

Behavior-based Feedback Loop for Attentive E-reading (BFLAe): A Real-Time Computer Vision Approach

Lee, Yoon; Migut, Gosia; Specht, Marcus

Publication date

2023

Document Version

Final published version

Published in

Proceedings of IJCAI-2023 Workshop&Challenge on Micro-gesture Analysis for Hidden Emotion Understanding (MiGA 2023)

Citation (APA)

Lee, Y., Migut, G., & Specht, M. (2023). Behavior-based Feedback Loop for Attentive E-reading (BFLAe): A Real-Time Computer Vision Approach. In G. Zhao, B. W. Schuller, E. Adeli, T. Zhu, & H. Chen (Eds.), *Proceedings of IJCAI-2023 Workshop&Challenge on Micro-gesture Analysis for Hidden Emotion Understanding (MiGA 2023)* Article 12

Important note

To cite this publication, please use the final published version (if applicable). Please check the document version above.

Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights. We will remove access to the work immediately and investigate your claim.

Behavior-based Feedback Loop for Attentive E-reading (BFLAe): A Real-Time Computer Vision Approach

Yoon Lee^{1,*}, Gosia Migut^{1,‡} and Marcus Specht^{1,§}

¹Faculty of Electrical Engineering, Mathematics, and Computer Science, Delft University of Technology

Abstract

This study is built upon a behavior-based framework for real-time attention evaluation of higher education learners in e-reading. Significant challenges in AI model developments for learning analytics have been 1) defining valid indicators and 2) connecting the analytics results to interventions, balancing the generalization and personalization needs. To address this, we utilized a public multimodal WEDAR dataset and trained a neural network model based on real-time features of learners, aiming at predicting learners' moment-to-moment distractions. Real-time features for model training include 30 learners' attention regulation behaviors annotated every second, reaction times to blur stimuli, and page numbers indicating various reading phases. Our preliminary model based on a neural network has achieved 66.26% accuracy in predicting self-reported distractions. Based on the model, we suggest a framework of a Behavior-based Feedback Loop for Attentive e-reading (BFLAe). It has text blur as feedback, a mechanism responsive to learners' distractions that also works as data for next-round feedback. The general feedback implementation rules are established on a statistical analysis conducted on all learners. In addition, we propose a strategy for personalizing feedback using a quartile analysis of individual data, promoting learner-specific feedback. Our framework addresses the high demand for an automated e-learning assistant with non-intrusive data collection based on real-world settings and intuitive feedback provision. The feedback system aims to help learners with longer attention spans and less frequent distractions, leading to more engaging e-reading.

Keywords

Behavior-based Learning Analytics, Neural Networks, E-reading Application, Multimodal Feedback Loop

1. Background

With recent quantitative and qualitative growth in data and computing availability, machine learning approaches are becoming more prevalent in learning analytics and educational data mining [1]. Behavior-based learning analytics is one approach that utilizes cameras and wearable sensors (e.g., eye tracker [2, 3]) to investigate human needs and necessities from their lifestyle, habits, abnormal patterns, and conditions [4]. In learning analytics, machine learning models are often used to predict learning

MiGA@IJCAI23: International IJCAI Workshop on Micro-gesture Analysis for Hidden Emotion Understanding, August 21, 2023, Macao, China.


*Corresponding author.

†These authors contributed equally.

✉ y.lee@tudelft.nl (Y. Lee); m.a.migut@tudelft.nl (G. Migut); m.m.specht@tudelft.nl (M. Specht)

🆔 0000-0002-1913-5672 (Y. Lee); 0000-0001-7116-9338 (G. Migut); 0000-0002-6086-8480 (M. Specht)

© 2023 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)

performances and specific internal states of learners from their affective (e.g., arousal, valence [2, 5]) and cognitive states (e.g., mind-wandering [6, 7], switches of internal thoughts [8]) that are associated with learners' performances and experiences. These approaches are applied to individual-level and group levels [9, 10] for various learning scenarios. Based on real-time action recognition and assessment, most systems aim to form an intervention loop and fundamentally aid learning [11, 12].

