Expertise estimation based on simple multimodal features. In *Proceedings of the 15th ACM on International conference on multimodal interaction*, pages 583–590. ACM, 2013.
8. S. Oviatt, A. Cohen, and N. Weibel. Multimodal learning analytics: description of math data corpus for icmi grand challenge workshop. In *Proceedings of the 15th ACM on International conference on multimodal interaction*, pages 563–568. ACM, 2013.
9. A. Radermacher and G. Walia. Gaps between industry expectations and the abilities of graduates. In *Proceeding of the 44th ACM technical symposium on Computer science education*, pages 525–530. ACM, 2013.
10. M. Wong-Villacrés, R. Granda, M. Ortiz, and K. Chiluiza. Exploring the impact of a tabletop-generated group work feedback on students' collaborative skills. volume 1601, pages 58–64, 2016.
11. M. Wong-Villacrés, M. Ortiz, V. Echeverría, and K. Chiluiza. A tabletop system to promote argumentation in computer science students. In *Proceedings of the 2015 International Conference on Interactive Tabletops & Surfaces*, ITS '15, pages 325–330, New York, NY, USA, 2015. ACM.
12. M. Worsley. Multimodal learning analytics: Enabling the future of learning through multimodal data analysis and interfaces. In *Proceedings of the 14th ACM International Conference on Multimodal Interaction*, ICMI '12, pages 353–356, New York, NY, USA, 2012. ACM.