ENHANCING PILOT COGNITIVE RESILIENCE THROUGH NEUROERGONOMIC-INFORMED ADAPTIVE HUMAN-MACHINE INTERACTION FOR SPATIAL DISORIENTATION MITIGATION

YUAN XIN

PhD

The Hong Kong Polytechnic University

2026

The Hong Kong Polytechnic University

Department of Aeronautical and Aviation Engineering

Enhancing Pilot Cognitive Resilience through Neuroergonomic-informed Adaptive Human-Machine Interaction for Spatial Disorientation Mitigation

Yuan Xin

A thesis submitted in partial fulfilment of requirements for the degree of Doctor of Philosophy

August 2025

CERTIFICATE OF ORGINALITY

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no materials previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

______________________________ (Signed)

YUAN Xin (Name of student)

Abstract

Spatial Disorientation (SD) remains a critical threat to aviation safety, contributing to a significant percentage of fatal accidents. Existing SD mitigation strategies are often static and fail to adequately address the neurocognitive underpinnings of SD, such as Loss of Attention Control (LoAC) and a breakdown in cognitive resilience. Current approaches, including instrument training and auditory alerts, often fall short in dynamically adapting to pilots' evolving cognitive states. This thesis addresses this critical gap by establishing a novel neuro-ergonomic framework designed to enhance pilot cognitive resilience through adaptive Human-Machine Interaction (HMI).

The research unfolds across three interconnected stages. **First,** a comprehensive neuro-ergonomic investigation employs electroencephalography (EEG) and eye-tracking to decode pilots' neural and ocular signatures across four distinct SD types (No SD, Type I: unrecognised; Type II: recognised; Type III: incapacitating) during simulated flight phases. Based on these defined signatures, advanced machine learning models were developed and applied to classify a pilot's cognitive state and predict SD onset with an accuracy of 96.8% up to 20 seconds prior to symptom manifestation.

Second, to further refine predictive capabilities, we developed Hidden Markov Models (HMMs) to identify phase-specific scanning patterns, including Areas of Interest (AOI) transitions and scanning sequence strategies. Additionally, a hybrid approach combining Conditional Random Fields (CRF) with LightGBM, utilizing a sliding window technique, enabled fine-grained, real-time prediction of visual scanning anomalies. Leveraging these predictions of SD occurrence and visual scanning anomalies, the Adaptive Attention-Awareness Pilot-Aircraft Interaction System (3A-

PAI) was proposed. The successful integration of these models into a closed-loop HMI provides a concrete solution, offering a new HMI framework for pilot support by shifting the focus from reactive assistance to proactive, real-time intervention.

Third, to explore effective interaction paradigms, this phase investigated how enhanced display technologies could naturally improve pilot cognitive resilience, especially SD. The emergence of technologies like Head-Up Displays (HUDs) offers promising opportunities to revolutionise the cockpit experience by providing pilots with more intuitive and contextualised information. Two enhanced HUD prototypes—a holographic visor and a projection-based system—were engineered to optimise the future pilot-aircraft interaction paradigm within the cockpit. Among these, AR-glasses based HUD could broad the interaction boundary of cockpit, providing multimodal interaction. The projection-based HUD demonstrated more natural interaction approach and a superior user experience with lower learning costs. Finally, a closed-loop collaborative HMI system for SD mitigation was prototyped, where real-time SD predictions trigger context-aware HUD interventions to provide adaptive support to pilots.

This work collectively presents a novel, neuro-ergonomic framework that not only advances the understanding of four distinct SD types but also delivers a pragmatic, adaptive HMI solution, thereby setting a new trajectory for pilot-centric cockpit design and significantly enhancing aviation safety.

Supervisor: Prof. Kam K.H. NG

Co-supervisor: Prof. Gangyan XU

Publication arising from the thesis

** corresponding author; # Top 10% in the field; $ Top 5% in the field.*

Journal articles (in reverse chronological order)

Yuan, X., Ng, K. H.*, Li, Q., Yiu, C. Y., Zhou, Q., Bu, Y. (2025). From Unrecognised to Incapacitating: Neuro-ergonomic Differentiation of Spatial Disorientation Typologies Induced by Visual Illusions for Pilot Cognitive Training and Cockpit Design, Chinese Journal of Aeronautics, 103992. # *# Top 10% in the field*

Yuan, X., Ng, K. H.*, Li, Q., Yiu, C. Y., Lau, C. K., Fung, K. H., and Ng, L. H. (2024). Exploring the Human-Centric Interaction Paradigm: Augmented Reality-Assisted Head-Up Display Design for Collaborative Human-Machine Interface in Cockpit. Advanced Engineering Informatics, 62, 102656. *$ Top 5% in the field*

Manuscript under review (in reverse chronological order)

Yuan, X., Ng, K. H.*, Li, Q., Yiu, C.Y. (2025). Towards Pilot's Spatial Awareness Enhancement: An Adaptive Attention-Awareness Pilot-Aircraft Interaction System in Future Cockpit (Under 1st Review)

Yuan, X., Ng, K. H.*, Trapsilawati, F., Li, Q., Yiu, C. Y., Bu, Y., and Zhou, Q. (2025). Empowering Future Human-Machine Interaction Paradigms Considering Human Factors based on Ambient intelligence for Transportation Safety: A Review of Current Status, Issues, and Opportunities. (Under 1st Review)

Yuan, X., Ng, K. H.*, Li, Q., Yiu, C.Y., and Jin, Z. (2024). Recongising and Predicting the Interference Effects of Processing Visual and Auditory Information on Working Memory Capacity: A Preliminary Neuro-ergonomics Study. Accident Analysis and Prediction. (Under 1st Review)

Conference paper (in reverse chronological order)

Yuan, X., Ng, K.H.*, Li, Q., Lau, C.K., Fung, K.H. and Ng, L.H. (2025). Enhancing Pilot Engagement and Situational Awareness with Augmented Reality: A Neuro-Ergonomic Evaluation of an Innovative AR-HUD. AIAA AVIATION Forum 2025, Las Vegas, USA.

Yuan, X., Ng, K.H.*, Yiu, C.Y., and Li, Q. (2025). Optimising Pilot-Aircraft Interaction: A Low-Cost Projection-Enhanced Head-Up Display (PE-HUD) with Neuro-Ergonomic Validation. 27th International Conference on Human-Computer Interaction. Gothenburg, Sweden.

Yuan, X., Ng, K.H.*, Lau, C.K., Fung, K.H. (2024). Enhancing Single-Pilot Operations with Augmented Reality HUD: Improving Situational Awareness and Performance in a Human-Centric Cockpit Design. AIAA AVIATION Forum 2024, Las Vegas, USA.

Acknowledgements

I would first like to express my sincere gratitude to my supervisor, Prof. Kam. K.H. Ng. His kindness, understanding, and constant willingness to consider things from the students' perspective have always supported me throughout my PhD journey. His knowledge and guidance were invaluable, especially during the most difficult moments when he continued to find ways to help me move forward. I am also deeply grateful to my co-supervisor, Prof. Xu Gangyan, for providing timely support and guidance whenever I needed it. My thanks also go to Prof. Liu Wei, who, despite his busy schedule, kindly served as the Chair for both my confirmation and PhD oral defence, and offered me important advice and encouragement.

My sincere appreciation goes to the external reviewers of my dissertation, Prof. Han Ting and Prof. Wang Lei. Their professional comments greatly improved the quality of my thesis. I am also thankful to both professors for travelling from outside Hong Kong to attend my oral defence in person.

I would like to thank my lab mates - Li Qinbiao, Zhang Chenliang, Chu Nana, Liu Ye, Yao Senna, Zhu Yipeng, Jin Zhongyi, Du Congcong, Zhu Changxin, Tang Rong, Zhou Qinyu, Bu Yuanyuan, James, Dave, Alan, Terrence, Anson, Dickson, and Haowen Liang. Your laughter and companionship carried me through my entire PhD life. I truly hope our friendship lasts for many years to come. Thanks also to my FYP team members - Franky, Jeff, Loch, Jacky, Edwin and Alex, for your important contributions to this thesis. I would also like to thank my two soulmates, Xie Luyang and Kong Yunlu. Over the past three years, we have each gone through difficult times, yet we were fortunate to have one another's support. Although we may not keep in touch as frequently as before, our hearts have always remained connected.

My heartfelt gratitude goes to my parents. This was my first time studying so far away from home, and I know it was even harder for you than for me. Thank you for supporting me and always being there from afar. Please take good care of your health — there are still many beautiful places I want to take you to in the future. I am also deeply grateful to my boyfriend, Liu Huanxi, who stayed by my side through many nights of tears. This thesis is also dedicated to my grandparents in heaven — every step I have taken has been witnessed by you.

No matter how difficult the path, every challenge can be overcome. *Dream it, do it*, and *keep climbing*!