Regardless of their accurate prediction capabilities, sensor-based approaches are often criticized for being intrusive [12], changing the nature of learning experiences. Thus, various computer vision-based approaches [13, 8] have been suggested to make learning and system design more seamless for real-world applications. Especially behavior-based analytics is valuable in that particular behavior that machines recognize is also observable and semantically interpretable to humans to some extent [14, 15]. Common challenges in behavior-based machine learning applications in learning analytics have been 1) to find valuable features for model training [14] and 2) to specify the implementation conditions and parameters that best support the accurate recognition of targeted signals [16]. 3) Also, closing the feedback loop, considering generalization and personalization [12] in the analytics phases, and implementing the feedback has been difficult.

In this regard, our objective is to suggest a Behavior-Based Feedback Loop for Attentive e-reading (BFLAe) framework, which involves 1) webcam-based video data collection, 2) computer vision-based learning analytics, 3) blur feedback implementation in text, and 4) further cognitive&behavioral changes of learners as consequences of feedback loop implementation. The framework is built upon a multimodal WEDAR dataset, which provides valuable insight into learners' behavior during e-reading activities. Our approach involves training a neural network model on real-time features that reflect learner behavior, including attention regulation behaviors, reaction times to blur stimuli, and page numbers that reflect different reading phases from the public WEDAR dataset [17]. These features provide a basis for predicting learners' perceived distractions and form a foundation for implementing feedback mechanisms. By implementing the blur feedback on the screen-based e-reader, we aimed to close the feedback loop that enables the further loops, which is not obstructive to the primary reading task and is semantically intuitive. Feedback could potentially help learners reflect on their current state and strategize for future reading [1], which may not be subjectively noticeable to them. The objectives of the behavior-based real-time feedback loop have been 1) extending the overall attention span of learners and 2) reducing the frequency of distractions.

We believe that this personalized, behavior-based feedback loop offers a practical solution to the challenges faced by the fields of Technology-Enhanced Learning (TEL) and Multimodal Learning Analytics (MMLA), promoting more engaging, effective, and individually tailored learning experiences [18]. This article contributes to the ongoing discussion of how best to use technology and learning analytics to support learners. By presenting an innovative framework for an attention regulation behavior-based feedback loop in e-reading, we hope to inspire further research and practical applications of behavior-based models in education.

Our contributions are as stated follows:

- According to our best knowledge, it is the first framework to introduce a real-time feedback loop for attentive e-reading. Our webcam-based behavioral framework is non-obstructive and applicable to diverse e-learning scenarios which involve e-reading as a major learning activity. Our BFLAe framework with increasing digital reading in formal and informal learning with prevalent digital technologies will be more valuable.
- It is a framework built upon WEDAR, a multimodal public dataset collected in an e-reading scenario. It offers more relevant data specified for attention measurement for e-reading. With the implementation details depicted in our framework, the work can be reproduced and further elaborated for specific scenarios based on different tasks and implementation requirements.
- By specifying the statistical values of different behavior labels that represent attentive (i.e., neutral) and distractive (i.e., attention regulation behaviors) learner states, we provide researchers and instructional designers with options to make choices on thresholds for the feedback trigger. As feedback necessities vary depending on the system goals, our analysis result can provide valuable ground for the feedback rules for different systems.

2. Behavior-based Analysis on Multimodal WEDAR dataset

In this section, we briefly analyze the multimodal WEDAR dataset. By doing so, we tried to understand the dataset’s structure and attention regulation behaviors shown in e-reading and potential patterns that are shown together with the self-reported distractions.

2.1. Preliminary analysis on attention regulation behaviors

We used the multimodal WEDAR dataset in our investigation [17]. This dataset comprises human-labeled behavioral labels with five categories of attention regulation behaviors and a neutral behavior as the label, all annotated in every second of the video data. These videos were collected from 30 higher education learners. In particular, this study used real-time distraction reports as the ground truth for distraction instances [19]. As depicted in Figure 1, the distribution of attention regulation behaviors in the dataset is not even. The most common behaviors are body movements, which account for 18.5% of the behaviors, and hand movements, which contribute 12.1% to the duration of the video. The remainder consists of eyebrow movements (3.1%), mumbling (2.6%), and blinking (2.1%). Furthermore, neutral labels, indicating states of attention, constitute 90.9% of the behavioral labels. It is important to note that multiple attention regulation behaviors can co-occur within the same second, so the total proportions do not add up to 100%.

