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

Yulia G. Shevtsova[1, 2], Artem S. Yashin[1], Sergei L. Shishkin[1] and Anatoly N. Vasilyev[1, 2]

1 – MEG Center, Moscow State University of Psychology and Education, Moscow, Russia

2 – Department of Human and Animal Physiology, M.V. Lomonosov Moscow State University, Moscow, Russia

web: bci.megmoscow.ru   email: shevtsova.jg@gmail.com

For more details:

Shevtsova et al. IEEE Access, 9 May 2025

Introduction

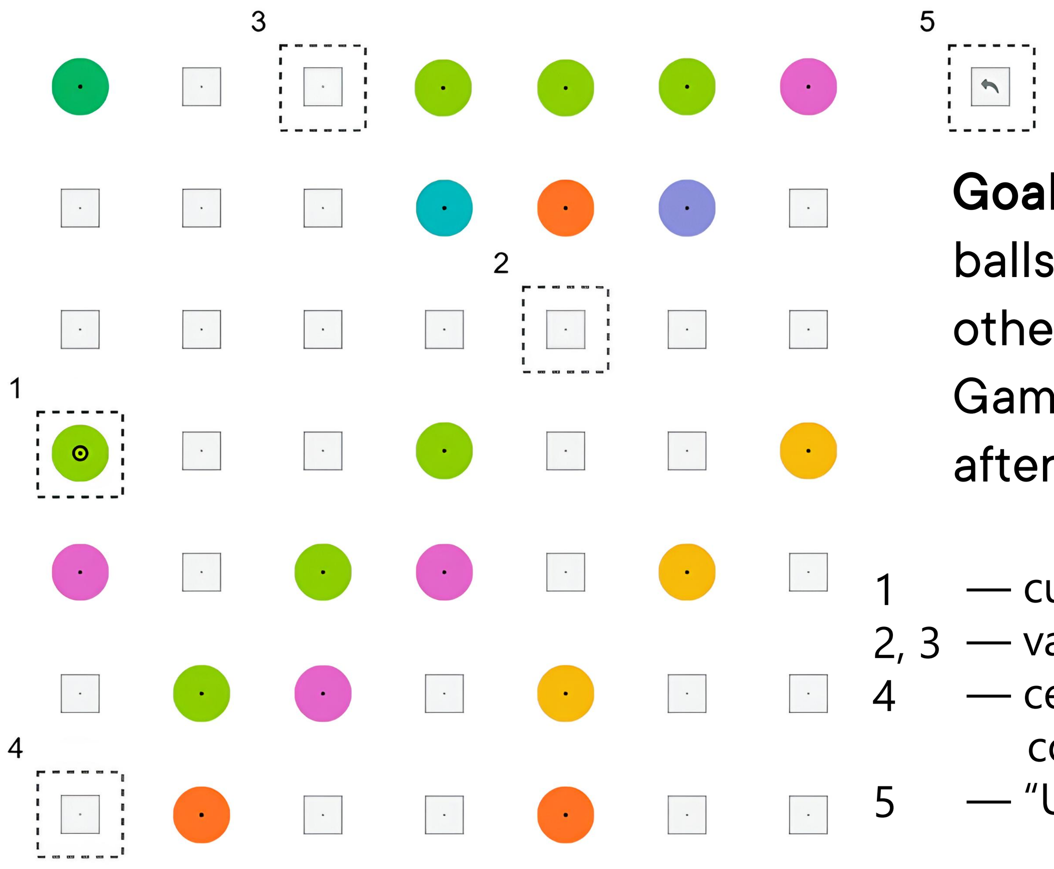
Gaze-based interfaces let users operate devices with their eyes, but they suffer from the classic Midas-Touch problem (Jacob 1990): whatever you look at gets selected.

This stems from **the dual role of gaze** as both visual input and a means of control. As a result, natural eye movements are often misinterpreted as commands, producing unintended selections and degrading usability.

A promising mitigation strategy **is to predict user intent** with machine-learning (ML) algorithms, yet such approaches have rarely been evaluated under conditions that allow fully natural gaze behavior.

Hypothesis: ML-assisted gaze control will be more efficient and preferred over the baseline dwell-time mode where every detected dwell triggers an action.

EyeLines Game (Shishkin et al. 2016) provided dynamic, rich environment and supported class labeling without disrupting gameplay.



Game rules

Goal: form same-colored lines of 4+ balls. Completed lines disappeared; otherwise, random balls were added. Game ended when the board filled or after 8 minutes.

1 — currently selected ball
2, 3 — valid spots that would complete a line
4 — cell blocked by the diagonal-move constraint
5 — "Undo last move" button

Methods

Participants: 15 naïve healthy volunteers.

Gaze-based control: EyeLink 1000 Plus at 1000 Hz.

