Jiang, B. et al. (Eds.) (2025). Proceedings of the 33rd International Conference on Computers in Education. Asia-Pacific Society for Computers in Education

The Use of Sensor Data for the Multimodal Analysis of Mathematical Proof Activity in a CSCL Environment

Masataka KANEKO[a*], Hironori EGI[b], Takuya KITAMOTO[c] & Takeo NODA[d]
[a]*Center for Higher Education Development, Toho University, Japan*
[b]*Graduate School of Informatics and Engineering,*
The University of Electro-Communications, Japan
[c]*Faculty of Education, Yamaguchi University, Japan*
[d]*Faculty of Science, Toho University, Japan*
*masataka.kaneko@phar.toho-u.ac.jp

Abstract: Recent technological developments have greatly enhanced the power of multimodal learning analytics to help researchers better understand how students learn in a CSCL environment in a scalable manner. In this study, we try to use depth sensor data to identify the temporal transition of learners' behaviors and investigate the relationship between it and the progress of their mathematical thinking. Human-coded behavioral data are compared with the log data of learner movements derived from a depth sensor to evaluate the extent to which the sensor data can be used to infer the progress of learners' thinking during mathematical proof activity in a CSCL environment.

Keywords: Depth sensor data, mathematical proof activity, dynamic geometry, multimodal learning analytics

1. Introduction

Many previous studies in cognitive neuroscience have illustrated that mathematical thinking emerges from the interplay of the cerebral networks for language processing and those for visuospatial processing (Dehaene, et al., 1999). Moreover, mathematical cognition is grounded in the mental simulation of dynamic imagery supported by the interlocking network of the visual system and the motor system (Lakoff and Nunez, 2001). Regarding these characteristics of mathematical cognition, it is natural that many CSCL studies in mathematics have been conducted from a Multimodal Learning Analytics (MMLA) perspective. For instance, one of the pioneering studies used a process data stream derived from high school students' collaborative work for solving algebra and geometry problems and extracted features of their behavioral patterns to predict their expertise with automatic algorithms (Ochoa, et al., 2013). The data used in the study was the Math Data Corpus (Oviatt, et al., 2013) which is time-synchronized multimodal data composed of video recordings of learners' physical movements including their calculator use, audio data from their conversations, and digital pen stroke data derived from their writing. It also includes human-coded data on each group's problem-solving correctness and the representational content of each student's writing. While automatic algorithms can give good predictors in these cases when students are engaged in well-structured tasks, extra elaboration for combining hand-annotated data and system-collected data is needed to track and interpret learners' thinking during project-based tasks such as programming, electronics, and mathematical proof (Spikol, et al., 2016, Blikstein, & Worsley, 2016). One previous study (Spikol, et al., 2017) illustrated how sensor-based learning analytics systems can help us identify the MMLA features relevant for collaborative problem solving even in the case of complex engineering tasks, by the regression analysis using human-evaluated scores of learners' collaborative problem solving as output and data from learners' hand movements and face directions collected through low-cost sensing technologies as input, together with the machine-learning based estimation of those regression methods.

In this study, we investigate the verbal and nonverbal behaviors of several pairs of university students as they were engaged in a mathematical proof activity. Using HTML-based dynamic content collaboratively, they were asked to prove the addition theorem for trigonometric functions when the related angles were not necessarily acute. The analysis was conducted by aligning the behavioral data derived from video recordings, the sound data of their conversations derived from audio recordings, and the log data of body movements derived from a depth sensor camera on a single timeline. Based on the theory of mathematical cognition, it is assumed that the progress of learners' mathematical thinking might be linked to the appropriate pattern of temporal transition involving their linguistic, visual, and motor functions. Moreover, a recent CSCL study investigating the quality of collaborative group engagement in designing a floor plan of an office while using a 3D modeling tool has demonstrated that the quality rated by using learners' verbal communication data is highly correlated to their nonverbal behaviors (Paneth, et al., 2023). This knowledge gave the authors the idea that the qualitative analysis of verbal communication data can be complemented by an analysis of the change in behavioral patterns observed from the depth sensor data. As a preliminary step to examine this idea, we pose the following two RESEARCH QUESTIONS in this study: (1) Is the progress of learners' thinking identified in their verbal communication data interrelated to any change in their behavioral pattern observed in videotaped images? (2) Can the change in learners' behavioral patterns be portrayed on the log data derived from the depth sensor?

