Estimation of Difficulty When Reading VR-based Educational Comics Using Gaze, Facial Movement, Heart Rate, and Electroencephalography*

Hiroyuki Ishizuka[1], Kenya Sakamoto[2], Shizuka Shirai[2], Yutaro Hirao[1], Monica Perusquia-Hernandez[1], Hideaki Uchiyama[1] and Kiyoshi Kiyokawa[1]

[1]*Nara Institute of Science and Technology, Japan*
[2]*Osaka University, Japan*

Abstract

Using the metaverse in education can improve motivation, presence, engagement, immersion, interest, and performance. However, automated recognition of progress or learning difficulties is still a challenge. In this study, we search for relevant biometric features that can be used to estimate the user's self-reported difficulty level of educational comic books (manga) materials in Virtual Reality (VR). Educational manga has been an effective learning tool for a wide range of fields. It has unique spatial features that are unavailable in traditional approaches such as browsing history or log data analysis. Our approach uses facial expressions, electrocardiogram (ECG), and electroencephalogram (EEG) data, in addition to eye gaze, which has been used previously, to estimate difficulty when reading VR-based educational manga. We measured learners' data while reading manga materials in a VR space and estimated the learner's self-reported difficulty at two levels (challenging and accessible) using Support Vector Machine (SVM) and Random Forest (RF) algorithms. As a result, the RF approach using facial expressions and ECG features achieved an accuracy of 0.97 and an F1 score of 0.94 on the two-level difficulty estimation task.

Keywords
Difficulty Estimation, Biosignals, Virtual Reality, Affective Computing

1. Introduction

The metaverse, gaining global attention recently, holds promise for education, known as the Edu-Metaverse [1, 2]. The immersive experience of Virtual Reality (VR) space in the Edu-Metaverse is being utilized to develop virtual laboratories [3] and simulation materials [4], enhancing educational performance [1].

Moreover, manga (comic book) learning materials have been growing remarkably and have the potential to make learning enjoyable. Previous studies have shown that manga materials motivate learners to learn and that the story's presentation is effective for long-term memory retention [5, 6].

Furthermore, advancements in biosensing and machine learning technologies have opened up possibilities for adaptive learning within the Edu-Metaverse, an approach that tailors learning materials and processes to individual learners' abilities and needs. However, realizing adaptive learning in the Edu-Metaverse requires accurately estimating the learner's state, which has traditionally been done using log data [7].

Therefore, this study proposes a method for estimating the learner's self-reported difficulty level while reading manga materials in VR using eye gaze [8], facial expression, heart rate, and brain wave information. The proposed method introduces facial expression, heartbeat, and EEG features for psychological state estimation. The effectiveness of the features was investigated using SVM and RF.

2. Related works

Cognitive load, affected by various factors, is treated as a subjective perception of task load in this study [9]. Additionally, log data, including clickstream data from learning management systems and results of online tests, have been used to estimate the learner's state [7]. We aim to estimate the learner's self-reported difficulty level from physiological data, which is suitable for real-time support.

Eye-tracking technology has been employed to assess cognitive load and cognitive processes related to

APMAR'24: The 16th Asia-Pacific Workshop on Mixed and Augmented Reality, Nov. 29-30, 2024, Kyoto, Japan
*Corresponding author.
ishizuka.hiroyuki.id2@is.naist.jp (H. Ishizuka); sakamoto.kenya@lab.ime.cmc.osaka-u.ac.jp (K. Sakamoto); shizuka.shirai.cmc@osaka-u.ac.jp (S. Shirai); yutaro.hirao@is.naist.jp (Y. Hirao); perusquia@ieee.org (M. Perusquia-Hernandez); hideaki.uchiyama@is.naist.jp (H. Uchiyama); kiyo@is.naist.jp (K. Kiyokawa)
https://www.monicaperusquia.com/ (M. Perusquia-Hernandez); https://carelab.info/ (K. Kiyokawa)
0009-0005-2249-1562 (K. Sakamoto); 0000-0003-2676-3404 (S. Shirai); 0000-0003-3546-3454 (Y. Hirao); 0000-0002-0486-1743 (M. Perusquia-Hernandez); 0000-0002-6119-1184 (H. Uchiyama); 0000-0003-2260-1707 (K. Kiyokawa)
© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).
CEUR Workshop Proceedings (CEUR-WS.org)

CEUR-WS.org/Vol-3907/paper4.pdf

CEUR Workshop Proceedings ceur-ws.org ISSN 1613-0073

visual exploration and reading [10]. Eye gaze information helps assess learners' comprehension in educational settings [11]. Sakamoto et al. used eye-tracking information to estimate the difficulty level of reading manga materials in a VR environment [8].

