aerospace

MDPI

Article

Investigating XR Pilot Training Through Gaze Behavior Analysis Using Sensor Technology

Aleksandar Knežević [1,*], Branimir Krstić [1,*], Aleksandar Bukvić [1], Dalibor Petrović [1] and Boško Rašuo [2]

[1] Military Academy, University of Defence in Belgrade, Veljka Lukića Kurjaka 33, 11042 Belgrade, Serbia; aleksandar.bukvic@mod.gov.rs (A.B.); daliborpetrovic140@gmail.com (D.P.)
[2] Faculty of Mechanical Engineering, University of Belgrade, Kraljice Marije 16, 11120 Belgrade, Serbia; brasuo@mas.bg.ac.rs
* Correspondence: aleksandar.knezevic@va.mod.gov.rs (A.K.); branimir.krstic@va.mod.gov.rs (B.K.)

Abstract

This research aims to characterize extended reality flight trainers and to provide a detailed account of the sensors employed to collect data essential for qualitative task performance analysis, with a particular focus on gaze behavior within the extended reality environment. A comparative study was conducted to evaluate the effectiveness of an extended reality environment relative to traditional flight simulators. Eight flight instructor candidates, advanced pilots with comparable flight-hour experience, were divided into four groups based on airplane or helicopter type and cockpit configuration (analog or digital). In the traditional simulator, fixation numbers, dwell time percentages, revisit numbers, and revisit time percentages were recorded, while in the extended reality environment, the following metrics were analyzed: fixation numbers and durations, saccade numbers and durations, smooth pursuits and durations, and number of blinks. These eye-tracking parameters were evaluated alongside flight performance metrics across all trials. Each scenario involved a takeoff and initial climb task within the traffic pattern of a fixed-wing aircraft. Despite the diversity of pilot groups, no statistically significant differences were observed in either flight performance or gaze behavior metrics between the two environments. Moreover, differences identified between certain pilot groups within one scenario were consistently observed in another, indicating the sensitivity of the proposed evaluation procedure. The enhanced realism and validated effectiveness are therefore crucial for establishing standards that support the formal adoption of extended reality technologies in pilot training programs. Integrating this digital space significantly enhances the overall training experience and provides a higher level of simulation fidelity for next-generation cadet training.

Keywords: flight simulator; pilot training; extended reality; gaze behavior; sensor technology

Check for updates

Academic Editors: Miroslav Kelemen, Peter Korba and Wenjiang Yang

Received: 25 November 2025
Revised: 12 January 2026
Accepted: 13 January 2026
Published: 16 January 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

The application of extended reality (XR) simulators, encompassing Virtual Reality (VR), Augmented Reality (AR), and Mixed Reality (MR), has expanded rapidly across various fields, including aviation, medicine, industry, and education, [1–3]. These technologies provide highly immersive training environments, but the extent to which skills learned within XR settings transfer to real-world performance remains an active area of investigation. Nevertheless, the effectiveness of XR-based training and the extent to which skills acquired in such environments transfer to real-world performance remain insufficiently established.

The Flight Training Device (FTD) for the Lasta aircraft (Figure 1), a trainer used by the Serbian Air Force and manufactured in Serbia, was developed by the research team of the University of Defence in Belgrade as part of the VA-TT/2/24-26 project "Application of the Concept of Digital Training and Pilot Selection" [4]. The Lasta FTD (Figure 2) fully satisfies all technical and functional criteria required for classification as a FTD Level 3 in accordance with current European and international regulatory standards [5].



Figure 1. Lasta aircraft: in flight (**left**) and real cockpit layout (**right**).



Figure 2. Lasta FTD.

According to EASA CS-FSTD(A) Issue 2 [5], an FTD Level 3 device must accurately replicate the cockpit layout, instrumentation, avionics, and flight dynamics of the actual aircraft and must include the following capabilities:

1. A fully functional avionics suite, including GNSS-based equipment;
2. Realistic simulation of lateral and vertical navigation guidance;
3. Accurate aerodynamic modeling and faithful response to pilot control inputs;
4. A visual system providing at least 150° horizontal and 40° vertical fields of view per pilot;
5. Simulation of lighting and environmental conditions representing day, night, dusk, and poor-visibility scenarios;
6. An integrated system for monitoring and recording all flight parameters and pilot inputs for performance evaluation.

A body of research highlights the importance of real-time monitoring and recording of all flight parameters and pilot control inputs as a critical component of simulator-based training. Micunovic et al. [4] emphasize that such capabilities enable exact performance evaluation, allowing instructors to review every control input, navigation decision, and procedural action executed during a training session. The monitoring and data-logging system was also examined in early work conducted by the University of Defence in Belgrade research team [6]. In that study, Knezevic et al. developed methodological approaches for assessing cadet pilots' training processes within a digitally enhanced cockpit environment, noting that these data-driven insights support informed decisions regarding a candidate's suitability for further training on a specific aircraft type. For systematic data acquisition,

the Flight Simulator Recorder module (developed by Matthias Neusinger, 2007–2010) was used to capture real-time flight parameters, including heading, airspeed, altitude, flap position, and key instrument readings. Each recorded parameter was precisely time-stamped, enabling detailed post-flight analysis and the identification of deviations from standard procedures.

The free software programming environment, such as R and Python, widely applied in aviation research, enabled the efficient analysis and visualization of large-scale simulator datasets. Its automated processing capabilities supported the statistical comparison of pilot performance across flight segments, providing a basis for evidence-based debriefing, progress assessment, and the identification of individual training needs [7]. In the context of the programming environment, it is also essential to consider a distinct category of aviation training devices: Personal Computer-Based Aviation Training Devices (PCATDs). This type of device relies on a computer platform for flight simulation and requires a high level of fidelity in both the simulated aircraft model and its visualization system. Such software add-ons must accurately replicate aircraft flight dynamics and cockpit representations to ensure their training value [6].

As part of the VA-TT/2/24-26 project, the research team developed an innovative add-on for the Lasta training aircraft, implemented using the Prepar3D V6 flight simulation platform (Figure 3), a commercially available product from Lockheed Martin. The software introduces significant advancements, including advanced customization of aircraft models, sensors, camera systems, and environmental parameters; enhanced lighting and rendering functionalities; and a comprehensive Earth model based on the WGS-84 coordinate system.



Figure 3. New Prepar3D V6-based Lasta add-on screenshot.

Notably, its virtual cockpit capabilities and compatibility with a wide range of VR and MR headsets, including devices from HTC, Vrgineers, Varjo, Oculus, and HP, as well as other SteamVR-compatible systems, substantially enhance simulator fidelity and immersion. Similar simulation platforms have been employed in previous studies by Rizvi et al. [8] and Sheets and Elmore [9]. In contrast, other researchers have adopted alternative platforms, such as X-Plane, as reported by Oberhauser and Dreyer [10].

Another important component of aircraft simulator fidelity is the degree of procedural similarity and task difficulty. Typically described through cognitive and physical dimensions, these elements enable novice trainees to develop foundational motor skills, while more advanced users can engage in complex training scenarios, such as removal from irregular positions, commonly referred to as Upset Prevention and Recovery Training (UPRT) [11], or the application of Performance-Based Navigation (PBN) procedures, as demonstrated in [4].

The only way to achieve these essential fidelity characteristics is to incorporate a physical cockpit with operational switches, levers, and engine controls, as well as a fully functional avionics suite, including GNSS-based systems such as the Garmin GNS 430 W. Figure 4 shows the integration of the Lasta cockpit mock-up into the existing PCATD trainer, previously described by Knezevic et al. [6].



Figure 4. Lasta cockpit: the positioning inside the projection cylinder (**center**), left cockpit panel (**left**), and right cockpit panel (**right**).

In alignment with previous studies [6], the research team at the University of Defence in Belgrade investigated the potential to enhance the fidelity of an existing FTD without significantly increasing logistical costs, which would typically be expected when employing a full flight simulator. Notably, the acquisition and annual maintenance costs of a full flight simulator are similar to those of procuring a Lasta training aircraft, largely due to the platform's straightforward design.

Oberhauser et al. [12] employed flight tasks based on a standard traffic pattern, a scenario comparable to one used in the present research. Similarly, Sheets and Elmore [9] utilized a modified USAF visual box pattern for an overhead break, a procedure commonly practiced at Undergraduate Pilot Training (UPT) bases. The diversity of tasks and monitoring demands associated with takeoff, landing, and level flight procedures within a traffic pattern make such scenarios particularly well suited for examining pilot expertise through EM behavior. Deviations from prescribed parameters, including flight path, airspeed, or pitch angle, can be readily identified and used to assess task performance.

Furthermore, the integration of sensor-derived EM data enables the assessment of examinees' cognitive load in a manner consistent with the methodology proposed by Ahlstrom et al. [13], thereby supporting reliable comparative performance evaluations. Accordingly, changes in both flight performance and gaze behavior were used to evaluate two hypotheses: (1) pilots would achieve equivalent flight-task performance in an XR flight simulator and a conventional simulator; (2) EM metrics obtained from an XR HMD-integrated sensor using open-source software would be comparable to those recorded by external, cockpit-mounted eye-tracking systems.

2. Literature Review

2.1. Engineering and Learning Aspects

Recent studies indicate that the use of XR/MR Head-Mounted Displays (HMDs) can enhance human–machine interaction in a cost-effective manner. Macchiarella et al. [14] examined the application of AR in aviation and aerospace training and reported superior short- and long-term memory performance among learners using AR-based instructional methods compared with conventional approaches. Similarly, Ross and Gilbey [15] evaluated XR simulation as a more efficient and economical alternative to traditional flight simulators or training aircraft, while Rizvi et al. [8] investigated the integration of AR/VR technologies within ground-training modules.