2. Methods

2.1 The Content Used and Task Design

The learning task is to prove the addition theorem for the trigonometric function

$$\sin\,(\alpha + \beta) = \sin\,\alpha\,\cos\,\beta + \cos\,\alpha\,\sin\,\beta$$

when the relevant angles α, β, and $\alpha + \beta$ are not necessarily acute. In these general cases, the positional relationship between geometric elements becomes complex compared to the rudimentary case where the relevant angles are acute. Figure 1 shows the HTML-based dynamic geometry content used in this study. The left and right sides display the geometric figure and the mathematical expression involved in the proof for the rudimentary case respectively. The central part displays the dynamic geometry screen on which learners can move the angles α and β by manipulating points A and B.

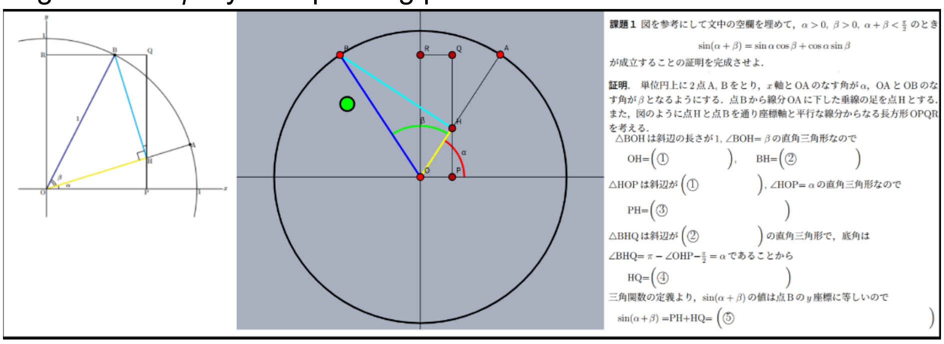

Figure 1. The HTML-based content used in this study

Before the collaborative learning session, participants were asked to review the proof in the rudimentary case individually using printed material including the left and right parts in Figure 1. After the instructor confirmed the completion of this preliminary step, they were asked to work on the proof activity in pairs using an iPad onto which HTML content had been implemented. While manipulating the points A and B, learners can discuss why the same formula holds although the positional relationship between geometric elements changes accordingly. The focal point of discussion is the signs of the trigonometric ratios for supplementary and complementary angles. Since identifying those signs in specific situations was assumed to burden heavy cognitive load on the participants (Sweller, et al., 2011), they were allowed to use a worksheet to write down their ideas and discuss them with their partners.

2.2 The Flow of Experiment and Data Analysis

Participants were first-year students in a Japanese university majoring in engineering or information technology. In high school, they learned the proof for the theorem in the rudimentary case. The members of each dyad had been friends since before the experiment. For each group, the participant on the left side was named A and the one the right was named B. While an iPad was delivered to each group, a worksheet was prepared for each participant to record their ideas during the collaborative learning session. The position of these items was fixed as shown in Figure 2 (left).

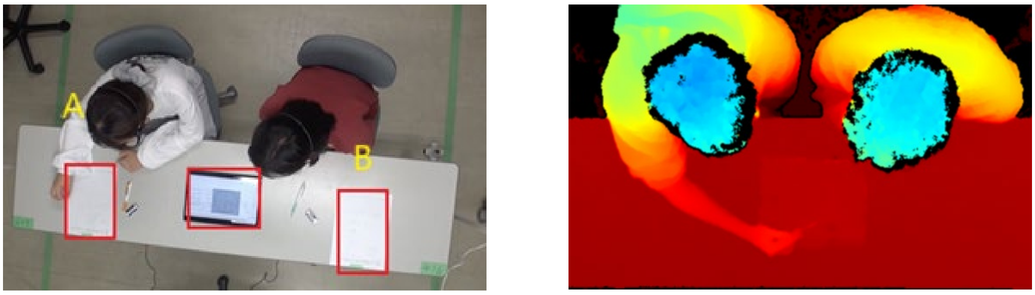

Figure 2. The setting of iPad and worksheets on a table (left) and the sample of the colored depth image derived from Intel RealSense depth camera (right)