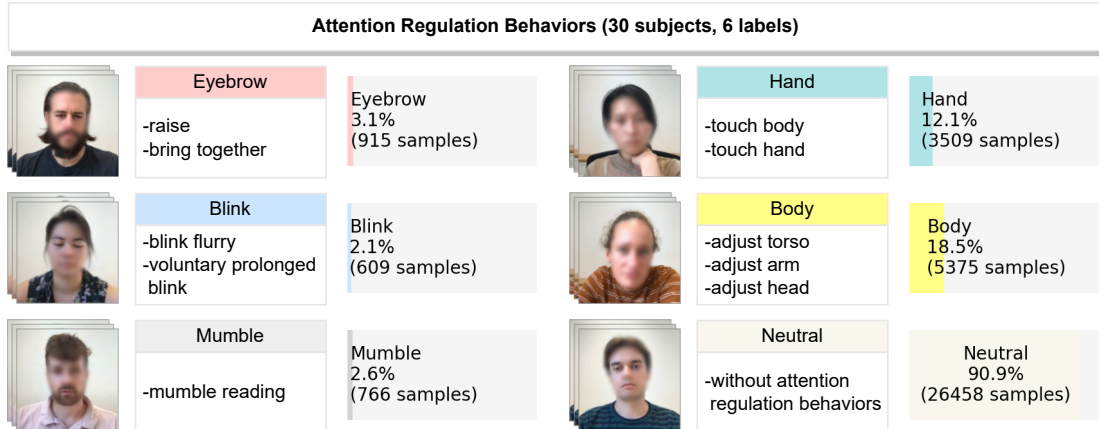


Figure 1: The multimodal WEDAR dataset contains second-to-second annotation labels of attention regulation behaviors and neutral behavior, consisting of varied label proportions.

2.2. Unobservable patterns between attention regulation behaviors and self-reported distractions

We graphically represented the five categories of attention regulation behaviors and neutral behaviors along with distraction reports to discern potential visual patterns between attention regulation behaviors and self-reported distractions. As is evident in Figure 2, participants exhibited a wide range of reading speeds, ranging from 461 seconds (7.7 minutes) to 1661 seconds (or 27.7 minutes). Moreover, we noticed substantial variation in the use of attention regulation behaviors, as well as in the patterns of perceived distractions and the reporting of these distractions. Given this unobservability, the integration of machine learning becomes crucial. It also represents the limitations of human educators in detecting complex patterns hidden within the behavioral patterns of learners.

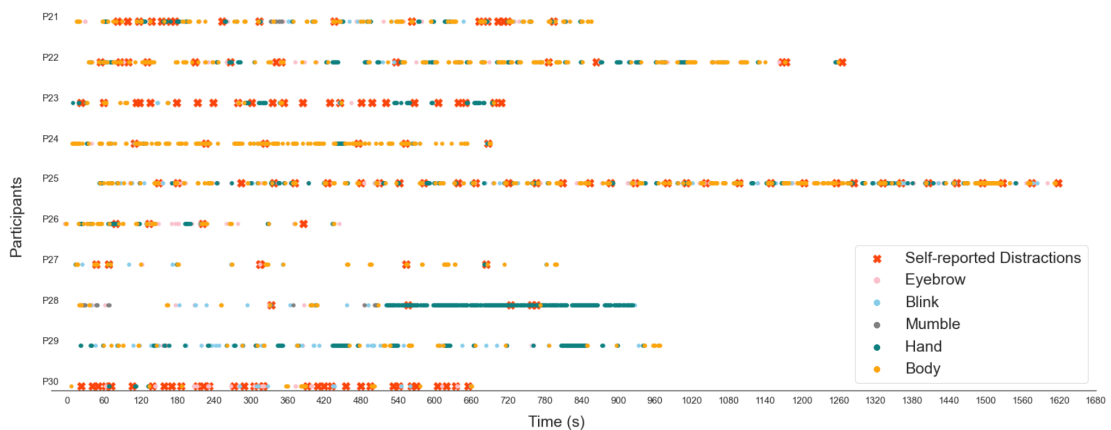


Figure 2: Self-reported distractions and five attention regulation behaviors visualized in time for one-third of all participants (P21 - P30)

