



Measuring Multimodal Data in an XR-Based Training Simulation Environment

Kukhyeon Kim¹, Jongho Kim¹, Daeun Kim¹, Hongyu Xian¹, Seoyeon Park¹ and Jeeheon Ryu¹

¹ Chonnam National University, Gwangju, Republic of Korea
jeeheon@jnu.ac.kr

Abstract. This pilot research aims to investigate the potential of multimodal data in predicting the user's performance in an XR-based training simulation environment. XR-based learning simulations have rapidly gained popularity in training education areas like medical and nursing, as well as STEAM, due to their ability to provide training opportunities in an immersive and authentic environment. In this study, multimodal data, such as eye-tracking and behavioral data (task completion time, accuracy, head movement, and hand movement) were collected, as well as the participant's perception of the simulation through a subjective survey questionnaire. We aim to investigate how and what kind of multimodal data could be collected in an XR-based learning environment to optimize the training curriculum, improve the user experience, and enhance the learning outcomes. To achieve this, we collected multimodal data for attention, cognitive load, and performance behavior: head movement, hand movement, and eye information from 22 medical residents who participated in the XR-based simulation experience focused on strabismus diagnosis for resident training. As a result, we can collect fifteen multimodal data types for attention and nine data types for performance behavior. These results could be the fundamental research for predicting the learner's performance states in XR-based training simulation.

Keywords: XR-based Simulation, HoloLens, Multimodal Data, Learner Performance.

1 Introduction

XR-based training simulations have gained popularity recently due to their ability to provide immersive and interactive learning experiences. XR-based training simulations allow for realistic and safe training environments that may be difficult or dangerous to recreate in the real world [1]. In addition, XR-based training simulations can be customized to the individual learner's needs and preferences, such as their own pace and level of expertise [2]. Additionally, XR-based training simulations can provide immediate feedback to learners, improving learning outcomes [3]. Another advantage of XR-based training simulations is that they can be more cost-effective and time-efficient [4]. Furthermore, XR-based training simulations can provide opportunities for remote learning and collaboration [5].

It is important to assess the learner's performance in XR learning environments. By understanding the learner's performance, immediate feedback can be provided, which can aid in predicting and evaluating learning outcomes and the learner's state. Multimodal information, such as brainwaves, eye gaze, facial expressions, heart rate, sweat, time, and performance behavior, provides learners with information they may not have been aware of during the learning process and offers insights into their cognitive and emotional state at specific moments.

Multimodal data is increasingly used in learning analytics to gain insights into learners' behavior and performance in educational settings. For example, eye-tracking and facial expression data have been used to identify patterns in learners' behavior that indicate engagement, interest, or understanding of the materials, as well as the user's cognitive load and attention [6-7]. In addition, mouse movement, clickstream, video recording, and

self-reported feedback can be used to recognize learners' behavior and performance in learning analytics [8-9]. Physiological data such as heart rate and skin conductance can be used to evaluate the user's stress level and emotional state [10], [11]. Head position data can be used to identify the user's level of engagement and attention [12], and hand movement data can be used to evaluate the user's motor skills and proficiency to evaluate learners' proficiency in surgical procedures [13]. Based on these previous studies, this study investigates various multimodal information sources that can be utilized to determine the learner's state and proposes future research directions based on these findings.

2 Method

2.1 Participants

Twenty-two medical residents (fourteen males and eight females) in South Korea participated in the experiment. The average age was 25.60 (SD=1.33). The within-subjects design allowed all participants to experience three scenarios of XR-based training simulation.

2.2 Experiments Contents: XR-based Simulation

The XR-based simulation aimed to diagnose pediatric patients with strabismus using holograms. Participants examined virtual patients and diagnosed strabismus by evaluating changes in the eye. The strabismus diagnosis simulation comprised three scenarios, and each process is shown in Figure 1. Firstly, the Ocular motility was performed with a stick to test the patient's eye movements. Second, the cover-uncover test was conducted to diagnose strabismus and types of strabismus while covering the eyes. Finally, the Prism test was performed to diagnose the degree of strabismus using a prism. In particular, the real prism tool could be practiced for holding the prism tool with the correct posture during the prism test. Each scenario lasted between 2 to 3 minutes to complete.