Sincerely,

Xin Yuan

Table of contents

List of Figures

List of Tables

Nomenclature

3A-PAI	Adaptive Attention-Awareness Pilot-Aircraft Interaction System
AI	Artificial Intelligence
AmI	Ambient Intelligence
AOI	Area of Interest
AR	Augmented Reality
AR-HUD	Augmented Reality Head-Up Displays
ATC	Air Traffic Controller
ATCO	Air Traffic Control officers
BCI	Brain-Computer Interfaces
CHMI	Collaboration Human-Machine-Interface
CRF	Conditional Random Fields
DL	Deep Learning
DPO	Dual-Pilot Operations
DT	Digital Twin
ECAM	Electronic Centralised Aircraft Monitoring
ECG	Cardiorespiratory Sensors
EEG	Electroencephalography
EMG	Electromyography
FAA	Federal Aviation Administration
FCU	Flight Control Units
FFT	Fast Fourier Transform
fNIRS	functional near-infrared spectroscopy
FTD	Flight Training Device
GUIs	Graphical User Interfaces

HAI	Human-Automation Interaction
HCI	Human-Computer Interaction
HFs	Human Factors
HITL	Human-in-the-loop
HMD	Head-Mounted Display
HMI	Human Machine Interaction
HMM	Hidden Markov Models
HR	Heart Rate
HUD	Head-up Display
IQR	Interquartile Range
IRU	Inertial Reference Unit
ISIS	Integrated Standby Instrument System
ITS	Intelligent Transportation Systems
LLM	Large Language Models
LMC	Left motor cortex
LOA	Level of Autonomy
LoAC	Loss of Attention Control
LOL	Left occipital lobe
LPFC	Left prefrontal cortex
LSTM	Long Short-Term Memory
MC	Motor Cortex
MCDU	Multifunction Control and Display Unit
ML	Machine Learning
MLP	Multilayer Perceptron
MSQ	Motion sickness questionnaire

NAMI	Naval Aerospace Medical Institute
ND	Navigation Display
OL	Occipital Lobe
OOB	Out-of-Bounds
OOTL	Out-of-the-loop
PE-HUD	Projection-Enhanced HUD
PF	Pilot Flying
PFD	Primary Flight Display
PM	Pilot Monitoring
PSD	Power Spectral Density
QRH	Quick Reference Handbook
RMC	Right motor cortex
ROL	Right occipital lobe
RPFC	Right prefrontal cortex
RR	Respiratory Rate
SA	Spatial Awareness
SAGAT	Situation Awareness Global Assessment Technique
SART	Situation Awareness Rating Technique
SD	Spatial Disorientation
SoP	Standard Operational Procedures
SPAM	Situation Awareness Present Assessment Method
SPO	Single-Pilot Operation
SSQ	Simulator Sickness Questionnaire
SVM	Support Vector Machine
SVS	Synthetic Vision Systems

TOI	Time of Interest
UNK	Unknown
USAF	U.S. Air Force
VR	Virtual Reality

Chapter 1. Introduction

This chapter, which consists of four main sections, introduces the thesis. Section 1.1 establishes the research context by detailing critical issues in aviation, namely the cognitive resilience, spatial disorientation (SD) and loss of attention control (LoAC), as well as adaptive human-machine interaction (HMI) within the cockpit, addressing human factors (HFs) considerations. This section thoroughly discusses relevant definitions and provides an overview of existing research. Following this contextualisation, Section 1.2 presents the specific research questions and their associated objectives. Subsequently, Section 1.3 addresses the significance of the present study. An outline of the thesis's structure is then provided in Section 1.4, the conclusion.

1.1. The Necessity of Pilot Cognitive Resilience in Modern Aviation

The aviation domain represents a high-risk and safety-critical environment where pilots must maintain exceptional cognitive performance under dynamic and frequently stressful conditions to ensure flight safety. Operating within complex, dynamic environments, pilots are required to process a large volume of rapidly changing information (Haslbeck & Zhang, 2017). Human error, particularly errors associated with impaired cognitive states, remains a primary cause of aviation accidents Qingli Liu et al. (2025). Consequently, sustaining optimal cognitive function, encompassing attention and spatial orientation capabilities, is paramount for flight safety.

1.1.1. Definition of Cognitive Resilience

Aviation operations require operators to sustain effective performance in environments where conditions evolve rapidly and unpredictably. Within this context, it is important to distinguish several related but conceptually different constructs. Cognitive flexibility concerns the capacity to shift strategies when facing novel or unexpected situations, allowing individuals to generate alternative solutions and avert hazards (M. Miller & Holley, 2023). Psychological resilience, in contrast, refers to the ability to maintain stable functioning and recover from setbacks during prolonged periods of stress or challenge (Ceken, 2025). At the neurophysiological level, such resilience is supported by mechanisms of neural plasticity that allow rapid re-stabilisation of cognitive processes under perturbation (Qingjin Liu et al., 2020). Cognitive resilience builds upon both concepts but occupies a distinct role in aviation. It captures an individual's ability to preserve coherent reasoning, resolve conflicts, and regain functional cognitive organisation when unexpected events disrupt ongoing mental models or impose excessive cognitive load (Liao et al., 2025). For commercial pilots, this resilience manifests in the capacity to diagnose and resolve technical inconsistencies or digital flight-deck anomalies without losing situational clarity (M. Miller & Holley, 2023).

1.1.2. Impact of Human Factors on Pilot Performance and Safety

About 75% of aviation accidents are attributed to human error, which is commonly mentioned as the main cause of these mishaps (Le Coze, 2022). These errors often originate from HFs and a lack of cognitive flexibility (Yiu et al., 2025c). The increasing digitalisation and automation in modern cockpits, ironically, can lead to a deficit in cognitive resilience. While automation aims to reduce workload, system malfunctions or unexpected events can quickly overwhelm pilots, leading to cognitive overload (Liao et al., 2025; Yiu et al., 2025b).

A notable phenomenon is that while automation is designed to reduce workload and enhance safety, it can, under specific failure conditions, introduce new and severe cognitive challenges that human pilots struggle to cope with without additional support (Papadimitriou et al., 2020). This reveals an "automation paradox," where an increase in automation, when the system deviates from normal operation, can paradoxically lead to a decrease in human cognitive resilience. This observation suggests that traditional pilot training and resilience-building strategies, which might focus more on manual flying skills or general stress management, may be insufficient to meet the demands of modern highly automated cockpits. Therefore, research must address this contemporary challenge within the context of advanced human-machine collaboration, shifting from a purely human-centric perspective to one of human-machine system integration. Future cockpit designs must proactively manage HMI to prevent cognitive overload and enhance resilience, rather than merely assuming automation inherently resolves HFs issues.

1.2. Research Background

1.2.1. Spatial Awareness and Loss of Attention Control in Aviation

1.2.1.1. Theoretical Concepts of Spatial Awareness (SA)

In flight operations, spatial awareness denotes the pilot's capacity to accurately interpret aircraft location, orientation, and dynamic state in relation to the Earth and other environmental references. The cognitive structure underlying this construct is commonly interpreted through Endsley's tripartite model of situation awareness, which conceptualizes awareness as progressing from perception to comprehension and ultimately to projection (Mica R Endsley, 1999). In dynamic flight contexts, SA enables pilots to integrate sensory inputs with mental models, thereby guiding effective decision-making and control actions.

The formation of SA relies on the integration of three core sensory modalities, namely visual information, vestibular cues from the inner ear, and proprioceptive signals generated by receptors embedded in the musculoskeletal system (Newman & FAICD, 2007). When these visual, vestibular, and proprioceptive inputs differ or are inconsistent in intensity, direction, and frequency, sensory conflicts arise, leading to illusions and triggering spatial disorientation (SD). Vision is thought to be the largest contributor to orientation (Gil-Cabrera et al., 2021; Ledegang & Groen, 2018). This is especially true in conditions of low visibility or monotonous environments, where visual cues are degraded. Vision is typically the dominant sense, providing approximately 80% of the sensory input required for orientation. However, in conditions of poor visibility, such as at night, in fog or on monotonous terrain, non-visual senses can be misled, leading to illusions and disorientation. Visual illusions caused by inconsistent instrument display also make confusion of the pilot perception, like Figure 1-1 shows. Even though pilots are aware of SD, it continues to be one of the primary reasons for aviation mishaps. According to statistical statistics, SD is responsible for 5% to 10% of general aviation accidents, with up to 90% of these being deadly illusions (Gibb et al., 2011). It causes catastrophic consequences, loss of aircraft control, and improper control inputs. The root cause is typically due to operator control loop failures resulting from distraction, interrupted cockpit scans, or task saturation (Newman & FAICD, 2007).

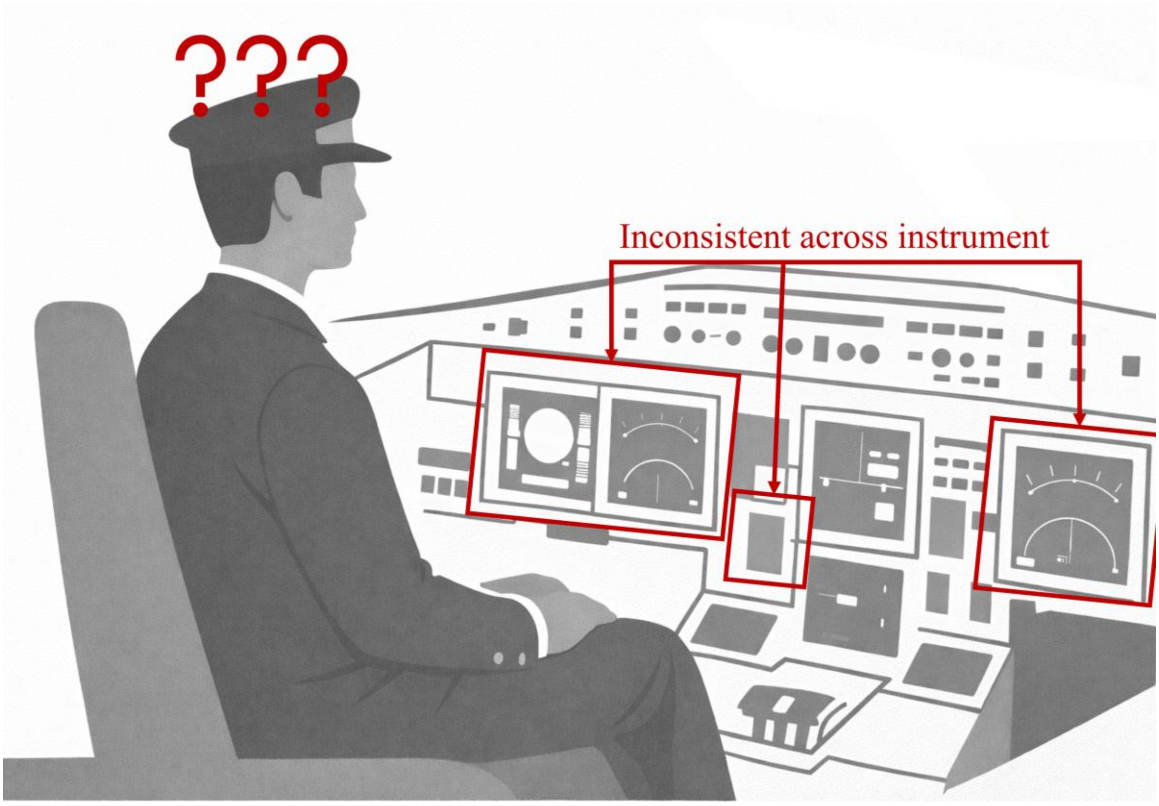


Figure 1-1. Sample of spatial disorientation occurrence caused by inconsistent instrument display.

