GTK: An Open-Source Toolkit for Gaze-based Interaction in XR

László Kopácsi*,[1] Michael Barz[1] Daniel Sonntag[1,2]

[1]Interactive Machine Learning, German Research Center for Artificial Intelligence (DFKI), Saarbrücken, Germany
[2]Applied Artificial Intelligence, University of Oldenburg, Oldenburg, Germany

Abstract

Gaze interaction is becoming increasingly prevalent in Extended Reality (XR) due to the widespread integration of eye-tracking sensors in modern head-mounted displays. However, implementing robust gaze-based interfaces requires substantial development effort, often leading to reimplementation of established interaction patterns. We present the Gaze Interaction Toolkit (GTK), an open-source Unity package designed to facilitate the integration of active and passive gaze-based interaction in XR environments. GTK provides modular components including gaze-aware interactables, gaze-contingent information displays, and spatial context menus, as well as tools for onboarding, visual attention monitoring, and eye-tracking accuracy assessment. We demonstrate these features through an industrial sorting task scenario that systematically introduces users to multimodal gaze-assisted interaction techniques, illustrating the toolkit's capabilities for developing gaze-responsive user interfaces in XR training applications.

Index Terms: Gaze-based Interaction, Unity Toolkit, User Onboarding, Eye Tracking, Virtual Reality.

1 Introduction

While hardware integration of eye tracking in Head-Mounted Displays (HMDs) has matured, software support for rapid prototyping remains fragmented. Developers typically rely on commercial SDKs like Tobii[1] or general-purpose frameworks such as the Mixed Reality Toolkit (MRTK)[2] and Unity's XR Interaction Toolkit (XRI)[3]. While these tools provide foundational access to gaze data and basic dwell-time selection, they often lack high-level, modular components for complex interaction patterns. Consequently, researchers often face the burden of reimplementing complex features such as gaze-contingent displays, robust multimodal selection, and attention analytics [2].

To address this gap, we introduce the Gaze Interaction Toolkit (GTK)[4], a modular and extensible framework for the Unity game engine. GTK is built on OpenXR and supports a wide range of eye-tracking HMDs. It extends Unity's XRI with pre-configured prefabs for gaze-responsive UI elements, validation tools, and an integrated onboarding system, streamlining the development of XR training applications utilising multimodal interaction methods [4]. We demonstrate the core capabilities of GTK through an industrial quality-control sorting task (Figure 1) designed to systematically introduce participants to various active gaze-based interaction methods, while the system passively monitors visual attention for post-hoc analysis.

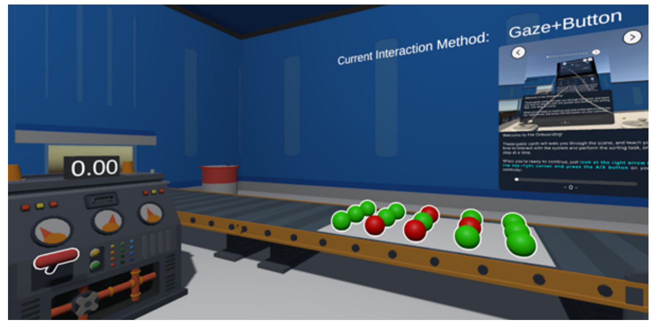

Figure 1: The sorting task demonstration features the Guide Card (right) for onboarding, a conveyor belt delivering trays of sortable objects (centre), and the control station (left). Interactive elements are highlighted to guide user attention during the training sequence.

2 Gaze Interaction Toolkit (GTK)

GTK adopts a component-based architecture (Figure 2), ensuring flexibility and ease of integration into existing Unity projects. The core of the system is the *Global Gaze Manager*, a central controller for gaze-related functionalities that manages global interaction states. GTK provides gaze-interactable components that extend standard XRI interactables to support gaze-specific behaviours, including object highlighting, gaze-contingent information displays based on [7], and multiple selection modalities to accommodate different task requirements. Developers may configure objects for gaze-only selection methods, like *dwell-based selection*, where prolonged fixation triggers an action, or *border-crossing selection* by setting a zero-millisecond dwell threshold on specific item selection zones [5]. Alternatively, in addition to gaze-only and controller-only interaction, the interactables also support multimodal interaction methods [1], where gaze is used for pointing and a physical button confirms the selection. This approach combines the speed of saccadic eye movements with the deliberate nature of manual input, thereby mitigating the Midas Touch problem that plagues pure dwell-based systems.