Fig. 1. Three simulation scenarios: left: Ocular motility test, middle: Cover uncover test, right: Prism cover test.

2.3 Variable

Table 1 shows the types of multimodal data and variables collected. Gaze information and the head position of the user were used to measure the user's attention and cognitive load [7, 12, 14]. The hand position and interaction with the 3D object were used to check the accuracy of the diagnostic performance and skills, as shown in Figure 2 [8,13]. Figure 3 shows the process of data collection.

Table 1. Multimodal data types.

Dependent variable	Multimodal data
Attention / Cognitive Load	Head
	Gaze
Diagnostic Correct Position	Hand
	Object

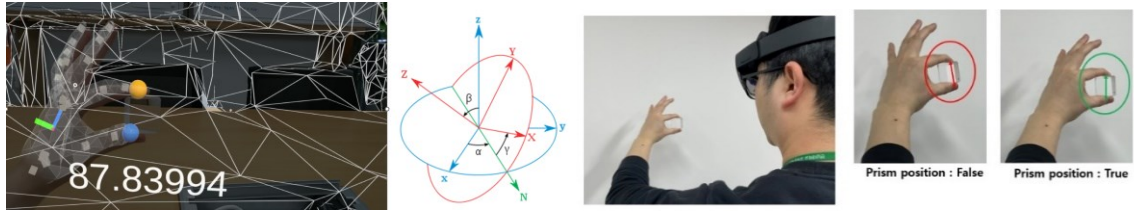


Fig. 2. The diagnostic performance of grabbing Prism.

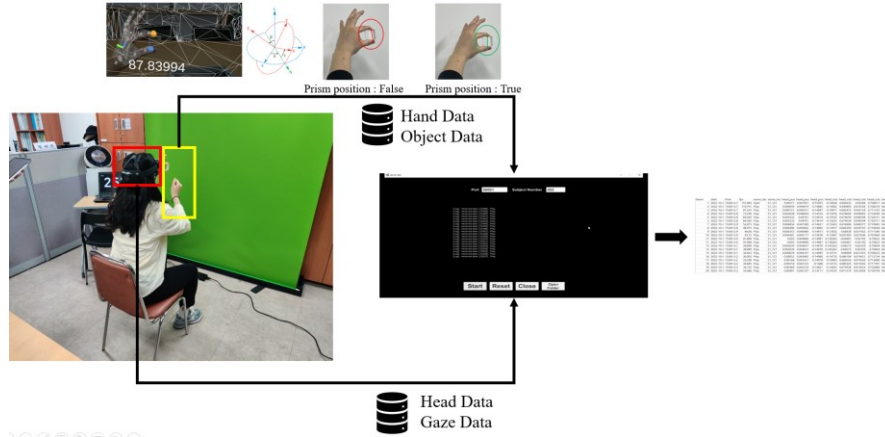


Fig. 3. The process of data collection.

2.4 Data Collection

Hand, head, and gaze data was extracted using HoloLens Pro 2 devices. In addition, object interaction data was extracted using data from participants interacting with real objects. Fig. 2 shows the data extraction flow from the interaction phase to the Excel spreadsheet. When using the prism, the back side of the prism had to be perpendicular to the patient’s visual axis. Otherwise, the patient’s angle of view could be diagnosed incorrectly. All the multimodal data were extracted at a rate of 30 times per second.

3 Results

Table 2. Structure of Multimodal Data.