Facial expression information has been used to evaluate learner engagement using machine learning techniques [12]. Nakamura et al. estimated the difficulty of an English word test using a combination of facial features and system operation logs [13].

Heart rate information obtained from ECG has been used as a physiological method for internal boundary conditions [9, 14, 15]. Heart rate increases with task demands and memory load [9]. HRV, the variability of the RR interval of the heartbeat, is used to assess stress [16]. RMSSD of HRV, unaffected by respiration [17], correlates with parasympathetic activity and stress. The 20th and 80th percentiles of the HRV data are used to estimate the state of stress and relaxation. pNN20 and pNN50 reflect vagus nerve [17]. These indices are often adequate for estimating a learner's self-reported difficulty when reading manga materials in a VR environment.

Previous studies on EEG-based cognitive load assessment have focused on θ and α frequency bands. θ wave activity is linked to frontal lobe activity and working memory capacity, while α wave activity is related to cognitive attention, arousal, and memory capacity [18, 19]. Studies measuring EEG in VR environments have reported increased cognitive load in such environments [20, 19].

3. Data set

We conducted a data collection experiment in a VR space using educational manga about immunology to develop a model for difficulty estimation.

3.1. Participants

Nine students (eight males, one female, 22-24 years old, mean 23 years old, standard deviation 0.3) with no prior knowledge of immunology were recruited. Two participants with missing data were excluded from the analysis. This study was approved by our institution's IRB.

3.2. Task

There were three phases: baseline, reading, and annotation (see Fig 1). Initially, a dark screen was displayed for three minutes to establish a baseline, during which the participants were asked to relax. Then, they learned the contents by reading the immunology manga material in a VR space during the reading phase. After that, they went through the same material and labeled the

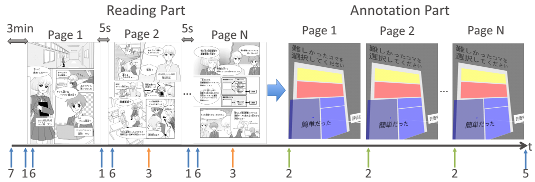

Figure 1: The experiment consists of three phases: baseline, reading, and annotation. Note that the manga sample in this figure is a dummy; it was not used in the actual experiment.

perceived difficulty level of each panel in the annotation phase.

3.3. Procedure

Participants were briefed on the experiment's purpose, the VR environment, and the data collection process. They were informed that eye gaze, facial expression, heart rate, and EEG information would be recorded during the learning process. A questionnaire was administered to determine the participants' prior knowledge of immunology and VR experience. Afterward, participants put on the HMD and practiced operating the interface in the VR environment using manga materials different from those used in the experiment.

Next, the participants wore the ECG, followed by the EEG. Gel was used to adjust the impedance to 30 kΩ or lower. Afterward, the participants were asked to put on the HMD again and confirm that eye gaze and face tracking were working correctly.

3.4. Apparatus

We used an HTC VIVE Pro Eye and an HP VR Backpack G2 with an i7-8850H processor and NVIDIA GeForce 2080 GDDR6 8GB in the experiment (Fig. 2). The software consisted of Unity version 2019.2.3.f1. The HMD's eye-tracker and SRanipal 1.3.2.0 SDK were used to collect eye-tracking information. The VIVE Facial Tracker collected facial expression information as 38 different blend shapes, including lips, chin, and cheeks, at 60 Hz with a response time of 10 ms. Shimmer3[1] was used to collect ECG. Two Shimmer terminals were used for the measurement and synchronization with Unity. An EEGo sports[2], a portable mobile EEG machine, was used for the EEG measurement. The 32 wet electrodes of the EEG cap were configured in the 10-20 system. EEG data were recorded via the EEGo sports software.

[1]https://shimmersensing.com/product/shimmer3-ecg-unit-2/
[2]https://www.ant-neuro.com/products/eego_sports

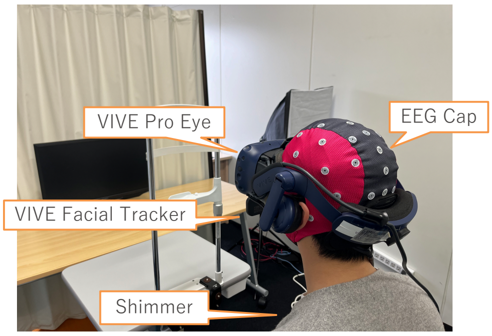

Figure 2: Image showing an experiment setup. Participants wore the VIVE Pro Eye, the VIVE Facial Tracker, the EEG Cap, and Shimmer3.