3. Framework of Behavior-based Feedback Loop for Attentive E-reading (BFLAe) and its architecture

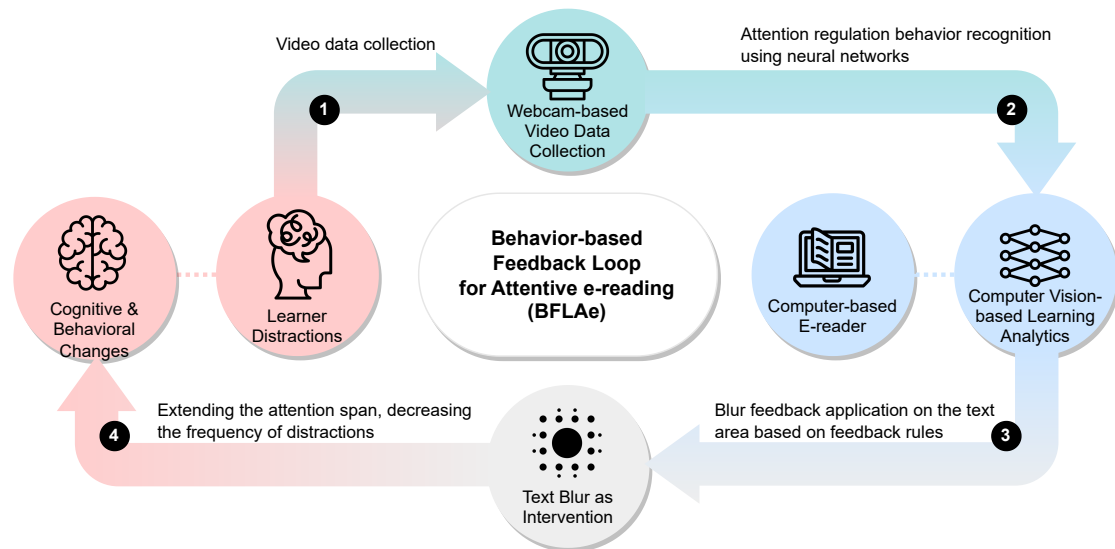


Figure 3: The overall architecture of the Behavior-based Feedback Loop for Attentive e-reading (BFLAe) framework includes four stages: 1) webcam-based video data collection, 2) computer vision-based learning analytics, 3) text blur as intervention, and 4) cognitive&behavioral changes aimed by the feedback.

This section presents the system’s architecture, as shown in Figure 3. Drawing on previous research in the realm of multimodal learning analytics [16, 12], critical factors in forming a multimodal feedback loop for learning include 1) the alignment and integration of data streams, 2) the identification of learning requirements, 3) informed design decisions for multimodal feedback, and 4) the observation of implications within specific learning scenarios. Consequently, we propose a four-stage approach to BFLAe.

3.1. Framework of BFLAe: four stages in system architecture

In the first stage, webcam-based video data is collected during e-reading. This method offers an unobtrusive approach compared to other sensor-based strategies. The second stage involves learning analytics, which is based on a model developed from attention regulation behaviors and self-reported distractions. The following section will detail the specific features used in model training and the rules for triggering system feedback. In the third stage, a blur effect is applied to the reader’s screen for the feedback generation condition, which was decided in the previous phase. The blur effect can be deactivated by the learner clicking on the reading area. This stage not only aids learners by increasing arousal but also serves as additional data for further learning analytics since the reaction time provides crucial cues about the learners’ cognitive states. The final stage of the loop aims to induce cognitive and behavioral changes in learners. Specifically, the system’s

objectives are: 1) extending the attention span between distractions and 2) decreasing the frequencies of distractions, as measured by attention regulation behaviors, reaction speed to the blur stimuli, and self-reported distractions.

4. Behavior-based attention predictions based on Neural Network

This section introduces the features and computational model that we have established to predict attention levels: a prerequisite step integral to the subsequent feedback generation.

4.1. Feature engineering of real-time features

The WEDAR dataset provides behavioral attributes in real-time from 30 higher education learners engaged in e-reading. As referenced in Table 1, eight distinctive features have been harnessed for model training. Five attention regulation behaviors were used as binary features (feature 1) and independent features (features 2-6). Reaction times to secondary blur stimuli, activated at random intervals, have been implemented as another feature (feature 7). Reaction time is a classical measure used to assess learners' arousal levels [8, 20]: shorter reaction time is often interpreted as higher arousal, while a longer reaction time is often considered an indicator of more distractions. The last feature is the specific page number (ranging from 1 to 10) that the learners were on, which represents the reading phases of the learners. For feature engineering, this data was one-hot-encoded (feature 8). It is important to note that we have only extracted real-time features from the dataset. This decision aligns with the feedback loop's objective of a real-time approach.