Each participant wore a microphone through which their utterances were recorded. The recorded discourse was subsequently transcribed with time stamps. The transcribed text data was analyzed to identify when and how their thinking progressed. Also, the participants' behaviors were videotaped and the recorded image was imported into the behavioral analysis system Sportscode (https://www.hudl.com/products/sportscode). Their behaviors were classified into several categories including "manipulating dynamic content", "making gestures", "pointing to mathematical expressions", and "writing on the worksheet". The temporal transition of each group's behaviors was coded on a Sportscode time line. Moreover, the Intel RealSense depth camera (https://www.intelrealsense.com/stereo-depth), based on stereo vision technology, was used to measure the depth from the sensor to each picture element on the table and depth data were logged with time stamps. Figure 2 (right) shows a sample of colored depth image derived from RealSense. Using this log data, the movements of the participants' hands and heads intercepting light from the depth sensor over the specific regions including two worksheets and the iPad (outlined in red on the left of Figure 2) were visualized on a time line respectively. The time intervals during which progress in participants' thinking was identified from the transcribed text data were plotted on the Sportscode timelines, which were examined for changes in their behavioral patterns across these intervals. Additionally, RealSense timelines were compared with the Sportscode timelines to determine if any of those changes were portrayed on the RealSense data.

3. Results

The participants were assumed to argue about the following three cases: (1) $\alpha > \frac{\pi}{2}, \alpha + \beta < \pi$, (2) $\beta > \frac{\pi}{2}, \alpha + \beta < \pi$, (3) $\alpha < \frac{\pi}{2}, \beta < \frac{\pi}{2}, \frac{\pi}{2} < \alpha + \beta < \pi$. Here we analyze the data from the CSCL of 8 groups I - VIII. Due to the confusion about the range of angles, groups II and VII reached an impasse during their session and completed the task for only one of these cases. Also, groups VI and VIII became confused due to their insufficient understanding of the trigonometric ratios for supplementary and complementary angles, though they finally accomplished the task for at least two cases. While group IV took as long to accomplish the task as these groups, the participants' struggle was caused by their difficulty in reconciling the case when $\alpha = \frac{\pi}{2}$ in which some relevant triangle "collapses". Regarding the utterance of Participant B, "It is enough to choose the calculation method (addition and subtraction) in accordance with the positional relationship between PH and HQ" while pointing to the mathematical expression in rudimentary case (34:29), it can be seen that group IV got the whole picture of the proof at the last stage. However, except for the limited time intervals involving this sort of discussion, the participants' communication was seemingly at an impasse. The remaining 3 groups I, III, and V completed the task for the above mentioned three cases and recognized their similarities and differences. While a previous gesture study by Nathan (Nathan, et al., 2014) illustrated that task-relevant actions together with pedagogical language can facilitate students' gaining mathematical insight, another previous study by Alibali et al. (Alibali, et al., 2012) clarified that gestures manifest embodied mathematical knowledge in three distinct ways. In the case of this study, most of the observed gestures were pointing to the geometric elements in the dynamic geometry screen and the rest were mainly conceptual metaphors depicting the trigonometric ratios for supplementary and complementary angles.

Figure 3 shows timelines derived from Sportscode and RealSense in which the horizontal axes pointing to the right represent the passage of time and are synchronized for each group. In the Sportscode timelines, the behaviors – "manipulation of the dynamic content", "making gestures", "pointing to the symbolic expression", and "writing on the worksheet" - of each participant were coded from next to the top row to the bottom row. The time intervals of the progress in thinking were identified by analyzing the transcribed text data and are highlighted in yellow. It can be observed that, through these time intervals, participants' behaviors shifted their weight from manipulation and gesture (acquisition of visual information) to pointing at and writing mathematical expressions (conversion to formal deduction). In the RealSense timelines, the fluctuations of the depth from the sensor to the iPad region, left worksheet region (participant A), and right worksheet region (participant B) are visualized by the bar graphs in blue, green, and red respectively. As observed in the videotaped image, while writing behaviors usually occurred over the worksheet region, other ones occurred mainly over the iPad region which can be regarded as "the space for collaborative work". In accordance with this observation, red and green bars become conspicuous in RealSense timelines when participants A and B were engaged in writing on the worksheet respectively.

The analysis of videotaped images also indicates that the role of writing behavior changed as participants' thinking progressed. At the earlier stage, writing behavior and pointing to geometric objects often occurred alternately and participants wrote down some pieces of their observations on the worksheets. In contrast, at the later stage, they were solely engaged in writing behavior to describe the whole picture of proof. Regarding this change, the case of group V is exceptional. In fact, considerable weight was placed on writing compared to other groups and participants began to describe the whole process of proof at a very early stage with the utterance by B (02:06), "Writing may lead to our understanding". Judging from this utterance, the participants might have already recognized that a similar outline of proof can be applied to these general cases with minor changes in a few pieces.