3.5. Stimuli

We used a manga entitled "Understanding Immunology through Manga", published by Ohmsha [21], which granted permission for us to use the manga material. Ohmsha's Manga "de Wakaru series" is widely recognized for its easy-to-understand explanations and manga introducing various fields such as science, electricity/-electronics, machinery, and architecture. Chapters 1-3 (76 pages) were used in this experiment.

3.6. Measurements

3.6.1. Difficulty Annotations

The time participants spent per panel within one page was not constant due to the diverse difficulty levels of the panels. Thus, we asked them to annotate their perceived difficulty of each panel at three levels: "easy," "difficult but understandable," and "difficult and not understandable." They could record the difficulty levels using a ray-casting interaction to mark difficult vignettes while reading. To avoid differences in the definitions of each label among the participants, before the experiment, participants were asked to rate a panel as "difficult but understandable" if they understood the manga after reading it; and "difficult but not understandable" if they did not understand the manga contents. If they were unsure, the participants were asked to rate it as "difficult but understandable." This paper reports the results of estimating the user's self-reported difficulty based on the "easy" and "other" classifications. After each chapter, the NASA TLX Questionnaire was used to assess general workload.

3.6.2. Physiological Data

Thirty-eight different BlendShapes corresponding to the lips, chin, cheeks, and other body parts were collected for facial expression information [22]. ECG measurements were taken with electrodes attached following the standard five-lead placement method. The electrodes used for the EEG measurements were Fz, F3, F4, Cz, C3, C4, POz, Pz, P3, P4, M1, and M2, arranged according to the international 10-20 method. After the caps were attached, the gel was inserted, and the impedance was confirmed to be less than 30 kΩ before measurement began.

4. Analysis

Features were extracted from sliding windows separately per modality [23]. The window slid in 0.5-second steps for all modalities.

Facial Expression Features. BlendShapes associated with face region movements were used. The resting state was set to zero. We calculated the minimum, average, and maximum values for each five-second window by summing the values of 38 BlendShapes. We also calculated the change in the total value and used the minimum, average, and maximum values as the feature values. The reason for using this method is that learners are likely to make lip thrusting or mouthing movements when reading a difficult passage, even if they do not have a visibly sad or frustrated expression.

ECG Features. NeuroKit2 [24] was used to read the ECG data, clean the signal, and perform peak detection. The window size was 20 seconds. Features were output for each window. Based on previous work, the features used were mean heart rate, HRV minimum, maximum, mean, standard deviation, RMSSD, 20th and 80th percentiles, pNN20, and pNN50.

EEG Features. The EEGo sports has an output of 0.5$\sim$30Hz band-pass filter. The output data included Trigger information sent from Unity. Then, the noise was removed using Independent Component Analysis (ICA) in MNE-Python [25], and Morlet wavelets were used to perform time-frequency analysis at 3$\sim$13Hz. The average frequency power of the θ wave (4$\sim$8Hz) and the α wave (8$\sim$12Hz) were calculated. The window size was then set to two seconds, and the power of the θ and α waves within the window was calculated. We hypothesize that θ activity is higher and α activity is lower when the learner perceives difficulty.

Classification. To validate the proposed method, we constructed a model with two classes, "easy" and "difficult," employing Support Vector Machine (SVM) and Random Forest (RF) algorithms. We fitted subject-independent models using leave-one-subject-out cross-validation. Sixteen different combinations were compared in the subject-independent model, each consisting of either the presence or absence of eye gaze, facial expression, electrocardiogram, and EEG features. Accuracy, Precision, Recall, and F1 scores were used to evaluate performance.

5. Results

Self-reported workload. Table 1 shows the results of each questionnaire, including the difficulty reported in each chapter. For simplicity, the Likert scale is considered an interval scale. It was observed that the difficulty level of Chapter 1 was the lowest. The difficulty level of Chapters 2 and 3 increased because the chapters contained more content and specialized vocabulary as they progress. The level of fatigue was also rated higher in the last chapters.

Number of Labels for Learner's self-reported Evaluation. Table 2 shows the percentage of the learner's self-reported evaluation labels per panel and per window of the collected manga teaching materials. While only 8.12% of the panels were labeled as "difficult," this proportion increased to 30.0% when the time unit was taken into account. This suggests that participants spent more time on panels that were perceived as "difficult."