Dependent variable	Multimodal data	Extracted data	Number of subfactors
Attention / Cognitive Load	Head	Head_positon(x,y,z),Head_rotation(x,y,z,w), Head_tracking(state)	8
	Gaze	Gaze_origin(x,y,z),Gaze_direction(x,y,z), Gaze_hit_position(x,y,z), Gaze_hit_normal(x,y,z)	12
Diagnostic Correct Position	Hand	Hand_joint left, Hand_joint right	98
	Object interaction	Object_grab(stick, cover, prism), Object_try(stick, cover, prism),	9
		Object_time(stick, cover, prism)	

Table 2 shows the extracted multimodal data and the number of categories. Head data comprised eight sub-factors: head position (x,y,z), head rotation (x,y,x,z), and head tracking (state). Gaze data included a total of twelve sub-factors: gaze origin (x,y,z), gaze direction (x,y,z), gaze hit position (x,y,z), and gaze hit normal (x,y,z). Hand data consisted of ninety-eight hand joint positions. Object data included nine sub-factors: stick grab valid, stick try number, stick try time, cover grab valid, cover try number, cover try time, prism grab valid, prism try number, and prism try time.

4 Conclusion & Discussion

This study proposes using multimodal data extracted from head, gaze, hand, and object interactions to identify cognitive load, attention, and performance levels in XR-based training simulations. Multimodal data can confirm information such as how much the user pays attention to the learning task during the training process, the level of cognitive load experienced by the user, and the accuracy of the user's performance in skill training with the target skill object. The results of this study are expected to provide baseline data for the development of a program that predicts learners' performance state and provides appropriate feedback in the context of learning analytics based on artificial intelligence technology like deep learning and machine learning [14].

Acknowledgments

This work was supported by the Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korean government (MSIT) (NO.2022-0-00137, XR User Interaction Evaluation and Employing Its Technology)

References

1. A. D. Kaplan, J. Cruik, M. Endsley, S. M. Beers, B. D. Sawyer, and P. A. Hancock, *Human Factors: The Journal of the Human Factors and Ergonomics Society* **63**, 706 (2020).
2. S. K. Jagatheesaperumal, K. Ahmad, A. Al-Fuqaha, and J. Qadir, *ArXiv.org* (2022).
3. S. Doolani, C. Wessels, V. Kanal, C. Sevastopoulos, A. Jaiswal, H. Nambiappan, and F. Makedon, *Technologies* **8**, 77 (2020).
4. S. L. Farra, M. Gneuhs, E. Hodgson, B. Kawosa, E. T. Miller, A. Simon, N. Timm, and J. Hausfeld, *CIN: Computers, Informatics, Nursing* **37**, 446 (2019).
5. P. S. Mathew and A. S. Pillai, *SpringerLink* (2022).
6. J. K. Olsen, A. Guneyosu Ozgur, K. Sharma, and W. Johal, *International Journal of Child-Computer Interaction* **31**, 100447 (2022).
7. I. Dubovi, *Computers & Education* **183**, 104495 (2022).
8. J. L. Felipe, J. Garcia-Unanue, D. Viejo-Romero, A. Navandar, and J. Sánchez-Sánchez, *Sensors* **19**, 4113 (2019).
9. S. Praharaj, M. Scheffel, H. Drachsler, and M. Specht, *IEEE Transactions on Learning Technologies* **14**, 367 (2021).
10. *Proceedings of the 2nd International Conference on Physiological Computing Systems* (2015).
11. D. P. Salgado, F. R. Martins, T. B. Rodrigues, C. Keighrey, R. Flynn, E. L. Naves, and N. Murray, *Proceedings of the 9th ACM Multimedia Systems Conference* (2018).
12. S. Rossi, E. Leone, and M. Staffa, *AI*IA 2016 Advances in Artificial Intelligence* 89 (2016).
13. P. Churi, S. Joshi, M. Elhoseny, and A. Omrane, *Artificial Intelligence in Higher Education: A Practical Approach* (CRC Press, Taylor & Francis Group, Boca Raton, 2023).
14. F. Ouyang, M. Wu, L. Zheng, L. Zhang, and P. Jiao, *International Journal of Educational Technology in Higher Education* **20**, (2023).