Iyer, S. et al. (Eds.) (2022). Proceedings of the 30th International Conference on Computers in Education. Asia-Pacific Society for Computers in Education

Development of a Learning Companion Robot with Adaptive Engagement Enhancement

Bowei YAO[a], Koichi OTA[b,a], Akihiro KASHIHARA[c], Teruhiko UNOKI [d] & Shinobu HASEGAWA[b,a*]
[a]*Division of Advanced Science and Technology, JAIST, Japan*
[b]*Center for Innovative Distance Education and Research, JAIST, Japan*
[c]*Graduate School of Informatics and Engineering, The University of Electro-Communications, Japan*
[d]*College of Foreign Studies, Kansai Gaidai University, Japan*
*hasegawa@jaist.ac.jp

Abstract: This research aims to develop a learning partner robot that can adapt its interaction to each learner's individual reactions to enhance learner engagement. Engagement is a mental state that positively influences learning and is an essential element that supports learner independence through immersion in the learning process. However, no established methodology exists to enhance learner engagement with proper support in the individualized and isolated learning process. This research focuses on a learning partner robot as a learning companion in self-directed learning. To realize the partner robot, we implemented the engagement estimation architecture from the learner's facial images during learning, then designed a robot interaction model to enhance learner engagement according to their engagement states, and extended the interaction model to update the robot's strategy according to the learner's response to the interactions. We conducted a comparative experiment with 20 graduate student participants with and without robots. The results indicated that the average engagement during learning was significantly higher in the with-robot condition, and the satisfaction with the robot interaction was highest in the last 1/3 of the learning period.

Keywords: Learner Engagement, Adaptation, Facial Images, Learning Companion Robot

1. Introduction

Self-regulated learning and learner engagement, defined as the involvement of the learners' cognitive and emotional energy to accomplish a learning task (Schunk & Mullen, 2012), is critical to the success of online learning (Schaeffer & Konetes, 2010). However, decreased interaction between instructors and friends and social isolation from the COVID-19 pandemic might easily demotivate and disengage their learning (Akuratiya & Meddage, 2020).

This research aims to develop a learning partner robot that enhances learner engagement to solve these problems. Leite et al. highlighted the increasing popularity of social robots in educational environments (Leite et al., 2013). Such robots have a potential to provide an adaptive learning experience and support learners in a new way that is not possible in today's resource-limited educational environments. However, Belpaeme et al. pointed out a couple of challenges, such as estimating the learners' abilities and progress and coordinating the timing of verbal and nonverbal action to select appropriate robot interactions (Belpaeme et al., 2018). In other words, no established methodology exists to enhance learner engagement with proper support in the individualized and isolated learning process.

This article proposes a learning partner robot that can adapt its interaction to each learner's individual reactions to enhance learner engagement during self-directed learning at online learning. The features of the developed robot are not to support learning in a specific domain but to support self-monitoring of engagement in self-regulated learning by providing feedback to the learners on their engagement status. We also try to realize an adaptive robot interaction for supporting engagement by each learner's preference.

2. Related Work

2.1 Learning Partner Robots

Lu et al. designed a new robot system called Smart Learning Partner, which used natural language processing and emotion recognition to provide learners with an enjoyable learning experience (Lu et al. 2018). Using self-determination theory as a guideline, they designed human-robot interaction in terms of autonomy, which refers to spontaneity in performing tasks, competence, a sense of challenge and effectiveness, and relatedness, a sense of connection with others. However, the effectiveness of this interaction has not yet been discussed.

Lubold et al. introduced a socially responsive speech interface for learning companions of embodied robots with voice adaptation based on acoustic-prosodic entrainment (Lubold et al. 2016). Through evaluation, they showed that social presence was significantly higher for social speech-adaptive speech interfaces than for purely social dialogues. Although these studies are closely related to our target engagement, this research is unique in effectively promoting engagement by integrating engagement estimation and interaction generation.

2.2 Sota as Communication Robots

Recent advances in robotics have led to the usage of various communicative robots in learning support, such as NAO (Softbank Robotics Co., Ltd., 2018), Saya (Hashimoto and Kobayashi, 2005), Robohon (Sharp Corporation, 2016), and PALRO (FUJISOFT Inc., 2010). This research adopts Sota, a table-top communication robot equipped with a camera, microphone, speaker, and network functions (Vstone Co. Ltd., 2010). Sota has advanced functions and an adorable design as a platform that can provide various robot services in conjunction with loT (The Internet of Things) devices and cloud AI. It has a built-in Intel(R) Edison with Linux, and applications using image recognition, speech recognition, and speech synthesis can be easily developed using Java.

