# Reanalyzing Effective Eye-related Information for Developing User's Intent Detection Systems

TOSHIYA ISOMOTO, University of Tsukuba, JAPAN
SHOTA YAMANAKA, Yahoo Japan Corporation, JAPAN
BUNTAROU SHIZUKI, University of Tsukuba, JAPAN

Studies on gaze-based interactions have utilized natural eye-related information to detect user intent. Most use a machine learning-based approach to minimize the cost of choosing appropriate eye-related information. While those studies demonstrated the effectiveness of an intent detection system, understanding which eye-related information is useful for interactions is important. In this paper, we reanalyze how eye-related information affected the detection performance of a previous study to develop better intent detection systems in the future. Specifically, we analyzed two aspects of dimensionality reduction and adaptation to different tasks. The results showed that saccade and fixation are not always useful, and the direction of gaze movement could potentially cause overfitting.

CCS Concepts: • **Human-centered computing → Human computer interaction (HCI)**; User models.

Additional Key Words and Phrases: Gaze-based interaction, eye information, machine-learning, feature selection

**ACM Reference Format:**
Toshiya Isomoto, Shota Yamanaka, and Buntarou Shizuki. 2023. Reanalyzing Effective Eye-related Information for Developing User's Intent Detection Systems. In *2023 Symposium on Eye Tracking Research and Applications (ETRA '23), May 30-June 2, 2023, Tubingen, Germany*. ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3588015.3589508

## 1 INTRODUCTION

Studies on gaze-based and gaze-combined interactions have utilized natural eye behavior to understand user intent. For instance, in addition to early research that focused on gaze coordinates, researchers have used eye-related information, such as saccade, fixation, pupillary, and vergence, which can be calculated from the data sampled by eye-trackers, to detect a user's intent to interact [Bednarik et al. 2012; David-John et al. 2021; Isomoto et al. 2022; Peacock et al. 2022; Sendhilnathan et al. 2022]. While eye-related information is useful for interactive systems, it can be challenging to determine which information to use because it varies between users and interaction situations. Many of these studies have used machine learning (ML) to develop user intent detection systems to minimize the cost of choosing the most appropriate eye-related information. Previous research has examined how different eye-related information can be used to improve detection performance by illustrating the difference of eye-related information between users with and without intent to interact [Isomoto et al. 2022] and investigating the effectiveness of saccade and fixation individually or in combination [David-John et al. 2021; Peacock et al. 2022; Sendhilnathan et al. 2022]. However, detailed discussions on whether a detection system could be effective in other interaction situations are lacking. For example, in real-world interactions, tasks would differ from those for which these detection systems were developed in terms of the experimental environment. Thus, these ML-based systems could potentially overfit the eye-related information from *experiments*. In the case of overfitting, a smaller number of eye-related information used for ML is preferred. Therefore, exploring how eye-related information can positively and negatively affect the detection system's performance is important.

Table 1. Eye-related information used as features of the ML model of [Isomoto et al. 2022].

| Features | | Numbers |
|---|---|---|
| plus, minus, absolute, and all (19) values of changes in $\mathbf{x}$, $\mathbf{y}$, $\mathbf{diff_x}$, and **pupil** | average, standard deviation (SD), amplitude, skewness, kurtosis | 80 (4×4×5) |
| durations of saccades, durations of fixations, distances of saccades, distances of fixations, velocities of saccades | average, first value, last value, last value minus first value, minimum value, max value, amplitude | 35 (5×7) |
| Changes in $\mathbf{x}$, $\mathbf{y}$, $\mathbf{diff_x}$, and **pupil** | 1st value, 19th value, difference between 19th and 1st values | 12 (4×3) |

In this paper, we reanalyze effective eye-related information for intent detection performance in different interaction situations. Due to the myriad of interaction situations, it is challenging to examine them all exhaustively. As a first step in understanding eye-related information for interactions, we use the dataset of [Isomoto et al. 2022], which comprises horizontal and vertical gaze movement, saccade, fixation, vergence distance, and pupillary response, to explore how detection performance changes when different eye-related information is used. Specifically, we focus on the perspective of dimensionality reduction and adaptation to different tasks. To our knowledge, no other research has used a more varied set of eye-related information than [Isomoto et al. 2022], so we decided to utilize their data for our analysis.

## 2 OVERVIEW OF PREVIOUS STUDY

We first summarize the eye-related information used in [Isomoto et al. 2022], including $\mathbf{x}$ and $\mathbf{y}$ gaze coordinates on the screen, **pupil** which is the diameter of the pupil; $\mathbf{diff_x}$, which indicates the focus distance; and saccade and fixation calculated using $\mathbf{x}$ and $\mathbf{y}$. This information was collected from 24 participants performing five different tasks: letter selection, word selection, sentence selection, image selection, and movie watching, and was used to detect a user's intent to select a target they were dwelling on.

By using the above eye-related information, they calculated features for the ML model (shown in Table. 2). Their ML model achieved an overall area under the curve (AUC) of the receiver operating characteristic curve of 0.903 and an AUC value of 0.898 for leave-one-participant-out cross-validation. While their model demonstrates generalizability, there is a lack of discussion on which eye-related information is important for intent detection. Therefore, a detailed discussion on how eye-related information affects detection is necessary, considering dimensionality reduction can help reduce prediction costs and prevent overfitting in an ML model. Additionally, leave-one-task-out cross-validation resulted in an AUC of 0.601, suggesting that the model overfitted the experimental tasks.

## 3 REANALYZING EYE-RELATED INFORMATION

We reanalyze effective eye-related information for interaction situations by using a dataset of [Isomoto et al. 2022].

