Supporting Gaze-Based Interaction in a Visually Rich and Dynamic Environment with Machine Learning: An Online Study

Yulia G. Shevtsova
Moscow State University of Psychology and Education
Moscow, Russian Federation
M.V. Lomonosov Moscow State University
Moscow, Russian Federation
shevtsova.jg@gmail.com

Artem S. Yashin
Moscow State University of Psychology and Education
Moscow, Russian Federation
yashinart1996@gmail.com

Sergei L. Shishkin
Moscow State University of Psychology and Education
Moscow, Russian Federation
sergshishkin@mail.ru

Anatoly N. Vasilyev
Moscow State University of Psychology and Education
Moscow, Russian Federation
M.V. Lomonosov Moscow State University
Moscow, Russian Federation
a.vasilyev@anvmail.com

Abstract

During gaze-based interaction, gaze provides both control and visual input. Although in simple tasks, like eye typing, these functions are separated, more complex scenarios can lead to misinterpretation of user intent. In our study, we explored if machine learning (ML) can aid in solving this problem. 15 participants played a gaze-controlled game, where they could freely select screen objects with a 500 ms dwell time. By applying ML to gaze features and contextual information, we achieved a threefold reduction in false positives. This study is the first to show how ML can enhance gaze-based interaction in visually demanding environments.

Keywords

Eye tracking, Gaze-based interaction, Midas touch, Machine learning, Intention recognition

ACM Reference Format:
Yulia G. Shevtsova, Artem S. Yashin, Sergei L. Shishkin, and Anatoly N. Vasilyev. 2025. Supporting Gaze-Based Interaction in a Visually Rich and Dynamic Environment with Machine Learning: An Online Study. In *2025 Symposium on Eye Tracking Research and Applications (ETRA '25), May 26–29, 2025, Tokyo, Japan.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3715669.3726799

1 Introduction

Gaze-based interaction is increasingly used but remains prone to the Midas touch problem [Jacob 1990], where using gaze for both perception and control results in unintended actions. Solutions like longer dwell times or additional confirmation via saccades or blinks reduce fluency and increase effort [Majaranta et al. 2019]. Machine learning (ML) has recently been used to predict user intent in gaze-based interaction [Isomoto et al. 2022]. Unfortunately, the reported performance may be biased, as the target search task used in this study typically induce longer dwells and greater pupil dilation than non-target viewing [Jangraw et al. 2014].

An effective test of gaze-based control should allow natural behavior in dynamic, visually rich environments and support class labeling without disrupting gameplay — criteria met by the EyeLines game [Shishkin et al. 2016]. We previously evaluated an ML approach for intent prediction during this game in offline simulation [Shevtsova et al. 2023]. In the current study, we enhanced intention recognition by incorporating a context-based algorithm, shown to improve interaction (e.g., autocomplete in gaze typing). We tested the combined classifier in a real-time application and evaluated whether the ML approach improves gaze-based interaction.

2 Methods

Participants: Data were collected from 15 naïve healthy volunteers (age: 25 ± 6 years, mean ± SD) who provided informed consent. All experimental procedures conformed to the Declaration of Helsinki and were approved by the local ethical committee.

Task: Participants played the gaze-controlled EyeLines game [Shishkin et al. 2016], selecting colored balls and setting their new positions on a 7×7 grid to form lines of the same color. Completed lines disappeared; otherwise, new balls were randomly added. The game ended when the board filled or after 8 minutes.

Gaze-based control: Eye tracking was performed at 1000 Hz using the EyeLink 1000 Plus. A selection was triggered when gaze remained within 2.3° for at least 500 ms – a behavior referred to as a "dwell," which could include multiple fixations – with the dwell center falling within 1.3° of the target's center. Two gaze-based control modes were tested: Mode D (all dwells triggered actions) and Mode C (a classifier decided if they were relevant).

Experimental design: Participants played three games in each mode on both days. On the first day (familiarization), Mode D was

ETRA '25, Tokyo, Japan

ACM ISBN 979-8-4007-1487-0/25/05
https://doi.org/10.1145/3715669.3726799

always given first. On the second day (performance evaluation and mode comparison), the order of modes was randomized.

Classification algorithm: In Mode C, dwells were classified based on the averaged probability p from two trained models: a gaze classifier and a contextual classifier. Dwells with p > threshold 1 were deemed as intentional and triggered ball selection, while those with p < threshold 2 were considered spontaneous and ignored. Approximately 30% of dwells fell between the thresholds and were labeled as "uncertain"; these triggered an action only if the dwell persisted until a 700 ms threshold was reached.

Classifier training: Support Vector Machine models with an RBF kernel were used. On the first day, models were trained on prior study data [Shevtsova et al. 2023]. Individual models were then trained on the collected data with randomly balanced classes and applied on the second day.

Gaze classifier: Gaze micro-behavior features, as described in [Shevtsova et al. 2023], were utilized. Features included coordinate variance and spread, microsaccade count and amplitude (all in overlapping 50 ms windows), and distance to the nearest ball (non-overlapping 50 ms). The Recursive Feature Elimination algorithm was employed to select the top seven features for each model.

Contextual classifier: Contextual classifier features captured a ball's position relative to others on the field. Based on 50,000 moves from another prior EyeLines study [Vasilyev et al. 2024], we manually identified features that could influence the likelihood of a player selecting a particular ball: some features (14) indicated a ball's role in forming or enabling same-colored lines, while others (7) reflected the general ball mobility on the board.

Ground truth: The true dwell labels were inferred from actions: if a selected ball was moved immediately after selection, the dwell was classified as intentional; if not, it was deemed spontaneous.