In addition, custom-built VR simulators, such as those described by Hebbar et al. [16] and Sheets and Elmore [9], have been employed to evaluate cognitive load and to leverage biometric data for improving both pilot training outcomes and organizational effectiveness. Clifford et al. [17,18] reported positive user feedback from firefighters trained using a VR-based simulator and a 270° cylindrical projection system (SimPit), a configuration comparable to that presented by Knezevic et al. [6]. Furthermore, Moesl et al. [19] explored

whether the most challenging elements of Type Rating (TR) courses could be addressed through AR-based applications. Finally, Oberhauser and Dreyer [10] introduced a VR flight simulator (VRFS) that combines the flexibility of desktop flight simulation with the immersive qualities typically associated with full flight simulators, focusing on human factors (HFs) engineering and the early-stage evaluation of flight deck designs.

2.2. Data Processing and Validation Methods

Eye movements (EMs), when combined with aircraft performance parameters such as airspeed, heading, and altitude deviations, can serve as reliable indicators of pilots' mental workload or potential overload. Eye-tracking recordings may also reveal differences between expert and novice pilots, as well as among pilot selection candidates under different workload conditions. Similar approaches have been applied to assess the cognitive load experienced by air traffic controllers [13]. When analyzed alongside flight performance metrics, eye-tracking data provides substantial insight into human–machine interaction processes [7]. Building on prior research on training methodologies, the research team at the University of Defence in Belgrade sought to address existing limitations by introducing continuous digital monitoring of flight performance into training evaluation procedures, thereby enabling the collection of a comprehensive dataset [6]. In a related context, Lawrynczyk [20] investigated the use of VR for pilot training, either as a supplement to or a replacement for traditional flight simulators. Oberhauser et al. [12] compared the functional fidelity of a VR flight simulator with that of a conventional simulator by examining pilot reaction times to cockpit controls, deviations from the ideal flight path, workload, and simulator-induced motion sickness. Software-based performance data acquisition and evaluation have also been reported by Ayala et al. [21], Yang et al. [22], McGowin et al. [23], Reweti et al. [24], and Le Ngoc et al. [25]. More recently, aviation research has increasingly employed machine learning and data augmentation techniques to address limitations associated with small datasets, as demonstrated by Zhou et al. [26]. Harris et al. [27] analyzed psychological fidelity and stimulus correspondence in training transfer using gaze-related metrics, including saccades, search rate, search behavior, and entropy. Similarly, Ahmadi et al. [28] developed a gaze-based intervention training approach to instruct student pilots on the visual flight rules of instrument cross-check procedures, with the objective of reducing the risk of loss of control following inadvertent entry into instrument meteorological conditions (IMC).

3. Materials and Methods

3.1. Participants

Eight military pilots, candidates for flight instructor (FI) training, were selected from various operational units during the theoretical phase of their program. The recruited pilots (1) were adults aged 30 years on average (SD = 5.01); (2) possessed the advanced expertise required for the FI program, with at least 200 flight hours, including 150 h as Pilot in Command (SD = 134.6); (3) had instrument flight rules (IFR) training on both actual airplane/helicopter and flight simulators; (4) had a minimum of 10 h of fixed-wing aircraft experience; (5) two pilots flew military airplanes with analog cockpits (AA group); (6) two pilots flew military airplanes with digital cockpits (AD group); (7) two pilots flew military helicopters with analog cockpits (HA group); and (8) two pilots flew military helicopters with digital cockpits (HD group). None of the pilots had prior experience with XR simulators. Given this diversity, the study aimed to create conditions that would elicit a broad range of results, thereby demonstrating the sensors' sensitivity in detecting differences in visual strategies and flight performances. In addition, consistent with the methodology described by Ahmadi et al. [26], an expert pilot and certified flight instructor

(CFI) for the Lasta aircraft was engaged to evaluate the system and provide a benchmark for comparison with participant performance.

3.2. Tools

3.2.1. Lasta Trainer Aircraft Mock-Up Cockpit

To replicate the parts of the aircraft and components with which the pilot has direct contact, such as the cockpit with its control units, instruments, and operating environment, the research team at the University of Defence in Belgrade employed high-end techniques and materials, including composites and additive manufactured parts. Supported by 3D modeling software Dassault Systèmes CATIA V5 for precision manufacturing and full material utilization, the research team assembled a fully functional mock-up cockpit of the Lasta trainer aircraft, Figure 5.

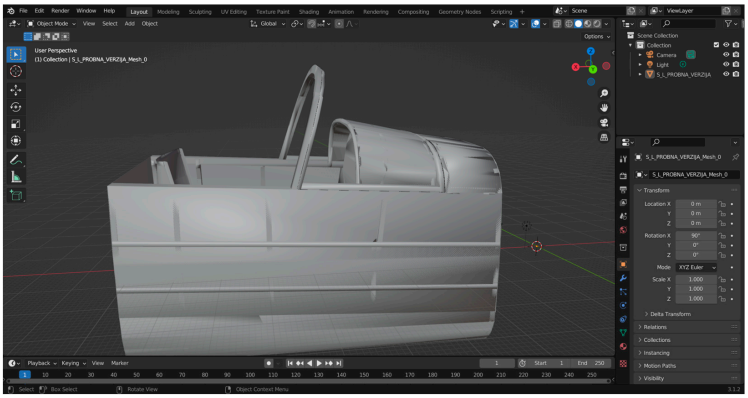


Figure 5. Lasta cockpit 3D model imported into Blender V3.1 software to enable easier manipulation.

Similarly to Kumar et al. [29], the research team employed an AR tool (instead of the VR) to evaluate the cockpit within an XR environment using Microsoft HoloLens 2 HMD before its physical production, Figure 6.

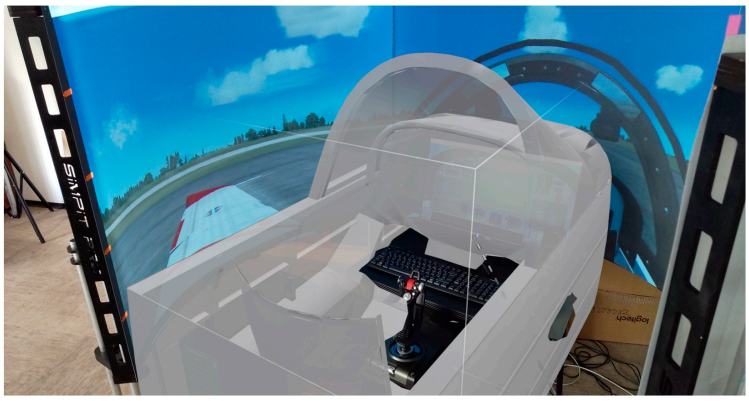


Figure 6. Lasta 3D model fitted into the 270° projection cylinder (SimPit), as seen through Microsoft HoloLens 2.

Switches shown in Figure 4 (right) were integrated into simulation software using the Leo Bodnar BBI-32 Button Box Interface (Leo Bodnar Electronics Ltd., Silverstone, UK), the Leo Bodnar BU0836X interface (Leo Bodnar Electronics Ltd., Silverstone, UK), an ARDUINO MEGA 2560-16U2 microcontroller (Qualcomm, San Diego, CA, USA) rotary encoders, and switches, along with third-party software for action assignments. The Thrustmaster HOTAS Warthog Flight Stick (Guillemot Corporation S.A., Carentoir, France) was used for the primary flight controls, as it closely resembles the Lasta trainer aircraft flight stick in both shape and size, Figure 4.

3.2.2. XR HMD and Supporting Hardware and Software

Many prior studies have employed various HMD hardware to test their hypotheses about the use of AR/VR in pilot training. Sheets and Elmore [9] used the following:

1. Intel Core i7 processor and Nvidia GeForce GTX 1060 graphics card;
2. The VR headset HTC Vive;
3. The gaming chair;
4. The Thrustmaster HOTAS Warthog Flight Stick and throttle system.

Since the HTC VIVE VR headset (HTC Corporation, Taoyuan City, Taiwan) is not intrinsically equipped with an eye-tracking sensor, First Pupil Lab cameras (Pupil Labs GmbH, Berlin, Germany) were installed around the two interior eyepieces to visually record the user's EMs with the fidelity required by Senseye Inc.(Senseye Inc., Austin, TX, USA) for cognitive load assessment. As already mentioned, the primary software utilized for flight simulation was Prepar3D V4, a commercially available product developed by Lockheed Martin.

Hebbar et al. [16] used the following HMD hardware:

1. The VR headset HTC Vive Pro Eye with a built-in eye tracker;
2. Emotiv 32-channel Electroencephalography (EEG) headset;
3. Thrustmaster HOTAS Warthog Flight Stick and throttle system.

The HMD provides a 110° diagonal field of view (FOV) and a resolution of 1440 × 1600 pixels. The HTC Vive Pro Eye VR headset includes built-in eye-tracking capabilities that record ocular parameters, such as gaze direction, pupil position, and size, and synchronize them with aircraft data at a rate of 120 Hz. Eye-tracking data were collected using the SRanipal SDK runtime, which interfaces with the Vive Facial Tracker (HTC Corporation, Taoyuan City, Taiwan) and other eye- and face-tracking hardware on Windows PCs [30], together with the Tobii XR SDK v1.8.0 [31].

Harris et al. [27] used the following:

1. The VR headset HTC Vive with a built-in Tobii eye tracker;
2. Intel Core i7 processor and Nvidia GeForce GTX Titan V graphics card.

The virtual environment was programmed in the Unity game engine using C#. Eye-tracking data were accessed through the Tobii Pro SDK. The eye tracker employed binocular dark-pupil tracking and sampled at 120 Hz across the full 110° FOV, with an accuracy of 0.5°.

Jusko et al. [32] presented research examining the effects of holographic visual cues (HVCs) on pilot handling-qualities ratings, workload, and task performance during piloted simulations. The HMD used in that study was the Microsoft HoloLens 2, an XR device developed and manufactured by Microsoft. HoloLens 2 combines waveguide technology with laser-based stereoscopic projection to create full-color XR smart glasses. It features a diagonal FOV of 52° and provides a resolution of 47 pixels per degree (PPD).