Table 1

Real-time features have been pre-processed from the multimodal WEDAR Dataset.

#	Feature name	Feature description	Categorical / Nominal
1	Attention_regulation_behavior_binary	Occurrences of any of attention regulation behaviors	0,1
2	Eyebrow_occurrence	Occurrences of movements from eyebrow as attention regulation behavior	0,1
3	Blink_occurrence	Occurrences of movements from blink as attention regulation behavior	0,1
4	Mumble_occurrence	Occurrences of movements from mumble as attention regulation behavior	0,1
5	Hand_occurrence	Occurrences of movements from hand as attention regulation behavior	0,1
6	Body_occurrence	Occurrences of movements from hand as attention regulation behavior	0,1
7	Reaction_time	Reaction time to randomly triggered blur stimuli	Continuous
8	Page_number (one hot encoded)	The page number that learners are currently on	1,2,3,4,5,6,7,8,9,10

4.2. Data pre-processing

We utilized eight real-time features described in Table 1 for our model training. We initially partitioned our dataset into training and testing sets, comprising 80% and 20% of the data, respectively. We balanced the data set, using the synthetic minority oversampling TEchnique (SMOTE) to prevent an imbalance between distracted and attentive states so that neither state would dominate the other in proportion and provide

sufficient data points for the training. Subsequently, we applied min-max normalization to confine the data distribution between 0 and 1. This process was implemented to mitigate any potential bias from different data ranges. Furthermore, min-max normalization is acknowledged for its ability to accelerate training. It is particularly advantageous for our approach, which will have many data points from second-to-second recognition.

4.3. Model training using neural network

As shown in Figure 4, we employ a sequential neural network model with its linear stack of layers. Our network architecture comprises three hidden layers with a rectified linear unit (ReLU) activation function. To mitigate the risk of overfitting, we incorporated a dropout layer into our model, which is widely used for randomly nullifying a fraction of the layer's output features during the training phase. In our case, the dropout layer is configured with a rate of 20%, omitting one-fifth of the input. The final layer of our model is a dense layer with a Sigmoid activation function, with an output range between 0 and 1. It is an optimal choice for our binary classification task. The loss function is designated as mean squared error (mse), the optimization algorithm is set as Adam, and the accuracy is selected as the metric for model evaluation during training. The model has reached an accuracy of 66.26%. This performance exceeds the 50.00% accuracy expected from random guess, which implies that the prediction capacity of the model is considerably better than the chance. The real-world implementation could be enhanced by integrating the feedback rules, which will be further elaborated on in the next section.

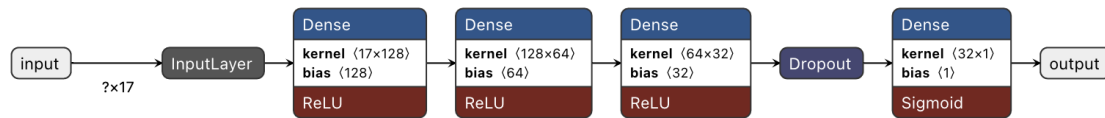


Figure 4: Our model structure, built upon a neural network, consists of one input layer, three hidden layers, one dropout layer, and one output layer.

5. Automatic feedback constructs with visual stimuli

This section introduces the rationale for implementing blur stimuli, feedback rules, and Human-Computer Interaction (HCI) architecture. See Figure 5 for descriptions of HCI, showing the functions of components and blur feedback applied in response to learners' distractions.