4. Discussion and Conclusion

Comparison of Sportscode timelines between groups gives some insight into the RQ. (1). While the transition from visual information acquisition to formal deduction is observed in the cases of groups with relatively short thinking processes, that transition is not so apparent in other cases including Groups II and IV. This contrast strongly indicates that monitoring the changes in learners' behavioral patterns in the CSCL environment can lead to a diagnosis of their impasse. A group-wise comparison between the timelines of Sportscode and RealSense gives clearer insight into the RQ. (2). The movements of participants' hands and heads over worksheet regions associated with their writing behaviors are clearly portrayed on RealSense time lines. These findings suggest that, as in many other related studies (Sharma and Gianna-

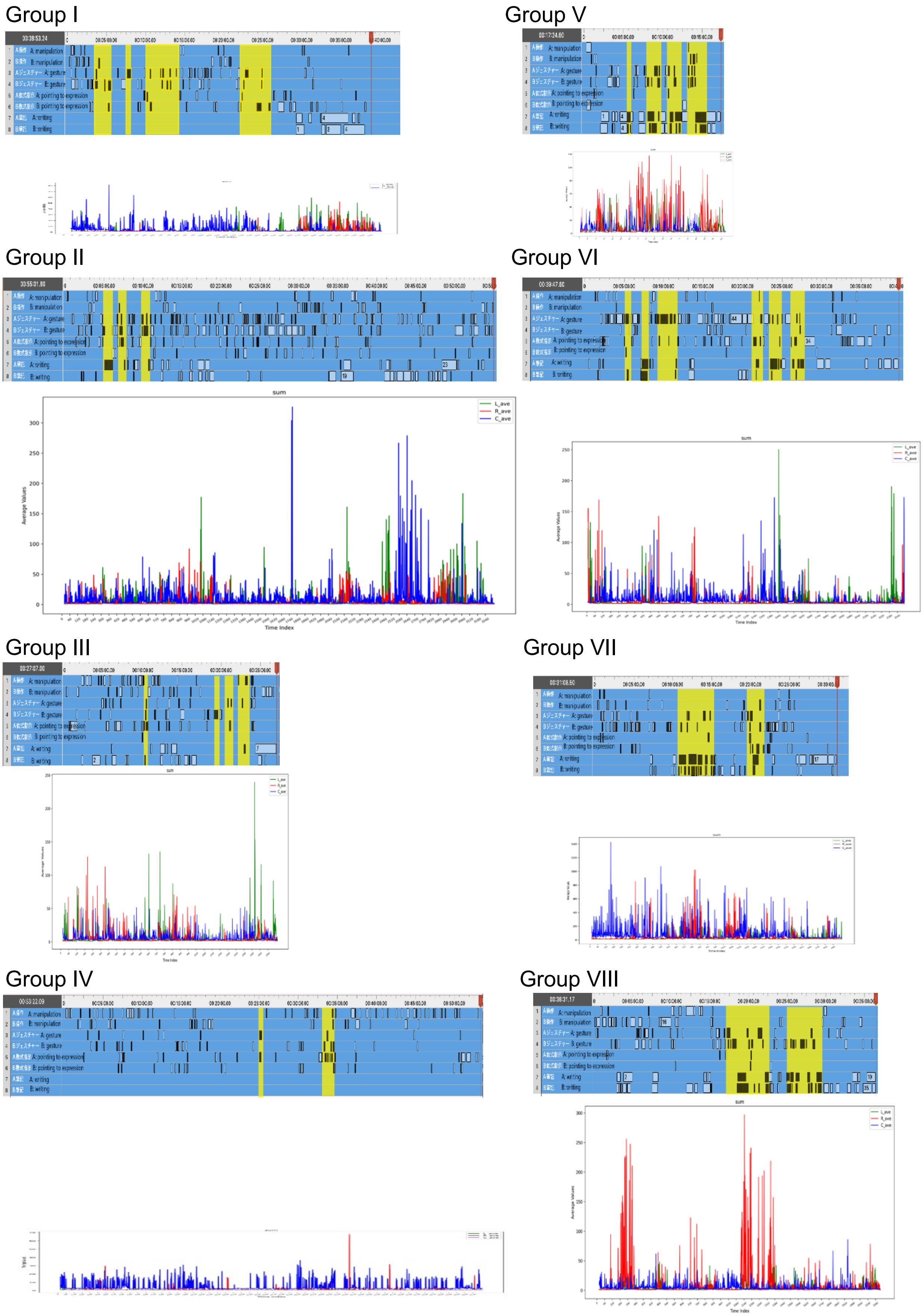

Figure 3. Time lines for each group derived from Sportscode (first figure) and RealSense (second figure). They are synchronized groupwise and their unit lengths are made uniform across groups. Writing behaviors are coded on the last two tiers in Sportscode time line.