For the purpose of this research, the research team at the University of Defence in Belgrade selected a VARJO XR-3 VR headset. Varjo headsets represent the first VR-based solutions for aviation training to be formally accepted for pilots' training in both the United States and Europe. On 26 April 2021, the European Union Aviation Safety Agency (EASA) officially qualified a VR-based training solution for the first time [33]. Similarly, on 8 August 2024, the Federal Aviation Administration (FAA) of the United States Department of Transportation officially qualified a VR-based aviation training system for the first time [34]. This qualification was granted for a helicopter VR simulator developed by Loft Dynamics (Dübendorf, Switzerland), which enables pilots to experience highly accurate flight visuals, movements, and scenarios.

Some of the key features of the Lasta XR trainer with the VARJO XR-3 headset include the following:

1. The XR concept incorporating the Varjo XR-3 HMD;
2. The built-in eye-tracking system;
3. The Lasta trainer cockpit mock-up equipped with a fully functional avionics suite, switches, levers, a replica of the engine controls, and primary flight controls;
4. Desktop PC configuration with Intel Core i7-14700K 3.40 GHz processor, Nvidia GeForce RTX 4080 16 GB AERO graphics card, and Windows 10 version 2H22 operating system.

3.2.3. VARJO Built-In Eye-Tracking System

The Varjo XR-3 features the industry's highest resolution across the widest FOV (115°). With a resolution exceeding 70 PPD at the center of the FOV, pilots can observe and read the smallest details in VR with exceptional clarity.

The Varjo XR eye sensor includes the following features:

- High-speed tracking: Operates at 200 Hz, providing fast and responsive gaze tracking.
- Sub-degree accuracy: Delivers precise tracking with sub-degree accuracy, essential for detail-oriented tasks.
- Foveated rendering: Uses eye-tracking data to render the user's focal area at the highest resolution, while reducing resolution in the peripheral area, improving performance without compromising visual quality.
- One-dot calibration: Offers a simplified calibration process that can be completed quickly.
- Data for analysis: Provides valuable data on user gaze, attention, and interaction, which can be used to analyze and optimize user experiences in applications such as training and research.
- Automatic interpupillary distance (IPD) adjustment: The system can measure and automatically adjust the IPD to enhance user comfort and reduce eye strain, with an IPD range of 58–72 mm.
- Gaze camera resolution: 640 × 400 pixels per camera.
- The system operates at a sampling rate of 1000 Hz, measuring the position of the eyes 1000 times a second. It also provides high spatial precision, with gaze estimates typically accurate to within 2–5 mm.

To create an XR flight simulator and integrate the cockpit mock-up into an immersive training environment, additional procedures are required. The first step involves generating masks to obscure specific areas of the real-world video feed using a designated mesh. The 3D solid model, created in Dassault Systèmes Catia V5 software and originally used for cockpit assembly, was also employed to generate the mesh for the cockpit interior. This mesh defines the areas where real or virtual elements should be visible. Additionally, the virtual object must be anchored and oriented relative to the real world to ensure that it remains correctly positioned as the user moves. Precision adjustments can be made using available VR controllers or the Varjo Lab Tools V1.5.7 (Varjo Technologies Oy, Helsinki, Finland) software. Figure 7 shows a screenshot of a mask used in the Lasta XR simulator. Specifically, the vertices and polygons of the front screen were deleted to allow the real world (or virtual content) to show through, even though these areas were initially defined as transparent material.



Figure 7. Anchored 3D model of the Lasta cockpit (Varjo Lab Tools screenshot).

After saving the configuration in Edit mode within the Varjo Lab Tools software, the mask setup is ready for deployment once the Use mode is enabled. Figure 8 illustrates the fully activated XR setup for the Lasta flight simulator within the Prepar3D scenery.



Figure 8. Lasta XR simulator setup (Batajnica Runway 30R).

Eye tracking must be recalibrated before each experimental session and every time the headset is removed and repositioned, even if used by the same participant. This process can be performed using a simple 1-dot calibration directly from the application, Figure 9.

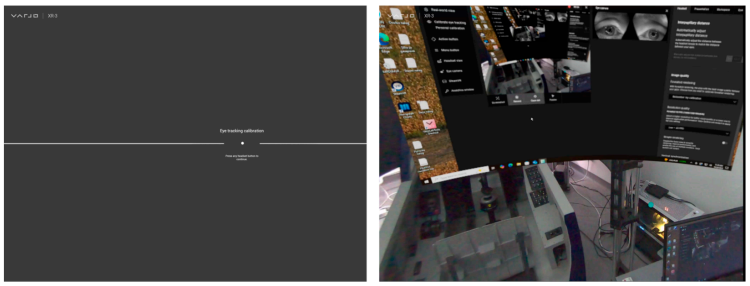


Figure 9. Calibration process displayed in the HMD view (**left**) and eye-tracking window with the user's eyes visible (**right**).

Figure 10 illustrates an example of a precise eye-tracking process for hand–eye coordination, demonstrated during a takeoff procedure involving landing gear retraction.



Figure 10. Eye tracking of hand–eye coordination during landing gear operation. The yellow dot indicates eye fixation.

The Varjo XR-3 base software enables the logging of gaze data by recording user views alongside eye-tracking information. The eye-tracking data are saved as a CSV file for subsequent analysis or visualization over video recordings. When eye-tracking logging is activated via the Record button in the Headset tab, an eye icon appears on the button to indicate that both video and eye data are being captured.

Each CSV file contains 46 columns of data, which can be categorized into two main types: generic gaze data and video capture-related data. Video capture-related data includes projected X-Y coordinates, used to map the gaze directly to video pixel coordinates, and a timestamp relative to the start of the video.

In addition to raw and Unix-epoch-relative timestamps, several columns in the generic gaze data provide information about the distance between the eye and the focus point, as well as the status and stability of the user's gaze. The status column indicates the validity of eye-tracking data, with a value of 2 representing valid data. The remaining columns describe pupil diameter and the position of the eyes in the coordinate system (X, Y, and Z). These coordinates can later be converted into fixations using free software tools, the Varjo SDK, or other spatial analysis software.

Testing the full performance capabilities of the Varjo XR-3 headset, it was assumed that gaze-tracking data were recorded at a 200 Hz sampling rate. For example, the takeoff segment of the traffic-pattern procedure, lasting 1 min and 58 s, produced over 23,700 entries in the "varjo_gaze_output.csv" file, reflecting the density of captured data.

There are several methods for handling such a large volume of gaze-tracking data. One approach is to convert the CSV file to XLSX using Get and Transform Data macro-enabled features, which allow for importing, connecting, and cleaning data from various sources such as files, databases, or websites. These tools can apply transformations, including removing columns, changing data types, or merging tables, before loading the cleaned dataset into Excel for generating reports and charts that can later be refreshed with new input data. Another approach is to use professional analytics software. The data format is compatible with eye-tracking analysis suites such as iMotions or GazePlotter, which provide advanced visualization options and automated metric calculation.

The most flexible, though also the most demanding, method involves analyzing the data using Python libraries such as pandas for data manipulation and visualization packages (e.g., matplotlib, seaborn) for creating custom outputs, including the following:

1. Heatmaps: Visualizing regions of the highest gaze concentration in synchrony with the corresponding screen or video recording.
2. Areas of Interest (AOI): Measuring how long participants fixate on specific elements of the virtual environment.
3. Scan Paths and Fixations: Identifying the sequence and duration of EMs.

Varjo Gaze Detector is an example of free software built using Python 3.9, Unity 2020.3.7f1, and Varjo Base 3.0.5.14 XR-1 developer edition firmware 2.5 by Thomas de Boer, incorporating the detection algorithm from "sp_tool" by Agtzidis et al. [35]. Similar methodologies have been applied using open-source software to evaluate digital training approaches and pilot selection procedures [6,7].

The Varjo Gaze Detector (Figure 11) uses the gaze vector within the headset's coordinate frame to detect EMs. Detection results can be visualized in real-time plots, and figures can be saved automatically. Metrics such as timestamps, amplitude, and mean velocity for each detected event are computed and stored in separate CSV files. Additionally, a classification column is appended to a copy of the raw eye-tracking dataset. The tool supports batch processing for multiple participants and trials, provides built-in sanity-check plots, and offers numerous options for subsequent data processing, if required.

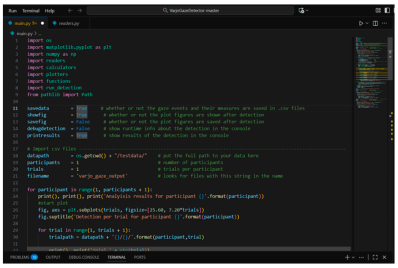


Figure 11. Python Varjo Gaze Detector project shown in Visual Studio Code V1.105.1.

3.2.4. Gaze Point Eye Tracker and Software for the Conventional Simulator

Pilot EMs were recorded using GP3 Desktop Eye Tracker (Gazepoint Research Inc., Vancouver, BC, Canada) (sampling rate: 60 Hz; visual angle accuracy: 0.5–1°; Gazepoint, 2019), as shown in Figure 12 (left). Gaze calibration was performed using Gazepoint Control software (Gazepoint Research Inc., Vancouver, BC, Canada), which maps pupil movements to visual stimuli presented on the selected display. EM data acquisition and visualization were conducted with Gazepoint Analysis software (Gazepoint Research Inc., Vancouver, BC, Canada) (Figure 12, right), which was also used to export EM data.



Figure 12. Lasta trainer cockpit equipped with GP3 Desktop Eye Tracker (**left**) and Gazepoint Control software during calibration (**right**).

The Gazepoint Analysis software exports two types of CSV files for each recording, {RECORDING_NAME}_all_gaze.csv and {RECORDING_NAME}_fixations.csv, where {RECORDING_NAME} denotes the corresponding user recording. The all_gaze file contains all recorded data points, enabling detailed temporal analysis, whereas the fixation file provides a reduced dataset, in which each fixation is represented as a single entry. These files can be readily imported into standard data-analysis tools, including Microsoft Excel.