Accuracy of Learner's Self-reported Difficulty Estimation. When all features were used for each participant, the results were as follows: accuracy of 77%, F1 score of 77% for SVM, accuracy of 94%, and F1 score of 94% for RF. All combinations of biometric information were analyzed to identify the most relevant features for difficulty estimation. Accuracy, Precision, Recall, and F1 scores are shown in Table 3. Accuracy, F1 score, and Recall were 96.5%, 94.0%, and 91.2%, respectively, when facial expression and electrocardiograms were combined, and precision was 97.8%, the highest accuracy when gaze, facial expression, and electrocardiograms were combined.

Table 1
Questionnaire results for each chapter rated on a Likert scale

	Chapter 1	Chapter 2	Chapter 3
Difficulty 1–4	1.29 (0.49)	2.43 (0.53)	3.29 (0.76)
Comprehension 1–4	4.00 (0.00)	3.14 (0.64)	2.86 (0.90)
Fatigue 1–4	1.71 (0.76)	2.86 (0.90)	3.57 (0.53)
Mental workload 1–10	4.28 (0.70)	4.43 (0.73)	4.57 (0.49)

Numbers in parentheses denote standard deviation *(SD)*.

Table 2
Per-panel and per-window comprehension level of the educational manga material

	Easy	Understandable	Difficult
Per-panel	89.9% (5.21%)	2.01% (1.83%)	8.12% (4.95%)
Per-window	70.0% (11.2%)		30.0% (11.2%)

Numbers in parentheses denote standard deviation *(SD)*.

6. Discussion and Conclusion

We investigated eye gaze, facial expression, electrocardiogram, and electroencephalogram (EEG) features to estimate the learner's self-reported difficulty level in reading educational manga. We collected data while reading educational manga in a VR environment and evaluated the estimation accuracy using time-based SVM and RF. Accuracy and F1 scores reached 0.965 and 0.940, respectively, using the RF approach with facial expressions and electrocardiogram features. The estimation accuracy was improved when the proposed method was used for facial expressions and electrocardiograms and decreased when EEG features were used.

Compared to the results obtained by estimating gaze features alone, the combination of gaze and facial expression and gaze and electrocardiogram improved the accuracy of difficulty estimation and F1 score. However, both estimation accuracy and F1 score were lower for the combination of gaze and EEG. Estimation accuracy and F1 score were higher when all features were used except EEG features. This is probably because the EEG signal-to-noise ratio was low when simultaneously wearing an HMD. Future work should improve head-movement-related noise reduction on the EEG. Combining facial expression and ECG features produced the highest estimation accuracy and F1 score. This indicates that the ECG

Table 3
Comparison of eye gaze, facial expression, electrocardiogram, and electroencephalogram information in RF (best scores in bold)

Gaze	Face	ECG	EEG	Acc.	Rec.	Pre.	F1
✓	✓	✓	✓	0.927	0.790	0.964	0.869
✓	✓	✓		0.961	0.892	**0.978**	0.933
✓	✓			0.927	0.792	0.961	0.868
✓				0.904	0.741	0.938	0.828
	✓	✓	✓	0.866	0.607	0.928	0.733
	✓	✓		**0.965**	**0.912**	0.970	**0.940**
	✓			0.879	0.672	0.904	0.770
		✓	✓	0.792	0.387	0.844	0.530
		✓		0.949	0.872	0.957	0.913
			✓	0.687	0.092	0.425	0.152
✓		✓		0.947	0.868	0.953	0.908
✓			✓	0.840	0.574	0.849	0.685
	✓		✓	0.716	0.132	0.656	0.220
✓		✓	✓	0.908	0.739	0.945	0.830
✓	✓		✓	0.860	0.617	0.890	0.728

features are important for estimating difficulty. These two features can be collected more quickly than others and are independent of learning materials. Therefore, they can be easily applied to adaptive learning systems in VR environments, not limited to educational manga materials.

References

[1] X. Chen, D. Zou, H. Xie, F. L. Wang, Metaverse in education: Contributors, cooperations, and research themes, IEEE Transactions on Learning Technologies (2023).

[2] M. Wang, H. Yu, Z. Bell, X. Chu, Constructing an edu-metaverse ecosystem: A new and innovative framework, IEEE Transactions on Learning Technologies 15 (2022) 685–696.

[3] R. Heradio, L. De La Torre, D. Galan, F. J. Cabrerizo, E. Herrera-Viedma, S. Dormido, Virtual and remote labs in education: A bibliometric analysis, Computers & Education 98 (2016) 14–38.

[4] R. Heradio, L. de la Torre, S. Dormido, Virtual and remote labs in control education: A survey, Annual Reviews in Control 42 (2016) 1–10.

