IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 7 February 2025, accepted 5 May 2025, date of publication 9 May 2025, date of current version 16 May 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3568650

RESEARCH ARTICLE

Disentangling High-Paced Alternating I/O in Gaze-Based Interaction

YULIA G. SHEVTSOVA [1,2], **ARTEM S. YASHIN** [1], **SERGEI L. SHISHKIN** [1], **AND ANATOLY N. VASILYEV** [1,2]

[1] MEG Center (Center for Neurocognitive Research), Moscow State University of Psychology and Education, 123290 Moscow, Russia
[2] Department of Human and Animal Physiology, Lomonosov Moscow State University, 119234 Moscow, Russia

Corresponding author: Yulia G. Shevtsova (shevtsova.jg@gmail.com)

This work was supported by the Russian Science Foundation under Grant 22-19-00528.

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Ethical Committee of MSUPE under Protocol No. 1, and performed in line with the Declaration of Helsinki.

ABSTRACT Gaze-based input to machines utilizes the ability of eye-gaze to serve as a user's ''output.'' However, gaze should also support information flow in the opposite direction, namely, ''input'' to the user's visual system from a machine's output. The two functions can be easily separated in some tasks, like eye typing, but more complex scenarios typically require users to perform additional actions to avoid misinterpreting their intent. In this study, we modeled a free-behavior interaction with rapid transitions between visual search, decision-making, and gaze-based input operations through an engaging game called EyeLines. When playing the game, 15 volunteers selected screen objects using a 500 ms dwell time without additional actions for intention confirmation. By applying machine learning algorithms to gaze features and action context information, we achieved a threefold reduction in false positives, improved the quality of in-game decisions, and increased participant satisfaction with system ergonomics. To our knowledge, this is the first study that demonstrates the effectiveness of machine learning applied to gaze features in enhancing gaze-based interaction within visually challenging environments.

INDEX TERMS Gaze tracking, gaze-based selection, Midas touch, gaze dwell, eye fixation, eye movements, gaze features, machine learning, intention recognition, user experience.

I. INTRODUCTION

Gaze-based interaction is a technology with almost a half-century history [1]. Its user, most typically, interacts with a computer by gazing at certain screen objects (buttons, web links, etc.) to select them. Gaze direction is captured by an eye tracker, so looking at an object basically works as pointing with a mouse cursor, and when a certain dwell time threshold is exceeded,[1] a command is issued, like when clicking the left mouse button. As looking at an object of interest is natural, the use of the technology can be relatively effortless. For decades, the technology was primarily developed to assist paralyzed individuals, but it also can be used by healthy persons, e.g., as a supplement to conventional input means [1], [2], [3], [4], [5].

However, the main function of eyes is, of course, vision. To fulfill it, our eyes constantly move and fixate to serve our vision without conscious intention, and we are not even aware of them [6], [7]. Unfortunately, this automatic behavior interferes with the ability to communicate via gaze.

The problem that arises at the intersection of the **visual** function of gaze direction control and the ability to use this control for **interaction** enabled by gaze-sensitive computer interfaces is called the Midas touch problem [8]: if certain gaze behavior issues a command to a computer, this command

The associate editor coordinating the review of this manuscript and approving it for publication was Nikhil Padhi.

[1] In this article we will use the word **dwells** for the events of relatively long stay of gaze in a certain area, which may result in gaze-sensitive interface response when the dwell time threshold is exceeded. Such events are called **fixations** in most of the literature on gaze-based interaction, but they do not fit strict definitions of an eye fixation used in the eye movement literature. A dwell may encompass more than one fixation and include some other kinds of eye movement.

will often be executed without the user's intention, because this gaze behavior is typically not used for intentional control. This problem was first formulated for systems which respond to simple gazing at an item [8], but it is inherent to any gaze-sensitive interface that cannot reliably differentiate gaze behavior used for interaction from any other behavior.

Many solutions for the Midas touch problem have been proposed (see [1], [9] for review). One group of solutions is making gaze behavior used for interaction less similar to natural, spontaneous eye movements. This is most commonly done by increasing the gaze dwell time threshold. Instead of intentional dwells, special learned sequences of saccades can be used. Another approach is to introduce additional behaviors to highlight intentional dwells. Most basically, this can be implemented by requiring the user to make a saccade to a special confirmation area. Gaze dwells or other eye movement patterns intended for interaction can also be marked by behaviors other than eye movements, such as blinks (typically more than one, to make the pattern less similar to spontaneous behavior), button presses, hand gestures (used in Apple Vision Pro), voice, and even imaginary movements detected by a brain-computer interface (BCI; e.g. [10], [11]). Each approach can be further enhanced by additional modifications; one interesting recent example was a dynamical adaptation of dwell time threshold to a task, which took into account previous performance and current gaze features [12]. However, existing solutions for the Midas touch problem require additional effort from a user, which worsens their experience and may make interaction less fluent.

Can the Midas touch problem be solved without requesting additional effort from the user? Intuitively, something in gaze behavior could be different in intentionally and spontaneously prolonged gaze dwells. Several studies showed that machine learning (ML) applied to **gaze features** can indeed recognize an intention-for-interaction with relatively high accuracy [13], [14], [15]. However, in these studies, interaction with a computer was at least partly manual (e.g., selection by combined gaze pointing and button pressing in [13]), so gaze patterns could be very different from pure gaze-based interaction.

Surprisingly, ML has been applied to gaze data collected during online gaze-based interaction **without manual input** only in recent studies. Reference [16] and [17] directly addressed the Midas touch problem, classifying gaze dwells intentionally used for interaction (which should trigger the selection command) and other (spontaneous) gaze dwells which should be ignored. Reference [12], who aimed at improving gaze-based interaction in VR, used ML to gradually adapt the dwell time threshold. This task can be considered as a more general version of accepting or rejecting a dwell with fixed time threshold, because very low/high thresholds lead to mandatory acceptance/rejection, respectively.

These studies demonstrated the significant potential of the ML-enhancement of gaze-based interaction. Nevertheless, they had a number of limitations. Reference [17] tested their algorithms only in offline simulations. Reference [16] made online tests, but their task included visual search, i.e., the participants had to search for a target prior to making a selection. Finding a target during visual search is associated with longer dwell times and stronger pupil dilations compared to looking at non-targets [18]. Effects of this kind can account at least for a portion of classifier performance when these or related features are used, as in [16] study. Therefore, this study could not model scenarios where the locations of the objects to be selected are well known to the user, a frequent case in real-life human-computer interaction. Most of the tasks used by [12] for online tests also included visual search, except for a sliding puzzle task (which was also used by [13]). In that task, however, the visual scene was very simple and was changing with time only slightly; consequently, the visual function of eye movements could be only slightly employed. Compared to many real-life situations, engagement of some participants (or even all of them) in the task could likely be lower due to relative simplicity of the task and its repeated use along the experiment, and this could also affect eye movement patterns. Finally, in the tasks used in experimental studies (including the sliding puzzle task) the number of available alternative selections was low, which also might lead to gaze patterns different from real-life conditions.

One more problem associated with modeling real-life interaction is that when participants are allowed to behave freely, the **ground truth** can be directly obtained only through introspection. Therefore, [12] and [13] assessed quality of online classification using overall performance in the task and responses to questionnaires. While the approach was generally relevant, it could provide only rough estimation of intention recognition quality, given the indirect relation of the indicators to classifier performance, high variability of these indicators' values and relatively small sample size. Classifier performance in these studies was directly estimated in offline modelling of classifier operation, but this was also a rough approximation, as gaze behavior should, most likely, change when the classifiers are applied online.

Thus, first steps toward ML-enhancement of gaze-based interaction were made, but experimental paradigms used to test the proposed algorithms did not sufficiently address the need to combine gaze use as an input tool and as a tool for vision in visually rich and changing environments.

To test **free and enganging online gaze-based interaction in a dynamically changing environment**, a gaze-controlled game *EyeLines* was proposed [19], [20], [21], [22]. In this game, a player makes moves on a game board with colored elements, some of which are added or removed after each move, so the visual environment is significantly changing along all the game; a move typically can be freely chosen from many existing variants, and players' significant engagement is normally observed (see a more detailed description below). The game was used to evaluate BCI-based selection intention recognition related to gaze dwells in offline [19], [20], [21], [22]. and online [23] BCI modes.

For intention recognition using only gaze features, this game was used by [17], but without online application of the ML algorithms. In most of these studies labelling of intentional and spontaneous dwells was made based on participants' gaze behaviors, without requiring them to mark their decisions explicitly (with a partial exception for the online BCI part of the study by [23]), where participants "semi-explicitly" labeled incorrect selections by a repeated dwell on the same object, thereby also cancelling selection). Labelling was still based on additional actions, namely switching on the controls prior to each move or by confirming a selection, in both cases using a dwell on a designated screen "button;" however, these additional actions seemed to automatize quickly and did not distract from the main task.

In the current study, we explored **ML-enhancement** of gaze-based interaction in the **online** mode.

Like in our previous studies, *EyeLines* game was used as a **testbed** with features which we considered important for relevant modeling of gaze-based interaction: (1) the need to combine intensive use of gaze both **for control and for vision**, (2) freedom of participants' behavior, (3) implicit labeling of intentional and spontaneous gaze dwells (not distracting a participant from the interaction); (4) engagement in interaction.

In addition to testing ML-enhancement in the online mode, the current study was different from our previous study of ML-enhancement of gaze-based interaction, where *EyeLines* was also used [17] in the following ways:

(1) Confirmation of selection was not used in this study in either ML-enhanced (C—classifier-enhanced) and baseline (D—based solely on dwell duration) modes, therefore moves in the game could be done significantly faster and interaction became more intensive. A somewhat different and more sophisticated labeling protocol was developed to solve the problem of the absence of explicit dwell type information.

(2) In addition to the ML-based algorithm based on gaze features (the **gaze classifier**), we further enhanced intention recognition with a context-based algorithm, hereafter refferenced to as the **contextual classifier**. This algorithm modified recognition of a dwell as intentional or spontaneous based on the current game context. Adding the contextual classifier was done to provide better approaching real-life use of gaze-based interaction, where similar algorithms are effectively used to improve interaction (especially, autocomplete and similar algorithms in gaze typing). Algorithms of this type are increasingly used in human-machine interaction in general [24]. We also aimed at achieving especially fluent experience of gaze-based interaction, to explore how gaze behavior can be modified in this case. To assess its contribution to system performance, we also simulated gaze-based interaction without this classifier in an offline study.

(3) Set of gaze features was improved compared to [17].

(4) To enable effective implementation of ML algorithms, the game was re-written in Python while preserving its game mechanics and most details of the visual appearance and interface behavior, with some differences highlighted in Methods.

The aim of this study was to test, in online (near real-time) mode and in a task where users smoothly alternate between making gaze-based actions and visual inspection, if ML indeed enables passive differentiation of gaze dwells used for control and spontaneous gaze dwells in a way that improves gaze-based human-computer interaction.

We tested several **hypotheses** about specifc aspects of the improvement:

H1 (control rate hypothesis): With the use of ML, a higher command rate (effective actions per minute, eAPM) will be observed.

H2 (efficiency hypothesis): With the use of ML, higher rate of ball removing will be observed (i.e., more balls will be removed from the game board per minute; we considered this as a key measure of successful progression in the *EyeLines* game).

H3 (UX hypothesis): With the use of ML, participants will rate their gaming experience higher, and/or they will prefer ML-enhanced interaction over the baseline interaction.

All these hypotheses were tested online for combined application of both the gaze classifier and the contextual classifier vs. baseline (no ML enhancement) dwell-time based control. In addition, offline modelling was used to assess contribution of the contextual classifier, i.e., to check which of the classifiers was primarily responsible for the observed results. We also explored, using data from the online experiment and from offline simulations, various aspects of ML-enhanced gaze-based interaction, to provide a better understanding of how it actually worked in near real-life conditions and how it could be further improved.