According to the Naval Aerospace Medical Institute (NAMI), SD is categorised into three main types based on the pilot's level of awareness (Lyons et al., 1994):

- **Type I: Unrecognised.** In this type, the pilot fails to recognise the need for corrective control inputs because they are unaware of the disorientation. An example is the "Leans illusion". This type is considered the most dangerous type of SD and a leading cause of fatal accidents.
- **Type II: Recognised but manageable.** The pilot is able to recognise the perceptual confusion but may still make corrective control inputs. An example is the "Graveyard Spin".
- **Type III: Incapacitated.** The pilot is unable to recover control of the aircraft due to the extreme degree of incapacitation. Extreme sensory conflicts prevent

them from correcting the disorientation, even if they are aware of it. For instance, the "Coriolis effect".

1.2.1.2. Theoretical Concepts of Loss of Attention Control (LoAC):

Cognitive load refers to the extent to which working memory resources are taxed when processing task-relevant information, particularly when cognitive demands approach or exceed an individual's processing capacity (A. Wang et al., 2024). Errors and possible flight risks result from high cognitive load, which is harmful to attention, decision-making, and other cognitive functions. It is frequently linked to ineffective pilot monitoring during the cognitively taxing, takeoff, approach, and landing stages of flight. When pilots experience attentional lapses or fail to maintain effective visual scanning, which is known as Loss of Attention Control (LoAC). Attentional tunnelling, also known as attentional narrowing, is a phenomenon of LoAC (Van der Burg et al., 2025). Pilots fail to adequately perceive all relevant information due to pre-existing expectations or an excessive focus on a single area of interest (AOI) (White & O'Hare, 2022). This reduces the breadth of visual attention and can lead to misjudgements, reduced aircraft control, and unstable approaches.

The cumulative effect of LoAC can severely reduce available relevant information and increasing cognitive load (Haslbeck & Zhang, 2017). Missing critical alerts in the cockpit due to high workload and impaired attention can lead to "out-of-the-loop" (OOTL) problems, thereby triggering accidents (Liang et al., 2021). A significant observation is that while the increasing digitalization and automation in modern cockpits are designed to simplify operations, they can paradoxically lead to cognitive overload during malfunctions or complex scenarios. High cognitive load, in turn, impairs attention. This creates a dangerous vicious cycle: the complexity introduced by automation, if improperly managed, directly triggers cognitive overload; this overload then manifests as deficits in attention control and OOTL, further compromising pilot performance and situational awareness, ultimately undermining the safety benefits that automation was intended to provide (Ohneiser et al., 2018). This emphasises the necessity of designing HMIs in a proactive, adaptive manner that can detect and react to a pilot's cognitive state in real time, avoiding cognitive disruption and attentional failures instead of only responding to obvious mistakes.

1.2.1.3. Limitations of Existing Mitigation Strategies

Current mitigation strategies primarily include pilot training (e.g., experiencing SD illusions in simulators like the Barany chair (Geva et al., 2025; H. Hu et al., 2021), VR dizzy training (S. Kim et al., 2025), trusting instruments over sensory perception during flight (Ledegang & Groen, 2018; Sánchez-Tena et al., 2018). However, despite these measures, the persistently high fatality and incidence rates of SD accidents suggest their limitations. The Federal Aviation Administration (FAA) acknowledges the need for "new technologies to help make flying safer" in this domain (Antunano, 2021).

This highlights that current mitigation strategies, while necessary, do not adequately address the underlying cognitive vulnerabilities that lead to SD (Authority, 2016). Therefore, there is a need for more fundamental, real-time, and adaptive interventions that can directly counteract these sensory mismatches or enhance pilots' cognitive abilities to overcome misleading perceptions. This necessitates technological solutions that go beyond traditional training.

1.2.2. Adaptive Human-Machine Interaction

Adaptive HMI represents a significant evolution in avionics, designed to dynamically adjust its behaviour and interface to accommodate changing operational environments and the pilot's cognitive needs and capabilities (W. Yu et al., 2023). Unlike traditional systems, adaptive HMI proactively adapts through either user-driven customisation or system-initiated modification. Adaptive HMI's fundamental idea is to "present appropriate information in an efficient and effective format at the suitable moment" (Yuan, Ng, Li, et al., 2024), so the pilot stays "in the loop" with (semi-)autonomous systems. The evolution of human-machine collaboration is progressing from human-machine isolation (functional coordination) to human-machine trust (task coordination), and ultimately towards human-machine fusion (deep integration) (Kirwan et al., 2024).

Adaptive HMI has the potential to significantly enhance pilot performance and survivability by dynamically adjusting to evolving threats and task parameters. For instance, they can prioritise critical data while deemphasising non-urgent information to mitigate cognitive overload. They greatly contribute to reducing pilot manual and mental workload, improving SA, and enhancing overall operational efficiency and safety. Research indicates lower workload level and higher SA with the use of adaptive HMIs (W. Yu et al., 2023).

1.3. Research questions and objectives

Recognising the challenges to pilot cognitive resilience, the persistent threat of SD, and the potential value of adaptive HMI, this study aims to address the fundamental question of "what is" to "how to implement," thereby providing effective support for pilots and enabling its practical application and validation. The core research question is to explore how cognitive resilience, primarily in the context of SD, can be integrated with adaptive HMI. This will be broken down into three research questions:

1) *Theoretical exploration:* How can EEG and eye-tracking data be integrated to quantitatively identify specific neurophysiological characteristics and typical behavioural patterns of pilots in normal states versus different types of SD (e.g., unrecognised, recognised, and incapacitating SD)?
2) *Framework construction:* How can a multimodal adaptive HMI framework be constructed based on real-time neurophysiological and behavioural data, and how can advanced machine learning (ML) models be utilised to achieve predictive assessment of pilot SD types?
3) *Interaction paradigm design:* Within the aforementioned adaptive HMI framework, how can interaction strategies be designed and validated for different types of SD?

This study seeks to accomplish the following three goals in conjunction with the research questions:

1) *Neurophysiological and behavioural characterisation:* To identify and quantitatively characterise the specific neurophysiological features and scanning behavioural patterns of pilots in different types of SD. Determining the neuro-ergonomic patterns of the SD types serves as the basis for further detection and prediction of SD.
2) *Adaptive Attention-Awareness Pilot-Aircraft Interaction System (3A-PAI) framework construction:* To construct a robust cognitive resilience detection and prediction joint model which integrates neurophysiological and behavioural

data streams within an adaptive HMI framework, culminating in context-sensitive HMI interventions.

3) *Interaction paradigm design and validation:* To design and validate a set of interaction paradigms via AR for mitigating different types of SD. These paradigms will be designed within the proposed adaptive HMI framework and evaluated for their effectiveness in enhancing pilot cognitive resilience in simulated flight environments.

1.4. Significance

The significance of this research is embodied in three main aspects, which collectively constitute its theoretical and practical value:

1) **Systematically analysing the cognitive mechanisms of SD from a neuro-ergonomics Perspective.** Current research on SD largely focuses on phenomenological descriptions and traditional HFs engineering analysis. This study will systematically delve into the cognitive and neurophysiological mechanisms underlying different types of SD from a neuro-ergonomics perspective. By leveraging neurophysiological data such as EEG and eye-tracking, this research will quantitatively identify the characteristic features of pilots' brain activity and behavioural patterns during the occurrence of SD. This will lay the theoretical foundation for future early detection and intervention strategies for SD, filling a research gap in the neuro-ergonomic understanding of this domain.
2) **Establishing an adaptive attention-awareness pilot-aircraft interaction (3A-PAI) Solution.** Addressing the limitation of exploring the boundaries between distinct SD types' signatures, this research will propose and construct

a complete solution for integrating SD and LoCA's detection and prediction. This solution aims to achieve a paradigm shift from passively reacting to SD to proactively preventing it. By intelligently managing the information flow from monitoring SD and LoCA, it can effectively help pilots maintain an optimal cognitive state, enhance their cognitive resilience, and thus enable rapid recognition and effective response when SD occurs.

3) **Exploring future cockpit interaction paradigm: providing a blueprint for an intelligent, pilot-centric cockpit:** This research will go beyond SD mitigation itself, further exploring how its findings will influence the interactive paradigm of future cockpits. Specifically, we will focus on the emerging medium of Augmented Reality Head-Up Displays (AR-HUDs) to investigate and compare different interaction modalities. This exploration will provide a concrete and feasible interaction blueprint for intelligent, pilot-centric cockpit design, and lay the foundation for the next generation of aviation human-factors interaction technology.