In addition, AOI statistics are summarized in the Data_Summary_export_{DATETIME_LABEL}.csv file, which reports AOI-specific metrics for each participant as well as aggregated averages across all users.

3.3. Experimental Design and Procedure

To assess the effectiveness of the XR-based flight simulator, participants were tasked with performing a takeoff, initiating the initial climb in the departure direction, and ascending to 200 m above ground level (AGL) within the traffic pattern for the Lasta trainer aircraft, Figure 13. The first scenario involved executing the takeoff procedure using the standard visual environment of the FTD Level 3 device, equipped with the GP3 Desktop Eye Tracker, Figure 12. The second scenario required performing the same task within the XR simulator, with VARJO HMD XR sensors (Varjo Technologies Oy, Helsinki, Finland) collecting gaze performance data.

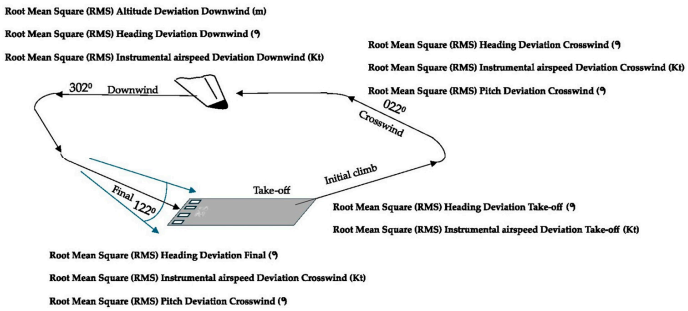


Figure 13. Traffic pattern used for comparing flight performance and instrument observation strategies.

Accordingly, the study featured a single independent variable: visual environment (visual fidelity level). Participants also practiced the traffic pattern illustrated in Figure 13, commonly used for takeoff and landing training on the Lasta aircraft, enabling evaluation of flight performance to test hypothesis 1 and comparison of instrument observation strategies across fidelity levels to assess hypothesis 2.

The diverse composition of the pilot group enabled a comprehensive evaluation, facilitating detailed identification of individual differences and demonstrating the sensitivity of the methodology. Following participant trials, an expert pilot completed a session in the XR environment to provide a benchmark for comparison. All experiments were conducted at the Air Force Department, Military Academy, University of Defense in Belgrade, within the Virtual Reality and Simulation Laboratory.

3.3.1. EM Metrics

Humans exhibit various types of EMs, including saccades, smooth pursuit movements, vergence movements, and vestibulo-ocular reflex movements [36]. Some of the EM parameters frequently analyzed in research include the follows:

- Saccades: Rapid, ballistic EMs that direct the gaze to another area of the visual field. Information processing is suppressed during saccades, a phenomenon known as saccadic suppression.
- Smooth pursuit movements: EMs that continuously align the gaze with a moving target (e.g., a passing aircraft). Masson and Stone [37] reported that visual perception of the target continues during smooth pursuit to update eye velocity and maintain tracking. This topic was also discussed by Agtzidis et al. [35].
- Fixations: Events in which the eyes remain focused on a point in the visual field, projecting a relatively stable image onto the retina. Visual information is primar-

ily extracted during these fixations. Following a saccade, the eyes fixate on a new point [38].

- Revisits: Lijing & Lin [39] consider that the EM transition starts, i.e., from outside the cockpit AOI, then shifts to the airspeed indicator AOI, and then returns to outside the cockpit. This sequence can start from any other AOI and constitute a returning scan path or "revisit" with any different AOI.
- Dwell: Defined as multiple fixations on a specific region of the visual field. For example, during scene perception, a viewer may inspect a particular area of the cockpit (e.g., a single instrument) through a series of small saccades before moving to another area via a larger saccade. A group of fixations on a specific piece of information is often referred to as a dwell [38], Figure 14.
- Blink (duration): Blink is a ubiquitous oculomotor action that lubricates and clears the corneal surface but has also been shown to correlate with mental workload in laboratory tasks [40].
- Diameter of the pupil: The pupil regulates the amount of light reaching the retina via smooth muscle adjustments in response to ambient luminance. Evidence indicates that the pupil diameter increases with rising cognitive workload [13].



Figure 14. Dwell captured on Lasta Garmin G500 PFD (screenshot from VA-TT/2/24-26 project trial).

3.3.2. Gaze Tracking

Modern digital aircraft cockpits present a highly complex visual environment in which the pilot's FOV is occupied by numerous instruments and digital displays. Within these displays, various indicators, tapes, symbols, and graphical elements convey information

related to the aircraft's attitude and performance parameters. These elements may dynamically change position, color, or shape or introduce additional visual cues (e.g., bugs), requiring the pilot to continuously search for and integrate information across multiple layers of VR or XR. Figure 9 (right) illustrates a conceptual representation of a "reality within a reality".

In contrast to traditional analog cockpits, where information was primarily acquired from the relative position of needles on calibrated circular, semicircular, or linear scales, modern digital cockpits demand more frequent shifts in visual attention and impose a higher cognitive load on pilots. Consequently, recordings of EMs provide valuable insight into how pilots visually process cockpit display information and interpret performance-related feedback during flight tasks.

The first phase of the evaluation of piloting performance metrics and EM behavior during the takeoff procedure was conducted in the standard visual environment of an FTD Level 3 device equipped with a GP3 Desktop Eye Tracker. During a VFR takeoff within the traffic pattern (Figure 13), pilots are required to gather information from both inside and outside the cockpit. The GP3 Desktop Eye Tracker is calibrated to accurately capture EMs solely from the display to which it is attached.

Since the recording files provide the number of fixations and the percentage of gaze time within the predefined AOI, the remaining gaze time can be interpreted as time spent observing areas outside that AOI. Assuming adequate pilot training, it is reasonable to infer that the majority of this remaining gaze time was directed outside the cockpit. However, it cannot be conclusively determined whether some gazes were directed toward other cockpit regions not covered by the AOI. The revisit metric further quantifies the number of times the pilot returned their gaze to the instrument panel AOI after fixating on another area. Figure 15 illustrates a representative sequence of recorded gaze points within the instrument panel, treated as a single AOI.



Figure 15. Screenshot of the Gazepoint Analysis software showing the instrument panel defined as a single AOI (highlighted area). The FTD instrument panel consists of a single display mounted behind the physical cockpit surface, with purpose-designed cutouts and control elements arranged around the monitor to replicate the layout of the actual aircraft cockpit.

To enhance the differentiation of gaze strategies and more effectively identify distinctions among specific pilot groups, EM metrics were analyzed using multiple predefined AOI. Two AOI were considered: The Primary Flight Display AOI (PFD AOI) and the Engine Instruments AOI (EI AOI), as illustrated in Figure 16.



Figure 16. Screenshot of the Gazepoint Analysis software illustrating PFD AOI and EI AOI (highlighted areas). The FTD instrument panel consists of a single display mounted behind the physical cockpit surface, with purpose-designed cutouts and control elements arranged around the monitor to replicate the layout of the actual aircraft cockpit.

Within the PFD AOI, pilots acquired critical information related to pitch and bank attitude, heading magnetic (HDM), and indicated airspeed (IAS), all of which are essential for maintaining stable takeoff and initial climb conditions. The EI AOI provides information necessary for monitoring and adjusting engine parameters, including revolutions per minute (RPM) and manifold pressure (MP), required during takeoff and the initial climb phase.

The number and percentage of fixations within these two AOI, relative to those recorded for the entire instrument panel AOI, quantify both the absolute and proportional allocation of visual attention. Revisit metrics were computed exclusively for the defined AOI, with the revisit ratio serving as an indicator of the underlying gaze strategy employed by the pilots.

3.3.3. Piloting Metrics

Piloting metrics were defined based on the flight objectives. During takeoff and initial climb, participants aimed to maintain a heading aligned with the runway by adjusting the roll attitude and to sustain a constant airspeed by controlling the pitch attitude after retracting the landing gear and flaps. The specific metrics were a heading deviation around HDM of 302° and airspeed deviation around IAS of 90 kt following landing gear and flaps retraction. The trajectory during takeoff and initial climb was also recorded. Tacview (Raia Software Inc., Mirabel, QC, Canada, 2006–2025) is a cross-platform flight analysis tool used to process all flights. After exporting flight data from Tacview, flight parameters were stored in a CSV file organized in rows and columns. Each row represents a single data sample, while each column corresponds to a specific parameter, such as latitude, longitude, altitude, pitch, bank, heading, flaps, IAS, and TAS (true airspeed). Values within a row are separated by whitespace (tabs or spaces).

The "timestamp" column is particularly crucial for synchronizing eye-tracking and flight performance recordings, as it records the time (in seconds) from the start of the recording for each parameter. This allows for precise comparison of simulation events with the eye-tracking software, which begins recording almost simultaneously, differing by only a single click.

3.4. Data Manipulation

The procedures for test sample preparation, experimental execution, and data acquisition for Step 1 and Step 2 are illustrated in Figure 17. The subsequent processing workflow for piloting performance metrics and EM metrics is presented in Figure 18.

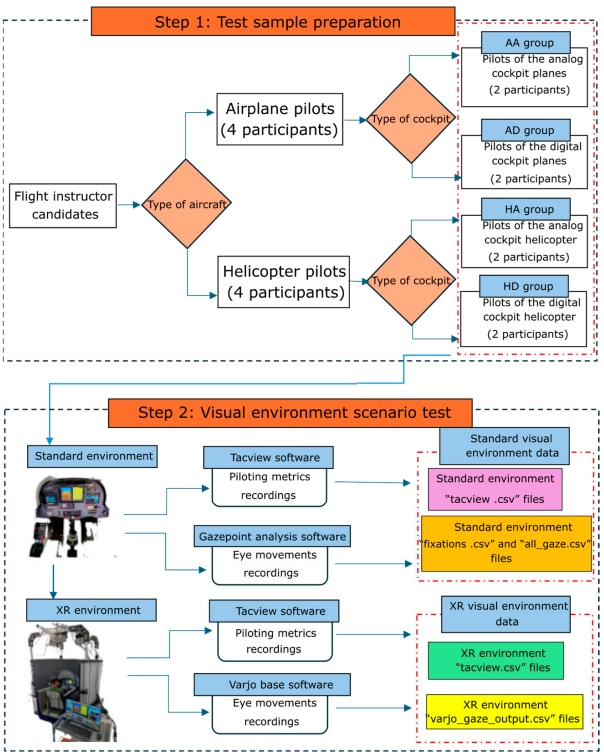


Figure 17. Flowchart illustrating participant selection and implementation of the experimental procedure.