3. Proposed Method

3.1 Engagement Estimation

This research implemented the engagement estimation model proposed by Huynh et al. (Huynh et al. 2019). Their work demonstrated the effect of face and gaze related features on engagement estimation in the "Wild" environment. This research extracted the following two feature sets from the learning videos based on Huynh's method.

F1 feature set: 60-dimension data with the average, standard deviation, maximum, and minimum values of gaze direction, eye landmarks, distance from eye to the camera, and head pose at video segments with regular intervals extracted by OpenFace, an open-source toolkit for facial expression analysis (Baltrusaits et al. 2018).

F2 feature set: 128-dimension data extracted by SE-ResNet-50 (Hu et al. 2018), ResNet-50 with a Squeeze-and-Excitation (SE) block, for face regions extracted by OpenFace. SE-ResNet-50 has been pre-trained on MS-Celeb-1M (Guo et al. 2016) and VGGFace2 (Cao et al. 2018).

Learner engagement changes over time. Based on Huynh et al., LSTM-FC and FCLSTM-FC (fully connected right-front LSTM), which combine the LSTM (long short-term memory) and FC (fully connected) layers, train each feature set to capture the change. The ensemble layer receives these results and outputs the averaged learner engagement in each video segment. According to the original model by Huynh et al., the output engagement intensities were the following four levels:

0: Disengaged: 0 <= Engagement-Intensity < 0.4
1: Barely Engaged: 0.4 <= Engagement-Intensity < 0.6
2: Engaged: 0.6 <= Engagement-Intensity < 0.83
3: Highly Engaged: 0.83 <= Engagement-Intensity <= 1.00

3.2 Interaction Network

The components of partner robot interaction depend on verbal and nonverbal robot behavior. Generally speaking, emotions involve verbal and nonverbal domains: verbal expression refers to speeches or languages (Fussell, 2002), and nonverbal expression includes speed, intonation, posture, gesture, body movement, and facial expression (Dael et al., 2012). Based on these references, we designed verbal and nonverbal expressions for robot interactions to enhance learner engagement.

For the verbal expression, we referred to several feedback techniques that instructors use with their learners. "Support," "care," and "praise" positively influence learners' performance and behavior (Brophy 1981, Jiang 2019). In addition, Cutumisu et al. showed that instructors' critical feedback is more effective on learners' memory than confirmatory feedback (Cutumisu et al. 2018). Thus, we prepared three short verbal speeches in each of the four interaction categories shown in Table 1.

Table 1. *Verbal Speeches as Interaction Components*

Types	Short Verbal Speeches		
Support	I'm supporting you.	Please don't lose.	I'll always be there for you.
Criticize	Have you done something?	Are you wasting time?	When will you start learning?
Care	Take it easy.	Please take a rest.	It's tough for you.
Praise	Good job!	I'm proud you.	Keep it up!

For nonverbal expressions, we decided to use gestures and body movement to emphasize the intention of robot speeches and employed the four motion types, "Call," "Presen (presentation)," "Talk," and "Bye," as presets by the MotionAsSotaWish class, Sota's automatic motion generation function. Each motion type defines several motions of a few seconds duration, and Sota acts simultaneously with its speeches. In addition, Sota uses the TextToSpeechSota class, a cloud speech API, which generates .wav files for utterances. As another nonverbal interaction component, we adopted speech rate, and three types were selected: Slow, Medium, and Fast.

To provide the support that reflects each learner's interaction preferences, we propose the interaction network model. The right bottom of Figure 1 shows the initial state of the model, a fully connected network consisting of four layers of interaction components (speech type, motion, and speech rate), with the estimated results of engagement intensity in Section 3.1 as input. We set the initial weights among each path based on previous research (Fried, 2012). The sum of the path weights for each node to the next layer is 1. These weights are used to select the next node when determining interactions probabilistically. It allows different interactions to be generated even in the same situation. When the learner's engagement is "Highly Engaged," the robot does not interact with the learner.

The proposed method uses an engagement estimation model to compare the learner's engagement changes before and after the interaction with Sota. It determines whether the interaction was appropriate for the learner and updates the interaction network. If the learner's engagement remains the same or decreases after the interaction, the weights of all paths where the interaction took place are reduced, and the weights of other paths are increased to the sum of the weights to 1. This makes Sota implement other interaction strategies when a similar situation occurs. The interaction network would slowly find the most suitable weights to achieve good interactive experiences tailored to different learners' personalities. There would be inappropriate interactions while adjusting the weights, but it would improve as the update progresses.