### 3.1 Re-Evaluating Eye-Related Information for Dimensionality Reduction

We investigated the use of eye-related information to improve the model of [Isomoto et al. 2022] from the perspective of dimensionality reduction. While many studies use saccade and fixation information as a primary feature indicating a user's intent, [Isomoto et al. 2022] found that gaze movement and pupil changes were important to detect user intent in their model. Among the top 30 feature gains (reported in the supplemental material of [Isomoto et al. 2022]), only peak-to-peak saccade distance was raised as an important feature in saccade information. This suggests that, for their model, saccade and fixation information may not be more important than other eye-related features.

Thus, we re-trained an ML model with features that excluded saccade and fixation information and found that the overall AUC improved from 0.903 to 0.904 while reducing the number of features from 127 to 92. While it is difficult to clearly explain the specific reasons behind the performance improvement due to the use of an ML approach, we speculate that the saccades behavior did not differ significantly between the negative and positive classes. As one example, the distance of saccades was 4.5° and 3.9° in the negative and positive classes, respectively. 1.0° was equal to 1.1 mm in their experimental environment, and 0.6° of difference may account for the small difference. Note that [Isomoto et al. 2022] used under 5.0° as the target size, which may account for both under 5.0° and small difference (approx. 0.6°) in the distance. For an ML model detecting a user's intent that is unrelated to dwelling, such as [Sendhilnathan et al. 2022], the saccade and fixation counts may be more varied than those in [Isomoto et al. 2022]. Therefore, despite many studies using saccade and fixation information as indicators of a user's intent, it is important to carefully consider the inclusion of such features.

### 3.2 Adaptation to Different Task

[Isomoto et al. 2022] found that their model suffered from overfitting to the tasks, as evidenced by the inadequate results of their leave-one-task-out cross-validation (AUC=0.601). In our study, we examine the use of eye-related information in relation to tasks. [Isomoto et al. 2022] utilized the gaze movement direction was used as a feature, represented by plus and minus values of the $\mathbf{x}$ and $\mathbf{y}$, to indicate gaze movement directions. However, these values may be influenced by various factors, including the type of content being viewed, the aspect ratio of the interface, its size, and the arrangement of content. Therefore, it is essential to carefully consider the use of gaze movement direction in relation to tasks.

Consequently, we re-trained an ML model by excluding plus and minus values of changes in $\mathbf{x}$ and $\mathbf{y}$. As a result, the AUC improved to 0.627 in the leave-one-task-out cross-validation, up from the initial AUC of 0.601; the overall AUC increased to 0.913 from 0.903. Although this improvement is still insufficient, excluding eye-related information that depends on the task is one potential solution for overfitting that we should consider while developing an ML model.

## 4 CONCLUSION

In this paper, we conducted a reanalysis of effective eye-related information to develop a user's intent detection system, utilizing the dataset from [Isomoto et al. 2022]. Our analysis yielded two significant findings that contribute to the consideration of which eye-related information is best suited for further development of gaze-based interaction. Firstly, saccade and fixation, which are commonly used as primary features to indicate a user's intent, may be subject to dimensionality reduction in their system. Secondly, the direction of gaze movement has the potential to cause overfitting in the presence of different tasks. While we focused on reanalyzing eye-related information in this paper, there remains a concern about the system's applicability developed with the information presented in Section 3 to real-world interaction scenarios. Nevertheless, as a first step in understanding eye-related information for interactions, we are confident that our discussions will assist future studies in developing an intent detection system using eye-related information.

## ACKNOWLEDGMENTS

This work is supported by JSPS, KAKENHI Grant Number 21J10301, Tateisi Science and Technology Foundation.

## REFERENCES


Roman Bednarik, Hana Vrzakova, and Michal Hradis. 2012. What Do You Want to Do Next: A Novel Approach for Intent Prediction in Gaze-based Interaction. In *Proceedings of the 2012 ACM Symposium on Eye Tracking Research & Applications* (Santa Barbara, California) *(ETRA '12)*. Association for Computing Machinery, New York, NY, USA, 83–90. https://doi.org/10.1145/2168556.2168569

Brendan David-John, Candace Peacock, Ting Zhang, T. Scott Murdison, Hrvoje Benko, and Tanya R. Jonker. 2021. Towards Gaze-Based Prediction of the Intent to Interact in Virtual Reality. In *ACM Symposium on Eye Tracking Research and Applications* (Virtual Event, Germany) *(ETRA '21 Short Papers)*. Association for Computing Machinery, New York, NY, USA, Article 2, 7 pages. https://doi.org/10.1145/3448018.3458008

Toshiya Isomoto, Shota Yamanaka, and Buntarou Shizuki. 2022. Dwell Selection with ML-Based Intent Prediction Using Only Gaze Data. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 3, Article 120 (Sep 2022), 21 pages. https://doi.org/10.1145/3550301

Candace E. Peacock, Ben Lafreniere, Ting Zhang, Stephanie Santosa, Hrvoje Benko, and Tanya R. Jonker. 2022. Gaze as an Indicator of Input Recognition Errors. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, ETRA, Article 142 (May 2022), 18 pages. https://doi.org/10.1145/3530883

Naveen Sendhilnathan, Ting Zhang, Ben Lafreniere, Tovi Grossman, and Tanya R. Jonker. 2022. Detecting Input Recognition Errors and User Errors Using Gaze Dynamics in Virtual Reality. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology* (Bend, OR, USA) *(UIST '22)*. Association for Computing Machinery, New York, NY, USA, Article 38, 19 pages. https://doi.org/10.1145/3526113.3545628