II. MATERIALS AND METHODS

A. PARTICIPANTS

17 naïve healthy volunteers (7 males, 10 females; age 25 ± 6 years, $M \pm SD$) participated in this study. All participants had normal or corrected-to-normal vision. Five volunteers had previously participated in our study where they played the *EyeLines* game using gaze-based control without ML-enhancement, while the rest had no prior experience with gaze control. All participants were introduced to the procedure and provided informed consent. The experimental procedures were approved by the Ethical Committee of MSUPE (protocol №. 1, dated 12.03.2015) and were conducted in accordance with institutional and national guidelines for research involving human participants, as well as the Declaration of Helsinki. One of the participants was excluded from the analysis due to poor playing, and data from another one were lost due to technical issue. Thus, data from a total of 15 participants were analyzed.

B. EXPERIMENTAL DESIGN

The experiment comprised two sessions conducted on separate days, with an interval of up to two weeks between them. During each session, participants played the *EyeLines* game. This game is similar to the conventional Lines game, but gaze dwell time was used instead of mouse clicks or touches to a touchscreen (see section II-C for details). To familiarize themselves with the task, participants were asked to practice playing the traditional Lines game using either a mouse or a touchscreen before the first experimental session.

On the first day, participants were introduced to the game and gaze control techniques, while the second day focused on performance evaluation and comparing the two game modes. Participants showed noticeable improvement in game and/or gaze control skills during the first session, and by the second day, these skills had stabilized, as will be discussed in the results section (III.B.). The sequence of game modes on the first day was fixed to help participants become familiar with gaze control mechanics, such as spontaneous selections. This fixed order was necessary because the data from the baseline mode (D) was used to adjust the classifier parameters for the classifier-enhanced mode (C). On the second day, the sequence of modes was randomized across participants, as the classifier parameters had already been adjusted based on the first day's data. The experimental design was as follows:

- Day 1: test, D, D, D, S1, rest, C, C, C, S1, S2
- Day 2: test, D, D, D, S1, rest, C, C, C, S1, S2
  or test, C, C, C, S1, rest, D, D, D, S1, S2

test: On day 1, this section involved explaining the rules and demonstrating the game using a special mode with a mouse. On both days, participants played the game for 3 minutes in D mode. Additionally, on day 1, they played for 3 minutes in the first mode of that day.

rest: On both days, participants rested for 5–10 min between the two modes.

D: One game in the mode with control using only 500 ms dwell time threshold.

C: One game in the mode with control based on dwell time and classifier decisions.

S1: Survey regarding the perceived mode qualities.

S2: Survey regarding comparison of the two modes.

The participants were unaware of the differences between the modes. Typical duration of sessions was 1 h 30 min on Day 1 and 1 h 20 min on Day 2.

C. EYELINES GAME

All variants of the gaze-based control were tested on the eye-controlled game *EyeLines*, designed earlier [21], [23] based on a popular computer game Lines (with some adaptation for gaze-based control) and implemented here in Python to support more flexible online data processing. In *EyeLines*, the player selects a colored "ball" randomly placed on the game board and moves it to form lines of four or more balls of the same color. These lines can be arranged vertically, horizontally, or diagonally. When a line is successfully formed, it disappears, freeing space on the board.

The scoring system awards the player 40 points for forming a line of four balls. Longer lines result in higher scores: a line of five balls earns 100 points, six balls yield 180 points, and seven balls—the maximum due to the board's size—earn 280 points. A ball can only be moved if there is a path of orthogonally adjacent empty cells between the starting and destination cells. In *EyeLines*, each "ball" is represented as a simple colored circle with a small dot in the center to aid precise and stable gaze fixations.

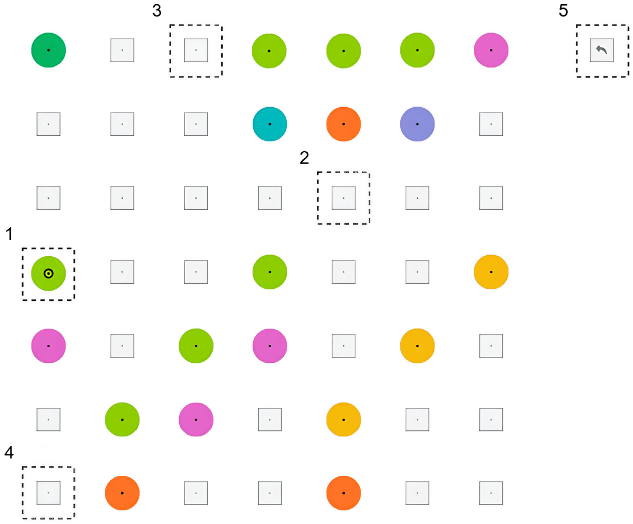


FIGURE 1. An example of the EyeLines display. The dotted areas indicate the following: 1–the selected ball; 2, 3–empty cells where the selected ball can be moved to complete a colored line; 4–empty cell that is inaccessible due to the restricted diagonal move rule; 5–the "undo last move" button.

In this study, game board size was 16.0° × 15.8° and included a 7 × 7 grid, and each "ball" subtended about 1.2° (Fig. 1). Ball selection and moving were both made based on dwell time threshold with or without additional decision rules (for details, see the II.E. Classifiers section below). Unlike in previous *EyeLines* versions [19], [21], [23], the selected ball was marked by enlarging the central dot in it from 0.1° to 0.3°. Also, to facilitate gaze control, participants received visual feedback on their gaze position through a small dark blue cross, displayed at the median of the gaze coordinates (Fig. 1). If moving a ball to the selected cell was not possible according to the rules, a red cross temporarily appeared in the cell.

If a ball was selected unintentionally, participants could deselect it by gazing at it again or by selecting a different ball. They also had the option to cancel a completed move (ball selection and placement) by gazing at a designated area on the screen, located a few cells away from the top-right corner of the playing field. Participants were instructed to cancel a move only in cases of errors, such as incorrect ball movement due to a wrong classifier decision, unintended selection, or calibration issues, but not to undo an intentional move to choose a better option.

Each game was limited to a duration of 8 minutes, though it could end earlier if the board was filled with balls or due to occasional issues with eye-tracker calibration. In cases where the game ended prematurely, additional games were offered to ensure a total playtime of 20–24 minutes per mode. After each game, the player's score was displayed.

D. EYE TRACKING AND GAZE DWELL DETECTION

Eye tracking was conducted in binocular mode using the EyeLink 1000 Plus eye tracker, with a sampling rate of 1000 Hz. Calibration was performed before each game, adhering to the following validation accuracy requirements (in degrees): average < 0.5, maximum < 1.0.

The dominant eye was identified using the Miles and Porta tests (10 participants had a dominant right eye, 5 had a dominant left eye), and its coordinates were used for real-time control. Real-time (online) data processing was achieved by collecting coordinates in batches of 20 samples using the Resonance software system [25]. Saccades and fixations were identified using the EyeLink online software algorithm, with the following criteria for saccades: acceleration $> 8000°/s^2$, speed $> 30°/s$, and a minimum duration of 6 ms.

Dwells were identified using a spatiotemporal criterion: the dispersion of coordinates over 500 ms did not exceed 2.3° for both axes. After reaching the time threshold, a dwell "continued" if the following additional criteria were met:

- The spread (max–min) of coordinates in the last 500 ms remained within 2.3° for OX and OY axes;
- The distance between the center of gaze X and Y coordinate distributions in the last 500 and the initial 500 ms did not exceed 1.8°.

Time threshold of 500 ms was chosen as the most comfortable for users based on our previous study with the *EyeLines* game [21].

When these conditions were not met, the current dwell was discarded, allowing detection of the next dwell when the dwell criteria were met again. This approach accounted for gaze drift while preventing the repeated selection of the same ball by maintaining gaze in one location.

The square hitbox (sensitive area) for each ball and cell were ± 1.3° relative to their center, ensuring there were no blind spots on the playing field.

During the experiment, the start time of a gaze dwell was corrected to the first fixation start time in both modes if the difference between these time points did not exceed 50 ms. Otherwise, the gaze dwell was adjusted forward by 20 ms, which was the average difference between the start of the fixation and the dwell based on our previous experiments with the same game.

E. CLASSIFIERS

1) CLASSIFIER APPLICATION

We first describe application of the classification algorithm to make clear its functioning and therefore to clarify what was the goals of classifier training, before explaining how this training was organized.

In Mode C, for each dwell falling within the hitbox of a ball, a classifier decided whether the ball had to be selected (or deselected, if it has been already been selected). Selection of an empty cell for moving a ball into it or the undo button did not employ the classifier and relied solely on 500 ms gaze dwells.

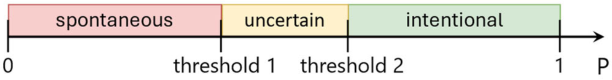


FIGURE 2. Classifier decisions based on thresholding the probability of assigning gaze dwells to the intentional class.

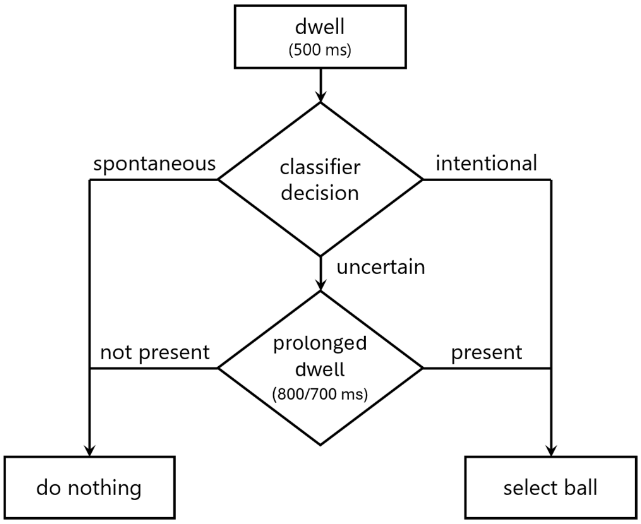


FIGURE 3. Dwell to selection translation based on 2-stage decision pipeline of in mode C.

The classification algorithm consisted of two components: a gaze classifier, which used features of gaze micro-behaviors, and a contextual classifier, employing features describing ball position in the context of other balls' arrangement on the board. The predicted probabilities from both classifiers (P_{gaze}, P_{ctx}) were averaged, resulting in the following classifier prediction (estimated probability of intention to select, P):

$$P = \frac{1}{2}(P_{\mathrm{gaze}} + P_{\mathrm{ctx}}) \tag{1}$$

The classifier decision was based on thresholds of P value: a gaze dwell was deemed "spontaneous" (below threshold 1), "intentional" (above threshold 2) or "uncertain" (between threshold 1 and 2) (for graphical explanation see Fig. 2).

Thresholds were individually adjusted for each participant and each session (for day 1: based on Mode D data on day 1; for day 2: based on day 1 data from both modes), ensuring that the proportion of uncertain gaze dwells remained below 30% of all classifier decisions. This optimization aimed to achieve the highest balanced accuracy while minimizing the false positive rate (FPR) on the first day and the false negative rate (FNR) on the second day. The switch to FNR on the second day was due to participants' feedback gathered on the first day.

Fig. 3 further clarifies the decision-making scheme. Gaze dwells classified as intentional led to the selection of a ball. Gaze dwells classified as spontaneous did not result in any action. For uncertain gaze dwells, a ball was only selected after a prolonged dwell—an uninterrupted dwell with a duration of 800 ms (on the first day) or 700 ms (on the second day). If the dwell was interrupted before reaching this threshold, the ball was not selected.

2) CLASSIFIER TRAINING

The classifiers utilized the SVM RBF model. On the first day, gaze and contextual classifiers were trained on data from 21 participants of our previous *EyeLines* experiment [17], where gaze-based control was mostly like in the B mode in the current study. Classes were individually balanced by removing random dwells from the class where their number was highest. For threshold tuning, the group classifier was tested on data collected in the D mode on the first day, and the selected threshold was used online in mode C further on the first day.

On the second day, individual gaze and contextual classifiers were constructed using data from the first day. For testing the individual classifiers and setting their thresholds, the dataset was split into two parts—1/3 dwells were used for feature selection (for the gaze classifier) and model hyperparameter tuning, and 2/3 dwells entered cross-validation performance assessment. Classes were balanced randomly beforehand. The model with the tuned hyperparameters was then trained, using the selected features, on all class-balanced data from the participant, and used online in mode C on the second day.