For menu navigation, GTK provides gaze-optimised *Context Menus* in Radial, Pie, List, and Grid layouts. These designs draw on principles for efficient gaze-based menu navigation [5] to enable rapid selection with minimal head or hand movement, while supporting various input modes. The toolkit also includes a *Navigable Card* system designed to structure user progression. It includes *Guide Cards* for formation-specific active onboarding [3]; *Quiz Cards* for assessing conceptual understanding; and *Questionnaire Cards* for collecting user feedback via Likert scales. The *Visual Attention Monitoring* module captures telemetry data on gaze and controller interactions, renders real-time visualisations of the attention sequence, and exports CSV logs for post-hoc analysis. Additionally, GTK integrates tools for verifying the eye-tracking signal quality, including a standardised *Gaze Accuracy Test* and a

*e-mail: laszlo.kopacsi@dfki.de

[1]https://developer.tobii.com/xr/ (Accessed 6 Jan 2026)
[2]https://github.com/microsoft/MixedRealityToolkit-Unity (Accessed 6 Jan 2026)
[3]https://docs.unity3d.com/Packages/com.unity.xr.interaction.toolkit@3.0 (Accessed 6 Jan 2026)
[4]https://github.com/DFKI-Interactive-Machine-Learning/de.dfki-iml.xr-gaze-interaction-toolkit

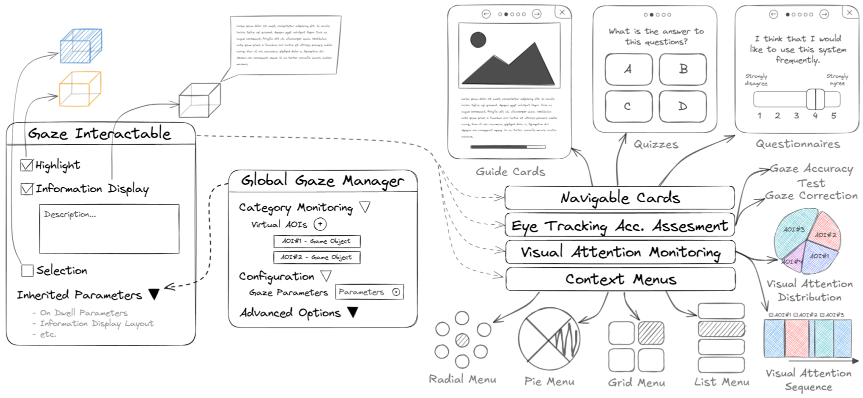

Figure 2: Module design of GTK. GTK provides modular components, including gaze-aware interactables, spatial context menus, as well as tools for onboarding, visual attention monitoring, and eye-tracking accuracy assessment.

correction mechanism based on the linear transformation approach described by Padikal et al. [6].

3 Demonstration Scenario

The accompanying industrial sorting task onboarding demonstration situates the user within a simulated industrial quality-control environment (Figure 1). The demonstration is implemented in Unity 2022.3.40f1, using the XR Interaction Toolkit v3.0.3, and deployed on the HTC Vive XR Elite headset.

The onboarding process is orchestrated by the Guide Cards. Following the framework by Chauvergne et al. [3], we designed the Guide Cards to provide a formation-specific context, i.e., a dedicated environment *dedicated to learning* interaction techniques before the main task. The system enforces active learning by requiring users to perform specific gaze-based tasks to progress. To support this, we employ multimodal assistance combining visual aids (embedded video demonstrations) and verbal instruction (audio narration). Furthermore, the system manages complexity through partially enabled interactivity, where interaction capabilities are restricted to relevant objects (e.g., highlighted sorting items) to avoid unexpected behaviours and focus user attention.