3.3 Implementation

Figure 1 shows the overall architecture of the developed system. The engagement estimation and interaction network modules were implemented on a PC with Intel i7-9900k, 32G RAM, Nvidia GeForce RTX 2080 (6GB), ubuntu 20.04 OS, python 3.8, and TensorFlow 2.0. First, the system takes the learner's face image from a USB camera as input and estimates the engagement intensity at one-minute intervals using a pre-trained model described in section 3.1. Next, the interaction network selects Sota's interaction from the estimated engagement intensity explained in section 3.2. Then, Sota obtains the output of the interaction network in JSON format at one-minute intervals and performs the selected interaction. Finally, the system updates the interaction network according to the estimated results of the learner's engagement after the interaction discussed in section 3.2.

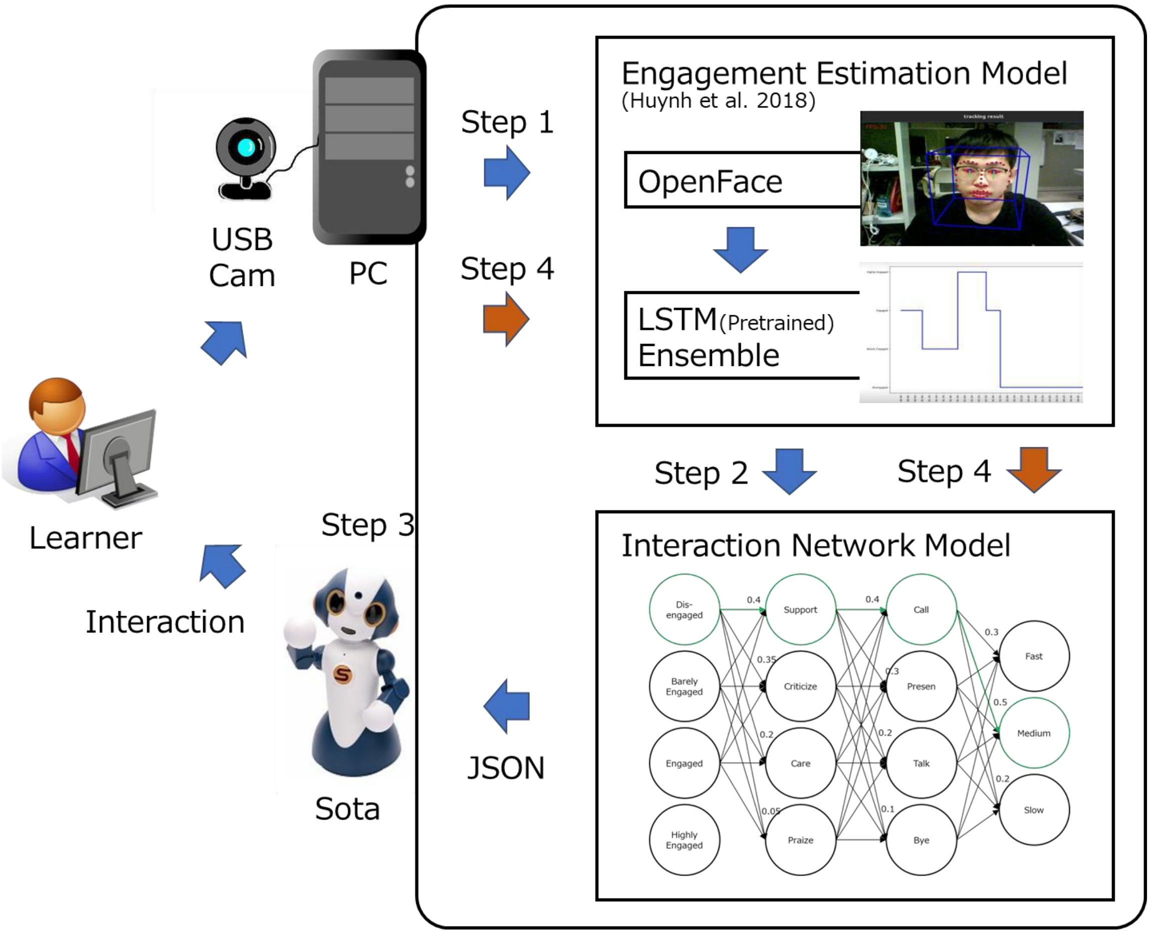

Figure 1. System Architecture

4. Experiment

To evaluate the validity of the proposed method, we conducted a within-subjects design experiment with 20 graduate students. In particular, we evaluated the effect of Sota's interaction on learner engagement and the effect of the interaction network update algorithm. In the experiment, subjects studied 30-minute learning videos, undergraduate level physics and high school level chemistry, with and without Sota conditions. The order of the experiment was counterbalanced concerning the learning videos and the with and without Sota. After learning each video, subjects rated themselves on a 4-point scale (0: Disengaged, 1: Barely Engaged, 2: Engaged, and 3: Highly Engaged). for each 5-minute engagement while watching their own facial videos and on a 5-point scale (1: Very Bad, 2: Bad, 3: Neither, 4: Good, 5: Very Good) for the satisfaction of timing and content with each interaction in the condition with Sota.