Training procedure was the same for both days. All features were standardized. Hyperparameters were tuned using sklearn.model_selection.RandomizedSearchCV with the following adjustable parameters [$n = 3$, C =(0.05..10), gamma =(0.005, 1), $n_{\text{iter}} = 50$, score='balanced_accuracy']. The seven most significant features for the gaze classifier were selected using the RFE algorithm (Recursive Feature Elimination, sklearn.feature_selection.RFE), with feature ranking performed on the SVM Linear model.

3) FEATURES FOR THE GAZE CLASSIFIER

The features used for the gaze classifier (see Table 1) were derived from raw eye-tracker data and gaze positions relative to the gameboard elements.

Features marked with * were calculated in overlapping 50–ms windows. e.g., 0–500, 50–500, . . . , 450–500 ms, while those marked with † were calculated in non-overlapping windows, e.g., 0–50, 50–100, . . . , 450–500 ms

Features based on pupil data (C1), binocular data (D1), and initial fixation (A1) were not used in the online experiment but were considered for offline simulations. A total of 84 gaze features were analyzed in this study; however, only 7 features, selected using the Recursive Feature Elimination (RFE) algorithm, were used in each model of the gaze classifier. The most frequently used features across participants are presented in Table 5.

TABLE 1. Features used for classification.

Symbol	Name	Definition
A	*Saccades and fixation within the dwell*	
A1	Initial fixation duration	duration of the first fixation within the dwell
A2	*Count of saccades	number of saccades during the dwell
A3	*Sum of amplitudes of saccades	sum of saccade amplitudes saccades during the dwell
B	*Gaze point during dwell*	
B1	Variance of coordinates: X, Y, Average	variance within 500ms window
B2	*Spread of coordinates: X, Y, Average	difference between min and max coordinate of corresponding axis withing the time window
B3	†Distance to the center of the closest ball	distance in pixels from the center of the gaze center within the time window and center of the closest ball
C	*Pupil size*	
C1	*Median area	unit-normalized to the values within the same game
D	*Binocular vergence*	
D1	*Vergence	Median difference between left and right eye gaze points

4) FEATURES FOR THE CONTEXTUAL CLASSIFIER

The features of the contextual classifier were the properties of ball position in the context of other balls' arrangement on the game board. We manually selected features which could influence the probability that the ball will be chosen by the player for the current move, based on the analysis of a total of 50,000 moves from games in a previous study (not published yet) with the same *EyeLines* variant. The features were categorized into two main types: those indicating the potential for forming same-colored lines (either by creating a line or by freeing up space for a ball(s) that could complete a line) and those describing the overall visibility and mobility of the balls on the game board.

The first category included the change in the number of lines of same-colored balls with all possible movements of the selected ball. We examined the change in the number of lines of the following types: bb, bbb, bbbb, bbbbb, bbbbbb, bbbbbbb, b0b, b0bb, bb0b, bb0bb, b0bbb, bbb0b (where 'b' represents balls of the selected ball's color, and '0' represents an empty cell). For example, if moving a ball would change a line of two same-colored balls into a line of four balls, the features would be: bb = -1, bbbb = 1, and all others = 0.

To assess the potential of a move for freeing up space for line construction, we considered the change in the number of unblocked lines of three balls: c0cc, ccc0 (where 'c' represents balls of a different color from the selected ball, and '0' represents an empty cell). For instance, if moving a ball freed up space for an existing line of three balls, the feature ccc0 = 1.

Features in the second category, related to the overall visibility and mobility of the balls, included: the number of empty cells to which a ball could be moved (noting that the presence of other balls could block paths to certain cells), the Euclidean distance of the selected ball from the center of the board, the total number of balls of that color on the board, the overall number of balls on the board, the percentage of balls of that color relative to all balls on the board, and the change in the size of the cluster of empty cells (if moving a ball frees up a path, the cluster size increases).

In total, 21 features were used for the contextual classifier—14 from the first category, and the remaining 7 from the second.

5) APPROXIMATE GROUND TRUTH LABELLING

To maintain motivation and engagement in the participants, we used the *EyeLines* game paradigm intentionally designed [21], [23] without constrains specific to an experiment. This game rules and gameplay mostly follow its predecessor, the Lines game, so that the players freely choose each move. As often happens in free behavior experiments, however, this comes at a price that the ground truth for dwell classes, intentional (voluntary) vs. unintentional (spontaneous), could not be approached directly. We did not ask the participants to label selections as intentional or spontaneous, as this could strongly distract them from the game and made their gaze behavior different from normal gaze interaction. However, we could approximately determine the true classes for gaze dwells based on the participant's actions. The primary criterion was if the selected ball was moved (intentional dwells) or not (spontaneous dwells). Additional labeling criteria included canceling a move after a ball was moved, deselecting a ball with a repeated gaze dwell, and attempting to move a ball to a prohibited area on the game board (a "no path" situation, where the move violated the rules). In a small portion of cases (approximately 2%; see Section III-A), it was not possible to determine the class of a given gaze dwell, and these instances were excluded from the further analysis of classifier performance. A detailed summary of the labeling rules can be found in Table 2.

Note that this procedure is inherently approximate and may introduce label noise. Typical mislabeling errors include tagging an intentional selection as spontaneous when a participant aborts a planned move at the last moment, or conversely, labeling a spontaneous selection as intentional when a participant happened to move a ball that was selected without prior intent. This noise affected both training and testing data (online and offline), potentially reducing classifier discriminability and degrading performance metrics (e.g., overall accuracy). We estimate that data from inexperienced participants (Day 1) contained higher levels of label noise; as participants became more familiar with the task (see II-A), their action repertoire stabilized and more consistently conformed to our labeling criteria.

F. PERFORMANCE ASSESSMENT

1) CLASSIFIER PERFORMANCE

Several metrics were used to assess classification performance, including balanced accuracy, area under the curve of receiver operating characteristics (ROC AUC), and the P4 metric proposed by [26]. The latter was employed due to the balanced approach to accounting different types of errors, which is important when tuning classifier in imbalanced classes scenario. The P4 metric is the harmonic mean of recall (true positive rate—TPR), specificity (true negative rate—TNR), precision (positive predictive value—PPV) and negative predictive value (NPV):

$$\mathrm{P4} = \frac{4}{\frac{1}{\mathrm{TPR}} + \frac{1}{\mathrm{TNR}} + \frac{1}{\mathrm{PPV}} + \frac{1}{\mathrm{NPV}}} \tag{2}$$

where $\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{TP+FN}}$; $\mathrm{TNR} = \frac{\mathrm{TN}}{\mathrm{TN+FP}}$; $\mathrm{FPR} = \frac{\mathrm{FP}}{\mathrm{FP+TP}}$; $\mathrm{FNR} = \frac{\mathrm{FN}}{\mathrm{FN+TP}}$

We also introduced a modified version of P4 to account for the varying impact of false positives (FP) on gameplay. In this weighted metric (P4*w*), the number of FPs in the formula was adjusted based on their actual impact on game performance. Not all false positives were noticed by users; according to the questionnaires, on average, participants only noticed about one-fifth of unintended selections, and not all of these had significant consequences for gameplay. Unintentionally selected balls disrupted gameplay only when they led to unintended ball movements (assessed by counting the use of the undo button followed by moving another ball) or when users had to spend extra time deselecting the ball with a repeated gaze dwell. As a result, the FP*w* variable was calculated as the total number of move cancellations and ball deselections (corresponding to cases №. 8 and №. 12 from Table 2):

$$\mathrm{FP}_w = n_{\mathrm{cancelled}} + n_{\mathrm{deselected}} \tag{3}$$

After substituting the formulas for TPR, TNR, FPR, and FNR into (2), the P4 formula becomes more convenient for incorporating the modified version of FP, allowing P4w to be calculated as follows:

$$\mathrm{P4}_w = \frac{4 \times \mathrm{TP} \times \mathrm{TN}}{4 \times \mathrm{TP} \times \mathrm{TN} + (\mathrm{TP} + \mathrm{TN}) \times (\mathrm{FN} + \mathbf{FP_w})} \tag{4}$$

To evaluate classifier performance during offline simulations, a special coefficient was calculated for each participant based on the ratio of FPw to FP during online performance:

$$\mathrm{coef} = \mathrm{FP}_w/\mathrm{FP} \tag{5}$$

This coefficient was used to convert the number of FPs from the offline simulation into those that would have resulted in negative gameplay consequences. Therefore, in offline simulations, FP*w* was calculated as:

$$\mathrm{FP}_w = \mathrm{FP} \times \mathrm{coef} \tag{6}$$

TABLE 2. Dwell Labelling* Based on Observed Action Sequences.

Class labels (online)	Observed action sequence	Interpretation	Error type	NO
Intentional— *intentional ball selection*	The ball was moved, and the move was not cancelled.	Proper of intentional selection.	TP	1
	The ball was moved; the move was canceled, but the same ball was moved on the next move.	The move was canceled not because of the ball's incorrect selection but due to incorrect translation destination.	TP	2
	The selected ball was deselected, then selected again and moved.	A repeated dwell on an already selected ball, resulting in accidental deselection.	TP	3
	An attempt was made to move the ball to form a line of three or more balls of the same color but failed due to the absence of available path to the destination.	The attempt failed due to the absence of a path that was not noticed.	TP	4
	Mode C only: the ball was not selected due to classifier's negative decision, but then there was an immediate repeated dwell on it, after which the ball was moved.	The classifier likely made an error, and initial selection was intentional.	FN	5
Undefined— *cannot determine if selection was intentional or spontaneous*	An attempt was made to move the ball to a place beside a ball of the same color is, but there was no valid path.	Two situations are equally likely. Either an intentional attempt to form a line of two balls of the same color, but the absence of a path was not noticed; or the move attempt was accidental after an unintentional ball selection.	–	6
	Mode C only: the ball was not selected due to classifier's negative decision but selected after a dwell on another ball.	The classifier likely made an error, but it is unclear if the participant tried to select the ball or was just scanning the field.	–	7
Random— *spontaneous ball selection*	The ball selection was canceled, and the next move was with another ball.	Highly likely, the ball was initially selected unintentionally.	FP	8
	Immediately after selection of the ball selection another ball was selected.	A common situation of unintentional selection during field inspection.	FP	9
	An attempt was made to move the ball to a cell without adjacent balls of the same color but failed due to the absence of available path to the destination.	Purposeless action that can't be explained by gaze positioning errors or similar reasons, hence, the ball was selected unintentionally.	FP	10
	Mode C only: the ball was not selected, and later a move was made with another ball.	The classifier correctly categorized this ball selection as unintentional.	TN	11
	The move was canceled, and the next move was with another ball.	The move was canceled because the ball was selected and moved to a new position unintentionally.	FP	12

*Note that only ball selections were labeled, as an empty cell to which the ball was moved was always selected without any classifier, based solely on gaze dwell time.

2) EFFICIENCY OF THE GAZE-BASED CONTROL

Two major aspects of control efficiency were considered: the effectiveness of playing the game and the efficiency of performing gaze-based actions. The objective of the game was to form lines of same-colored balls, which would result in their removal, clearing the game field. The game continued until the end of the allotted time, with the goal of scoring the highest number of points by removing as many balls as possible. *EyeLines* players generally found it unpleasant when the game ended before the 8-minute limit, as this usually resulted in a lower score, which was a key indicator of success for participants.

Two performance metrics were used to assess these aspects:

- Game playing time (as a percentage of the game's 8-minute duration, averaged across all games);
- Rate of ball removal (in balls per minute).

In both modes, D and C balls could be selected without the user's intention due to classifier's false positives (FP). The rate of unintentional ball selection was calculated as a ratio of the number of TP to the number of FP. For example, TP/FP = 3 means that on average every 4th ball is selected unintentionally. (Note that the ground truth could be approached only approximately, so such computations also lead to approximate values).

As mentioned above, only unintentional selections that resulted in gameplay consequences hindered the gameplay, forcing participants to take additional actions, such as deselecting balls or canceling moves. Therefore, we also calculated the ratio of true positives (TP) to weighted false positives (FP*w*), which is inversely proportional to the frequency of issues participants encountered during the game due to false positives (FP).

False negatives (FN) appeared in gaze control mode with a classifier (Mode C) when gaze dwells were mistakenly classified as spontaneous, making it impossible to select the ball. The ratio of TP to FN showed how rarely a ball cannot be selected (TP/FN = 10 means that every 11th ball cannot be selected).