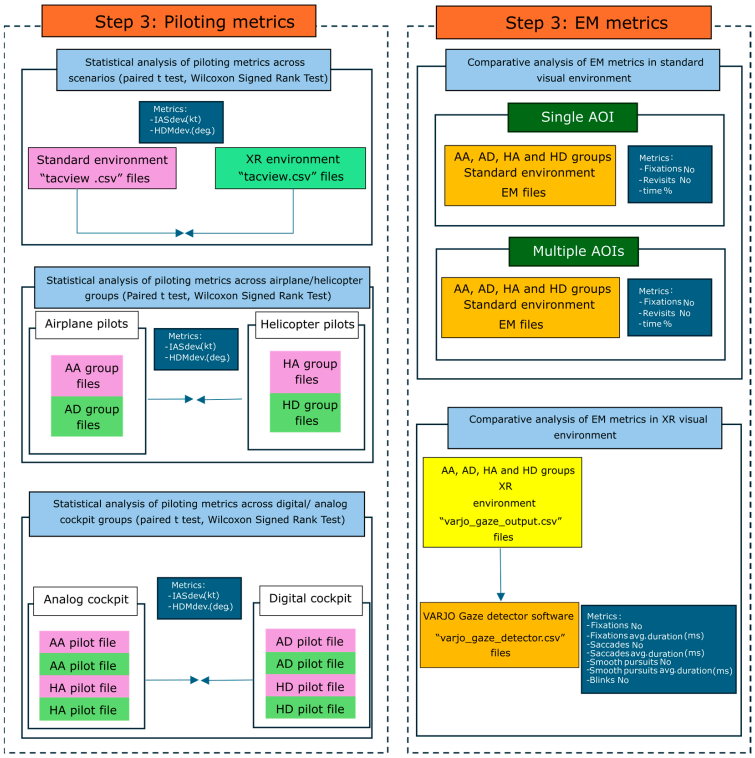


Figure 18. Schematic overview of the data processing workflow, including statistical evaluation of piloting performance metrics and comparative analysis of EM metrics.

3.4.1. Step 1: Test Sample Preparation

In the first step, four airplane pilots and four helicopter pilots were selected from a pool of FI candidates with comparable levels of flight experience. To further emphasize diversity related to cockpit configuration, additional selection criteria were applied based on the

type of cockpit operated by each pilot. This resulted in the formation of two additional subgroups, yielding a total of four pilot groups.

This grouping strategy was employed specifically for Step 3, where statistical analyses and comparative evaluations of piloting performance metrics and EM metrics were conducted.

3.4.2. Step 2: Visual Environment Scenario Test

In the second step, all pilots from the four groups were tested sequentially in two simulation environments. First, each participant performed the task in the standard visual environment of the FTD Level 3 device, equipped with the GP3 Desktop Eye Tracker. Subsequently, the same pilots completed an identical task in the XR simulator, configured with VARJO HMD eye-tracking sensors.

Two categories of software were employed for data acquisition. Tacview software was used in both scenarios to record piloting performance metrics. EM data were captured using dedicated recording software corresponding to each sensing device: the external GP3 Desktop Eye Tracker in the standard environment and the integrated VARJO HMD eye-tracking system in the XR environment. All datasets were exported in comma-separated values (CSV) format.

In Figure 17, the color coding of the data-file blocks denotes the simulation environment in which each dataset was recorded, which is essential for understanding the data processing workflow described in Step 3.

3.4.3. Step 3: Data Processing of Piloting Metrics and EM Metrics

In the third step, statistical analyses of piloting performance metrics were conducted using data from all eight participants, comparing results obtained in the standard visual environment with those from the XR environment. The evaluated metrics are summarized in Figure 18 and Table 1.

Table 1. The values of the outcomes of the flight data across scenarios.

		Comparison Between Scenarios		
		Standard FTD scenario	XR scenario	Pairwise comparison
Metric	Group	$\mu \pm SD$	$\mu \pm SD$	*p*-value
IAS Dev. (kt)	AA, AD, HA, HD	4.76 ± 1.78	4.75 ± 0.99	0.60
Heading Dev. (deg.)	AA, AD, HA, HD	10.36 ± 3.88	8.55 ± 5.98	0.23

Subsequently, the same statistical procedures were applied to subgroup comparisons, contrasting data from four airplane pilots with those from four helicopter pilots across both simulation environments. Finally, piloting performance data from pilots operating analog cockpit aircraft were compared with those from pilots flying digital cockpit aircraft, again considering both standard and XR conditions.

EM metrics obtained from the standard and XR visual environments were analyzed and interpreted separately within each environment. EM datasets recorded using VARJO HMD required additional post-processing using the VARJO Gaze Detector software prior to analysis.

3.5. Statistical Analysis

Following both trials, performance metrics from the standard visual and XR environment groups were compared using paired *t*-tests for normally distributed data or the Wilcoxon Signed Rank Test for non-normal data.

4. Results

The performance of eight pilot FI candidates in terms of piloting and EMs was compared. Only an FTD Level 3 simulator was available, used by a diverse group including military combat and transport pilots as well as helicopter pilots. All participants had at least 10 h on a fixed-wing piston aircraft. No equipment issues arose.

4.1. Flight Performance

4.1.1. Piloting Performance Across Scenarios

Eight pilots operating in the standard FTD visual environment exhibited an average Root Mean Square IAS deviation of 4.76 kt (SD = 1.78) compared to 4.75 kt (SD = 0.99) in the XR environment, with no statistically significant difference observed between the groups, Table 1.

For the HDM objectives, pilots demonstrated an average Root Mean Square HDM deviation of 10.36° (SD = 3.88) in the standard FTD environment and 8.55° (SD = 5.98) in XR, again with no statistically significant difference between conditions, Table 1.

4.1.2. Piloting Performance Across Airplane/Helicopter Groups

IAS results indicated that the four fixed-wing pilots (AA and AD groups) exhibited an average Root Mean Square Deviation (RMSD) of 4.38 kt (SD = 1.71) across both standard and XR environments. The helicopter pilots (HA and HD groups) recorded a mean RMSD of 5.13 kt (SD = 0.97) under the same conditions. No significant differences were observed between groups, Table 2.

Table 2. The values of the outcomes of the flight data across groups.

		Comparison Between Airplane/Helicopter Groups			**Comparison Between Analog/Digital Groups**		
		Airplane pilots (AA, AD)	Helicopter pilots (HA, HD)	Pairwise comparison	Pilots flying an analog cockpit (AA, HA)	Pilots flying a digital cockpit (AD, HD)	Pairwise comparison
Metric	Scenario	μ ± SD	μ ± SD	*p*-value	μ ± SD	μ ± SD	*p*-value
IAS Dev. (kt)	Standard and XR	4.38 ± 1.71	5.13 ± 0.97	0.57	5.09 ± 1.71	4.42 ± 1.25	0.28
Heading Dev. (deg.)	Standard and XR	11.23 ± 4.81	7.69 ± 4.74	0.08	9.26 ± 4.16	9.65 ± 4.97	1

For the HDM objectives, the fixed-wing pilots demonstrated an average RMSD of 11.23° (SD = 4.81), while the helicopter pilots exhibited an average RMSD of 7.69° (SD = 4.74) across both environments. Again, no statistically significant differences were found between groups, Table 2.

4.1.3. Piloting Performance Across Digital/Analog Cockpit Groups

IAS results indicated that pilots operating analog cockpit aircraft (AA and HA groups) exhibited an average RMSD of 5.09 kt (SD = 1.71) across both standard and XR environments. Pilots flying digital cockpit aircraft (AD and HD groups) recorded a mean RMSD of 4.42 kt

(SD = 1.25) under the same conditions. No significant differences were observed between groups, Table 2.

For HDM objectives, analog cockpit pilots demonstrated an average RMSD of 9.26 (SD = 4.61), while digital cockpit pilots showed an average RMSD of 9.65 (SD = 4.97) across both environments. Again, no statistically significant differences were found between groups, Table 2.

4.2. EMs

4.2.1. EMs in a Standard Visual Environment

The number of fixations across the entire instrument panel, treated as a single AOI, exceeded the average of 126 for six pilots from the AA, AD, and HA groups, while two pilots in the HD group recorded 101 fixations, 25 below the average, Table 3.

Table 3. The values of EM metrics in a standard visual environment.

		Pilot Groups								
		AA		AD		HA		HD		
Single AOI										
AOI name	Metric									μ
Instrument panel	Fixations No	136	∧	131	∧	136	∧	101	∨	**126**
	Revisits No	21	<>	22	<>	15.5	∨	25.5	∧	**21**
	time %	45.5	∨	53	<>	59.5	∧	41	∨	**49.75**
Multiple AOI										
PFD + EI	Fixations No	124	∧	98.5	<>	48.5	∨	50	∨	**80.25**
	Revisits No	24.5	∨	40.5	∧	44	∧	27	∨	**34**
PFD	Fixations No	121	∧	93.5	∧	24	∨	49	∨	**71.875**
EI	Fixations No	3	∨	5	∨	29.5	∧	1	∨	**9.625**
PFD	Revisits No	22.5	∨	38.5	∧	19	∨	27	<>	**26.75**
EI	Revisits No	2	∨	2	∨	25	∧	0	∨	**7.25**
PFD	Time %	42.5	∧	37.5	∧	6	∨	16.5	∨	**25.625**
EI	Time %	0.5	∨	1	∨	6	∧	0.35	∨	**1.9625**

Legend: ∧—above average; ∨—below average; and <>—average.

