See Through Them: A Framework for Inferring The Cognitive States of Puzzle Game Players via Eye Gaze

Hao He
haohe@link.cuhk.edu.cn
The Chinese University of Hong Kong, Shenzhen

Yingying She
yingyingshe@xmu.edu.cn
Xiamen University

Yu Jiang
220019020@link.cuhk.edu.cn
The Chinese University of Hong Kong, Shenzhen

Wei Cai
caiwei@cuhk.edu.cn
The Chinese University of Hong Kong, Shenzhen

ABSTRACT

In this paper, we investigate the potential of eye gaze modality in understanding the continuous cognitive states of the players when they are involved in an episode-based puzzle game and how this phenomenon can contribute insights to the game designers to achieve a better QoE of the game. We selected a puzzle game called "Machinarium" as the experimental interface to experiment. We collected the gaze data of the players and inferred their cognitive states in each intersection of decision-making from a game level. The inferred cognitive states were compared to the ground-truth experiences from the players via questionnaire and the official visual guidance extracted from the walkthrough of the game level. The results showed that the implemented framework could infer the cognitive states of the players in a guaranteed accuracy. Besides, the similarities and differences between the players' actual performance and the game level's visual guidance could be the feedback to impact the further optimization of the game design.

CCS CONCEPTS

• **Human-centered computing → Walkthrough evaluations**.

KEYWORDS

user analysis, eye tracking, game design

ACM Reference Format:

Hao He, Yingying She, Yu Jiang, and Wei Cai. 2022. See Through Them: A Framework for Inferring The Cognitive States of Puzzle Game Players via Eye Gaze. In *2nd edition of the Game Systems Workshop (GameSys '22), June 14, 2022, Athlone, Ireland*. ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3534085.3534338

1 INTRODUCTION

Inspired by the features of eye gaze modality, the related applications that attempt to understand the psychophysiological conditions of humans are gaining more and more attention from different domains. Though the psychophysiological research driven by eye gaze modality covers a wide range of domains, the studies about the QoE (Quality of Experience) [Tekinbas and Zimmerman 2003] of puzzle games are still in an early state. Puzzle games usually have rules, where players manipulate game pieces on a grid, network, or other interaction space. The approaches to measuring the QoE in puzzle games can be either objective or subjective. However, both approaches lack the intermediate processes of how the players achieve their outcomes. If the entire experience of the players when they are making decisions can be perceived, it is possible to conduct a more reasonable QoE for the game itself.

Based on the above facts, we intend to investigate the potential of eye gaze modality in understanding the continuous cognitive states of the players in the intersections of decision-making under the context of an episode-based puzzle game. In our study, an *intersection* represents a single puzzle-solving case at a game level. The cognitive states of a player within an intersection may affect the development of the gameplay. Similar to the work done by [Rivu et al. 2019], we assume that the cognitive states of the players can be inferred by analyzing the eye gaze modality. In addition, we expect to extract insightful conclusions from this procedure and offer feedback to the game designers so that a better QoE will be provided in the further game iteration. In order to fulfill our expectations, we propose a framework to address three functions: **a) Puzzle game visual guidance extraction:** extract the visual guidance from the official walkthroughs of the game levels; **b) Player cognitive state inference:** collect the gaze data of the players and infer their cognitive states in each intersection of decision-making; **c) Game level QoE estimation:** compare the inferred cognitive states of the players to their actual game experience as well as the visual guidance of the game level for the generation of QoE conclusions.

2 RELATED WORK

Based on the features of eye gaze, a great number of applications have been brought out with the matching implementation methodologies within multiple fields. In the domain of psychological investigation, gaze data is collected and analyzed to verify a theory or an idea that the human factor is the studied object, including the behavioral patterns and human states [Shimonishi 2016]. Chen [Eckstein 2011] exploited that eye tracking can help to learn the habit of the elder to interact with the textures and improve the performance of texture messaging. Liu et al. [Lim et al. 2013] developed a system to examine how partners with mismatched visual perceptual capabilities collaborate to accomplish joint tasks. The