kos, 2020), aligning different sort of data (sensing data and manipulation log in this case) can help instructors to assess learning performance or diagnose learners' struggle. In particular, the strong interrelation between interaction patterns from a touchscreen and high levels of cognitive workload illustrated in some previous research (Mock, et al., 2016) seems to be consistent with the findings of this study that participants' persistence in the acquisition of visual information indicates their struggle.

While the inexpensiveness of the depth camera offers a great potential for making the framework of this study scalable, there are several challenges associated with this research method, as noted in other related studies (Sharma and Giannakos, 2020; Ochoa, 2022). First, generalizing findings of this study to other learning themes and contexts is not straightforward. In fact, presumed thinking processes and behavioral patterns of learners heavily depend on the specific characteristics of these themes and contexts. Therefore, further refinement is necessary for the design and configuration of learning tools and sensing devices. Second, the participant population in this study is limited, and additional data are required to accurately analyze the thinking processes of learners from different populations. For instance, in Group IV of this study, the quality of participants' thinking was noticeably higher compared to other groups, despite their proof activity appearing stagnant. This suggests that aligning other types of learning data is essential to fully understand their learning processes. Further study across a wider range of learning themes and populations is needed.

Acknowledgements

This work was partially supported by JSPS Kakenhi 21K02752 and 22K02951.

References

Alibali, M. W., & Nathan, M. J. (2012). Embodiment in mathematics teaching and learning: Evidence from learners' and teachers' gestures. *The Journal of the Learning Sciences*, *21*, 247-286.

Blikstein, P., & Worsley, M. (2016). Multimodal learning analytics and educational data mining: Using computational technologies to measure complex learning tasks. *Journal of Learning Analytics, 3(2),* 220-238.

Dehaene, S., Spelke, E., Pinel, P., Stanescu, R., Tsivkin, S. (1999). Sources of mathematical thinking: Behavioral and brain-imaging evidence. *Science, 284*, 970-974.

Lakoff, G., & Nunez. R. E. (2001) *Where Mathematics Comes From*. Basic Books.

Mock, P., Gerjets, P., Tibus, M., Trautwein, U., Moller, K., & Rosenstiel, W. (2016). Using touchscreen interaction data to predict cognitive workload. *Proceedings of ICMI'18*, 349-356.

Nathan, M. J., Walkington, C., Boncoddo, R., Pier, E., Williams, C. C., & Alibali, M. W. (2014). Actions speak louder with words: The roles of action and pedagogical language for grounding mathematical proof. *Learning and Instruction, 33*, 182-193.

Ochoa, X. (2022). Multimodal learning analytics – Rationale, Process, Examples and Direction, in *Handbook of Learning Analytics* 22, 54-65.

Ochoa, X., Chiluiza, K., Mendez, G., Luzardo, G., Guaman, B., & Castells, J. (2013). Expertise estimation based on simple multimodal features. *Proceedings of ICMI'13*, 583-590.

Oviatt, S., Cohen, A., & Weibel, N. (2013). Multimodal learning analytics: Description of math data corpus for ICMI grand challenge workshop. *Proceedings of ICMI'13*, 563-568.

Paneth, L., Jeitziner, L. T., Rack, O., & Zahn, C. (2023). A multi-method approach to capture quality of collaborative group engagement. *Proceedings of CSCL'23,* 91-98.

Sharma, K., & Giannakos, M. (2020). Multimodal data capabilities for learning: What can multimodal data tell us about learning? *British Journal of Educational Technology*, 51-5, 1450-1484.

Spikol, D., Avramides, K., & Cukurova, M. (2016). Exploring the interplay between human and machine annotated multimodal learning analytics in hands-on STEM activities. *Proceedings of LAK'16,* 522-523.

Spikol, D., Ruffaldi, E., & Cukurova, M. (2017). Using multimodal learning analytics to identify aspects of collaboration in project-based learning. *Proceedings of CSCL'17,* 263-270.

Sweller, J., Ayres, P., & Kalyuga S. (2011). *Cognitive Load Theory.* Springer.