1.5. Thesis organisation

Figure 1-2 shows the organisation and relationship between the work completed for this thesis and the three primary research phases of the broader framework. The remainder of this thesis is structured as follows, following a brief introduction in Chapter 1:

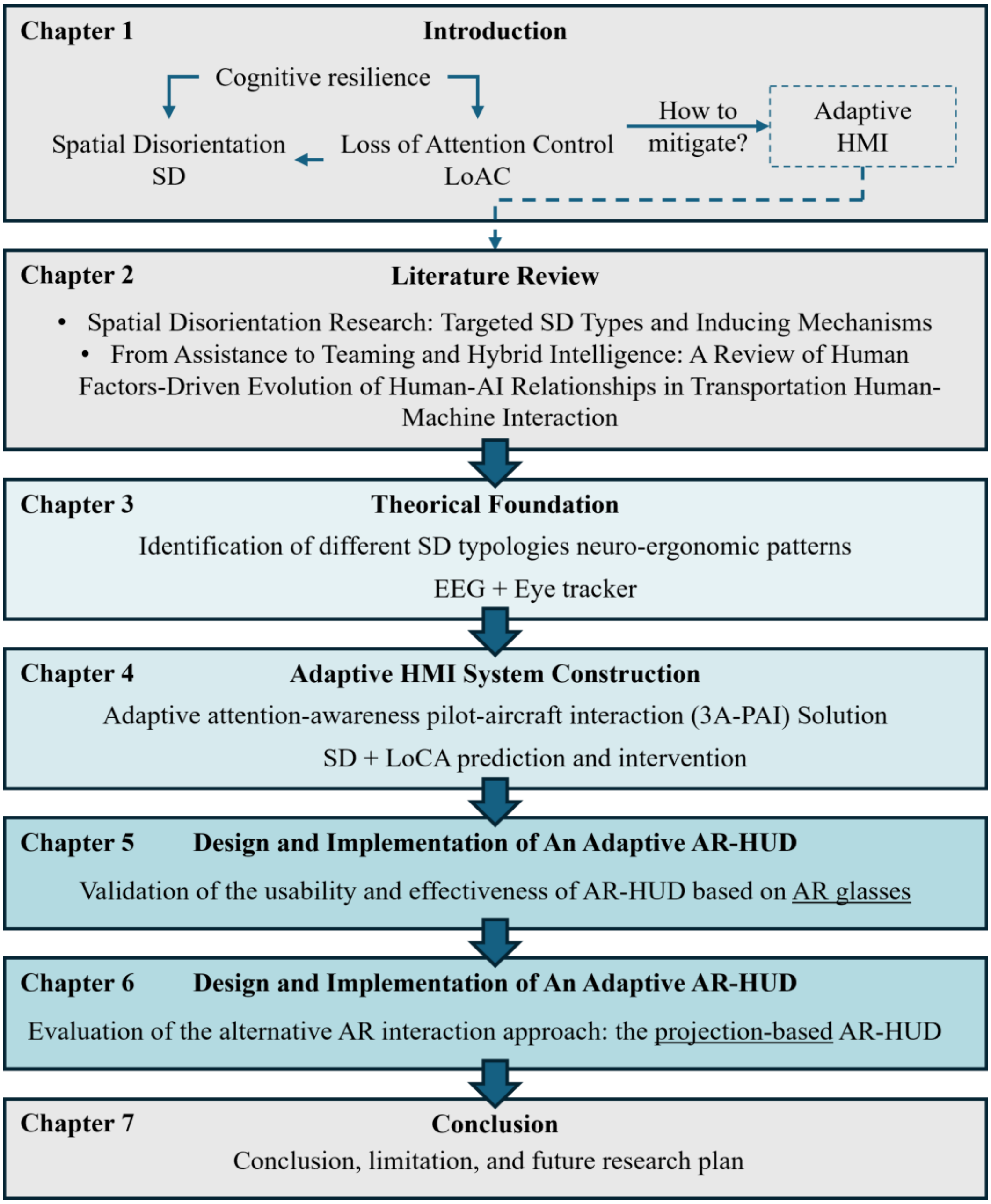


Figure 1-2. The structure of the thesis.

Chapter 2 presents a thorough and in-depth examination of the literature on 1) SD related research, including the mechanism investigation and mitigation strategy, and the 2) HMI paradigms, with a focus on the evolution of human-AI collaboration. This chapter seeks to highlight the critical role that HFs play in effective HMI design by critically analysing existing research. It also explores future pilot-aircraft interaction paradigm as the next-generation cockpit in aviation.

Chapter 3 details an initial empirical study, serving as a crucial precursor to the comprehensive research framework. The central aim of this investigation is to explore the distinct neuro-ergonomic signatures characteristic of four defined SD classifications (three types of SD and a normal, non-SD state). To achieve this, Electroencephalography (EEG) and eye-tracking data were utilised to evaluate pilots' cognitive states during SD triggered by simulated visual illusion scenarios. The extracted neuro-ergonomic patterns were then analyzed to delineate the boundaries and unique characteristics distinguishing these four distinct SD classifications.

Chapter 4 introduces the 3A-PAI system. Building upon the neuro-ergonomic signatures of the four SD classifications established in Chapter 3, the 3A-PAI system is developed to enhance pilot SA and attention stability through multimodal sensing and robust recognition of cognitive states. This proposed system is specifically designed to detect early signs of SD and LoAC, predict their occurrence, and subsequently deliver context-relevant interventions to support safe and effective pilot performance.

Chapter 5 presents the design and implementation of an adaptive Augmented Reality Head-Up Display (AR-HUD). A crucial initial step involved verifying and evaluating the usability of AR glasses as a viable platform for AR-HUD development. The AR-HUD interface was subsequently designed by applying the interaction strategies outlined in the previous chapter, with a specific investigation into its potential for assisting pilots during subtle emergency situations.

Chapter 6 introduces and validates an alternative AR presentation format: the Projection-Enhanced HUD (PE-HUD). This creative strategy seeks to get beyond the obstacles to adoption that conventional AR glasses solutions frequently face. This study used functional Near-Infrared Spectroscopy (fNIRS), a novel neuro-ergonomic technique, to assess the influence of the PE-HUD during particular flight phases, specifically takeoff and landing, in contrast to the methods used in Chapter 5. This study advances affordable HMI design in aircraft by establishing neuro-ergonomic evidence for phase-sensitive HUD validation.

The research is finally brought to a close in **Chapter 7**, which offers recommendations for future research paths as well as an explanation of the contributions.

Chapter 2. Literature review

2.1. Spatial Disorientation Research: Targeted SD Types and Inducing Mechanisms

SD has increasingly been recognised not as a unitary phenomenon but as a family of distinct yet interacting cognitive–perceptual states, each associated with different operational risks. Recent research has therefore shifted from descriptive accounts of disorientation to systematic investigations of which SD types are elicited, how they are induced, and how they degrade perception, cognition, and control during flight. Two organising dimensions dominate the contemporary literature: 1) the type of SD under investigation, 2) the sensory mechanisms through which SD is triggered and 3) SD mitigation strategies.

2.1.1. SD Types Addressed in the Literature

Across the selected literature, most studies treat spatial disorientation (SD) primarily as an induced operational hazard rather than as a rigorously classified state across Type I (unrecognised), Type II (recognised but manageable), and Type III (incapacitating) SD. In practice, many experiments infer SD from task manipulations (e.g., vestibular cues, degraded visuals) and behavioural consequences (e.g., control errors, scan disruptions), but do not formally map episodes to Type I–III. This creates an interpretive gap: the same induced condition may reflect different SD types depending on whether the pilot recognises the conflict, can resolve it, or experiences functional incapacitation.

Only a small subset of the literature explicitly isolates a particular SD type, and most of literatures only investigated the most prevailing one, Type I SD. Several simulator and motion-cue studies implicitly align with Type I SD because performance degradations occur even when pilots do not reliably report or recognise the SD event. Evidence includes inappropriate control inputs, disrupted instrument cross-checking, and altered scanning patterns under unanticipated SD events (Ledegang & Groen, 2018). These findings support the operational concern that the most hazardous SD may unfold **without subjective awareness**, undermining self-correction. (Landman et al., 2024) demonstrated that Type II SD can itself impair ongoing cognition, via surprise and increased mental workload. This work reframes SD awareness as a double-edged factor: recognition may be necessary for recovery, but it can temporarily destabilise cognitive performance and attention allocation. Within the reviewed set, Type III SD (incapacitating) is not a primary experimental target, likely due to ethical and safety constraints. Instead, the literature concentrates on non-incapacitating but operationally significant episodes that nonetheless produce measurable control errors, attention tunnelling, and impaired decision-making—mechanisms that can escalate risk if not detected and managed early (Webb et al., 2012).

2.1.2. Mechanisms for Inducing SD: Visual and Vestibular Pathways

The reviewed studies converge on two primary SD triggering pathways: (i) **vestibular system disturbances** (sensory conflict driven by motion cues) and (ii) **visual or information-based illusions** (degraded, misleading, delayed, or inconsistent external cues and display information). Importantly, the choice of trigger is strongly shaped by experimental platform constraints: in non–full-motion simulators or fixed-base contexts, **visual/information illusions** become the most practical and controllable method for eliciting SD-like effects.

A core vestibular mechanism is illustrated by leans-related paradigms, where subthreshold or misleading roll cues bias internal estimates of bank angle. Evidence shows that such vestibular cues can induce systematic roll-reversal errors when participants/pilots attempt to recover wings-level using the attitude indicator, implying that expectation and motion-based priors can dominate over instrument information in time-critical corrections (Landman et al., 2019). At a more physiological level, somatogyral-illusion work has begun to characterise candidate neurophysiological markers under rotational disorientation conditions, including reported decreases in frontal theta power during illusion periods alongside oculomotor signatures such as nystagmus. Although these studies are not framed around Type I–III SD, they strengthen the argument that vestibular SD has identifiable neural correlates that could support detection (Geva et al., 2025).