4.1 Effect of Sota's Interaction on Learner's Engagement

The average engagement with Sota was 2.07 (S.D. = 0.47), and the average engagement without Sota was 1.71 (S.D. = 0.47). The t-test result $t(19)=2.48$, $p=.022$ showed that the average engagement was significantly higher in the With Sota condition. Although this result is not sufficient due to the short experiment time, it suggests that the interaction by Sota may have a certain effect on enhancing engagement.

Table 2 shows the questionnaire results regarding the number of satisfactions with the timing and content of each interaction by Sota in the With Sota condition. The average satisfaction with the interaction timing was 3.43 (S.D. = 0.608), and the average satisfaction with the interaction content was 3.24 (S.D. = 0.728), indicating that about half of the interactions were positively accepted. The time lag for estimation itself and model update might cause low satisfaction and the accuracy limitations of the engagement estimation model.

Table 2. *Number of Satisfaction of Interaction Timing and Content with Sota*

	1 (Very Bad)	2 (Bad)	3 (Neither)	4 (Good)	5 (Very Good)
Timing	0	22	79	79	13
Content	0	39	72	72	8

4.2 Effect of Interaction Network Update Algorithm

Table 3 compares the averages of the questionnaire results for each subject's satisfaction with the interaction content when divided into 1/3 of the time order. We tested the data with the unpaired ANOVA to compare the three groups, which showed that $F=18.9$ ($p=<.00001$). Tukey's HSD was used to compare each group. The results showed $Q=3.34$ ($p=.0498$) between the first 1/3 and the middle 1/3, $Q=5.22$ ($p=.0085$) between the first 1/3 and the last 1/3, and $Q=8.56$ ($p=.0000$) between the middle 1/3 and the last 1/3, respectively. These results show that the subjects' satisfaction with the interaction decreased in the middle 1/3 of the experiment but improved as the weights were adjusted in the last 1/3 experiment.

Table 3. *Average Satisfaction of Interaction Content among Time Order Groups*

	# of Answer	Average	S.D.
First 1/3	63	3.17	0.814
Middle 1/3	61	2.84	0.879
Last 1/3	63	3.70	0.706

5. Conclusion

This research developed a learning partner robot based on Sota that can adapt its interaction to each learner's individual reactions to enhance learner engagement during self-directed learning at online learning. We also conducted a within-subjects design experiment (with/without Sota) with 20 graduate students to evaluate its effectiveness.

We constructed an initial interaction network to generate interactions that promote engagement based on learner engagement. The network is a fully connected network consisting of four layers representing the learner engagement intensity, the robot speech type, and motion and speech rate. The experiment results showed that the average engagement with Sota condition was higher than those without Sota condition. Since each session in this experiment was short (30 minutes) and the subjects interacted with Sota for the first time, we need additional experiments for a long learning period.

We also proposed an interaction updating model in which the robot changes the interaction content by judging the previous interaction based on the estimated learner engagement after the interaction. The experimental results showed that the satisfaction of the last 1/3 of the interaction content was significantly higher than that of the previous interaction content. The current interaction network has only a four-layer structure, which does not satisfy some learners' preferences for Sota because the interaction patterns are small. We should increase the number of layers in the network and add more content and interaction methods. Furthermore, since the weights in the early stages affect the satisfaction of early learner interactions, the speed and rules for updating the weights affect the satisfaction of later interactions.

Acknowledgements

This work was supported by JSPS KAKENHI Grant Number 20H04294 and Photron limited.

References

Akuratiya, D. A. & Meddage, D. N. R. (2020). Students' Perception of Online Learning during COVID-19 Pandemic: A Survey Study of IT Students. International Journal of Research and Innovation in Social Science. 4(9). 755-758.

Baltrusaitis, T., Zadeh, A., Lim, Y. C., and Morency, L., P. (2018). Openface 2.0: Facial behavior analysis toolkit. In 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018). pp.59-66.

Belpaeme, T., Kennedy, J., Ramachandran, A., Scassellati, B. and Tanaka, F. (2018). Social robots for education: A review. Science Robotics. 3(21). DOI: 10.1126/scirobotics.aat595.