As both types of errors hindered performance and subjective assessment of interaction, a unified metric, Mean Time Between Failures (MTBF) was used, together with MTBF*w* that accounted not for all FPs but only for FP*w*:

$$\mathrm{MTBF} = \frac{\mathrm{TP}}{\mathrm{FP} + \mathrm{FN}} \tag{7}$$

$$\mathrm{MTBF}_w = \frac{\mathrm{TP}}{\mathrm{FP}_w + \mathrm{FN}} \tag{8}$$

3) QUESTIONNAIRES

We used questionnaires to evaluate the users' experience with gaze-based control while playing *Eyelines*. Participants completed a questionnaire after each condition (presented as condition 1 and 2) on both the first and second days of the experiment. Additionally, a final questionnaire was provided at the end of each day to compare the two modes. The questionnaires were printed on paper, and participants filled them out with a pen. For the open-ended questions, participants verbally provided their answers, which the experimenter typed on a computer.

In the post-condition questionnaires, participants rated various control features of the control mode, including the rate of ball selection and the frequency of unintentional ball selections or movements. Additionally, participants provided feedback through open-ended comments.

Meanwhile, our primary focus was on the final questionnaires, where participants compared the two control modes. We aimed to determine whether users found the controls in the C condition superior to those in the D condition. In these final questionnaires, participants rated different mode properties or expressed their agreement with various statements using continuous scales. The length of the scales equaled 100 mm. For each scale, participants placed two marks, one for each mode (either mode 1 or mode 2). Below, we provide a table with the exact wording of the questionnaire items and the labels used for the scale anchors (see Table 3). At the end of the questionnaire, participants provided open-ended comments on the differences between the modes and explained their preferences for one mode over the other.

III. RESULTS

A. GAZE-CONTROL EFFICIENCY

The eye tracker calibration quality, validated at the beginning of each game, averaged 0.26 ± 0.07 degrees across 9 points, with a maximum deviation of 0.54 ± 0.15 degrees for the dominant eye. No link to game performance was detected in relation to calibration quality metrics.

Group statistics for gaze dwell classes labelled according to the rules from the Table 2 showed a predominance of intentional dwells (see Table 4). The proportion of this dwell class varied among participants, ranging from 25% to 50%. The proportion of intentional gaze dwells increased significantly on the second day (Student's t-test, $t(14) = -3.525$, $p = 0.003$), likely due to improved gameplay skills and enhanced gaze control efficiency.

A two-way ANOVA with repeated measures was conducted to evaluate the effects of Day and Mode on the rate of ball removal. The results indicated a significant main effect for Day only: $F(1,14) = 15.53$, $p = 0.0015$, highlighting the clear progression of game skills across days (Fig. 4). This progression is also evident when the removal rate is plotted against the game number within a day (Fig. 5). Consequently, to compare the efficiency of gaze control between modes, we used game performance metrics solely from the results of the second day.

TABLE 3. Contents of the final questionnaires.

#	Statements or mode properties	Left anchor	Right anchor
Q1	Command input speed	Too slow	Too fast
Q2	Unwanted/unintentional ball selections	Did not occur	Occurred too frequently
Q3	Unwanted/unintentional ball movements	Did not occur	Occurred too frequently
Q4	The game controls interfered with thinking and decision making, distracted from the game	Did not interfere	Severely interfered
Q5	Game controls were annoying for one reason or another	Not annoying at all	Severely annoying
Q6	Duration of individual games or of the experiment as a whole	Did not affect my well-being and effectiveness	Were severely tiring, by the end I felt like I couldn't play.
Q7	Overall effort required to control the game	Did not require effort	Required a great deal of effort
Q8	Overall comfort with gameplay	Extreme discomfort	Enjoyed playing the game
Q9	If I often played "Lines" on my computer, I would like to use this control mode instead of the mouse	Strongly disagree	Strongly agree
Q10	When performing everyday tasks on my computer, I would not mind using this control mode instead of the mouse	Strongly disagree	Strongly agree

TABLE 4. Average number of Dwells of each class per participant.

label	Day 1		Day 2	
	n	%	n	%
Intentional	526±153	58.6±10.3	638±170	65.7±8.5
Spontaneous	346±92	39.1±9.4	310±74	32.7±7.9
Undefined	20±10	2.3±1.1	15±6.0	1.6±0.8

Mean ± SD for N = 15

The ball removal and movement rates did not differ between D and C modes on Day 2 (Student's t-test: for removal rate $t(14) = 0.05$, $p = 0.96$ and for move rate $t(14) = 1.08$, $p=0.30$). However, Mode C allowed participants to play longer (i.e., fewer games ended prematurely due to board filling with balls, Wilcoxon test: $W(13) = -73.0$, $p = 0.0078$) and perform less actions to remove the same number of balls ($W(15) = 110.0$, $p = 0.0006$) (Fig. 6).

In terms of error frequency during gaze control, Mode C shows a higher ratio of true positives (TP) to both false positives (FP) and false positives weighted (FP*w*) compared to Mode D (Wilcoxon test, TP/FP: $W(15) = -120.0$, $p = 0.00006$; TP/FP*w*: $W(15) = -118.0$, $p = 0.0001$),

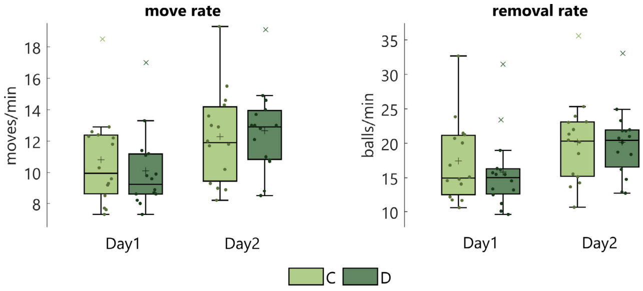


FIGURE 4. Group average (N = 15) rate of ball movement (left) and removal (right) across days and game modes. Note that the rate of ball removal directly characterizes perceived success in the game (game scores).

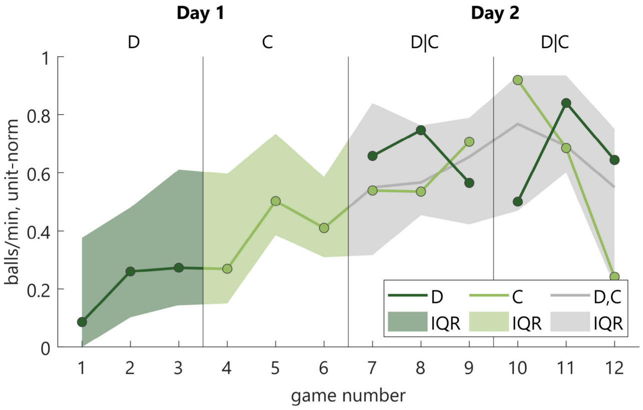


FIGURE 5. Group average (N = 15) participant-normalized rate of ball removal as a function of game number. The gray line represents the overall data, while the data for Day 2 are split into Mode D and Mode C, shown by dark and light green lines, respectively. The shaded areas on the graphs represent the interquartile range (IQR).

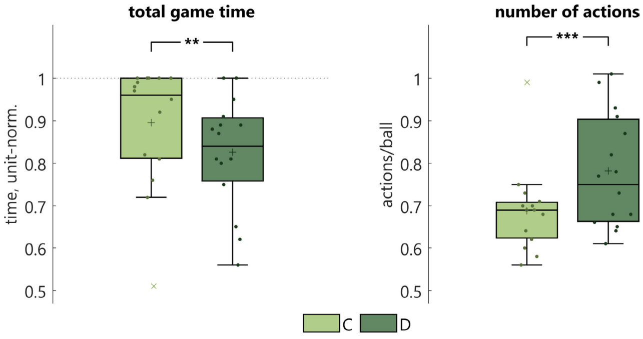


FIGURE 6. Group average (N = 15) for total game time (left; higher values indicate better performance) and the number of actions per ball removed (right; lower values indicate better performance) across game modes on Day 2. Asterisks indicate statistical significance: $**p < 0.01$, $***p < 0.001$ (Wilcoxon test).

as illustrated in Fig. 7. Although false negatives (FN) caused by incorrect classifier decisions contribute to errors, Mode C remains superior to Mode D when all FP are considered (MTBF: W(15) = −120.0, p = 0.00006). However, if only FP*w* are considered, the addition of FN results in an equal number of failures across both gaze-control modes (MTBF*w*: W(15) = 2.0, p = 0.98).

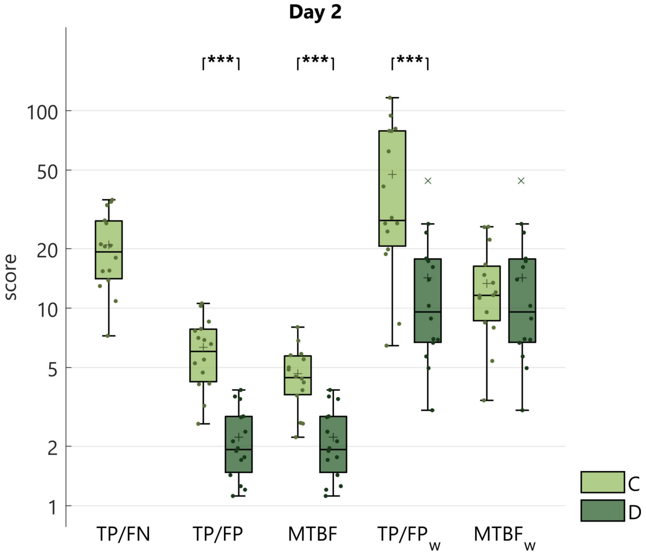


FIGURE 7. Group average (N = 15) ratio of true positives to false negatives (TP/FN), true positives to false positives with and without weighing (TP/FP_w, TP/FP), mean time between failures: conventional variant (MTBF) and with weighed FP (MTBFw) for experimental modes D and C on Day 2. The scores represent the logarithmic values of each variable. Asterisks indicate statistical significance: $***p < 0.001$ (Wilcoxon test).

B. QUESTIONNAIRES

Analysis of the questionnaires revealed that only one out of the 10 questions (Fig. 8) showed significant differences across days (Q6, Wilcoxon test: W(11) = 66.0 p = 0.001). The statistical significance of the differences between the two modes was assessed using a permutation test (n = 10,000 permutations, with the t-test as the statistical criterion, df = 14) conducted separately for each day. On the first day, participants slightly preferred Mode D when asked about the influence of game duration on their well-being and effectiveness. However, by the second day, their preferences had shifted, with Mode C receiving higher ratings (Q6, p = 0.02).

Mode D was perceived as slower than Mode C (Q1: day 1, p = 0.002; day 2, p = 0.016). In terms of the other questions, either Mode C was favored, or no significant preferences were expressed at the group level. According to the questionnaires, Mode C was perceived as less annoying on both days (Q5: day 1, p = 0.03; day 2, p = 0.04). Additionally, on the first day, participants reported fewer unintended selections (Q2, p = 0.006) and unintended moves (Q3, p = 0.009) in Mode C.

C. CLASSIFIER PERFORMANCE

1) FEATURE IMPORTANCE

The feature analysis revealed that the most reliable indicator was the distance from the center of the object (B3, see Fig. 9, Table 5), with a characteristic monotonic increase up to 500 ms. This was likely because, during intentional dwells, participants tried to keep their gaze centered on

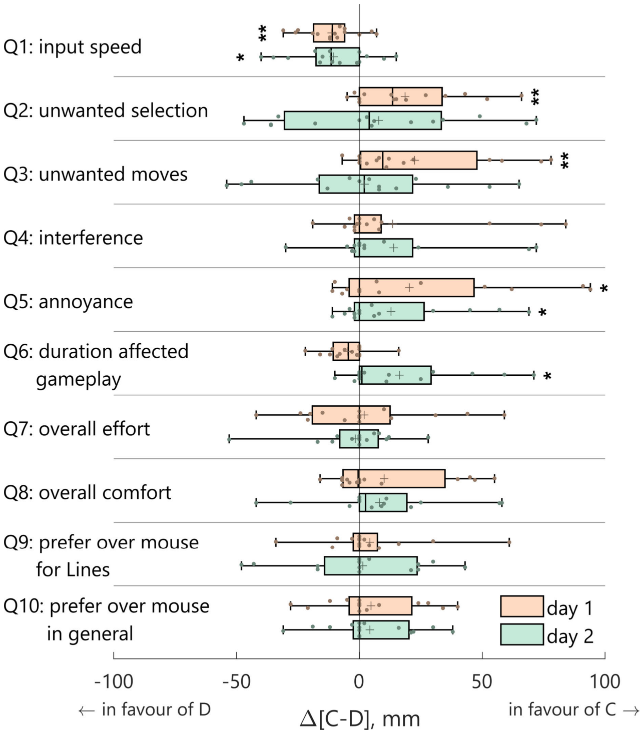


FIGURE 8. The results of comparing the two modes based on questionnaire responses across days. The values were obtained by subtracting the scores for Mode D from those for Mode C. The axes for each question were adjusted so that when Mode C was rated more favorably than Mode D, the values appear on the right side of the plot. Asterisks indicate a statistically significant difference from zero: $*p < 0.05$, $**p < 0.01$ (permutation test).