A second dominant pathway is visual illusion and information distortion, which is particularly common where motion cue fidelity is limited. The black hole illusion (BHI) is a prototypical example: by removing reliable external visual references during approach/landing, BHI environments impair pilots' perception of altitude and descent profile, producing phase-dependent degradations in operational performance (M. Chang et al., 2022; L. Huang et al., 2023). Crucially, visual-triggered SD does not require "pure" external visual deprivation; display-mediated inconsistencies can also produce SD-relevant breakdowns in spatial awareness. For instance, delayed data-linked weather radar images introduce a temporal mismatch between displayed hazards and the evolving environment, degrading pilots' spatial awareness and increasing the likelihood of misjudging hazard proximity and dynamics. This is a particularly relevant modern mechanism because it reflects information latency rather than classic sensory loss (Hua et al., 2022). More generally, simulator studies linking SD to visual scanning degradation show that unanticipated SD events—whether vestibular or visual—can reduce scan efficiency and distort cross-check patterns, reinforcing the view that SD is often expressed behaviourally as attention misallocation and scanning destabilisation (Ledegang & Groen, 2018).

In summary, vestibular triggers emphasise motion-cue conflict (leans/somatogyral mechanisms), while visual/information triggers emphasise degraded external cues and inconsistent/latent display information. In non–full-motion environments, visual or information illusions are widely used and demonstrably effective for inducing SD-relevant performance breakdowns.

2.1.3. Mitigation strategies for SD

Across the reviewed literature, mitigation is still predominantly training-centred, consistent with established aviation practice. A recurring theme is that SD events disrupt the instrument cross-check, and therefore training often targets (i) improving instrument scan discipline under conflict, (ii) increasing SD awareness, and (iii) rehearsing recovery procedures under representative scenarios.

Simulator-based evidence specifically supports using gaze tracking as feedback in SD training, because SD episodes measurably alter visual scanning and can evoke inappropriate control inputs even when not recognised. This implies that training effectiveness may be improved by objectively monitoring and coaching scan strategies rather than relying on subjective debrief alone (Ledegang & Groen, 2018). However, the literature also signals an important nuance: awareness training must be designed carefully because becoming aware of disorientation (Type II SD) can transiently impair cognition through surprise and increased workload. This suggests that effective mitigation should not only "raise awareness," but also train pilots to manage the immediate cognitive cost of recognition—e.g., by reinforcing task prioritisation, stabilisation heuristics, and structured instrument cross-check sequences during the recognition window (Landman et al., 2024).

Beyond training, a smaller but growing set of studies points toward technology-enabled mitigation, primarily through detection and monitoring. Machine-learning approaches for SD detection in rotary-wing contexts illustrate how flight data can be used to discriminate SD-related patterns, positioning SD identification as a component of automated safety monitoring (Sehoon et al., 2024). Likewise, neurophysiological work suggests that EEG (and eye movement signatures) may provide additional markers of vestibular SD states, supporting future multimodal recognition frameworks (Geva et al., 2025).

Mitigation in the current literature is mostly pilot training (scenario-based SD exposure, instrument cross-check coaching, and awareness/recovery training). Yet, emerging detection research (ML on flight data; EEG/oculomotor markers) provides a pathway toward closed-loop mitigation, where state inference can trigger timely support rather than relying solely on the pilot's self-recognition.

2.2. From Assistance to Teaming and Hybrid Intelligence: A Review of Human Factors-Driven Evolution of Human-AI Relationships in Transportation Human-Machine Interaction

2.2.1. Introduction

2.2.1.1. The Challenge of Evolving HMI and Cognitive Failures

In safety-critical transportation, Human-Machine Interaction (HMI) serves as the vital link between operators and intelligent systems (Joo & Shin, 2019). Rapid advancements in Artificial Intelligence (AI), sensing, and data-driven modelling have transformed HMI from static interfaces into multi-dimensional, highly automated interaction terminals (F. Li et al., 2022). Yet increasing system intelligence alone does not guarantee improved human performance or safety (Yiu et al., 2025b). Highly automated systems operated in dynamic, information-dense environments fundamentally shift the operator's role, demanding intensified cognitive processing to manage (J. Zhang & Edwards, 2017). However, human cognitive limitations under high time pressure and information load make purely automation-centric designs insufficient. This increased complexity often precipitates critical human factor (HF) issues, such as degraded situation awareness (SA), fluctuating workload, and miscalibrated trust (Mengtao et al., 2023). When HMI design fails to account for these cognitive constraints under high-pressure scenarios, it leads to operational failures, mode confusion, and misaligned autonomy (Agrawal & Peeta, 2021). Thus, traditional automation-centric HMI is no longer sufficient for the cognitive demands of modern transportation.

2.2.1.2. The Shift Toward Human-Centric HMI

To address these cognitive bottlenecks, the HMI design paradigm has shifted from superficial interface-level optimization toward a deeper integration of human cognitive processes, giving rise to **human-centric HMI** (W. Xu et al., 2023) (Dekker & Hollnagel, 2004). Rather than treating humans as passive recipients of system outputs, this paradigm emphasizes active human involvement and adaptive support that is aligned with operators' cognitive states and task demands (Kant et al., 2025). This goal can only be met if HMI is designed as an adaptive feedback system that continuously integrates contextual information from the environment with ongoing assessments of human cognitive states (Yuan, Ng, Lau, et al., 2024). By leveraging emerging AI, such as adaptive automation and generative AI—human-centric HMI enables dynamic function allocation (Lyu & Li, 2025) and context-sensitive support (Causse et al., 2025). Ultimately, the core challenge of the next-generation HMI is no longer just increasing algorithmic intelligence, but ensuring that AI is seamlessly integrated into the human cognitive loop, prioritizing HF capabilities as the foundational constraint for safety and performance

2.2.1.3. The Evolutionary Stages of Human-AI Relationships

The realization of this human-centric paradigm depends on the depth of integration within the human-AI relationship, which has undergone a fundamental HF-driven evolution (Xia et al., 2025). Rather than a static configuration, this relationship manifests in three progressive stages that redefine the HMI mediating layer. Initially, **AI-assisted human** systems functioned as external tools **(**Z. Li et al., 2024**),** providing warnings or recommendations without directly engaging the operator's internal cognitive loop. **As autonomy increased,** the relationship evolved into **human-AI teaming (HAT) (**McNeese et al., 2021**),** where HMI facilitates coordination between two autonomous agents. Most recently, the field has moved toward **human-AI hybrid**

(HAH) systems ([Ali et al., 2025](#)**).** In this stage, AI is no longer a separate peer but is embedded within the human cognitive process itself, enabling seamless offloading and proactive support during cognitive degradation. This evolutionary trajectory transforms HMI from a mere display interface into a dynamic cognitive bridge linking human reasoning, AI intelligence, and system behavior. Figure 1 illustrates the evolution of human-AI relationships and the role of HMI in three stages.

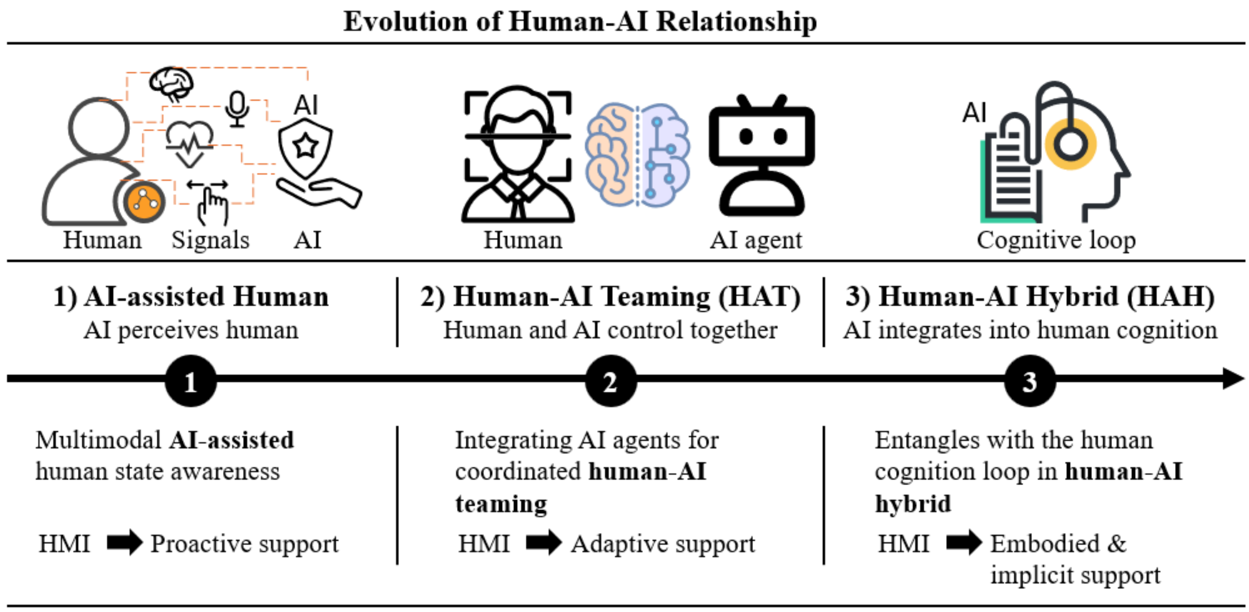


Figure 2-1. Evolution of human-AI relationships

Although many studies address each of these areas individually, existing reviews typically examine one facet at a time—for example, focusing on human state monitoring (Lohani et al., 2019; T. Zhang et al., 2020), automation strategies (Xing et al., 2021), or interface design (Amditis et al., 2010). Notably, recent surveys (Ali et al., 2025; Xia et al., 2025) likewise concentrate on specific stages or aspects rather than examining how HMI should evolve across the AI assistance-HAT-HAH continuum from a HFs perspective. As a result, the progression of the human-AI relationship and its implications for HMI design has been largely overlooked. In particular, how HFs should guide the transition from basic AI assistance to true HAT and ultimately HAH remains an open question. To address this gap, this review adopts a HFs–driven perspective, framing AI-enabled HMI as a closed-loop interaction between human and machine rather than a collection of independent components. The review is guided by the following research questions (RQs):

- **RQ1:** How can multimodal sensing enable **AI-assisted** HMI to move beyond passive HFs monitoring toward human state awareness in transportation contexts?
- **RQ2:** How can HFs inform the integration of generative AI and adaptive automation into the AI agent to enable coordinated **HAT**, rather than isolated automation?
- **RQ3:** How can future interaction modalities be evolved to realize **HAH** systems, where seamless cognitive coupling transcends explicit communication to achieve symbiotic intelligence?