GameSys '22, June 14, 2022, Athlone, Ireland

ACM ISBN 978-1-4503-9381-2/22/06…$15.00
https://doi.org/10.1145/3534085.3534338

result showed that, partners with mismatched perceptions were willing to collaborate. Bal et al. [Bal et al. 2010] tried to figure out the relations between children with autism spectrum disorder and their emotion recognition abilities by analyzing their eye gaze patterns while observing different emotional expressions. Bader and Beyerer [Bader and Beyerer 2013] revealed that analyzing natural gaze behavior can infer the user's intention or experience for designing proactive or adaptive intelligent user interfaces. Moreover, gaze-based intention estimation is valuable for compensating for the inaccuracy of imprecise hand gestures. Steichen et al. [Salous et al. 2018] presented that gaze data which mainly was a user's saccade angles and fixation durations, can be used to infer a user's cognitive style during information visualization usage with up to 86% accuracy. Based on the features of eye gaze modality, we believe that it is possible to apply it to infer the cognitive state of the player during puzzle gameplay.

3 TECHNICAL APPROACH

As shown in Figure 1, to realize the vision of offering the player a better QoE, we will first acquire the visual guidance of the game level along with the gaze data of the player. Next, the actual game experience of the players is acquired to verify the effectiveness of the cognitive state inference method to confirm our hypothesis. Finally, the feedback to the designers is generated based on QoE estimation that combines all the acquired information. In the following subsections, we will discuss the details of the framework.

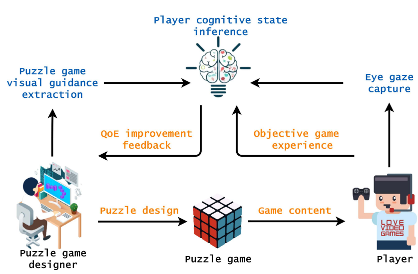

Figure 1: An eye gaze-based framework for inferring the cognitive states of puzzle game player.

3.1 Puzzle game visual guidance extraction

The visual guidance in a puzzle game conveys the instructions and clues that the designers intend to show to the players so that he/she can pass through the game level in a predesigned way. The conspicuousness of the visual guidance can directly affect the QoE of the game level. A game level usually requires more than one puzzle to be solved to pass. Each puzzle can be treated as an intersection of decision-making for the players to perform the correct operations with the correct items. Visual guidance is the collection of intersections set in a specific order. In our framework, we study the pattern of episode-based puzzle gamesconclude the key components of an intersection as **goal**, **key item**, and **operation**.

3.2 Player cognitive state inference

The key problem of this function is to collect the gaze data and formulate a method to map the data to specify the cognitive state [Putnam 2013]. When involved in the context of the episode-based puzzle game, the cognitive states of the players are typically driven by the activity of problem-solving, which is meant to find out the solutions for each puzzle. By referring to the theory of visual problem-solving [Goldschmidt 1992], the appear categories of the eye movement are usually **fixation** and **saccade**. By referring to the key components of visual guidance, we define four categories of cognitive states for the players as **focus**, **distracted**, **aware** and **confused**.

During each intersection of puzzle-solving, all the collected gaze data is treated as a judging unit to infer the cognitive states of the players. In the inference process, we apply a 1D CNN-BLSTM model [Startsev et al. 2018] to classify the eye movements of the players based on the collected gaze data, and infer their corresponding cognitive states. There are four input features for the eye movement prediction, including time (in microseconds), x and y (the on-screen coordinates), and the confidence for the tracking of the subjects' eyes. The inference steps can be summarized as follows: we first acquire the coordinates of the gaze data e, then input them to the 1D CNN-BLSTM model to determine the category of eye movement m (0 as **fixation** and 1 as **saccade**). After we obtain m, the offset o between the central point of the gaze point and the central point of the bounding box of the key item is calculated, t represents the threshold to determine the category of the basic cognitive state c_b (0 as **focus** and 1as **distracted**). Finally, the sum of the two cognitive states sum_f and sum_d during an intersection will be calculated. The greater one between sum_f and sum_d will determine the player's overall cognitive state c_o. When sum_f is greater, then c_o will be 2 as **aware**. In the other case, c_o will be 3 as **confused**.