TABLE 5. Most frequently used features (on day 2).

Feature, window	Used, N_{subj}	Used, $\%_{subj}$	r2, all	r2, where used)
B3, 450–500 ms	10	66	0.24	0.25
A3, 300–500 ms	5	33	0.18	0.13
B3, 350–400 ms	5	33	0.18	0.17
B2x, 350–500 ms	4	27	0.08	0.11
B1x	3	20	0.09	0.12
B2xy, 0–500 ms	3	20	0.11	0.19
A2, 400–500 ms	3	20	0.08	0.07
B3, 0–50 ms	3	20	0.02	0.02
B2x, 300–500 ms	3	20	0.11	0.18
A3, 450–500 ms	3	20	0.04	0.03
B3, 250–300 ms	3	20	0.12	0.16

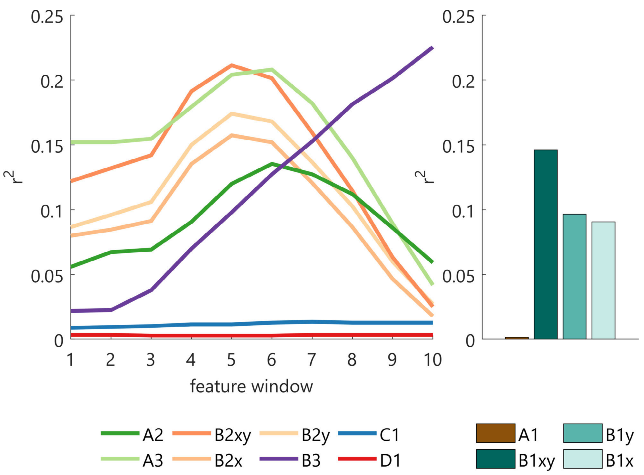


FIGURE 9. Coefficient of determination (r^2) for all features of gaze classifier (see II.E.3. for description of features A–C) on day 2.

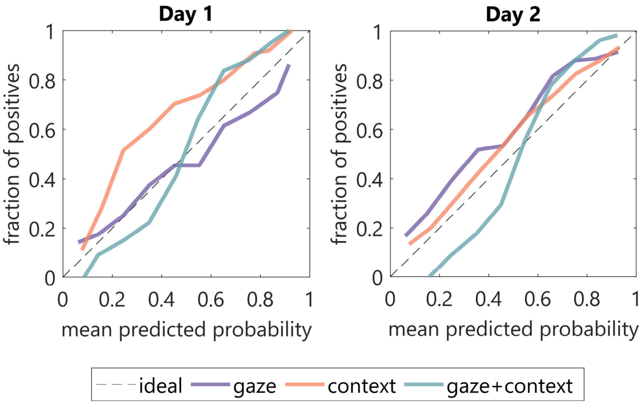


FIGURE 10. Reliability diagram for classifiers over two days. The diagram was created by averaging the predicted probabilities into 10 bins and comparing them with the corresponding fractions of positives obtained from each participant.

the element, whereas during spontaneous dwells, their gaze tended to drift away from the center toward the next point of interest. Gaze dispersion, assessed by various microsaccade metrics (A2–A3), as well as dispersion (B1) and spread (B2) of the gaze point coordinates, were also informative for distinguishing between intentional and spontaneous dwells, with the peak informativeness occurring around 250 ms. It appeared that at this point, a behavioral difference emerged: during spontaneous dwells, the gaze becomes more mobile to correct the foveation point, while during intentional dwells, the gaze freezes in anticipation of feedback.

It is also worth noting that pupil (C1) and vergence (D1) metrics were not informative.

2) ONLINE PERFORMANCE

Since the classification algorithm utilized probabilistic outputs, it is necessary to check their calibration. The reliability diagram in Fig. 10 indicates that the probability values are well-calibrated for both the gaze classifier and the context classifier.

Fig. 11 presents the online performance metrics for both classifiers and their combination on the second day of the experiment. A Friedman test with post-hoc Dunn's correction for multiple comparisons revealed that combining the classifiers with adjustable thresholds yielded the best results, outperforming the individual performance of each classifier in terms of the P4w metric (gaze classifier: $\chi^2(3) = -35$, $p < 1E-5$; contextual classifier: $\chi^2(3) = -35$, $p < 1E-5$), as well as the fixed 0.5 classifier threshold (without

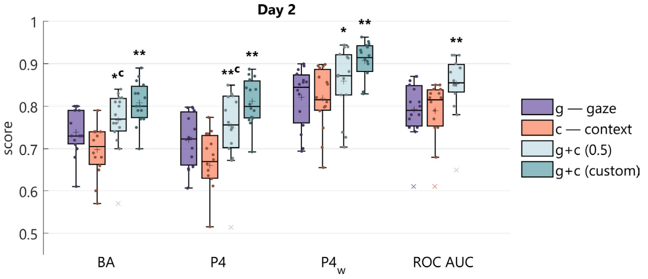


FIGURE 11. Group average (N = 15) online classifier performance metrics for Day 2 for gaze (g), context (c) classifiers alone and their combinations using flat (0.5) and custom thresholds for their average probability. Asterisks indicate statistical significance for post-hoc Dunn's test comparing the combined gaze + context classifiers (g+c) with the gaze and context classifiers: $*p < 0.05$, $**p < 0.01$.

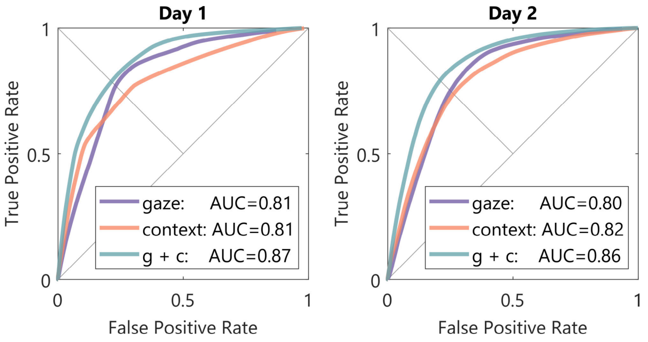


FIGURE 12. Averaged ROC curves (N = 15) for classifiers over two days. The median value of the data collected from all participants was calculated for each threshold value, ranging from 0 to 1 in increments of 0.01, using the threshold averaging method.

uncertain classifier decisions: gaze classifier: $\chi^2(3) = -19$, p = 0.04; contextual classifier: $\chi^2(3) = -19$, p = 0.04; see Fig. 11).

When all false positives (FPs) were considered, no preference was found for the gaze classifier over the combined classifiers with fixed 0.5 thresholds ($\chi^2(3) = -14$, p = 0.28). However, both classifiers performed worse than the combined classifiers with adjustable thresholds (gaze classifier: $\chi^2(3) = -30$, p = 0.0001; contextual classifier: $\chi^2(3) = -44$, p < 1E−6).

The ROC curves, shown in Fig. 12, demonstrate that the ROC AUC for the combined classifiers had significantly higher values compared to both the gaze classifier ($\chi^2(2) = -21$, p = 0.0004) and the contextual classifier ($\chi^2(2) = -24$, p = 0.00004). It is clear that combining the classifiers provided a significant improvement in accuracy across the entire range of values.

3) OFFLINE SIMULATIONS

Several offline simulations were conducted to evaluate the impact of various parameters on classifier performance. In these simulations, classifier models were trained using data collected on the first day and tested on data from the second day to more closely mirror real-world conditions.

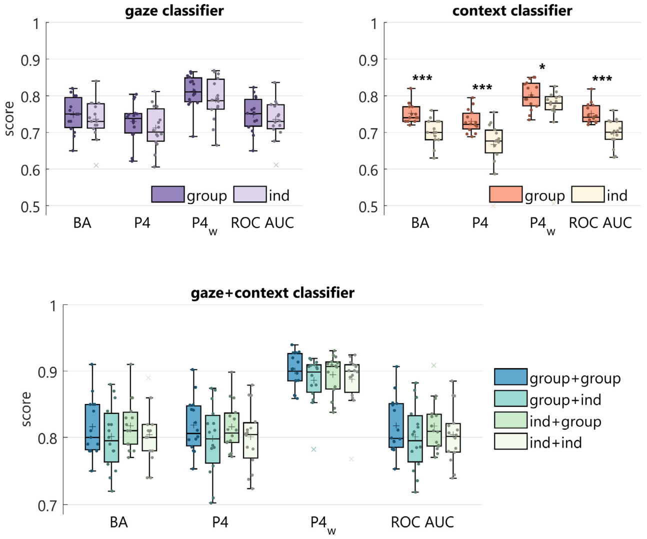


FIGURE 13. Group average (N = 15) offline performance metrics for group and individual classifiers when trained on Day 1 and tested on Day 2. Colors represent training dataset scope: "group" for all participants, "ind" for each individual participant. Asterisks indicate statistical significance: $*p < 0.05$, $***p < 0.001$ (Wilcoxon test).

A comparison of classifiers built on individual and group data (Fig. 13) revealed that the accuracy of the gaze classifier remained unchanged, whereas the context classifier performed better with group data (Wilcoxon test, p < 0.001 for BA, P4 and ROC AUC; p = 0.018 for P4*w*). This improvement is likely due to the context classifier using a relatively higher number of features, which require a larger dataset for effective training.

Another important parameter of the interface was the gaze dwell time threshold. To evaluate its impact on classifier performance, gaze coordinates were reprocessed in offline simulations using various dwell time thresholds, ranging from 200 to 700 ms. The lowest threshold corresponded to the typical fixation duration during natural gaze behavior, while the highest threshold was based on the dwell time threshold used in the online mode.

Our analysis revealed that higher dwell time thresholds improved the usefulness of gaze features for classifier performance. Fig. 14 shows an almost linear increase in the informativeness of gaze features as the analysis window lengthens, reaching its peak with a dwell time threshold of 700 ms. This suggests that the later segments of the dwell contribute more significantly to feature informativeness, which is defined as the classifier model's ability to distinguish between intentional and spontaneous dwells.

However, in a more realistic simulation, it is essential to consider that the proportion of spontaneous dwells varies significantly depending on the dwell time thresholds. As dwell duration increases, the number of spontaneous dwells decreases (Fig. 15, right panel). While the classifier is designed to reduce false positives (FP), when the percentage of spontaneous dwells is low, this issue is largely resolved on its own, rendering the classifier

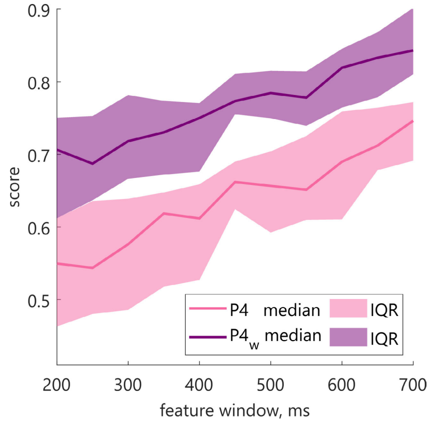


FIGURE 14. Group average (N = 15) results of a simulation of feature window size effect on P4 and P4$_W$ metrics. Same trials used for all feature window size. The shaded areas on the graphs represent the interquartile range (IQR).