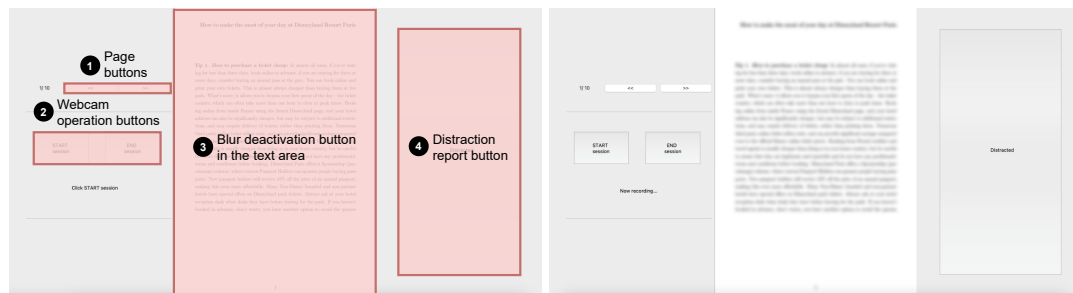
5.1. Type of feedback: blur stimuli

We suggest the implementation of blur on text area as automatic visual feedback, which has also been used to measure reaction time in previous studies [8, 21]. In the following, we introduce the advantages of introducing blur stimuli as part of a feedback loop.

1) The blur stimuli serve a dual function: they trigger the learner's arousal and simultaneously work as data points for future feedback loops. Different reaction times,

behavioral features, and self-distraction reports are incorporated into the screen-based reader as next-round feedback, enabling more precise predictions and personalized feedback.

2) Critics often suggest that feedback interrupts the primary task by adding secondary tasks to learners, inducing cognitive overload [3]. In this context, the interaction between the learner and the system is semantically intuitive and actionable by having a prominently placed deactivation button, where the learners naturally focus during the reading task.



(a) HCI components and functions: page, webcam operation, blur deactivation, and distraction report buttons. (b) Blur feedback is applied to the text area as an intervention triggered by recognized distractions.

Figure 5: Various button functions and blur feedback have been suggested for the screen-based e-reader, which assists the feedback loop. Note that our framework has focused on the feedback mechanisms and the following HCI architecture, not the specific design choices of Graphical User Interface (GUI).

5.2. Feedback implementation rules: statistical analysis on learner behaviors indicating different attentional states

The window size in machine learning refers to the number of data points that are considered to capture information and contexts at each step, which is especially crucial for sequential data processing [4]. We propose tailoring different window sizes to different attention regulation behaviors to enhance the prediction of self-reported distractions. As evidenced in Table 2, derived from the WEDAR dataset, the minimum, maximum, average, median, standard deviations, and quartiles of behaviors exhibit variability of the duration of each state. The current distraction prediction model was designed based on second-to-second labeling for all attention regulation behaviors. However, incorporating different behaviors and applying a range of sliding windows could potentially improve the accuracy of the learners' distraction predictions.

The system's feedback mechanisms can be varied according to its specified objectives. For example, some may apply a window size spanning the third quartile to maximum values of specific behavior for attention prediction. On the contrary, those who require stricter self-regulation among learners may opt to utilize a window size between medium and maximum values for the same task. By establishing specific ranges that act as a foundation for feedback implementation, researchers and educational practitioners will benefit from devising their intervention rules, drawing on general learning behavior.

Table 2

Statistical analysis conducted on durations of each behavior label, collected from 30 participants.

¹Behavior labels have been annotated second-to-second, making the minimum, maximum, median, Q1, Q2, and Q3 values integers.

		Durations (s) ¹							
Attentional States	Behavior Labels	Min	Max	Mean	Median	SD	Q1	Q2	Q3
Attention	Neutral	1.0	124.0	9.44	5.0	12.73	2.0	5.0	12.0
	Eyebrow	1.0	5.0	1.20	1.0	0.52	1.0	1.0	1.0
	Blink	1.0	5.0	1.14	1.0	0.45	1.0	1.0	1.0
Distraction	Mumble	1.0	35.0	3.15	2.0	5.11	1.0	2.0	3.0
	Hand	1.0	62.0	3.58	2.0	4.88	1.0	2.0	4.0
	Body	1.0	44.0	3.20	2.0	2.85	1.0	2.0	4.0

Please note that our analysis has been performed on the WEDAR dataset. Thus, the predefined ranges may undergo further refinement with the accumulation of additional sample data in future studies.