Fixation time on the full instrument panel was above the mean of 49.75% for pilots in the HA group, at average levels for the AD group, and below average for the HD and AA groups, with 41% and 45.5%, respectively, Table 3.

Revisit frequency to the full instrument panel exceeded the average of 21 for the HD group, was below average for the HA group, and matched the average for the AA and AD groups, Table 3.

For the two specific AOI, PFD and EI, fixation counts were below the overall average of 80.25 for helicopter pilots (HA and HD groups), at the average for AD pilots, and above average for the AA group, Table 3.

Within the PFD AOI, fixation counts were below the average of 71.87 for helicopter pilots and above average for airplane pilots, whereas fixation counts in the EI AOI were below the mean of 9.62 for all groups except HA pilots, Table 3.

Fixation time in the PFD AOI was below the average of 25.62% for helicopter pilots and above average for airplane pilots, while fixation time in the EI AOI exceeded the average of 1.96 for HA pilots only and was below average for all other groups, Table 3.

Revisit counts to the PFD AOI were above the average of 26.75 for AD pilots, at the average for HD pilots, and below average for AA and HA pilots. For the EI AOI, revisits exceeded the average of 7.25 for HA pilots only, with all other groups remaining below average, Table 3.

4.2.2. EMs in an XR Environment

Data collected from the XR environment, together with the evaluation of gaze behavior by an experienced FI, are presented in Table 4. In addition to these tabular results, the Varjo Gaze Detector software provides graphical representations of additional metrics. Figures 19 and 20 illustrate the saccade duration and mean velocity of the smooth pursuit, respectively. Based on the characteristics of the observed variables, the probability density function of the Rayleigh distribution was determined to be the most appropriate model for fitting the dataset. Estimation of the distribution parameters allows for the prediction of future probabilities and the forecasting of event frequencies.

Table 4. The EM metrics in an XR environment.

		Pilot Groups									
		AA		AD		HA		HD			
The whole visual field as a single AOI											
	Metrics									μ	Expert
Fixations	No.	98.5	∧	116.5	∧	101.5	∧	59	∨	93.875	181
	Avg. duration (ms)	125.5	∨	183.5	∨	200.5	<>	310.5	∧	205	93
Saccades	No.	198	∧	233	∧	177	<>	123.5	∨	182.875	431
	Avg. duration (ms)	77.5	∧	75	∧	66.5	∨	72	<>	72.75	58
Smooth pursuits	No.	265	∧	277.5	∧	118	∨	81	∨	185.375	1118
	Avg. duration (ms)	239.5	∨	327	∨	636	∧	701.5	∧	476	44
Blinks	No.	42.5	∧	22	∨	15.5	∨	12.5	∨	23.125	61

Legend: ∧—above average; ∨—below average; and <>—average.

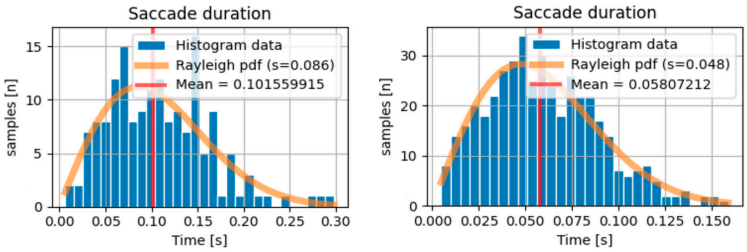


Figure 19. Saccade duration distribution, shown in plots generated by the Varjo Gaze Detector of a participant from the AD group (**left**) and an expert (**right**).

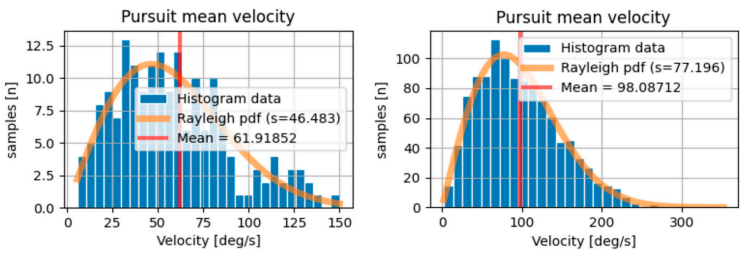


Figure 20. Smooth pursuits mean velocity distribution, shown in plots generated by the Varjo Gaze Detector.

Finally, all detected gaze velocities and angles are computed and presented over the duration of the selected sample, Figure 21.

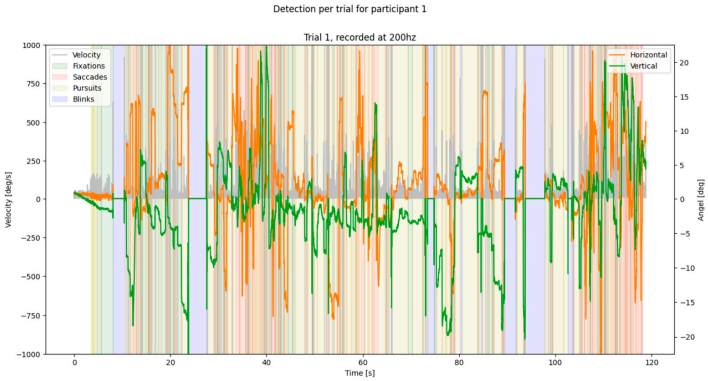


Figure 21. Gaze velocities and gaze-angle metrics during the experiment, shown in a plot generated by the Varjo Gaze Detector.

Recorded flight trajectories from both the standard and XR visual environments are presented in Figure 22, plotted using the "ggplot" library in the R programming environment.

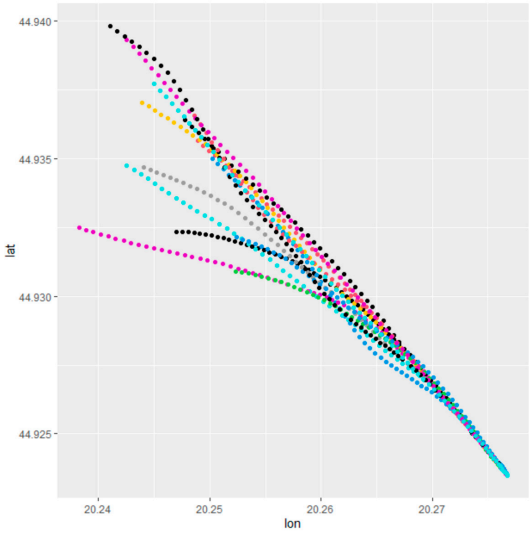


Figure 22. Two-dimensional representation of takeoff and initial climb trajectories across all analyzed scenarios (lon: eastward longitude coordinate; lat: northward latitude coordinate).

5. Discussion

5.1. The Piloting Performance Outcomes of the Changes in the Visual Environment Setup

In the present study, a custom-made FTD Level 3 cockpit for the Lasta advanced training aircraft was employed to assess the effectiveness of XR flight training. Additional procedures were required to build the XR flight simulator and integrate the cockpit mock-up into a fully immersive environment using the VARJO XR-3 headset. Tacview V1.9.5 software captures the moment of landing gear and flaps retraction by changing the corresponding variable from one to zero. Filtering out non-zero rows ensured that relevant IAS data were analyzed. As the Lasta's initial climb procedure requires fixed power settings (throttle and propeller levers set at 26 inHg and 2600 rpm), the only method to maintain the target IAS of 90 kt is via pitch attitude adjustments, which is why IAS was the primary parameter considered. These aspects of power setting and aircraft control are consistent with Stojakovic et al. [41–43]. Maintaining the roll attitude is also critical to ensure consistent HDM during takeoff and the initial climb. The sequence concludes when pilots reach 200 m

AGL prior to the crosswind turn. Due to variability in pitch control among pilots, sequence durations differed, as illustrated in Figure 22, where trajectories show participants reaching 200 m AGL at varying points. The takeoff and initial climb scenario were selected because XR recordings generate large datasets in both video and performance files, presenting data-handling challenges.

Given the diversity of the pilot group, the study aimed to examine the effect of scenario variations across different experience levels. While testing environmental changes in the exact simulators used by each pilot would have been ideal, the primary goal was to compare XR performance to that of standard simulators (Figure 23) across diverse pilot groups.



Figure 23. Lasta standard FTD visual environment (**left**) and XR environment (**right**).

The key finding regarding pilot behavior is that pilots maintained a consistent performance despite changes in the simulator's visual environment, Table 1. No statistically significant differences were observed in maintaining the initial HDM of 300° or the IAS of 90 kt, even when comparing performances across standard and XR environments for all pilot groups, Table 2.

5.2. The EM Metrics Outcomes of the Changes in the Visual Environment Setup

EM sensors provide insightful data on pilots' gaze behavior during the instrument cross-check and cockpit procedure. This study investigated gaze metrics from two types of sensors. The first, the GP3 Desktop Eye Tracker, was mounted on the existing FTD cockpit (Figure 12, left) and captured EMs within the instrument panel, as it is calibrated for displays up to 24 inches (the Lasta instrument panel has a 23-inch diagonal) (Figure 12, right). Visual areas outside the instrument panel (Figure 23, left) were not directly detected but could be indirectly inferred through fixation and revisit counts. The first three rows of Table 3 show the number of fixations, revisits, and the percentage of time pilots spent focusing on the instrument panel as a single AOI, averaging approximately 50%, which is consistent with expectations, since takeoff and initial climb rely heavily on visual cues outside the cockpit and information from PFD and EI displays. The remaining 50% of gaze was directed outside the instrument panel, though it could not be determined whether this was inside or outside the cockpit. Pilots in the HD group recorded fewer fixations, but their revisit counts and time percentages were comparable to other groups. By implementing XR HMD built-in eye-tracking sensors, the full FOV could be monitored. Previous studies by Ayala et al. [21], Young et al. [22], and Ahmadi et al. [28] used eyeglass-based trackers in standard visual environments, allowing for the partial observation of EMs both inside and outside the cockpit. Figure 23 (right) shows a VARJO video export, illustrating the actual FOV captured by the XR sensor, which synchronizes with head movements to record fixations within the cockpit and external environment. HD pilots continued to show below-average fixation counts (Table 4) and longer average fixation durations, indicating suboptimal gaze strategies. Analysis of two separate AOI within the instrument panel

(PFD and EI) revealed that helicopter pilots had substantially fewer fixations in the PFD AOI, while HA pilots displayed more fixations and repeated glances in the EI AOI than other groups, though such extensive focus is unnecessary for effective power management during airplane takeoff, Table 3.