2.2.2. AI perceives human: Multimodal AI-assisted human state awareness for proactive HMI

This section addresses how multimodal sensing enables AI-assisted HMI to move beyond passive HFs monitoring toward human state awareness in transportation contexts. It is structured around three tightly connected perspectives. Traditional HF topics, including SA, workload, attentional distribution, and human-automation interaction are no longer treated solely as surface-level performance measures. Instead, they are reframed as latent cognitive conditions that intelligent systems seek to model and interpret (Olaverri-Monreal & Jizba, 2016). Second, commonly adopted sensing modalities and measurement techniques are reviewed as the foundations for capturing observable signals related to these human states. Third, AI-enabled approaches are discussed in terms of how multimodal signals are interpreted, integrated, and transformed into meaningful representations of underlying human states. Together, these three perspectives illustrate the conceptual and technical pathway through which AI begins to "understand" the human operator.

2.2.2.1. Human factor issues underlying transportation HMI

Workload and fatigue management

Physical and mental fatigue is widely recognised as a major safety concern in transportation, accounting for an estimated 15-20% of transport-related accidents (Crestelo Moreno et al., 2022). Fatigue may arise from multiple sources, including excessive mental workload, prolonged task demands, insufficient sleep, circadian disruption, and physical exhaustion (Martins et al., 2021). With the increasing adoption of automation, operators are gradually shifting from active control roles to supervisory monitoring, which introduces new forms of fatigue associated with cognitive underload rather than overload (Matthews et al., 2019). In highly automated aviation and maritime operations, routine and repetitive tasks often require limited active engagement, increasing the risk of reduced alertness and passive fatigue, particularly during long-duration or monotonous operations (Lampe & Deml, 2022).

Recent advances in HMI enable the continuous assessment of workload and fatigue by integrating data from physiological sensors, behavioural measures, and operational information. By supporting real-time monitoring of operator states, such systems allow early identification of fatigue-related risks and provide a basis for timely intervention before performance degradation leads to safety-critical outcomes (Martins et al., 2021).

From a human-state perspective, workload and fatigue are therefore not only performance factors but key indicators of an operator's capacity to maintain effective interaction with automated transportation systems.

Human-automation interaction issues

Automation has fundamentally transformed transportation, applying to road, rail, and aviation domains. Automation is commonly implemented to mitigate workload and optimise system performance; however, human factors concerns increasingly centre on the dynamics of human–automation coordination (Yiu et al., 2024). In safety-critical domains such as aviation and road transport, automation does not eliminate the human role but restructures it, requiring carefully calibrated interaction to sustain effective joint performance (Janssen et al., 2019). However, the increasing level of automation has introduced new interaction demands, including challenges related to system transparency, supervision, and timely intervention (W. Yu et al., 2023).

When interaction between humans and automated agents is poorly designed, operators may drift into an out-of-the-loop (OOTL) condition, marked by declining awareness of system dynamics and limited readiness to assume control when necessary (Lyu, Li, Lee, et al., 2024). OOTL conditions are often associated with reduced situational awareness (SA), excessive reliance on automation, and delayed or inappropriate responses during unexpected events. Although human-in-the-loop (HITL) concepts emphasise active human involvement in automated systems (Papadimitriou et al., 2020), high perceived automation reliability often encourages reduced vigilance (Ayas et al., 2023) and over-trust (Gegoff et al., 2023). As a result, operators may struggle to regain SA and perform safe take-over actions after extended periods of automation use, particularly in safety-critical situations (Berberian et al., 2017). From a human state perspective, these interaction issues reflect dynamic changes in operator awareness, engagement, and readiness to act, which are central to safe human-automation cooperation.

2.2.2.2. Multimodal sensing foundations for human state inference

Recent advances in multimodal sensing and AI techniques have substantially improved the ability of transportation systems to perceive and interpret human-related information, creating new opportunities for enhanced HMI (Tan et al., 2022). Recognition technologies in HMI aim to capture observable human, system, and environmental signals that support the understanding of operator behaviour and internal states (Yuan, Ng, Li, et al., 2025). Rather than focusing on individual devices, this section organises existing recognition technologies around sensing modalities and their roles in human state inference. Wearable-based sensing techniques enable continuous and on-site data collection during real-world operations. By capturing neural and physiological signals, neuro-ergonomic approaches make it possible to infer underlying cognitive demands, alertness fluctuations, and fatigue-related changes. Visual behaviour-sensing approaches capture overt indicators of attention allocation and SA. Together, these multimodal sensing foundations enable AI-assisted systems to move beyond passive monitoring toward meaningful interpretation of human states.

Understanding human states in transportation HMI requires access to observable signals that reflect operators' internal cognitive and perceptual conditions. This section reviews multimodal sensing approaches that provide the empirical foundations for human state inference. Rather than focusing on individual devices, the discussion is organised around sensing modalities and their roles in capturing complementary aspects of human behaviour and physiology, as Figure 2-2 shows. Wearable-based sensing platforms support continuous and in-situ data collection during real-world operations. Physiological and neuro-ergonomic techniques offer insight into internal cognitive states such as workload, fatigue, and vigilance. Visual behaviour-sensing approaches capture overt indicators of attention allocation and situational awareness. Together, these modalities form the basis for AI-enabled interpretation of human states by linking observable signals to underlying cognitive and behavioural processes.

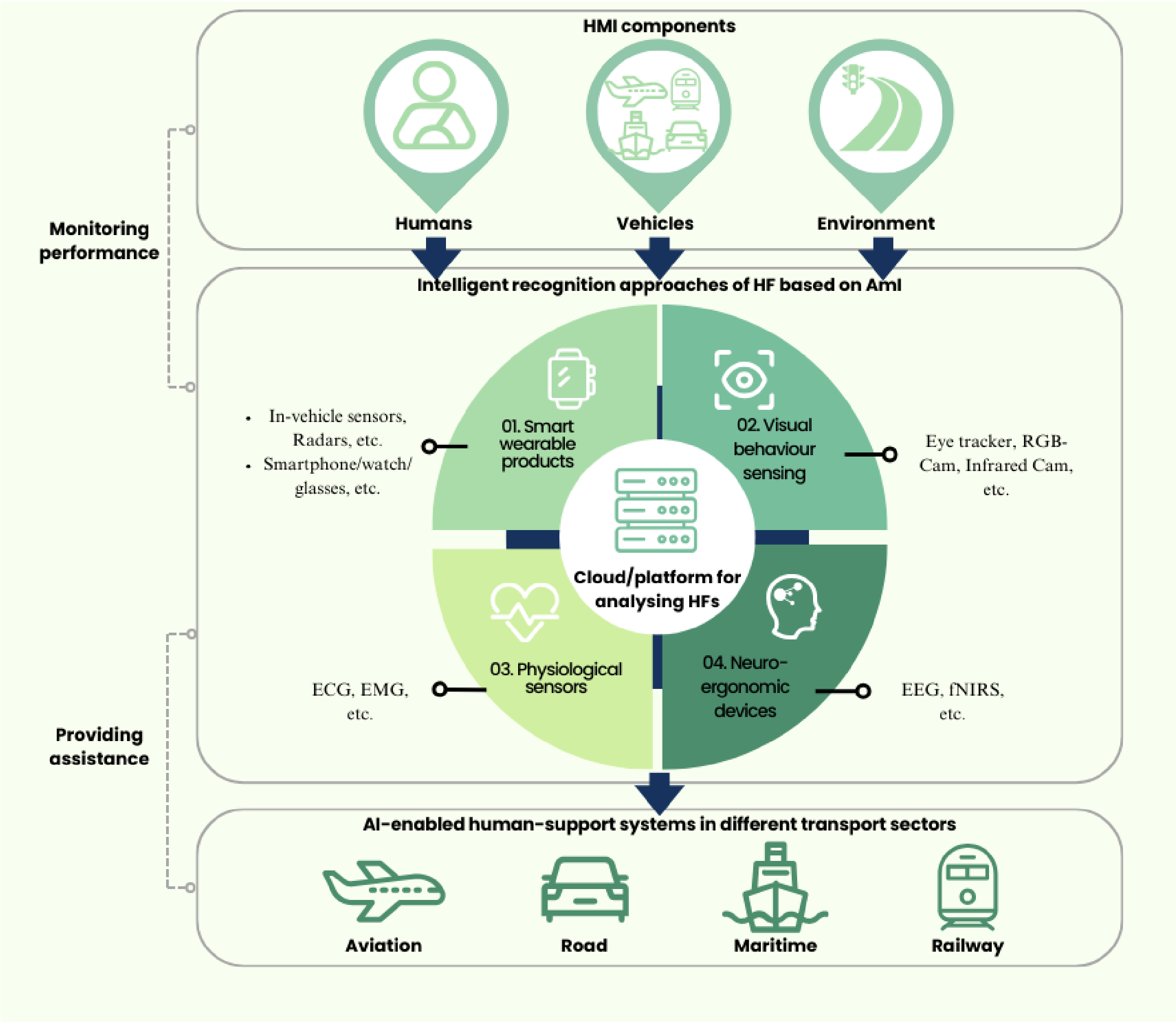


Figure 2-2. A Review of recognition technologies empowered by AI in transportation industry to identify HF issues.