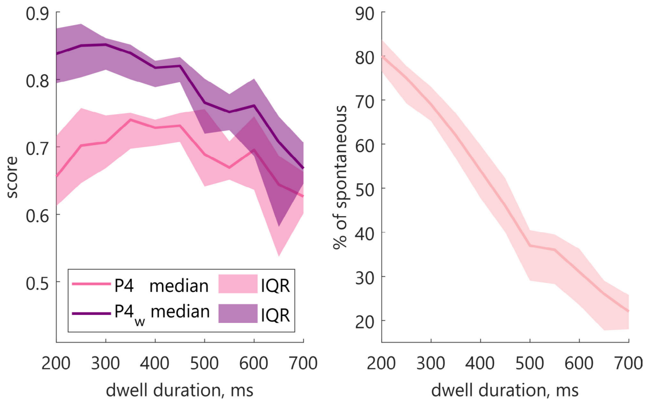


FIGURE 15. Group average (N = 15) offline classification performance assessed with P4 and P4$_W$ metrics as a function of dwell duration (left panel). Number of spontaneous dwells changes freely with dwell duration (right panel). The shaded areas on the graphs represent the interquartile range (IQR).

less useful. Additionally, the classifier introduces false negatives (FN), and the number of FN errors increases as the imbalance between spontaneous and intentional dwells grows.

As shown in Fig. 15, the best P4 score is achieved with a dwell time threshold of 350–450 ms, where the percentage of spontaneous dwells is slightly above 50%. Beyond this point, as the imbalance increases, the P4 score deteriorates. It's important to note that the P4 score takes into account all types of errors, meaning that an increase in FP (at lower dwell time thresholds) and FN (at higher dwell time thresholds) negatively impacts the overall score.

Examining the MTBF (Mean Time Between Failures) metric, it is clear that the classifier enhances performance by reducing false positives across the simulated range of dwell durations (Fig. 16). However, the relative improvement enabled by the classifier over the basic dwell-time based selection diminished as dwell duration increased, likely due to the decrease in the number of spontaneous dwells.

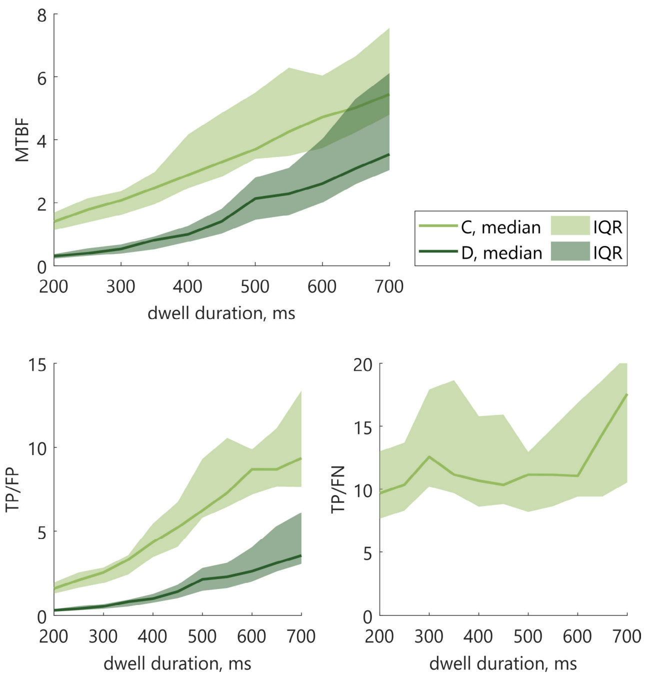


FIGURE 16. Group average (N = 15) simulation results for mean time between failures (MTBF) and its components, TP/FN and TP/FP, as a function of dwell duration. The shaded areas on the graphs represent the interquartile range (IQR).

IV. DISCUSSION

In the present study, we demonstrated that machine learning can improve effectiveness and efficiency of gaze-based interaction in a context requiring rapid transitions between visual exploration, decision-making, and making actions.

Such a context was modelled using the *EyeLines* game, where new objects entered the game board along the game, some objects remained at their positions (so memory of these positions can be exploited by the player), some objects were removed by the player successful actions, so the visual field was dynamically changing. This required significant work of vision and provoked a variety of automatic gaze behaviors, including those that mimicked gaze dwells intentionally used by the player to make moves. Moves were basically made using a relatively short, 500 ms dwell time threshold, which enabled fluent and engaging interaction but further increased the rate of false responses due to multiple spontaneous gaze dwells exceeding the threshold. We therefore considered *EyeLines* as a testbed well suited for exploring ML solutions for the Midas touch problem. In addition, players typically could freely choose from many options, so the game was also fit to test the ability of algorithms to deal with the challenges of free user behavior.

Machine learning (ML) algorithms were employed to improve the discrimination of intentional and spontaneous gaze dwells. Participants reported higher gameplay satisfaction with fewer actions required to achieve gameplay goals, reflecting the success of the proposed solutions in addressing the "Midas touch" problem. Notably, the number of unintended actions was reduced by nearly threefold compared to the baseline gaze-only input, without compromising the command rate (effective actions per minute, eAPM).

Furthermore, we overcame the challenge of the lack of ground truth data in online free-behavior experiments by developing a methodology for approximate offline labeling of gaze dwells as intentional and spontaneous. This made possible detailed offline analysis of the classification performance data from these experiments without limiting participants' freedom of decision making and, at the same time, without requiring from them explicit labeling of their intentions or interface errors. The analysis provided additional insights which may help to further improve the gaze-based interaction with a more targeted tuning of the ML-based support.

A. BEHAVIOR AND INTERACTION EFFICIENCY

The results indicate that the ML-enhanced mode (Mode C) generally allowed participants to interact more efficiently with the game, as demonstrated by longer game times and fewer unintended actions compared to the dwell-only mode (Mode D). Initially, we assumed that efficiency could be measured using the removal rate metric, since success in the *EyeLines* game is determined by the points earned from removing balls. However, no significant differences in removal rate were observed between the modes (Fig. 4, right). It seems that the removal rate is more reflective of a player's skill level, as it increases with the number of games played, particularly on Day 1 (Fig. 5).

Other metrics of efficiency, such as total game time and the number of actions required to remove a ball (Fig. 6), provide stronger evidence of Mode C's advantage. The significant differences in total game time between the modes suggest that Mode C enabled players to make more deliberate and effective moves. This was likely because players felt more comfortable when searching for the best ball, selected and moved it, without the fear of incidental events. Indeed, without a clear vision for the next move, unintentional ball selection served as a suggestion for the player, who likely accepted it to avoid spending extra time and having the risk to get further (likely worse) false selections while searching for a better target. Consequently, the nearly threefold reduction in false positives (Wilcoxon test for FP/(FP+TP) in Mode C vs. Mode D: $W(15) = 120.0$, $p = 0.00006$) reduced the number of actions needed per ball removal while simultaneously indirectly improved the quality of decisions. Interestingly, while the reduction of false positives improved higher-level performance measures, it did not affect the command rate. The command rate, measured as eAPM (effective actions per minute), represents the intended ball movements in our game (move rate, Fig. 4, left). Despite the improvements in other areas, the eAPM remained unaffected.

A possible reason for this appears when we consider metrics of time between failures, i.e., MTBF, MTBF*w* and their components. Fig. 7 shows that MTBF increased significantly, most likely due to the reduction of false positives in Mode C. The small number of false negatives introduced by the imperfect classifier did not affect the failure period. However, when failure is interpreted as the need for additional user action (as estimated by MTBF*w*), the impact of false positives is drastically reduced, making false negatives a more pronounced factor. This equalizes the MTBF*w* metrics between C and D modes. From this perspective, the task's low sensitivity to FPs was likely the main reason we did not observe an improvement in eAPM for Mode C. This suggests that the "Midas Touch" problem is only worth addressing for tasks where the cost of FPs is high (which is usually considered in gaze-based interaction design, starting at least as early as from [8] study). However, further improvement of our algorithms for the *EyeLines*-like tasks could focus on reducing FNs when developing machine learning models. Another approach could be adjusting both FP and FN rates for specific tasks and for individual users, considering task requirements and user experience.

It is important to note that the actual number of false positives (FPs) that impacted users fell between the total number of FPs (in the worst case) and FP*w* (in the best case). However, it is difficult to determine the exact value. This is likely a highly individual parameter, influenced by various factors such as the user's gaze-control strategy (e.g., avoiding looking at objects during screen observation) and their decision-making speed in this specific task.

Another reason for the low eAPM effects is the nature of the Lines game, which is a puzzle-type game where a significant amount of time is spent searching for and planning each move. However, this combination is quite typical for real-life tasks, where interaction typically constitutes only a portion of the total time spent using a computer.

It is worth noting that [12] observed a far greater improvement in command rates with their machine learning-enhanced system, likely due to the much higher dwell time in their no-ML mode. In their study, the baseline condition required extended dwell times (except for one oversimplified task), which naturally limited interaction speed and efficiency, making the improvement in the machine learning-enhanced mode more pronounced. This contrasts with our design, where the dwell time in Mode D was already optimized for quicker selections, resulting in a less dramatic, but still significant, improvement in Mode C. The subjective assessments from the questionnaires indicate that while participants appreciated the speed and efficiency of the ML-enhanced control mode C, they also reported that it occasionally interfered with their decision-making process. This feedback underscores a key tension in gaze-based interaction design: the trade-off between system automation and user control. Participants seemed to prefer the enhanced efficiency of Mode C but were also sensitive to the instances where the system's decisions conflicted with their intentions. Therefore, error awareness possibly played a major role in the subjective perception of Mode C, where the machine learning algorithm filtered out many unintentional selections before they could affect gameplay. However, participants might not have been fully aware of these "prevented errors," as they never materialized into noticeable mistakes. In contrast, when errors did occur in Mode C, they might have been more noticeable because participants expected the system to be more accurate. This could make the errors in Mode C seem more salient, leading to a perceived similarity in error rates between the two modes.

This finding is consistent with previous studies that have shown user preference for gaze-sensitive systems that offer a high degree of control and predictability, even at the cost of some efficiency [3]. The challenge for future research and development lies in further refining the systems to better align with user expectations without sacrificing the benefits of automation.

B. MACHINE LEARNING

Similar to our previous report [17] on offline classification of the eye movement data recorded during gaze-based interaction, the most useful features for distinguishing between intentional and spontaneous dwells were the distance from the center of the dwell to the center of the selected ball, the spread of gaze coordinates, and the total amplitude of microsaccades during the dwell. These features became more effective when calculated over specific segments of the analysis window, rather than the entire window. Moreover, increasing the size of the analysis window improved classifier performance, as longer dwell times allowed for the extraction of more informative gaze features.

However, the need for a classifier at higher dwell time thresholds becomes questionable. Based on the collected data, we were able to simulate performance up to a dwell time threshold of 700 ms (due to the online dwell time of 500 ms). Even at this threshold, the proportion of spontaneous dwells decreased to around 20%, which significantly reduced the occurrence of false positives, effectively mitigating the Midas Touch problem on its own. With such a significant class imbalance, false negatives should become a more serious issue for users when a classifier is applied. As a result, the benefits of longer dwell windows—despite their increased informativeness—are outweighed by diminishing returns in classifier utility. Nonetheless, in critical applications such as medical or assistive technologies, extending dwell thresholds to 700–800 ms may provide a simple and robust alternative to potentially unreliable machine learning components by naturally reducing false activations. In contrast, consumer-oriented interfaces may favor shorter, more responsive dwell times (400–500 ms), where higher false positive rates can be effectively controlled through the use of a trained classifier to maintain fluid interaction.

Features based on pupil and binocular data, which were not used in the online experiment, were explored in offline simulations (not included in the Results) but did not improve classifier performance. This was not unexpected, as pupil data are very sensitive to many factors that vary in different directions and may result in a low signal-to-noise ratio for this feature (see also comments on [16] in Introduction). Eye vergence (reflected in the binocular features) was shown to be very sensitive to mind wandering in some tasks [27], but the engaging nature of our task likely made mind wandering rare.