3.3 QoE estimation

We attempt to seek clues from the comparisons among the inferred cognitive states, the players' experience, and the official visual guidance offered by the game designers. The first comparison is between the inferred cognitive states and the players' experience. A questionnaire asks players about their subjective experience for the game level. The players' actual game experience is utilized to verify our proposed cognitive state inference method. The comparison results can reveal whether eye gaze modality is potential for cognitive state inferring in episode-based puzzle games. The second comparison is between the actual game experience of the players and the official visual guidance offered by the designers. Based on the answers to the questionnaire, we can distinguish the effectiveness of the visual guidance in each intersection. The results can reveal whether the visual guidance impacts the players' game experience originally. This generated feedback can be the reference for the designer to improve the QoE in the further iteration of the game design.

4 EXPERIMENT

We employed an episode-based puzzle game called "Machinarium" as the experimental context in the experiment. "Machinarium" requires a series of continuous decision-making from the player to solve the puzzles. Players use a mouse to interact with the game element to help the character pass through the scene. In order to solve the puzzle, players have to click, drag and combine the interactive items.

We applied a Dell Workstation Precision T3630 PC equipped with a mouse and keyboard and a 1920*1080 resolution AOI monitor to display the game scene and perceive the input from the player. At the same time, a Tobii Eye Tracker 4C (60 HZ tracking frequency) was deployed to collect the eye gaze input from the player. We employed 11 volunteers, 4 males, and 7 females, from 18 to 21 years old, to naturally play the game level. All the participants had never played the game before.

The participants were asked to play level one of "Machinarium" by following the game's original instructions. The gaze data collection program was developed in the platform of Unity 3D with the support of Tobii Eye Tracking API. OBS Studio recorded the gameplay process to specify the duration of each intersection. After the gameplay, participants were asked to answer the prepared questionnaire about their actual game experience, including: **Q1:** *How do you rank the difficulty of each intersection?* **Q2:** *How do you solve the puzzle of each intersection?* **Q3:** *What is the experience when you are trying to solve the puzzles?* **Q4:** *Is the visual guidance clear enough to follow within each intersection? If not, where is the unclear part?*

We manually obtained key items of each intersection. As shown in Figure 2, the bounding boxes with different colors and numbers represent the visual guidance of the corresponding puzzle. There were five intersections on level one of "Machinarium." After the experiment procedure, we obtained the gaze data of the players as well as their corresponding screen records of the gameplay. The data was fed to the 1D CNN-BLSTM model first to classify the eye movements of the players. After the model implementation, the eye movement results and the visual guidance of the game level were adopted to infer the cognitive states of the players. The thresholds to determine the "aware" cognitive state for different intersections were set as the average values of the width and the length of the bounding boxes of the key items.

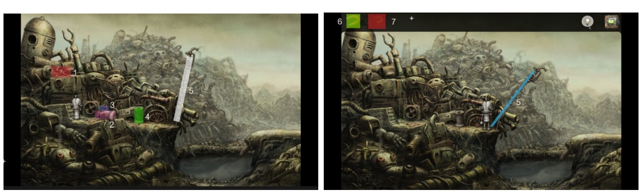

Figure 2: The key items of level one

5 RESULT AND DISCUSSION

5.1 Cognitive state inference result

For intersections 1 and 5, most of the players' cognitive states were inferred as **confused**. For intersections 2, 3, and 4, most players were inferred as **aware**. According to the results, we assumed that the visual guidance of intersections 1 and 5 were relatively "unclear" and the ones of intersections 2, 3, and 4 were relatively "clear;" The puzzle difficulties of intersections 1 and 5 were relatively "difficult" and the ones of intersection 2, 3 and 4 were relatively "easy." However, the final result had to be confirmed by comparing it to the actual game experience of the players.