Mode D | basic

A dwell was registered when gaze remained within 2.3° for ≥ 500 ms. It triggered selection if its centroid lied ≤ 1.3° from a ball's center.

Mode C | ML-enhanced

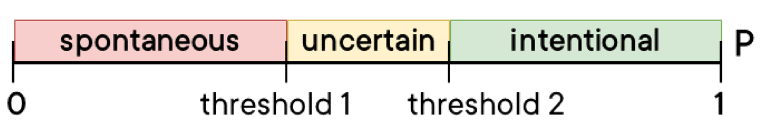
A selection was triggered only when the classifier marked the dwell as control-relevant.Two participant-specific SVM models with an RBF kernel were trained on datasets in which the classes were randomly balanced.

Gaze micro-behavior features: coordinate variance & spread, microsaccade count & amplitude (all in overlapping 50 ms windows), and distance to the nearest ball (non- overlapping 50 ms).

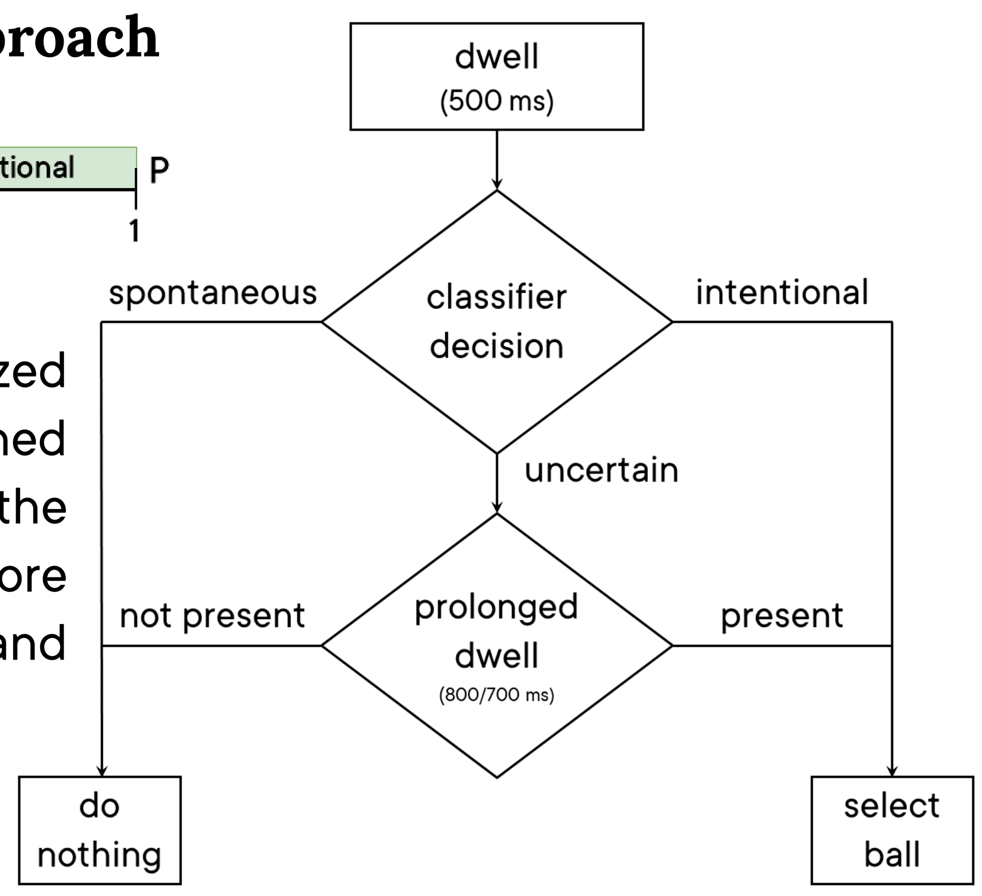
Contextual features encoded each ball's spatial context: 14 attributes captured its potential to extend same-colored lines, while 7 described its general freedom of movement on the field.

Ground-truth labeling primarily followed this rule: when the selected ball moved, the dwell was tagged intentional; when no movement followed, it was tagged spontaneous.