XR HMD-derived EM metrics were generally consistent with those from standard measurements across pilot groups. HD pilots underperformed in fixation numbers, saccade numbers, and smooth pursuit parameters, while HA pilots performed at or near average, and both airplane pilot groups performed well. Benchmark metrics from an expert FI provided an informative reference, highlighting differences in fixations, saccades, and smooth pursuits, Table 4. Graphical data (Figures 19 and 20) further confirm that expert saccade durations are consistently shorter and smooth pursuit velocities higher. These observations align with Robinski and Stein [44], who reported that experienced helicopter pilots employ target fixations (TFs) differently from novices, whose unintended TFs may cause them to miss critical flight cues. TF refers not to a single glance but to the total dwell time on specific targets or instruments.

5.3. Impact of XR Simulators

XR simulators provide safe and cost-effective training environments, enabling the exposure to diverse scenarios without real-world risk. Nevertheless, limitations in realism and fidelity may constrain the effective transfer of acquired skills to operational tasks. The findings of this study support the hypothesis that XR simulators can be effective training tools. Pilots adapted to the XR environment almost immediately, despite transitioning directly from the standard visual environment without a familiarization flight and having no prior experience with XR simulators. Moreover, EM data recorded using XR HMD-integrated sensors revealed performance patterns comparable to those obtained with standard eye-tracking systems across a diverse pilot cohort. These results highlight the potential for further spatial and behavioral analyses using open-source software tools and support continued scientific investigation in this domain.

5.4. Limitations

5.4.1. Sample Size

The number of candidate pilots enrolled in the theoretical part of the FI course was limited. In addition, the study design aimed to ensure participant diversity while maintaining balanced group sizes. For these reasons, the sample was restricted to eight participants. Future studies would benefit from the inclusion of a larger cohort of student pilots to improve statistical robustness and generalizability. Furthermore, the number of experimental sessions available for data collection was constrained by the intensive schedule of theoretical training. The decision to conduct a single XR simulator session without a prior familiarization flight was made deliberately in order to evaluate initial adaptation to the XR environment.

5.4.2. Flight Scenario

As none of the participants were expert pilots on the Lasta aircraft, a short takeoff and initial climb scenario was selected. Given the heterogenous experience levels within the pilot group, introducing more complex flight scenarios could have emphasized differences in aircraft-specific expertise rather than the characteristics and performance of the simulator itself. The chosen scenario, therefore, allowed for a more controlled assessment of simulator fidelity and pilot interaction with the XR environment.

5.4.3. Technological Issue

Overall, the equipment used in the study demonstrated reliable and stable performance. The primary technical challenge encountered was the storage and management of the large volumes of data generated by the VARJO XR3, particularly its high-resolution video recordings and gaze-tracking exports. This limitation further supported the decision to restrict the experiment to takeoff scenarios. Additionally, XR simulation environments demand substantial computational resources and advanced graphical processing capabilities, which must be carefully considered when designing extended training sessions or more complex flight scenarios.

The experimental setup was limited to a custom-built FTD specifically designed for advanced training on the Lasta military trainer aircraft. An XR solution based on an HMD was integrated into this FTD to enable the evaluation of trainer performance under XR conditions. Access to flight simulators representing civilian transport aircraft or helicopters would have further strengthened the study by allowing for a more comprehensive assessment of the sensitivity and generalizability of the proposed methodology. Nevertheless, the authors consider that the inclusion of a deliberately diverse pilot sample partially mitigates this limitation.

6. Conclusions

Flight simulators represent a prominent application of XR technology, supporting pilot training across multiple levels of complexity. Beyond aviation, XR-based simulators have been adopted in healthcare, athletic performance training, and educational environments, demonstrating their broad applicability. Despite this versatility, the effectiveness of transferring skills from simulated to real-world scenarios remains a subject of debate, and the capability of XR systems to fully replicate operational challenges is not universally accepted.

Through the careful selection of authentic aircraft features and the application of the proposed XR simulator development procedure, this study demonstrates that enhanced simulator fidelity can be achieved without compromising training effectiveness. An advanced experimental flight training platform was developed through the systematic engineering integration of several key components, which together form a robust environment for evaluating and comparing pilot training methodologies.

The platform integrates the following:

- XR Head-Mounted Display enabling immersive and realistic flight environments.
- High-Fidelity Physical Cockpit that reproduces the tactile and spatial characteristics of real aircraft controls.
- Synchronized Flight-Performance Logging that allows for precise temporal alignment between pilot inputs and aircraft responses.
- Eye-Gaze Sensing Technology for monitoring pilots' visual attention and scanning behavior during training tasks.

By integrating these elements, the proposed simulator enables controlled, repeatable, and objective comparisons between conventional simulator-based training and XR-enhanced training environments. This integrated framework constitutes the primary methodological contribution of the present work, offering a practical approach for simultaneously assessing piloting performance, gaze behavior, and training fidelity within a single experimental platform.

At present, the validation of eye-gaze metrics is performed internally through comparative analysis within the system. External validation across different eye-tracking devices and training platforms has not yet been conducted, which limits the generalizability of the findings. Addressing this limitation represents an important direction for future research.

Author Contributions: Conceptualization, A.K. and B.K.; methodology, A.K.; software, A.K., B.K. and B.R.; validation, A.K., B.K. and A.B.; formal analysis, A.K. and B.K.; investigation, A.K. and B.K.; resources, A.K., A.B. and B.R.; data curation, A.K., B.K. and D.P.; writing—original draft preparation, A.K. and B.K.; writing—review and editing, A.K., B.K., A.B., D.P. and B.R.; visualization, A.K. and B.K.; supervision, A.K. and B.K.; project administration, A.K., A.B. and D.P.; funding acquisition, A.K., B.K., A.B., D.P. and B.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research has been performed as part of activities within the project VA-TT/2/24-26 supported by the University of Defence in Belgrade.

Data Availability Statement: Data are contained within the article.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

XR	Extended Reality
VR	Virtual Reality
AR	Augmented Reality
MR	Mixed Reality
FTD	Flight Training Device
PCATD	Personal Computer-Based Aviation Training Device
UPRT	Upset Prevention and Recovery Training
PBN	Performance-Based Navigation
HMD	Head-Mounted Display
TR	Type Rating
VRFS	Virtual Reality Flight Simulator
HF	Human Factor
IMC	Instrument Meteorological Conditions
USAF	United States Air Force
UPT	Undergraduate Pilot Training
FI	Flight Instructor
SD	Standard Deviation
IFR	Instrument Flight Rules
CFI	Certified Flight Instructor
EEG	Electroencephalography
FOV	Field of View
SDK	Software Developer Kit
PPD	Pixels Per Degree
HVCs	Holographic Visual Cues
EASA	European Union Aviation Safety Agency
FAA	Federal Aviation Administration
IPD	Interpupillary Distance
CSV	Comma-Separated Values
AOI	Areas of Interest
EM	Eye Movement
AGL	Above Ground Level
HDM	Heading Magnetic
RMSD	Root Mean Square Deviation
IAS	Indicated Airspeed
TAS	True Airspeed
PFD	Primary Flight Display
EI	Engine Instruments