5.3. Considerations for Feedback Personalization: Quartile analysis in individual data

The creation of personalized models can be facilitated by conducting quartile analysis in individual data, considering individual differences in relation to their own unique behavioral status [22]. Quartile analysis offers a way to position specific learners within the broader learner population by distinguishing the first (0% to 25%), second (25% to 75%), and third (75% to 100%) quartiles. This study recommends applying quartile analysis to individual datasets for evaluating learner behaviors and performance. For example, in assessing the reaction time to blur stimuli, each reaction of a single individual can be classified as a fast (1st quartile), medium (2nd quartile), or slow (3rd quartile) response. These categories can also be correlated with high, medium, and low arousal states. Through the accumulation of such data as model features, we can enable the provision of more precise and personalized predictions and feedback provision.

6. Conclusion

We propose a framework of behavior-based feedback loops for attentive e-reading. As established in previous research, the challenge of closing the feedback loop has been a recurring issue in the fields of TEL and MMLA. We leverage the multimodal WEDAR dataset in this work, which aids in developing behavior-based predictions of learners' perceived distractions. Real-time features have been extracted to train a neural network that predicts learners' perceived distractions. These features encompass attention regulation behaviors, reaction time to blur stimuli, and reading phases derived from page numbers. Our approach involves the implementation of blur feedback in response to learners' distractions and establishing the foundation for feedback rules based on the statistical attention regulation behavior analysis derived from general data. Simultaneously, we propose a strategy for personalizing the feedback based on a quartile analysis of individual data. Our behavior-based model addresses the emerging need for an e-reader with automatic learning analytics and feedback mechanisms that can be applied to real-world scenarios.

7. Discussion and Future Work

Optimizing the window sizes of attention regulation behaviors for accurate distraction prediction A statistical analysis of learners' data in e-reading has been performed in the current framework. Broad ranges of learners' attention regulation behaviors have been derived, indicating learners' states of attention and distraction. In future work, several ranges of different behavior recognition technologies will be applied and tested. Doing so will provide practical insights into real-time recognition and feedback generation that can best assist our feedback objectives.

Testing the effects of the automated feedback from an intelligent e-reading system Though the overall behavior-based feedback loop framework has been suggested, the effects of implementing automated feedback still need to be tested: investigating the attention span and frequencies of distractions. Our intelligent system can be further evaluated for subsequent effects, such as learning outcomes and perceived learning experiences, with various qualitative and quantitative measures. Our next step involves comparing the intelligent feedback loop based on the current BFLAe framework and time-based feedback.

Exploring the effects of feedback types and modalities In this work, we suggested blur feedback due to its intuitive actionability and less cognitive load than other feedback. However, with the same feedback timing, we still need to validate whether different types and modalities (e.g., speech-based feedback from conversational agents) of feedback provide additional value in learning. We will further test the effects of varying feedback with various types and modalities built into our current attention recognition mechanisms.

Acknowledgments

This work has been supported by the Leiden-Delft-Erasmus Center for Education and Learning (LDE-CEL).

References

- [1] H. Khosravi, S. B. Shum, G. Chen, C. Conati, Y.-S. Tsai, J. Kay, S. Knight, R. Martinez-Maldonado, S. Sadiq, D. Gašević, Explainable artificial intelligence in education, *Computers and Education: Artificial Intelligence* 3 (2022) 100074.
- [2] J. Li, G. Ngai, H. V. Leong, S. C. Chan, Multimodal human attention detection for reading from facial expression, eye gaze, and mouse dynamics, *ACM SIGAPP Applied Computing Review* 16 (2016) 37–49.
- [3] Y. Lee, M. Specht, Can we empower attentive e-reading with a social robot? an introductory study with a novel multimodal dataset and deep learning approaches, in: *LAK 2023 Conference Proceedings-Towards Trustworthy Learning Analytics-13th*