Performance assessment: All analyses utilized second-day data. Classification performance was evaluated using the True Positive Rate (TPR), True Negative Rate (TNR), and Balanced Accuracy (BA). Gaze control effectiveness was assessed by the command rate (the number of intentional moves per minute), the number of actions required to remove one ball (including ball selection, movement, deselection, or cancellation of an unintended move) and the total game time (a percentage of the 8-minute game duration).

3 Results

Classifier performance: To assess whether the combined classifier outperformed individual models, a Friedman test with post-hoc Dunn's correction was applied to offline results. As shown in Figure 1 (a, b), both the gaze and contextual classifiers performed worse than the combined model (g+c), which achieved higher true positive rates, reduced false negative errors, and maintained a lower false positive rate, as evidenced by ROC curves (g+c vs. gaze: $\chi^2(2)=-21$, p<0.001; g+c vs. contextual: $\chi^2(2)=-24$, p<0.0001).

Efficiency of the gaze-based control: Mode C reduced false positives nearly threefold compared to Mode D (Wilcoxon test for FP/(FP+TP): W(15)=120.0, p<0.0001). Command rate remained similar across both modes (Student's t-test: t(14)=1.08, p=0.30). Mode C enabled participants to play longer (Wilcoxon test: W(13)=-73.0, p=0.0078) and required fewer actions to remove the same number of balls (W(15)=110.0, p=0.0006), as shown in Figure 1 (c).

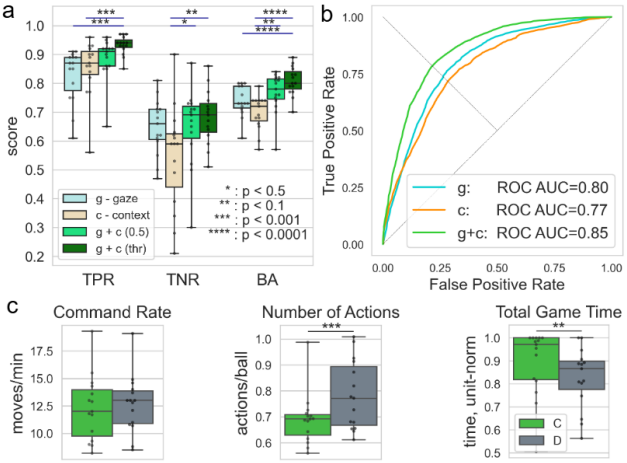

Figure 1: Classifier metrics (a), ROC curves (b) and Gaze control effectiveness (c). g+c (0.5): fixed thresholds; g+c (thr): adjustable thresholds. **p<0.01, ***p<0.001.

4 Discussion

This study demonstrates that ML applied online can enhance both the effectiveness and efficiency of gaze-based interaction in realistic settings. The EyeLines game served as a testbed where gaze was used intensively for both input and control, natural behavior was allowed, and performance was assessed without explicit labeling.

We introduced a combined algorithm that integrates gaze and contextual classifiers. While context features were game-specific, they are adaptable to other tasks where user behavior patterns can be observed and formalized. Averaging the classifiers' probabilities yielded better performance than either model alone. In low-confidence cases — opposing predictions or values near 0.5 — a longer time threshold was applied to reduce errors.

ML-enhanced Mode C improved interaction, with longer game durations and fewer actions per ball removal than dwell-only Mode D. Presumably, reducing unintended ball selections allowed players to make more deliberate moves. Although the command rate did not increase, likely because the dwell time in Mode D was already optimized for quick selections, the improvements in Mode C remain both significant and promising.

Gaze-based interaction systems pose potential privacy risks; however, all processing can be performed locally, and data may be discarded after classifier training, minimizing exposure.

Our findings highlight that ML can help address the Midas touch problem in gaze-based systems.

Acknowledgments

This work was supported by the Russian Science Foundation (grant number 22-19-00528).

References

Toshiya Isomoto, Shota Yamanaka, and Buntarou Shizuki. 2022. Dwell selection with ml-based intent prediction using only gaze data. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 3 (2022), 1–21.

Robert JK Jacob. 1990. What you look at is what you get: eye movement-based interaction techniques. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 11–18.

David C Jangraw, Jun Wang, Brent J Lance, Shih-Fu Chang, and Paul Sajda. 2014. Neurally and ocularly informed graph-based models for searching 3D environments. *Journal of neural engineering* 11, 4 (2014), 046003.

Päivi Majaranta, Kari-Jouko Räihä, Aulikki Hyrskykari, and Oleg Špakov. 2019. Eye movements and human-computer interaction. *Eye movement research: An introduction to its scientific foundations and applications* (2019), 971–1015.

Yulia G Shevtsova, Anatoly N Vasilyev, and Sergei L Shishkin. 2023. Machine Learning for Gaze-Based Selection: Performance Assessment Without Explicit Labeling. In *International Conference on Human-Computer Interaction*. Springer, 311–322.

Sergei L Shishkin, Yuri O Nuzhdin, Evgeny P Svirin, Alexander G Trofimov, Anastasia A Fedorova, Bogdan L Kozyrskiy, and Boris M Velichkovsky. 2016. EEG negativity in fixations used for gaze-based control: Toward converting intentions into actions with an eye-brain-computer interface. *Frontiers in neuroscience* 10 (2016), 528.

Anatoly N Vasilyev, Evgeniy P Svirin, Ignat A Dubynin, Anna V Butorina, Yuri O Nuzhdin, Alexei E Ossadtchi, Tatiana A Stroganova, and Sergei L Shishkin. 2024. Intentionally vs. Spontaneously Prolonged Gaze: A MEG Study of Active Gaze-Based Interaction. *bioRxiv* (2024), 2024.12. 11.627776.