Wearable-based multimodal sensing

Advances in communication technologies and embedded electronics of ITS have enabled the widespread adoption of wearable devices capable of continuous and real-time data collection (Tan et al., 2022). Wearable-based sensing platforms, including smartphones, smartwatches, smart glasses, and other body-worn devices, provide a practical means to capture human-related signals during naturalistic transportation operations (Stefana et al., 2021). By integrating multiple sensors such as cameras, accelerometers, gyroscopes, GPS, and physiological monitors, these platforms support the collection of behavioural, contextual, and physiological data relevant to human state assessment.

Smartphone-based applications have been widely used to collect mobility and behavioural data by combining onboard sensors with vehicle information (Bergasa et al., 2014). Compared with handheld devices, wearables such as smartwatches (Rodriguez-Paras et al., 2021) and smart glasses (W. C. Li et al., 2020) offer more stable and continuous monitoring, while also enabling timely feedback and alerts to operators. Physiological measures obtained from wearables, including heart rate (HR) (M. M. Dong, 2023) and oxygen saturation (Spaccarotella et al., 2022), have been used to support the assessment of operator fatigue and reduced alertness in safety-critical contexts (Çevik & Dahanayake, 2022). From a human state perspective, wearable-based multimodal sensing does not merely enable passive monitoring but provides continuous and context-aware observations that form a key foundation for AI-assisted inference of operator states in transportation systems.

Physiological and neuro-ergonomic sensing

Physiological and neuro-ergonomic sensing techniques provide direct access to internal human states that are not readily observable through behaviour alone. By capturing signals related to cardiovascular activity, muscular effort, and brain function, these techniques form a critical foundation for AI-assisted inference of workload, fatigue, vigilance, and cognitive engagement in transportation contexts.

Cardiorespiratory Sensors:

Cardiorespiratory sensors, particularly electrocardiography (ECG), are widely used to monitor physiological activity of the cardiovascular and respiratory systems, including HR, heart rate variability (HRV), respiratory rate (RR), blood pressure, and oxygen saturation (Chui, 2022). These measures are closely linked to autonomic nervous system activity and have been extensively applied to assess fatigue, workload, and stress in safety-critical operations (K. Lu et al., 2022). Continuous analysis of HRV patterns, especially during prolonged duty periods or repetitive task cycles, enables the identification of gradual physiological changes associated with accumulating fatigue. From a human state perspective, ECG-based measures provide temporally sensitive indicators that support real-time assessment of operator readiness and physiological capacity.

Electromyography (EMG):

EMG's value lies in its ability to detect neuromuscular fatigue, a direct consequence of prolonged activity and high-stress workloads (Sikander & Anwar, 2019). Surface EMG captures the neuromuscular system's time series signal from the skin's surface; its nonlinear dynamics and time-frequency domain are especially sensitive to the kind and degree of muscular activity (Cifrek et al., 2009). Monitoring muscle activities with EMG can identify early fatigue signs and assess fatigue's impact on operational performance. Such data are crucial for developing effective fatigue management strategies, including rest breaks, workload distribution and scheduling adjustments.

Neuro-ergonomic techniques:

Neuro-ergonomics primarily aims to identify neural correlates associated with perceptual, motor, cognitive processes and mental states, such as alertness, fatigue and workload by neuroimaging methods (Hernandez-Sabate et al., 2022). Situated measurements of brain activity in operational environments allow neuroergonomics to elucidate how cognitive mechanisms are dynamically influenced by interface configurations and task-imposed demands. Neuroimaging methods can be broadly categorised into direct and indirect approaches. Electroencephalography (EEG) directly measures electrical activity generated by neuronal firing, enabling high-temporal-resolution observation of cognitive dynamics and rapid state changes (Debener et al., 2005). Functional near-infrared spectroscopy (fNIRS), as an indirect method, assesses cortical blood oxygenation and provides spatially informative indicators of neural activation associated with sustained cognitive effort (Yuan et al., 2022). Neurological bases of behaviours and mental states in real-world tasks enable the management of adaptive automation through brain-computer interfaces (BCIs) (Dunbar et al., 2020). These BCIs utilised fewer EEG channels and featured a lightweight design. They can offer feedback to modify and adjust behaviour and brain activity, thereby improving learning, self-awareness, decision-making safety, and effectiveness. This approach encourages researchers to better understand how HMI elements affect cognitive and physiological responses by using neuro-ergonomic principles. Advances in wearable and portable neuroimaging technologies have enabled neuro-ergonomic approaches to be deployed in real-world operational environments rather than being confined to controlled laboratory contexts. When integrated with other physiological and behavioural signals, these neural measures enable AI-assisted systems to move beyond surface-level monitoring toward inference of latent cognitive states, supporting adaptive interaction and more human-aware HMI design.

Visual behaviour-sensing

Eye-tracking technology is now a commercialised instrument of visual behaviour-sensing technologies to analyse human visual attention and interactions by precisely capturing eye movements and gaze patterns. This technology employs infrared sensors, cameras or reflective markers to track and record eye movements, gaze points and pupil dilation, providing detailed insights into visual perception, attentional processes and cognitive behaviours (Kanaan et al., 2024). Compared with contact-based physiological sensing techniques such as EEG or EMG, eye-tracking offers a less intrusive and more user-acceptable means of data acquisition, making it particularly suitable for operational environments such as driving (F. Zhou et al., 2022) and aviation cockpits (Mengtao et al., 2023). The resulting eye-tracking data are commonly classified into two categories: gaze features and pupillometric measures. Unlike physiological sensors that must be attached to the body, such as EEG or EMG, eye-tracking does not need physical contact with the participant, offering a more comfortable and less obtrusive method of data collection. Eye-tracking datasets typically comprise two principal components: gaze behaviour and pupil-related measures. Parameters derived from gaze behaviour, including fixation patterns, saccadic transitions, and dwell duration, reflect the spatial and temporal allocation of visual attention (AB, 2024). Pupillometry focuses on eye closure, blink rate and pupil radius, providing insights into cognitive load, stress and attentional shifts (AB, 2024). From these, more complex measures such as visual entropy can be derived. Building upon these primary features, higher-level indicators, such as visual entropy, can be derived to capture the temporal dynamics and organisational structure of visual scanning behaviour, offering deeper insights into human cognitive states under varying task demands and environmental conditions.

Taken together, the multimodal sensing approaches reviewed in this section provide the empirical basis for human state inference in transportation systems by capturing complementary behavioural, physiological and neural information. However, these sensory signals do not map directly onto higher-level human states such as workload, situational awareness or disorientation, due to their nonlinearity, temporal dynamics and inter-individual variability. As a result, effective human state inference requires advanced data-driven methods capable of integrating heterogeneous signals and extracting latent cognitive states. This challenge has motivated the increasing adoption of AI-assisted approaches, which transform multimodal sensing data into actionable representations for adaptive and human-centric HMI design.

2.2.2.3. AI-assisted human state inference in transportation contexts

Building upon multimodal sensing foundations, AI-assisted human state inference has become a central mechanism for interpreting complex HFs in transportation systems. The real-time assimilation of multimodal data supports ML- and DL-based architectures in modelling state transitions and projecting performance vulnerabilities before observable behavioural failure occurs. These capabilities support a range of safety-critical applications, including continuous human state monitoring, timely alerts to external hazards, early warning of degraded cognitive performance and adaptive automation or takeover strategies in intelligent vehicles. Rather than focusing on algorithmic novelty or predictive accuracy alone, this section examines how AI methods operationalise multimodal sensing data for HFs recognition, assessment and future prediction within HMI frameworks.

Aviation

Table 2-1. Key literatures of the intelligent recognition approach in aviation sector.

Refs	HF issues	Sensing techniques	Key contributions
(Rodriguez-Paras et al., 2021)	SA, Workload	Smartwatch	Developing vibrotactile cuing to support pilot SA of updated weather information based on smartwatch.
(W. C. Li et al., 2020)	SA, Workload	Smart glasses	Merging the flight mode annunciator with flight data in the cockpit to realise augmented visualization cues, which facilitates pilot SA and workload.
(J. H. Yang et al., 2024)	Workload	ECG	Changes in cognitive demand across plane tracking, anti-gravity pedalling, and reaction tasks were quantified through HRV analysis. The analytical indicators consisted of LF power, HF power, normalised HF values, and the LF-to-HF ratio.
(J. Chen et al., 2022)	Workload	ECG, EMG, Eye tracker	A multimodal feature set comprising EMG signals, abdominal respiratory rate (RR), heart rate (HR), blood oxygen saturation, pupillary metrics, fixation duration, fixation frequency, and saccadic behaviour was processed through a hierarchical SVM classifier to support online estimation of mission-induced workload.
(E. Q. Wu et al., 2021), (Wu, Cao, et al., 2022)	Fatigue	EEG	Fatigue classification accuracy with an adversarial deep Bayesian neural network (BNN)

(Jiang et al., 2022)	Workload	EEG, Eye tracker	Mental workload was identified using an LSTM architecture augmented with an attention module, yielding a classification accuracy of 94%.
(Wu, Zhu, et al., 2022)	Fatigue	EEG	Brain fatigue dynamics were modelled using a nonparametric hierarchical hidden semi-Markov framework to enable state-level fatigue identification.
(P. Li et al., 2023)	Fatigue	EEG	Utilising global 3D brain power maps and local 1D rhythm-specific features from EEG to accurately detect driver fatigue with 88.5% accuracy
(C. R. Liang et al., 2023)	Workload	Eye tracker, ECG	Comparing touch screen and mouse, color and grayscale map, and information complexity to evaluate different HMI. Color map and mouse interaction modes aids in information processing and enhancing pilot SA.
(John et al., 2022).	Workload	Eye tracker, ECG, EEG	Unravelling several physiological features related to collision prediction in ATC task: Pupil size increases with the workload, but the blink rate decreases.
(Eisma et al., 2021)	Distraction	Eye-tracker	The Solution Space Diagram was proposed as an advanced feedback mechanism designed to guide ATCs' attentional distribution and enhance their ability to identify potential conflicts.
(Di Flumeri et al., 2019)	OOTL	EEG, Eye tracker	The "Vigilance and Attention Controller" was introduced as an adaptive framework capable of regulating automation authority based on the detection of OOTL signatures.
(J. Liu et al., 2016)	Workload	ECG, Eye tracker	A cognitive pilot–aircraft interface (CPAI) framework was proposed to enhance single-pilot operation through the coordinated consideration of physiological signals, operational demands, environmental conditions, cognitive states, and interface-level parameters.