In addition to the gaze classifier, a contextual classifier was employed to enhance interface performance. The features of the contextual classifier were specifically developed for the game environment but can be adapted to other gaze-based interaction tasks by redefining contextual cues relevant to another interface, for example, object visibility, task-relevant regions, or timing constraints. Since the classifier architecture is agnostic to the specific feature set, these contextual features can be re-derived in other domains without altering the underlying model or training procedure. Therefore, the use of task-specific context features does not limit the generalizability of our results, but rather illustrates how context-awareness can be flexibly integrated into intention inference across diverse applications.

In the current study, combining the outputs of both classifiers by averaging their predicted probabilities yielded significantly better results compared to using either classifier alone. The advantage of averaging probabilities likely came from cases where neither classifier was fully confident in its prediction. In such cases, the final decision was weighted toward the classifier with stronger confidence. When both classifiers were uncertain or predicted opposing classes, the average probability tended to be close to 0.5, indicating that the final decision cannot be made with high certainty. By introducing probability thresholds, we further improved the algorithm by flagging these uncertain cases and resolving them through alternative methods, such as applying a higher dwell time threshold.

In offline simulations, other approaches to combining classifier models were explored (beyond the scope of this paper), and the algorithm used here seemed to provide the best overall performance. The benefit of integrating the contextual classifier is evident in the ROC curves (Fig. 12), particularly in the region of lower false positive rates (FPR), especially on the second day. This indicates that the combined classifier achieved higher true positive rates (TPR), reducing false negative (FN) errors while maintaining a lower FPR. This is particularly important in situations with a relatively high proportion of spontaneous dwells.

Interestingly, the gaze classifier performed equally well when trained on group data compared to individual data, and the contextual classifier performed even better with group data. These findings are significant because they suggest that individual datasets are not necessary for training unique classifier models, simplifying the application of ML approaches to gaze-based interactions.

However, it is important to maintain consistency between the tasks in the training dataset and the tasks in the online mode where the classifier will be applied. As seen in the reliability diagram from the first day (Fig. 10), the contextual classifier performed poorly, resulting in a highly asymmetric ROC curve relative to the left diagonal (Fig. 12). This issue likely arose because the training dataset was collected from a slightly different game environment in our previous study.

C. VIEWING GAZE "OUTPUT" AS A SOCIAL FUNCTION

For future development of the ML-based enhancement of gaze-based interaction, it could be fruitful to learn more about the natural basis of human's ability to direct and hold their gaze intentionally. The origin of the ability to intentionally control the gaze direction possibly lies in social interaction. Gaze direction can be approximately estimated without any tools, just by visual observation, so it is natural that humans are not only aware of it but also can actively control it to hide the focus of their attention. Moreover, gaze is used to communicate attention (as in joint attention, [28], [29], [30]) and other social signals (e.g., direct gaze, [31], [32], [33]).

The Midas touch problem is minor in most scenarios of social interaction, where gaze is only rarely used and where its use is highly automatized. It does not seem to be severe in gaze typing, where gaze-sensitive area is limited to a keyboard: due to its fixed visual appearance the user does not need to use vision extensively while looking at it. However, spaces where vision and control should be used cannot be easily separated in many scenarios of human-machine interaction, e.g., web browsing, robot control, gaming. In such cases, solving the Midas touch problem is mandatory for effective and efficient gaze-based control. Fortunately, we may expect a difference in gaze features between gaze use for "input" and "output" in all such cases, so ML-based approaches could provide a solution.

D. COMBINATION WITH XR AND BCIS

In recent years, increasing attention has been attracted to gaze-based interaction for extended reality (XR), which encompasses virtual, augmented and mixed reality technologies (VR/AR/MR). For XR, input technologies developed for computers and smartphones are not well suited; dedicated XR controllers need to be carried in hands, and the use of gestures and voice is inconvenient in crowded environments, while input by gaze is free from such limitations [34]. XR helmets and glasses also provide a convenient platform for capturing and processing gaze data needed for gaze-based input. Release of a commercial Apple Vision Pro XR headset has demonstrated the feasibility and practical prospects of gaze-based selection in such an environment [35]. However, existing solutions still require using confirmatory hand gestures together with gaze-based control. On-the-fly recognition of the user intention based on gaze features, without confirmatory gestures, could significantly advance this technology.

Interestingly, first attempts to enhance gaze-based control with ML were undertaken using features extracted not from gaze data but from brain activity data [19], [20], [21], [22], [36], [37]. They were based on the idea of passive brain-computer interfaces (BCIs), which recognize patterns of brain activity without requiring the user to do anything specific to trigger these patterns [11]. The use of brain data in such "eye-brain-computer interfaces" (EBCIs) has many disadvantages, such as the need for additional equipment. However, dry EEG elctrodes can be easily placed within an XR helmet, which would help to make the combined technology more appealing to the users. One study already demonstrated recognition of gaze dwells used for control within an XR helmet [20]. It should be noted that most of such EBCI studies were run offline, while the only online test of this technology showed unsatisfactory performance, rising questions about possibly insufficient specificity of the brain markers used in these studies [23]. Nevertheless, the EBCI technology may become more effective when combined with the recognition of user's intent based on gaze features. Recently described new brain markers of the intentional gaze use during gaze-based interaction [38] may appear more specific to intentional dwelling and may provide additional opportunities for the development of such combined technologies.

E. LIMITATIONS AND FUTURE WORK

This study, while providing valuable insights into gaze-based interaction using the *EyeLines* game, is subject to several limitations that should be acknowledged.

1) SINGLE TASK FOCUS

The research concentrated exclusively on the *EyeLines* game as the model task. While this approach allowed for in-depth analysis, it also limits the generalizability of the findings. The performance and effectiveness of the proposed gaze-based interaction techniques may vary across different applications and user groups. Evidently, a broader range of tasks and environments should be explored to validate the generalizability of the approaches developed in the current study.

2) SMALL SAMPLE SIZE

The study involved a relatively small sample of participants, which may have constrained the robustness of the machine learning classifiers used. A larger sample size would likely provide more diverse data, enhancing the generalizability and reliability of the results. Higher volumes of data would also likely result in better classifier training. Moreover, only healthy young participants were involved in our study. For developing a practical technology, different groups of end users should be involved in testing; specific characteristics of their eye movements and their specific needs should be addressed in further algorithm development and parameter tuning.

3) NO SINGLE CLASSIFIER CONDITIONS IN THE ONLINE STUDY

Due to time constraints, we studied in the online tests only a combination of the gaze and contextual classifiers. This approach provided a better approximation of possible use cases, where both types of classifiers should be used whenever possible but made possible only approximate estimation of the contribution of each of them in the offline study. Nevertheless, this circumstance does not affect the main conclusions from this study. In the future, specific results can be estimated more accurately in tests where each classifier is applied alone.

4) LABEL NOISE

Our approximate ground-truth labeling method may introduce residual misclassifications, potentially biasing classifier training and inflating performance errors (e.g., balanced accuracy, P4). Although we mitigated this by excluding ambiguous events and balancing class distributions, future studies could employ external validation or participant self-reporting methods to improve the accuracy of intention annotations. Such enhancements may be particularly beneficial for participants exhibiting atypical interaction patterns.

5) SIMPLE DWELL-TIME ALGORITHM AS THE BASELINE

As we mentioned in the Introduction, many solutions to the Midas touch problem were proposed. A comprehensive test of a new solution should, of course, include at least a comparison with some of the best of them. In this work, we only compared the ML-enhanced intention detection with (1) the detection based on the most basic dwell time criterion and with (2) selection confirmation using an additional gaze dwell; in the second case, the data were taken from our previous study. However, for most of the existing solutions, it is evident that they would lead to significantly slower operation and could not be serious competitors. Other approaches could be combined with ML-based enhancement, resulting in even better performance.

6) USE OF A HIGH-GRADE EYE TRACKER

In this study, we used the high-precision EyeLink 1000 Plus eye tracker with a 1000 Hz sampling rate and head stabilization via chinrest. This setup enabled accurate analysis of eye movement data but limited the practicality of the setup for real-world gaze-based interaction. Most of the features we employed appear to be computable from data collected using more affordable and portable/wearable eye trackers, though this assumption requires further validation in future studies.

7) ABSENCE OF COMPLEX, ADAPTIVE CLASSIFIERS

The study did not employ advanced machine learning methods, such as neural networks, which are capable of adaptive learning and handling more complex data patterns. The use of simpler classifiers, while effective in this context, may not fully capture the intricacies of gaze behavior in more dynamic or varied environments [12]. Future work could explore the potential benefits of incorporating such advanced techniques to improve the adaptability and accuracy of gaze-based interaction systems.

We believe that these limitations do not undermine the significance of the current study but rather highlight areas for future research. Expanding the task scope, increasing the sample size, and exploring more sophisticated machine learning techniques could further enhance the applicability and impact of gaze-based interaction systems.

V. CONCLUSION

In this study, we explored the effectiveness of ML-enhanced gaze-based interaction within a high-paced gaming environment. Our approach combined traditional gaze dwell time with a machine learning algorithm designed to differentiate between intentional and spontaneous gaze actions. The results demonstrated a significant reduction in unintended actions and enhanced overall gameplay efficiency, without negatively impacting the command rate. Participants reported improved user experience and satisfaction, indicating the potential of this approach for more intuitive and ergonomic gaze-based interfaces.

The research gaze-controlled game used in this study, the *EyeLines*, proved to be an effective model for investigating the challenges of mixed gaze use for both interaction and vision, offering valuable insights into the practical application of gaze-based controls in real-time scenarios. Our findings contribute to the broader discussion on overcoming the Midas touch problem, a common challenge in gaze-based systems, by showing that machine learning can play a crucial role in refining the accuracy and usability of these interfaces.

ACKNOWLEDGMENT

The authors thank Yury Nuzhdin for his assistance with software modifications and valuable advice.

REFERENCES

[1] P. Majaranta, K.-J. Räihä, A. Hyrskykari, and O. Špakov, "Eye movements and human–computer interaction," in *Eye Movement Research: An Introduction to its Scientific Foundations and Applications*, C. Klein and U. Ettinger, Eds., Cham, Switzerland, Springer, 2019, pp. 971–1015, doi: 10.1007/978-3-030-20085-5_23.

[2] A. T. Duchowski, "A breadth-first survey of eye-tracking applications," *Behav. Res. Methods, Instrum., Comput.*, vol. 34, no. 4, pp. 455–470, Nov. 2002, doi: 10.3758/bf03195475.

[3] A. T. Duchowski, "Gaze-based interaction: A 30 year retrospective," *Comput. Graph.*, vol. 73, pp. 59–69, Jun. 2018, doi: 10.1016/j.cag.2018.04.002.

[4] P. Majaranta and A. Bulling, "Eye tracking and eye-based human-computer interaction," in *Advances in Physiological Computing*, Cham, Switzerland, Springer, 2014, pp. 39–65, doi: 10.1007/978-1-4471-6392-3_3.

[5] P. Majaranta and K.-J. Räihä, "Twenty years of eye typing: Systems and design issues," in *Proc. Symp. Eye Tracking Res. Appl.*, 2002, p. 15, doi: 10.1145/507072.507076.

[6] J. M. Findlay and I. D. Gilchrist, *Active Vision: The Psychology of Looking and Seeing*. London, U.K.: Oxford Univ. Press, 2003, doi: 10.1093/acprof:oso/9780198524793.001.0001.

[7] J. E. Hoffman, *Visual Attention and Eye Movements Attention*. London, U.K.: Psychology Press, 1998, pp. 119–153, doi: 10.4324/9781315784762.

[8] R. J. K. Jacob, "What you look at is what you get: Eye movement-based interaction techniques," in *Proc. SIGCHI Conf. Human Factors Comput. Syst. Empowering People*, 1990, pp. 11–18, doi: 10.1145/97243.97246.

[9] B. Velichkovsky, A. Sprenger, and P. Unema, "Towards gaze-mediated interaction: Collecting solutions of the 'Midas touch problem,'" in *Proc. 13th Int. Conf. Hum.-Comput. Interact.*, Sydney, NSW, Australia. 1997: Springer, 1997, pp. 509–516, doi: 10.1007/978-0-387-35175-9_77.

[10] B. J. Hou, P. Bekgaard, S. MacKenzie, J. P. P. Hansen, and S. Puthusserypady, ''GIMIS: Gaze input with motor imagery selection,'' in *Proc. ACM Symp. Eye Tracking Res. Appl.*, Jun. 2020, pp. 1–10, doi: 10.1145/3379157.3388932.