RPM	Revolutions Per Minute
MP	Manifold Pressure
FFS	Full Flight Simulator

References

1. Schuster-Amft, C.; Eng, K.; Suica, Z.; Thaler, I.; Signer, S.; Lehmann, I.; Schmid, L.; McCaskey, M.A.; Hawkins, M.; Verra, M.L.; et al. Effect of a Four-Week Virtual Reality-Based Training versus Conventional Therapy on Upper Limb Motor Function after Stroke: A Multicenter Parallel Group Randomized Trial. *PLoS ONE* **2018**, *13*, e0204455. [CrossRef]
2. De Pace, F.; Manuri, F.; Sanna, A.; Zappia, D. A Comparison Between Two Different Approaches for a Collaborative Mixed-Virtual Environment in Industrial Maintenance. *Front. Robot. AI* **2019**, *6*, 18. [CrossRef]
3. Chen, C.; Zhang, L.; Luczak, T.; Smith, E.; Burch, R. Using Microsoft HoloLens to Improve Memory Recall in Anatomy and Physiology: A Pilot Study to Examine the Efficacy of Using Augmented Reality in Education. *J. Educ. Technol. Dev. Exch.* **2019**, *12*, 17–31. [CrossRef]
4. Micunovic, V.; Knezevic, A.; Sibinovic, I. The Impact of Using Training Devices and Simulators in Pilot Training: A Case Study Based on PBN Performance. In Proceedings of the International Scientific Conference on Military Sciences "VOJNA 2025"; University of Defense: Belgrade, Serbia, 2025.
5. EASA (European Union Aviation Safety Agency). *Easy Access Rules for Aeroplane Flight Simulation Training Devices (CS-FSTD (A))*; EASA (European Union Aviation Safety Agency): Cologne, Germany, 2020.
6. Knezevic, A.; Bukvić, A.; Krstić, B. Application of Different Hardware and Software in the Concept of Digital Pilot Training and Selection. *Vojnoteh. Glas. (Engl. Mil. Tech. Cour.)* **2025**, *73*, 183–209. [CrossRef]
7. Knežević, A. *Analysis of Flight Parameters and Eye Movements under Altered Simulation Conditions Using Free Software—Experiments with the Application of the Concept of Digital Pilot Training and Selection*; Zenodo: Belgrade, Serbia, 2024. (In Serbian) [CrossRef]
8. Rizvi, S.; Rehman, U.; Cao, S.; Moncion, B. Exploring Technology Acceptance of Flight Simulation Training Devices and Augmented Reality in General Aviation Pilot Training. *Sci. Rep.* **2025**, *15*, 2302. [CrossRef]
9. Sheets, T.H.; Elmore, M.P. Abstract to Action: Targeted Learning System Theory Applied to Adaptive Flight Training. 2018. Available online: https://apps.dtic.mil/sti/trecms/pdf/AD1053015.pdf (accessed on 20 November 2025).
10. Oberhauser, M.; Dreyer, D. A Virtual Reality Flight Simulator for Human Factors Engineering. *Cogn. Technol. Work* **2017**, *19*, 263–277. [CrossRef]
11. Leland, R.; Rogers, R.O.; Boquet, A.; Glaser, S. *An Experiment to Evaluate Transfer of Upset-Recovery Training Conducted Using Two Different Flight Simulation Devices*; No. DOT/FAA/AM-09/17; Federal Aviation Administration: Washington, DC, USA, 2009.
12. Oberhauser, M.; Dreyer, D.; Braunstingl, R.; Koglbauer, I. What's Real About Virtual Reality Flight Simulation? *Aviat. Psychol. Appl. Hum. Factors* **2018**, *8*, 22–34. [CrossRef]
13. Ahlstrom, U.; Friedman-Berg, F. Using Eye Movement Activity as a Correlate of Cognitive Workload. *Int. J. Ind. Ergon.* **2006**, *36*, 623–636. [CrossRef]
14. Macchiarella, N.D.; Liu, D.; Gangadharan, S.N.; Vincenzi, D.A.; Majoros, A.E. Augmented Reality as a Training Medium for Aviation/Aerospace Application. In Proceedings of the Human Factors and Ergonomics Society 49th Annual Meeting—2005, Orlando, FL, USA, 26–30 September 2005; Volume 49, pp. 2174–2177. [CrossRef]
15. Ross, G.; Gilbey, A. Extended Reality (XR) Flight Simulators as an Adjunct to Traditional Flight Training Methods: A Scoping Review. *CEAS Aeronaut. J.* **2023**, *14*, 799–815. [CrossRef]
16. Hebbar, A.; Vinod, S.; Shah, A.; Pashilkar, A.; Biswas, P. Cognitive Load Estimation in VR Flight Simulator. *J. Eye Mov. Res.* **2023**, *15*, 1–16. [CrossRef]
17. Clifford, R.M.; Khan, H.; Hoermann, S.; Billinghurst, M.; Lindeman, R.W. Development of a Multi-Sensory Virtual Reality Training Simulator for Airborne Firefighters Supervising Aerial Wildfire Suppression. In Proceedings of the 2018 IEEE Workshop on Augmented and Virtual Realities for Good (VAR4Good), Reutlingen, Germany, 18 March 2018; School of Information Technology and Mathematical Sciences. IEEE: New York, NY, USA, 2018. [CrossRef]
18. Clifford, R.; McKenzie, T.; Lukosch, S.; Lindeman, R.; Hoermann, S. *The Effects of Multi-Sensory Aerial Firefighting Training in Virtual Reality on Situational Awareness, Workload, and Presence*; IEEE: New York, NY, USA, 2020; pp. 93–100. [CrossRef]
19. Mösl, B.; Schaffernak, H.; Vorraber, W.; Holy, M.; Herrele, T.; Braunstingl, R.; Koglbauer, I. Towards a More Socially Sustainable Advanced Pilot Training by Integrating Wearable Augmented Reality Devices. *Sustainability* **2022**, *14*, 2220. [CrossRef]
20. Lawrynczik, A. Exploring Virtual Reality Flight Training as a Viable Alternative to Traditional Simulator Flight Training. Master's Thesis, Carleton University, Ottawa, ON, Canada, 2018.
21. Ayala, N.; Zafar, A.; Kearns, S.; Irving, E.; Cao, S.; Niechwiej-Szwedo, E. The Effects of Task Difficulty on Gaze Behaviour During Landing with Visual Flight Rules in Low-Time Pilots. *J. Eye Mov. Res.* **2023**, *16*, 1–16. [CrossRef] [PubMed]
22. Yang, J.; Qu, Z.; Song, Z.; Qian, Y.; Chen, X.; Li, X. Initial Student Attention-Allocation and Flight-Performance Improvements Based on Eye-Movement Data. *Appl. Sci.* **2023**, *13*, 9876. [CrossRef]

23. McGowin, G.; Xi, Z.; Newton, O.; Sukthankar, G.; Fiore, S.; Oden, K. Examining Enhanced Learning Diagnostics in Virtual Reality Flight Trainers. *Proc. Hum. Factors Ergon. Soc. Annu. Meet.* **2020**, *64*, 1476–1480. [CrossRef]
24. Reweti, S.; Gilbey, A.; Jeffrey, L. Efficacy of Low-Cost PC-Based Aviation Training Devices. *J. Inf. Technol. Educ.-R* **2017**, *16*, 127–142. [CrossRef] [PubMed]
25. Le Ngoc, L.; Kalawsky, R.S. *Visual Circuit Flying with Augmented Head-Tracking on Limited Field of View Flight Training Devices*; American Institute of Aeronautics and Astronautics, Inc.: Boston, MA, USA, 2013. [CrossRef]
26. Zhou, Y.; Fu, C.; Wei, L.; Zhou, W.; Li, X.; You, Y. An Integrated Approach for Addressing Data Imbalance in Predicting Fatality of Helicopter Accident. *Reliab. Eng. Syst. Saf.* **2026**, *267*, 111921. [CrossRef]
27. Harris, D.; Hardcastle, K.; Wilson, M.; Vine, S. Assessing the Learning and Transfer of Gaze Behaviours in Immersive Virtual Reality. *Virtual Real.* **2021**, *25*, 961–973. [CrossRef]
28. Ahmadi, N.; Romoser, M.; Salmon, C. Improving the Tactical Scanning of Student Pilots: A Gaze-Based Training Intervention for Transition from Visual Flight into Instrument Meteorological Conditions. *Appl. Ergon.* **2022**, *100*, 103642. [CrossRef] [PubMed]
29. Kumar, S.P.; Kumar, S.K.; Prasad, N.N. An Approach to Develop Virtual Reality Flight Simulator for Combat Aircraft Cockpit Evaluation. *TAGA* **2018**, *14*, 3555–3560.
30. VIVE Developers. *SRanipal SDK*, Version 2.5.1; HTC Corporation: Taoyuan, Taiwan, 2022.
31. HTC. *Tobii XR Devzone*, Version 1.8.0; Tobii: Stockholm, Sweden, 2020.
32. Jusko, T.; Walko, C.; Kante, L. Holographic Visual Cues for Mission Task Elements. *CEAS Aeronaut. J.* **2025**, *16*, 1243–1264. [CrossRef]
33. EASA Approves the First Virtual Reality (VR) Based Flight Simulation Training Device. 2021. Available online: https://www.easa.europa.eu/en/newsroom-and-events/press-releases/easa-approves-first-virtual-reality-vr-based-flight-simulation (accessed on 11 November 2025).
34. Varjo Headsets Selected as Part of First FAA Approved VR Flight Simulation Training Device. 2024. Available online: https://varjo.com/news/varjo-headsets-selected-as-part-of-first-faa-approved-vr-flight-simulation-training-device (accessed on 3 November 2025).
35. Agtzidis, I.; Startsev, M.; Dorr, M. *Smooth Pursuit Detection Based on Multiple Observers*; Association for Computing Machinery: New York, NY, USA, 2016; pp. 303–306. [CrossRef]
36. Glaholt, M.G. *Eye Tracking in the Cockpit: A Review of the Relationships Between Eye Movements and the Aviator's Cognitive State*; Scientific Report DRDC-RDDC-2014-R153; Defence Research and Development Canada: Toronto, ON, Canada, 2014; p. 58.
37. Masson, G.; Stone, L. From Following Edges to Pursuing Objects. *J. Neurophysiol.* **2002**, *88*, 2869–2873. [CrossRef]
38. Rayner, K. Eye Movements in Reading and Information Processing: 20 Years of Research. *Psychol. Bull.* **1998**, *124*, 372–422. [CrossRef]
39. Lijing, W.; Ling, C. Eye Movement Comparison of Professional and Novice Pilots in Key Actions of Takeoff Phase. In Proceedings of the 2016 IEEE/CSAA International Conference on Aircraft Utility Systems (AUS), Beijing, China, 10–12 October 2016; Beihang University (BUAA), China: Beijing, China, 2016; pp. 703–706. [CrossRef]
40. Glaholt, M.; Reingold, E. The Time Course of Gaze Bias in Visual Decision Tasks. *Vis. Cogn.* **2009**, *17*, 1228–1243. [CrossRef]
41. Stojakovic, P.; Rasuo, B. Minimal Safe Speed of the Asymmetrically Loaded Combat Airplane. *Aircr. Eng. Aerosp. Technol.* **2016**, *88*, 42–52. [CrossRef]
42. Stojaković, P.; Velimirović, K.; Rašuo, B. Power Optimization of a Single Propeller Airplane Take-off Run on the Basis of Lateral Maneuver Limitations. *Aerosp. Sci. Technol.* **2018**, *72*, 553–563. [CrossRef]
43. Stojaković, P.; Rašuo, B. Single propeller airplane minimal flight speed based upon the lateral maneuver condition. *Aerosp. Sci. Technol.* **2016**, *49*, 239–249. [CrossRef]
44. Robinski, M.; Stein, M. Tracking Visual Scanning Techniques in Training Simulation for Helicopter Landing. *J. Eye Mov. Res.* **2013**, *6*, 1–17. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.