- International Conference on Learning Analytics and Knowledge, Association for Computing Machinery (ACM), 2023.
- [4] O. Banos, J.-M. Galvez, M. Damas, H. Pomares, I. Rojas, Window size impact in human activity recognition, *Sensors* 14 (2014) 6474–6499.
 - [5] L. Wang, Attention decrease detection based on video analysis in e-learning, *Transactions on Edutainment XIV* (2018) 166–179.
 - [6] R. Bixler, S. DMello, Automatic gaze-based user-independent detection of mind wandering during computerized reading, *User Modeling and User-Adapted Interaction* 26 (2016) 33–68.
 - [7] S. Hutt, K. Krasich, C. Mills, N. Bosch, S. White, J. R. Brockmole, S. K. DMello, Automated gaze-based mind wandering detection during computerized learning in classrooms, *User Modeling and User-Adapted Interaction* 29 (2019) 821–867.
 - [8] M. Xuelin Huang, J. Li, G. Ngai, H. V. Leong, A. Bulling, Moment-to-moment detection of internal thought from eye vergence behaviour, *arXiv e-prints* (2019) arXiv-1901.
 - [9] V. J. Flood, F. G. Amar, R. Nemirovsky, B. W. Harrer, M. R. Bruce, M. C. Wittmann, Paying attention to gesture when students talk chemistry: Interactional resources for responsive teaching, *Journal of Chemical Education* 92 (2015) 11–22.
 - [10] H. Chen, E. Tan, Y. Lee, S. Praharaaj, M. Specht, G. Zhao, Developing ai into explanatory supporting models: An explanation-visualized deep learning prototype, in: *The International Conference of Learning Science (ICLS)*, 2020.
 - [11] Y. Lee, H. Chen, E. Tan, S. Praharaaj, M. Specht, Flower: Feedback loop for group work supporter, in: *The International Learning Analytics and Knowledge Conference (LAK demo session)*, 2020.
 - [12] Y. Lee, G. Migut, M. Specht, Role of multimodal systems in technology enhanced learning (TEL): A scoping review (in press), in: *18th European Conference on Technology Enhanced Learning (EC-TEL23)*, 2023.
 - [13] S. DMello, T. Jackson, S. Craig, B. Morgan, P. Chipman, H. White, N. Person, B. Kort, R. El Kaliouby, R. Picard, et al., Autotutor detects and responds to learners affective and cognitive states, in: *Workshop on emotional and cognitive issues at the international conference on intelligent tutoring systems*, 2008, pp. 306–308.
 - [14] M. Cukurova, Q. Zhou, D. Spikol, L. Landolfi, Modelling collaborative problem-solving competence with transparent learning analytics: is video data enough?, in: *Proceedings of the tenth international conference on learning analytics & knowledge*, 2020, pp. 270–275.
 - [15] H. Chen, H. Shi, X. Liu, X. Li, G. Zhao, Smg: A micro-gesture dataset towards spontaneous body gestures for emotional stress state analysis, *International Journal of Computer Vision* 131 (2023) 1346–1366.
 - [16] D. Di Mitri, J. Schneider, M. Specht, H. Drachsler, From signals to knowledge: A conceptual model for multimodal learning analytics, *Journal of Computer Assisted Learning* 34 (2018) 338–349.
 - [17] Y. Lee, M. Specht, Multimodal WEDAR dataset for attention regulation behaviors, self-reported distractions, reaction time, and knowledge gain in e-reading , 2023. doi:10.4121/8f730aa3-ad04-4419-8a5b-325415d2294b.v1.

- [18] E. H. Mory, Feedback research revisited, in: Handbook of research on educational communications and technology, Routledge, 2013, pp. 738–776.
- [19] M. Faber, R. Bixler, S. K. DMello, An automated behavioral measure of mind wandering during computerized reading, *Behavior Research Methods* 50 (2018) 134–150.
- [20] J. van Kempen, G. M. Loughnane, D. P. Newman, S. P. Kelly, A. Thiele, R. G. O’Connell, M. A. Bellgrove, Behavioural and neural signatures of perceptual decision-making are modulated by pupil-linked arousal, *Elife* 8 (2019) e42541.
- [21] Y. Lee, H. Chen, G. Zhao, M. Specht, Wedar: Webcam-based attention analysis via attention regulator behavior recognition with a novel e-reading dataset, in: Proceedings of the 2022 International Conference on Multimodal Interaction, 2022, pp. 319–328.
- [22] R. D. Goffin, R. B. Jelley, D. M. Powell, N. G. Johnston, Taking advantage of social comparisons in performance appraisal: The relative percentile method, *Human Resource Management: Published in Cooperation with the School of Business Administration, The University of Michigan and in alliance with the Society of Human Resources Management* 48 (2009) 251–268.

8. Online Resources

The WEDAR dataset, which has been used in the work, is available via

- <https://doi.org/10.4121/8f730aa3-ad04-4419-8a5b-325415d2294b.v1>