The experience begins with an introductory Guide Card that welcomes the user, explains the session's objectives through audio narration and an embedded instructional video, and teaches the fundamental interaction methods required to navigate between cards. Then, the user is asked to complete a 3×3 gaze accuracy test in which targets appear sequentially, and the user fixates on each target until it disappears. This yields nine correspondences between the known target positions and the measured positions recorded during fixation. Then, we apply a linear transformation [6] to correct the gaze signal, and the system displays the corrected gaze indicators.

Having their gaze corrected, the user proceeds to an overview of the sorting task central to the demo. A tray carrying fifteen balls, four red and eleven green, arrives from a conveyor gate. The user's objective is to identify and mark each red ball as defective by selecting it using the specified interaction method. The demonstration then introduces the gaze-contingent information displays. Six objects in the scene are highlighted, and the user is asked to attend to each and read the presented descriptive information about its role in the workflow. Next, a short quiz is presented, designed to reinforce the user's understanding of the task.

Throughout the entire demonstration, the Visual Attention Monitoring system operates in the background. The recorded telemetry data is visualised after the quiz to provide participants with insights into their behaviour. Finally, the participants are asked to provide feedback through an optional questionnaire.

4 Conclusion and Future Work

GTK is an open-source, modular, and extensible framework to facilitate the integration of active and passive gaze-based interaction in XR environments. Future work will involve conducting a user study to validate the usability and efficiency of the incorporated active gaze-based interaction methods. We also plan to expand the toolkit's passive interaction capabilities by integrating real-time cognitive load estimation and adaptive interface behaviours.

Acknowledgments

This work was funded by the European Union under grant number 101093079 (MASTER), the Federal Ministry of Research, Technology and Space (BMFTR) under grant number 16IW23002 (NoIDLE), and the Endowed Chair of AAI at University of Oldenburg.

References

[1] S.-J. Bittner, R. A. Leist, L. Kopácsi, et al. Interactive Multimodal Photobook Co-Creation in Virtual Reality. In *Companion Proceedings of the 30th International Conference on Intelligent User Interfaces*, IUI '25 Companion, pp. 146–151. Association for Computing Machinery, New York, NY, USA, Mar. 2025. doi: 10.1145/3708557.3716355 1

[2] Y. Bondarenko, V. Kuts, S. Pizzagalli, et al. Universal XR Framework Architecture Based on Open-Source XR Tools. In *XR and Metaverse*, pp. 87–98. Springer Nature Switzerland, Cham, 2024. doi: 10.1007/978-3-031-50559-1_7 1

[3] E. Chauvergne, M. Hachet, and A. Prouzeau. User Onboarding in Virtual Reality: An Investigation of Current Practices. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23, pp. 1–15. Association for Computing Machinery, New York, NY, USA, Apr. 2023. doi: 10.1145/3544548.3581211 1, 2

[4] L. Kopácsi, P. Karagiannis, S. Makris, et al. The MASTER XR Platform for Robotics Training in Manufacturing. In *Proceedings of the 30th ACM Symposium on Virtual Reality Software and Technology*, VRST '24, pp. 1–2. Association for Computing Machinery, New York, NY, USA, Oct. 2024. doi: 10.1145/3641825.3689514 1

[5] L. Kopácsi, A. Klimenko, A. M. Selim, et al. Gaze-Based Menu Navigation in Virtual Reality: A Comparative Study of Layouts and Interaction Techniques. In *Human-Computer Interaction – INTERACT 2025*, pp. 520–543. Springer Nature Switzerland, Cham, 2026. doi: 10.1007/978-3-032-04999-5_31 1

[6] V. Padikal, A. Plonkowski, P. F. Lawton, et al. Gaze Error Estimation and Linear Transformation to Improve Accuracy of Video-Based Eye Trackers. *Vision*, 9(2):29, June 2025. Publisher: Multidisciplinary Digital Publishing Institute. doi: 10.3390/vision9020029 2

[7] R. Rivu, Y. Abdrabou, K. Pfeuffer, et al. StARe: Gaze-Assisted Face-to-Face Communication in Augmented Reality. In *ACM Symposium on Eye Tracking Research and Applications*, pp. 1–5. ACM, Stuttgart Germany, June 2020. doi: 10.1145/3379157.3388930 1