5.2 Discussion

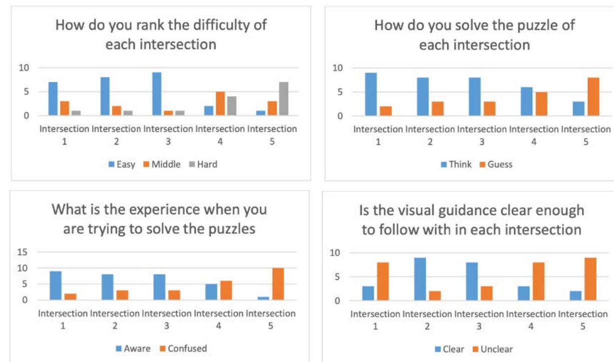

Figure 3: The statistical answers of the questionnaire

In Figure 3, we illustrated the answers collected from the players. The feedback shown that it revealed the same situations as the inferred cognitive states except intersection 1. To figure out the reason, we made a detailed interview to one of the players. From the interview, we figured out the reasons why the conflict happened. The first reason was overmuch visual content at the beginning of the scene, which might distract the player's attention; the second one was that there was no visual guidance about the "lift" operation for the robot. In conclusion, our framework can infer the cognitive states of the player in a preliminary manner compared to the actual user experience. When the inferred cognitive states and the official visual guidance match each other, then the clearness of the visual guidance can be assured. When conflicts are found between the inferred cognitive states and the official visual guidance, we can refer to the gaze data of the player to figure out the exact cognitive path they walk through before the final decision-making. The visualization of the gaze data can be the clue to identify the actual middle process of the players' experience towards the expression of the game level. Hence, the game designers can adjust the strategy to convey their ideas to provide better QoE for the players.

6 CONCLUSION

In this paper, we investigated the potential of eye gaze modality in inferring the cognitive states of the players in a puzzle game called "Machinarium." We design a framework that offers a series of approaches to extract the game level's visual guidance, infer the player's cognitive states, and estimate the QoE of the game design. The experiment results show that our proposed framework can infer the cognitive states of the player compared to his/her game experience. Besides, we successfully generated feedback to the designers based on estimation results upon the framework to create a better QoE for the players in the future.

7 ACKNOWLEDGEMENT

This work was supported by Shenzhen Science and Technology Program (Grant No. JCYJ20210324124205016).

REFERENCES

Thomas Bader and Jürgen Beyerer. 2013. Natural gaze behavior as input modality for human-computer interaction. In *Eye Gaze in Intelligent User Interfaces*. Springer, 161–183.

Elgiz Bal, Emily Harden, Damon Lamb, Amy Vaughan Van Hecke, John W Denver, and Stephen W Porges. 2010. Emotion recognition in children with autism spectrum disorders: Relations to eye gaze and autonomic state. *Journal of autism and developmental disorders* 40, 3 (2010), 358–370.

Miguel P Eckstein. 2011. Visual search: A retrospective. *Journal of vision* 11, 5 (2011), 14–14.

Gabriela Goldschmidt. 1992. Serial sketching: visual problem solving in designing. *Cybernetics and System* 23, 2 (1992), 191–219.

Kai Keat Lim, Max Friedrich, Jenni Radun, and Kristiina Jokinen. 2013. Lying through the eyes: detecting lies through eye movements. In *Proceedings of the 6th workshop on Eye gaze in intelligent human machine interaction: gaze in multimodal interaction*. 51–56.

Hilary Putnam. 2013. 17. The Nature of Mental States. In *Readings in Philosophy of Psychology, Volume I*. Harvard University Press, 223–231.

Sheikh Rivu, Yasmeen Abdrabou, Thomas Mayer, Ken Pfeuffer, and Florian Alt. 2019. GazeButton: enhancing buttons with eye gaze interactions. In *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications*. 1–7.

Mazen Salous, Felix Putze, Tanja Schultz, Jutta Hild, and Jürgen Beyerer. 2018. Investigating static and sequential models for intervention-free selection using multimodal data of EEG and eye tracking. In *Proceedings of the Workshop on Modeling Cognitive Processes from Multimodal Data*. 1–6.

Kei Shimonishi. 2016. Modeling user's decision process through gaze behavior. In *Proceedings of the 18th ACM International Conference on Multimodal Interaction*. 536–540.

Mikhail Startsev, Ioannis Agtzidis, and Michael Dorr. 2018. 1D CNN with BLSTM for automated classification of fixations, saccades, and smooth pursuits. *Behavior Research Methods* (08 Nov 2018). https://doi.org/10.3758/s13428-018-1144-2

Katie Salen Tekinbas and Eric Zimmerman. 2003. *Rules of play: Game design fundamentals*. MIT press.