Brophy, J. (1981). Teacher praise: A functional analysis. Review of Educational Research, 51(1), pp.5-32. https://doi.org/10.3102/00346543051001005

Cao, Q., Shen, L., Xie, W., Parkhi, O., M., and Zisserman A. (2018). Vggface2: A dataset for recognizing faces across pose and age. In 13th IEEE International Conference on Automatic Face & Gesture Recognition. pp.67-74.

Chiu, T.K.F., Lin, TJ. & Lonka, K. Motivating Online Learning: The Challenges of COVID-19 and Beyond. Asia-Pacific Edu Res 30, 187–190 (2021). https://doi.org/10.1007/s40299-021-00566-w.

Cutumisu, M. and Schwartz, D. L. (2018). The impact of critical feedback choice on students' revision, performance, learning, and memory. Computers in Human Behavior, 78(Jan.). pp.351-367. https://doi.org/10.1016/j.chb.2017.06.029

Dael, N., Mortillaro, M., Scherer, K.R. (2012). The Body Action and Posture coding system (BAP): Development and reliability. Journal of Nonverbal Behavior. 36. pp.97-121.

Fussell, S.R. (Ed.) (2002). The Verbal Communication of Emotion: Interdisciplinary Perspectives; Lawrence Erlbaum Associates: Mahwah, NJ, USA.

Fried, L. (2011). Teaching Teachers about Emotion Regulation in the Classroom. Australian Journal of Teacher Education, 36(3). http://dx.doi.org/10.14221/ajte.2011v36n3.1

FUJISOFT Inc. (2010). PALRO is A robot who cares. Retrieved May 31, 2022. https://palro.jp/en/

Guo, Y., Zhang, L., Hu, Y., He, X., and Gao. J., (2016). Ms-celeb-1m: A dataset and benchmark for large-scale face recognition. In European Conference on Computer Vision. pp.87-102.

Hashimoto, T. Kobayashi, H. (2005). Development of the receptionist system with an anthropomorphism face, Proceedings of the 5th Asian Symposium on Applied Electromagnetics And Mechanics, pp. 190-196.

Hu, J., Shen, L., and Sun, G. (2018). Squeeze-and-excitation networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp.7132-7141.

Huynh, V. T., Kim S. H., Lee G. S., Yang H. J. (2019) Engagement Intensity Prediction with Facial Behavior Features, 2019 International Conference on Multimodal Interaction, pp.567-571, https://doi.org/10.1145/3340555.3355714

Jiang, J., Vauras, M., Volet, S., and Salo, A. E. (2019). Teacher Beliefs and Emotion Expression in Light of Support for Student Psychological Needs: A Qualitative Study. Education Sciences. 9(68).

Lubold, N., Walker, E., and Pon-Barry, H. (2016) Effects of voice-adaptation and social dialogue on perceptions of a robotic learning companion. 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI), pp. 255-262, doi: 10.1109/HRI.2016.7451760.

Leite, I., Martinho, C., Paiva, A. (2013). Social robots for long-term interaction: A survey. International Journal of. Social Robotics. 5, 291–308.

Lu, Y., Chen, C., Chen, P., Chen, X., Zhuang, Z. (2018). Smart Learning Partner: An Interactive Robot for Education. In: , et al. Artificial Intelligence in Education. AIED 2018. Lecture Notes in Computer Science(), vol 10948. Springer, Cham. https://doi.org/10.1007/978-3-319-93846-2_84

Schaeffer CE & Konetes GD (2010). Impact of learner engagement on attrition rates and student success in online learning. International Journal of Instructional Technology & Distance Learning.7(5), pp..3-9.

Schunk, D. H., & Mullen, C. A. (2012). Self-efficacy as an engaged learner. In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), Handbook of research on student engagement (pp. 219–235). New York, NY: Springer. https://doi.org/10.1007/978-1-4614-2018-7

Sharp Corporation. (2016). Robohon. Retrieved May 31, 2022, from https://robohon.com/global/.

Softbank Robotics Co. Ltd. (2018). NAO the humanoid and programmable robot. Retrieved May 31, 2022 from https://www.softbankrobotics.com/emea/en/nao/.

Vstone Co. Ltd. (2010). Social communication robot Sota. Retrieved May 31, 2022 from https://www.vstone.co.jp/products/sota/.

Zheng, X., Hasegawa, S., Tran, MT., Ota, K., Unoki, T. (2021). Estimation of Learners' Engagement Using Face and Body Features by Transfer Learning. In: Degen, H., Ntoa, S. (eds) Artificial Intelligence in HCI. HCII 2021. Lecture Notes in Computer Science(), vol 12797. Springer, Cham. https://doi.org/10.1007/978-3-030-77772-2_36