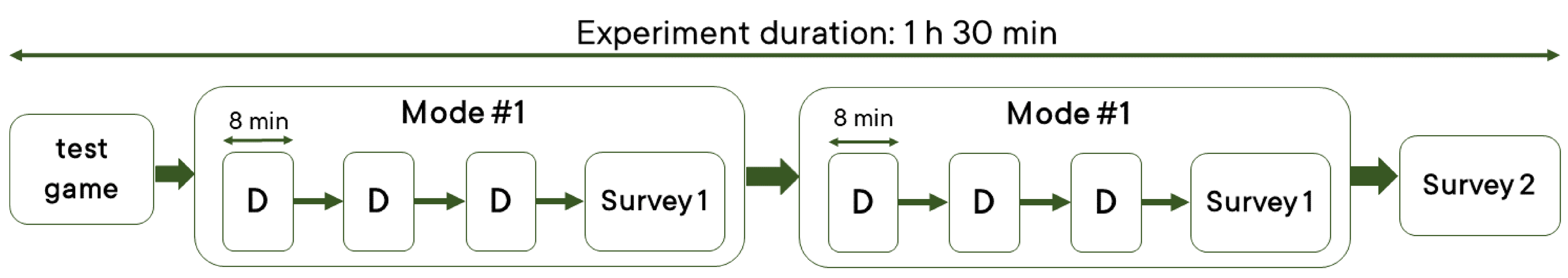
ML-approach



Dwells were categorized with two participant-tuned thresholds applied to the mean probability score output by the gaze and context classifiers.



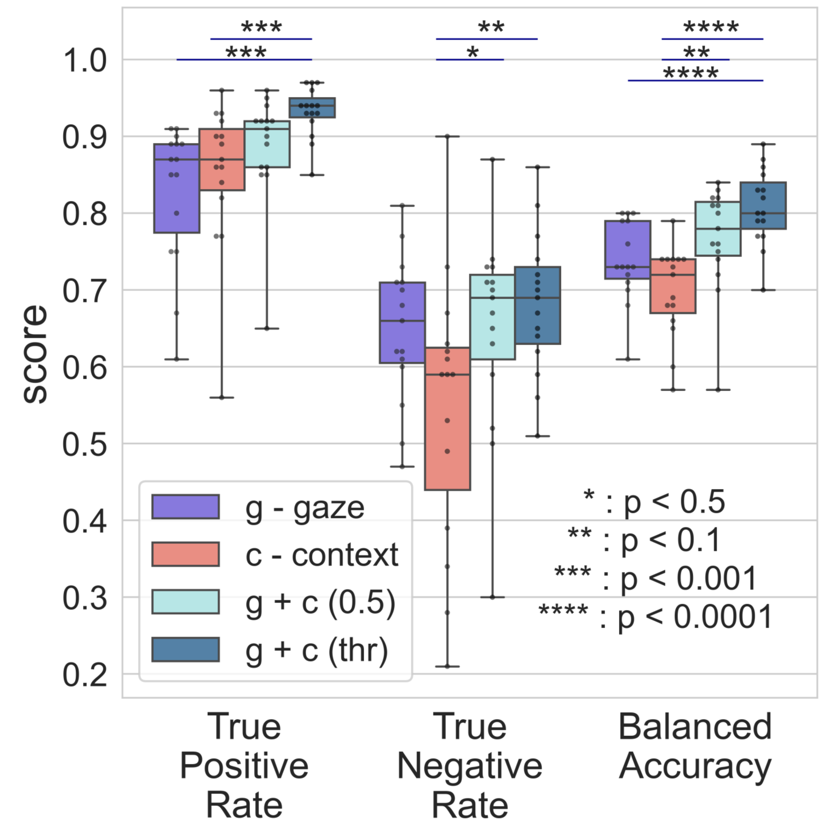
Experimental design:

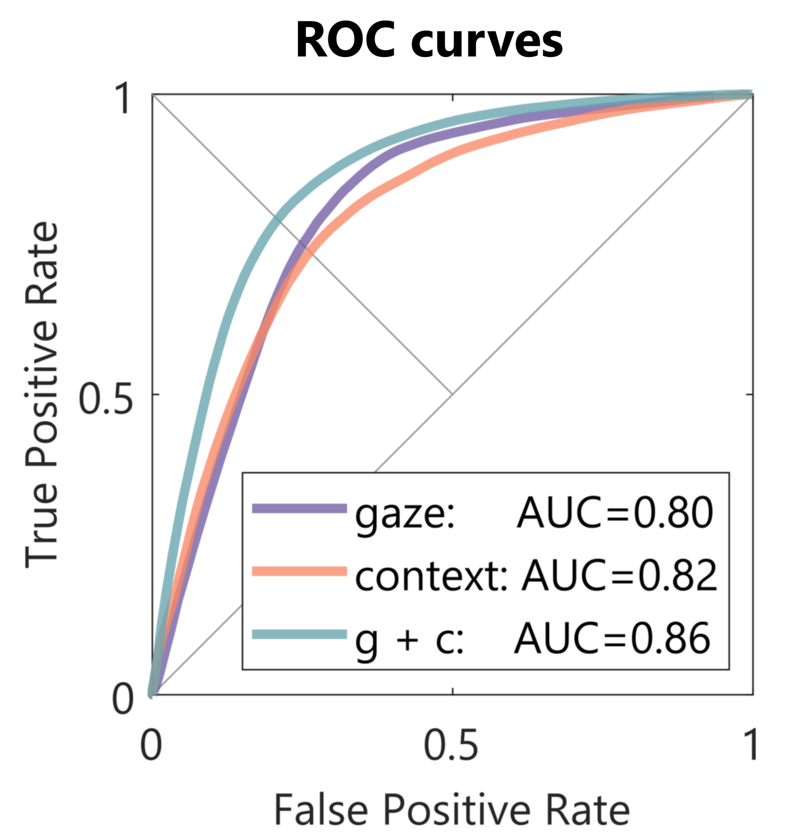


Survey 1: Perceived mode qualities; *Survey 2:* Mode comparison

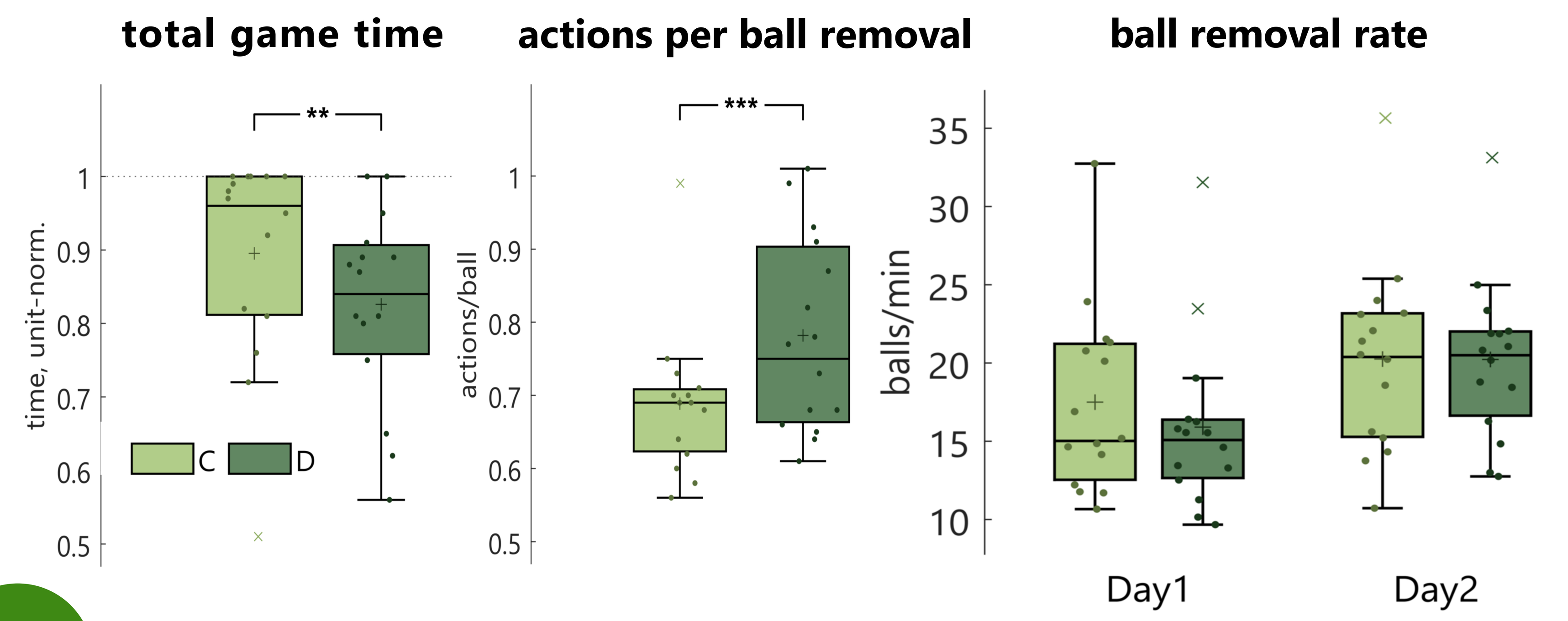
Results

Classifier performance





Game performance



Conclusions

- Online ML made gaze-based control both more accurate and more efficient, cutting unintended selections to one-third of their original number.
- Testing took place in the *EyeLines* game, providing realistic conditions.
- The joint gaze-plus-context (g + c) model outperformed either the gaze-only or context-only classifier.
- With fewer accidental clicks, players executed a higher proportion of deliberate moves.
- The overall removal rate stayed roughly constant across input modes, suggesting it was limited more by individual player skill than by the control method itself.
- In contrast to Isomoto et al. (2022), command rate did not rise with ML assistance – probably because Mode D's dwell-time threshold was already tuned for rapid selections.
- Taken together, these results show that ML can meaningfully alleviate the Midas Touch problem in gaze-controlled interfaces.

References

ISOMOTO et. al. 2022. Dwell selection with ml-based intent prediction using only gaze data. ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 6, 3. 1–21.

JACOB 1990. What you look at is what you get: eye movement-based interaction techniques. SIGCHI conference.

SHISHKIN et. al.. 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, 528.