[5] L. A. Mamolo, Development of digital interactive math comics (dimac) for senior high school students in general mathematics, Cogent Education 6 (2019) 1689639.

[6] I. Fuse, S. Okabe, Computer ethics education using video and manga teaching materials: learning effects and the order of using teaching materials, Transactions of Japanese Society for Information and Systems in Education 27 (2010) 327–336.

[7] P. García, A. Amandi, S. Schiaffino, M. Campo, Evaluating bayesian networks' precision for detecting students' learning styles, Computers & Education 49 (2007) 794–808.

[8] K. Sakamoto, S. Shirai, N. Takemura, J. Orlosky, H. Nagataki, M. Ueda, Y. Uranishi, H. Takemura, Subjective difficulty estimation of educational comics using gaze features, IEICE Transactions on Information and Systems E106.D (2023) 1038–1048.

[9] R. L. Charles, J. Nixon, Measuring mental workload using physiological measures: A systematic review, Applied ergonomics 74 (2019) 221–232.

[10] S. P. Marshall, The index of cognitive activity: Measuring cognitive workload, in: Proceedings of the IEEE 7th conference on Human Factors and Power Plants, IEEE, 2002, pp. 7–7.

[11] C. Lima Sanches, O. Augereau, K. Kise, Estimation of reading subjective understanding based on eye gaze analysis, PLOS ONE 13 (2018) 1–16.

[12] J. Whitehill, Z. Serpell, Y. C. Lin, A. Foster, J. R. Movellan, The faces of engagement: Automatic recognition of student engagement from facial expressions, IEEE Transactions on Affective Computing 5 (2014) 86–98.

[13] K. Nakamura, K. Kakusho, M. Murakami, M. Minoh, Estimating learners' subjective impressions of the difficulty of course materials by observing their faces in e-learning, The IEICE transactions on information and systems (D) 93 (2010) 568–578 (in Japanese).

[14] H.-G. Kim, E.-J. Cheon, D.-S. Bai, Y. H. Lee, B.-H. Koo, Stress and heart rate variability: a meta-analysis and review of the literature, Psychiatry investigation 15 (2018) 235.

[15] N. Meshkati, Heart rate variability and mental workload assessment, in: Advances in psychology, volume 52, Elsevier, 1988, pp. 101–115.

[16] S. Delliaux, A. Delaforge, J.-C. Deharo, G. Chaumet, Mental workload alters heart rate variability, lowering non-linear dynamics, Frontiers in physiology 10 (2019) 565.

[17] S. Laborde, E. Mosley, J. F. Thayer, Heart rate variability and cardiac vagal tone in psychophysiological research–recommendations for experiment planning, data analysis, and data reporting, Frontiers in psychology 8 (2017) 213.

[18] L. J. Castro-Meneses, J.-L. Kruger, S. Doherty, Validating theta power as an objective measure of cognitive load in educational video, Educational Technology Research and Development 68 (2020) 181–202.

[19] S. Baceviciute, A. Mottelson, T. Terkildsen, G. Makransky, Investigating representation of text and audio in educational vr using learning outcomes and eeg, in: Proceedings of the 2020 CHI conference on human factors in computing systems, 2020, pp. 1–13.

[20] G. Makransky, T. S. Terkildsen, R. E. Mayer, Adding immersive virtual reality to a science lab simulation causes more presence but less learning, Learning and instruction 60 (2019) 225–236.

[21] H. Kawamoto, The Manga Guide to Immunology, Ohmsha, 2014.

[22] P. Joshi, W. C. Tien, M. Desbrun, F. Pighin, Learning controls for blend shape based realistic facial animation, in: ACM Siggraph 2006 Courses, 2006, pp. 17–es.

[23] F. Dollack, K. Kiyokawa, H. Liu, M. Perusquia-Hernandez, C. Raman, H. Uchiyama, X. Wei, Ensemble Learning to Assess Dynamics of Affective Experience Ratings and Physiological Change, in: 2023 11th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW), 2023, pp. 1–8.

[24] D. Makowski, T. Pham, Z. J. Lau, J. C. Brammer, F. Lespinasse, H. Pham, C. Schölzel, S. H. A. Chen,

NeuroKit2: A python toolbox for neurophysiological signal processing, Behavior Research Methods 53 (2021) 1689–1696.

[25] A. Gramfort, M. Luessi, E. Larson, D. A. Engemann, D. Strohmeier, C. Brodbeck, R. Goj, M. Jas, T. Brooks, L. Parkkonen, M. S. Hämäläinen, MEG and EEG data analysis with MNE-Python, Frontiers in Neuroscience 7 (2013) 1–13.