[11] T. O. Zander and C. Kothe, ''Towards passive brain–computer interfaces: Applying brain–computer interface technology to human–machine systems in general,'' *J. Neural Eng.*, vol. 8, no. 2, Apr. 2011, Art. no. 025005, doi: 10.1088/1741-2560/8/2/025005.

[12] A. S. Narkar, J. J. Michalak, C. E. Peacock, and B. David-John, ''GazeIntent: Adapting dwell-time selection in VR interaction with real-time intent modeling,'' *Proc. ACM Human-Comput. Interact.*, vol. 8, no. ETRA, pp. 1–18, May 2024, doi: 10.1145/3655600.

[13] R. Bednarik, H. Vrzakova, and M. Hradis, ''What do you want to do next: A novel approach for intent prediction in gaze-based interaction,'' in *Proc. Symp. Eye Tracking Res. Appl.*, Mar. 2012, pp. 83–90, doi: 10.1145/2168556.2168569.

[14] X.-L. Chen and W.-J. Hou, ''Gaze-based interaction intention recognition in virtual reality,'' *Electronics*, vol. 11, no. 10, p. 1647, May 2022, doi: 10.3390/electronics11101647.

[15] B. David-John, C. Peacock, T. Zhang, T. S. Murdison, H. Benko, and T. R. Jonker, ''Towards gaze-based prediction of the intent to interact in virtual reality,'' in *Proc. ACM Symp. Eye Tracking Res. Appl.*, May 2021, pp. 1–7, doi: 10.1145/3448018.3458008.

[16] T. Isomoto, S. Yamanaka, and B. Shizuki, ''Dwell selection with ML-based intent prediction using only gaze data,'' in *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, Sep. 2022, vol. 6, no. 3, pp. 1–21, doi: 10.1145/3550301.

[17] Y. G. Shevtsova, A. N. Vasilyev, and S. L. Shishkin, ''Machine learning for gaze-based selection: Performance assessment without explicit labeling,'' in *Int. Conf. Hum.-Comput. Interact.*, Cham, Switzerland, Springer, 2023, pp. 311–322, doi: 10.1007/978-3-031-48038-6_19.

[18] D. C. Jangraw, J. Wang, B. J. Lance, S.-F. Chang, and P. Sajda, ''Neurally and ocularly informed graph-based models for searching 3D environments,'' *J. Neural Eng.*, vol. 11, no. 4, Aug. 2014, Art. no. 046003, doi: 10.1088/1741-2560/11/4/046003.

[19] A. O. Ovchinnikova, A. N. Vasilyev, I. P. Zubarev, B. L. Kozyrskiy, and S. L. Shishkin, ''MEG-based detection of voluntary eye fixations used to control a computer,'' *Frontiers Neurosci.*, vol. 15, Feb. 2021, Art. no. 619591, doi: 10.3389/fnins.2021.619591.

[20] G. S. R. Reddy, M. J. Proulx, L. Hirshfield, and A. Ries, ''Towards an eye-brain–computer interface: Combining gaze with the stimulus-preceding negativity for target selections in XR,'' in *Proc. CHI Conf. Human Factors Comput. Syst.*, May 2024, pp. 1–17, doi: 10.1145/3613904.3641925.

[21] S. L. Shishkin, Y. O. Nuzhdin, E. P. Svirin, A. G. Trofimov, A. A. Fedorova, B. L. Kozyrskiy, and B. M. Velichkovsky, ''EEG negativity in fixations used for gaze-based control: Toward converting intentions into actions with an eye-brain–computer interface,'' *Frontiers Neurosci.*, vol. 10, p. 528, Nov. 2016, doi: 10.3389/fnins.2016.00528.

[22] D. G. Zhao, A. N. Vasilyev, B. L. Kozyrskiy, E. V. Melnichuk, A. V. Isachenko, B. M. Velichkovsky, and S. L. Shishkin, ''A passive BCI for monitoring the intentionality of the gaze-based moving object selection,'' *J. Neural Eng.*, vol. 18, no. 2, Apr. 2021, Art. no. 026001, doi: 10.1088/1741-2552/abda09.

[23] Y. O. Nuzhdin, S. L. Shishkin, A. A. Fedorova, B. L. Kozyrskiy, A. A. Medyntsev, E. P. Svirin, O. V. Korsun, I. A. Dubynin, A. G. Trofimov, and B. M. Velichkovsky, ''Passive detection of feedback expectation: Towards fluent hybrid eye–brain–computer interfaces,'' in *Proc. GBCIC*, 2017, pp. 1–15, doi: 10.3217/978-3-85125-533-1-66.

[24] A. Carrera-Rivera, D. Reguera-Bakhache, F. Larrinaga, and G. Lasa, ''Exploring the transformation of user interactions to adaptive human-machine interfaces,'' in *Proc. 23rd Int. Conf. Human Comput. Interact.*, Sep. 2023, pp. 1–7, doi: 10.1145/3612783.3612807.

[25] Y. O. Nuzhdin, ''Resonance-a BCI framework for working with multiple data sources,'' in *Proc. GBCIC*, Graz, Austria, 2019, pp. 77–81, doi: 10.3217/978-3-85125-682-6-15.

[26] M. Sitarz, ''Extending F1 metric, probabilistic approach,'' *Adv. Artif. Intell. Mach. Learn.*, vol. 3, no. 2, pp. 1025–1038, 2023, doi: 10.54364/aaiml.2023.1161.

[27] M. X. Huang, J. Li, G. Ngai, H. V. Leong, and A. Bulling, ''Moment-to-Moment detection of internal thought during video viewing from eye vergence behavior,'' in *Proc. 27th ACM Int. Conf. Multimedia*, Oct. 2019, pp. 2254–2262, doi: 10.1145/3343031.3350573.

[28] P. Chevalier, K. Kompatsiari, F. Ciardo, and A. Wykowska, ''Examining joint attention with the use of humanoid robots–A new approach to study fundamental mechanisms of social cognition,'' *Psychonomic Bull. Rev.*, vol. 27, no. 2, pp. 217–236, Apr. 2020, doi: 10.3758/s13423-019-01689-4.

[29] B. Siposova and M. Carpenter, ''A new look at joint attention and common knowledge,'' *Cognition*, vol. 189, pp. 260–274, Aug. 2019, doi: 10.1016/j.cognition.2019.03.019.

[30] M. Tomasello and M. J. Farrar, ''Joint attention and early language,'' *Child Develop.*, vol. 57, no. 6, p. 1454, Dec. 1986, doi: 10.2307/1130423.

[31] R. B. Adams and R. E. Kleck, ''Effects of direct and averted gaze on the perception of facially communicated emotion,'' *Emotion*, vol. 5, no. 1, pp. 3–11, Mar. 2005, doi: 10.1037/1528-3542.5.1.3.

[32] A. Senju and T. Hasegawa, ''Direct gaze captures visuospatial attention,'' *Vis. Cognition*, vol. 12, no. 1, pp. 127–144, Jan. 2005, doi: 10.1080/13506280444000157.

[33] A. F. D. C. Hamilton, ''Gazing at me: The importance of social meaning in understanding direct-gaze cues,'' *Phil. Trans. Roy. Soc. B, Biol. Sci.*, vol. 371, no. 1686, Jan. 2016, Art. no. 20150080, doi: 10.1098/rstb.2015.0080.

[34] A. Plopski, T. Hirzle, N. Norouzi, L. Qian, G. Bruder, and T. Langlotz, ''The eye in extended reality: A survey on gaze interaction and eye tracking in head-Worn extended reality,'' *ACM Comput. Surv.*, vol. 55, no. 3, pp. 1–39, Mar. 2023, doi: 10.1145/3491207.

[35] J. O'Callaghan, ''Apple vision pro: What does it mean for scientists?'' *Nature*, vol. 2024, pp. 1–13, Feb. 2024, doi: 10.1038/d41586-024-00387-z.

[36] K. Ihme and T. O. Zander, ''What you expect is what you get? Potential use of contingent negative variation for passive BCI systems in gaze-based HCI,'' in *Proc. Int. Conf. Affect. Comput. Intell. Interact.*, Cham, Switzerland, Springer, Jan. 2011, pp. 447–456, doi: 10.1007/978-3-642-24571-8_57.

[37] J. Protzak, K. Ihme, and T. O. Zander, ''A passive brain–computer interface for supporting gaze-based human-machine interaction,'' in *Proc. Int. Conf. Universal Access Human-Comput. Interact.*, Las Vegas, NV, USA. Springer, 2013, pp. 662–671, doi: 10.1007/978-3-642-39188-0_71.

[38] A. N. Vasilyev, E. Svirin, I. A. Dubynin, A. Butorina, Y. O. Nuzhdin, A. Ossadtchi, T. A. Stroganova, and S. L. Shishkin, ''Intentionally vs. Spontaneously prolonged gaze: A MEG study of active gaze-based interaction,'' *bioRxiv*, Dec. 2024, doi: 10.1101/2024.12.11.627776.

YULIA G. SHEVTSOVA received the B.S. degree in biomedical engineering from Bauman Moscow State Technical University, Russia, in 2021, and the M.S. degree in applied physics and mathematics from Moscow Institute of Physics and Technology, Russia, in 2023. She is currently pursuing the Ph.D. degree in physiology with M. V. Lomonosov Moscow State University, Russia.

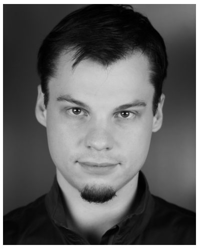
Since 2022, she has been a Junior Research Scientist with the Neurocognitive Interfaces Group, MEG Center, Moscow State University of Psychology and Education. Her research interests include the development of gaze-controlled interfaces using machine learning techniques and active brain-computer interfaces based on motor imagery and quasi-movements.

ARTEM S. YASHIN received the B.S. and M.S. degrees in applied physics and mathematics from Moscow Institute of Physics and Technology, Russia, in 2018 and 2020, respectively, and the Ph.D. degree in philosophy from M.V. Lomonosov Moscow State University, Russia, in 2024.

From 2020, he is a Junior Research Scientist at Neurocognitive Interfaces Group, MEG Center, Moscow State University of Psychology and Education. His research interests include the exploration of the sense of agency, particularly in the field of brain-computer interfaces (BCI). He has extensive experience in designing and conducting psychophysiological experiments. One of the recent topics of his research is the study of quasi-movements and their potential application in neurorehabilitation BCI.

SERGEI L. SHISHKIN received the diploma degree in physiology and the Ph.D. degree in physiology from Lomonosov Moscow State University, Moscow, Russia, in 1990 and 1997, respectively.

In 2003–2005, he worked at RIKEN Brain Science Institute, Japan. After returning to Russia, he joined the Laboratory for Neurophysiology and Neurocomputer Interfaces at M.V. Lomonosov Moscow State University. From 2011 to 2020, he worked at the National Research Center ''Kurchatov Institute'', where he established the Laboratory for Neuroergonomics and Brain-Computer Interfaces, later transformed into the Department for Neurocognitive Technologies. From 2020, he is a Leading Research Scientist at MEG Center, Moscow State University of Psychology and Education, where he heads the Neurocognitive Interfaces Group. His research interests include the development of advanced interfaces for rapid brain- and gaze-based control of computers and robotic systems, investigation of the neural mechanisms enabling such control, and the study of consciousness.



ANATOLY N. VASILYEV received the diploma degree in physiology and the Ph.D. degree in physiology from Lomonosov Moscow State University, Moscow, Russia, in 2013 and 2018, respectively.

Since 2019, he has been a Senior Research Scientist with the Laboratory for Neurophysiology and Neuro-Computer Interfaces, M. V. Lomonosov Moscow State University, and the Neurocognitive Interfaces Group, MEG Center, Moscow State University of Psychology and Education. His research interests include neurophysiological mechanisms underlying motor actions and their mental simulation, with applications in neurorehabilitation. He has developed several novel techniques for analyzing EEG and MEG data and holds patents related to the design of neurocomputer interaction protocols. He also has extensive experience with a range of neuroimaging and stimulation methods, including EEG, MEG, transcranial magnetic stimulation (TMS), functional near-infrared spectroscopy (fNIRS), and eye tracking.