Consumer-grade smart wearables, such as smartwatches and smart glasses, hold significant potential to enhance pilot-aircraft interaction in civil aviation. Smartwatches can continuously monitor pilots' HR-related data, limb movements, oxygen saturation, and sleep quality data, providing real-time detection on fatigue levels, stress, and cognitive workload (M. M. Dong, 2023). In addition, when the smartwatch detects an abnormal situation, it can remind the pilot through sound and tactile feedback and interaction. For example, Rodriguez-Paras et al. (2021) prompt the pilot to perceive weather conditions through vibration alerts from the smartwatch, as well as multitasking while processing weather information. This HMI design facilitates the recognition of weather information more frequently and faster, without affecting the time spent on other information. Further, with the development of augmented reality (AR), smart glasses can incorporate enhanced vision systems to serve as head-up displays (HUD) by projecting critical real-time flight information, including airspeed, altitude and navigation data. These technologies could enhance visibility in low-light or poor weather conditions (W. C. Li et al., 2020). The unobtrusive and hands-free nature of these wearable technologies allows for seamless integration into the cockpit, optimizing pilot-aircraft collaboration and improving overall flight safety and efficiency.

In addition to smart wearable products, ECG can provide more scientific monitoring of HR and HRV. Continuous monitoring of HRV patterns and extracting a composite HRV index using component analysis, enables the evaluation of high cognitive task load and provides a practical tool to assess cadet pilot training level (J. H. Yang et al., 2024). In addition, for the real-time workload monitoring, J. Chen et al. (2022) detected multiple physiological data, including EMG, HR, RR and blood oxygen saturation. The sensitivity of the method of integrating multiple physiological data is high and enables five levels of flight task workload classification.

Compared with conventional ML methods, DL methods based on neuroimaging have more advantages in HF identification and prediction. First, processing neuroimage has richer spatial information, and pure neural signal data lacks this rich spatial dimension (Craik et al., 2019). In addition, DL is good at integrating multiple neural signals to obtain a more comprehensive and accurate assessment of human status (Y. Y. Zhou et al., 2022). Deep learning models applied to neuroimaging data offer the additional advantage of linking learned representations to underlying neural structure and functional activity, thereby improving model interpretability and advancing the understanding of cognitive mechanisms (Y. Y. Zhou et al., 2022). Several studies have leveraged this potential in aviation contexts. For example, E. Q. Wu et al. (2021) introduced a deep contractive sparse autoencoder to detect pilot fatigue from EEG signals, and subsequently enhanced classification robustness using an adversarial Bayesian neural network framework (Wu, Cao, et al., 2022). Jiang et al. (2022) adopted a multimodal strategy that combined EEG, gaze behaviour, and flight performance indicators to estimate workload levels, employing an LSTM architecture to discriminate among five graded workload states. More broadly, probabilistic deep architectures such as Bayesian neural networks (Wu, Cao, et al., 2022) and hidden semi-Markov models (Wu, Zhu, et al., 2022) have been utilised to model fatigue dynamics and infer latent cognitive states across varying manoeuvre conditions. P. Li et al. (2023) proposes a 5-D DL network, E-5-D-BCDRNet, that utilises global 3D brain power maps and local 1D rhythm-specific features from EEG to accurately detect human fatigue with 88.5% accuracy, outperforming existing methods. The study reveals regional brain dynamics associated with fatigue, demonstrating the value of DL on neuroimaging data for neuro-ergonomics applications. However, wearing devices for monitoring brain activity such as EEG or fNIRS can trigger operator discomfort, which can interfere with task performance.

Visual behaviour sensing technologies have also proven valuable for enhancing pilot-aircraft interaction and flight desk design. Eye-tracking technology has also been instrumental in evaluating the effectiveness of cockpit displays, warning systems and flight simulators, facilitating the design of more intuitive and informative interfaces. As reported by C. R. Liang et al. (2023), touch screen and colour display modes can shorten fixation times and visual search efforts, thereby aiding information processing and enhancing pilot SA. Except for the pilot's study, the HF issues of air traffic controller (ATC) are equally important, as ATC is responsible for traffic coordination in complex airspace. Generally, pupil size increases with the workload, but the blink rate decreases (John et al., 2022). Eye-movement patterns have been leveraged to enhance attentional support in air traffic control. For instance, Eisma et al. (2021) introduced the Solution Space Diagram as an advanced feedback mechanism designed to guide controllers' attentional allocation and improve conflict detection performance. Given that high levels of automation may reduce sustained operator engagement and contribute to OOTL, Di Flumeri et al. (2019) proposed a "Vigilance and Attention Controller" integrating eye-tracking and EEG measures to dynamically monitor attentional states and adjust system support accordingly. The significance of a sensible information layout for maintaining human awareness and flying safety was demonstrated by these eye-tracking experiments.

Lastly, J. Liu et al. (2016) introduced the cognitive pilot–aircraft interface (CPAI) framework as an integrative architecture that synthesises physiological signals, operational context, environmental conditions, cognitive states, and system-interface attributes to enhance safety in single-pilot operations. The unique feature of the CPAI system is that the pilot's operator interface is driven by cognitive states, specifically defined by workload and fatigue levels, and finally by using varying levels of automation to redistribute task compliance to regulate the pilot to stay at optimal performance.

Road transportation

Table 2-2. Key literatures of the intelligent recognition approach in road transportation sector.

Refs	HF issues	Measurements	Key contributions
(H. Kim et al., 2023)	Others (Lack of communication)	Smartwatch	Collecting motion signals by smartwatch to recognise specific gesture, so as to enhance natural interaction between drivers and vehicles.
(W. J. Chang et al., 2018)	Fatigue/drowsiness	Smart glasses	Incorporating a miniature IR bandpass light sensor on the wearable smart glasses to detect fatigue.
(Hussain et al., 2023)	SA	EMG	Predicting the intended driving behaviour (left/right turn) by SVM to build the ADAS.
(Fan et al., 2021)	Others (Norms)	EMG	The system was designed to recognise five representative forms of anomalous driving behaviour in real time: turning back (TB), forward reaching (FF), object retrieval (PU), sudden steering adjustments, and sunroof interaction (TS).
(T. Xu et al.,	Fatigue	EEG	Proposing a unified framework to detect

2023)			driving fatigue with high performance based on CNN-attention structure.
(Perello-March et al., 2023)	SA, Trust	fNIRS	Using fNIRS to measure trust level in automation. Results indicated trust was related to monitoring and the reduction of working memory. Distrust was closely linked to emotional mechanism.
(Y. Yang et al., 2024), (Ramanathan Parthasarthy et al., 2021)	Distraction	EEG	Both roadside and in-vehicle visual distraction negatively affect speeding behaviours.
(Qin et al., 2022)	SA	Eye tracker	Proposing a real-time most safety-relevant object detection approach for the development of ADAS.

Wearable sensors are being used in more diverse applications in road traffic, and the HMIs developed are smarter and more user centred. Beyond as the physiological signal sensing device, smartwatches are emerging as promising input devices across disciplines due to advancements in wearable and ML technologies. H. Kim et al. (2023) introduced an automotive user-interface (UI) concept that transformed the driver’s lap into a touch-sensitive interaction surface through smartwatch-based motion sensing. To support gesture recognition, they formulated a supervised learning framework capable of extracting discriminative features from multivariate time-series motion data, achieving a classification accuracy of 94.0% across ten predefined gestures. Similarly, W. J. Chang et al. (2018) designed an integrated drowsiness–fatigue mitigation system that combined wearable smart glasses with in-vehicle telematics, onboard diagnostic modules, and rear-light alert mechanisms to monitor driver state and provide timely feedback. The driver's state of sleepiness or fatigue is monitored by infrared light sensors on the smart glasses, and when an abnormal state is detected, the vehicle's rear lights automatically flash to alert the driver. Hussain et al. (2023) designed the ADAS based on EMG monitoring, which was used to predict the intended driving actions by the conventional support vector machine (SVM). The perception of surrounding physical entities was realised by Lidar to measure the distance of tailing vehicles and camera to detect the lanes and vehicles. Therefore, ADAS is able to warn the driver about the risk before the potential collision happens.

In road transportation, the neuroergonomic studies underscore the importance of ergonomic evaluation in specific driving scenario, such as suburban areas. Fan et al. (2021) further applied multiple machine learning models to achieve a more fine-grained recognition of abnormal driving behaviours, expanding upon previously identified behavioural categories. Road traffic also use EEG to monitor driving fatigue online, which can be used as an on-board expert system for autonomous driving, showing great potential (T. Xu et al., 2023). A popular topic when it comes to autonomous vehicle is trust. Perello-March et al. (2023) innovatively demonstrated the strong relationship between SA and trust in automation during autonomous driving. The authors identified two distinct cortical mechanisms detected by fNIRS, one associated with trust, characterized by decreased monitoring and working memory, and another linked to distrust, which exhibited an event-related and emotion-based neural response.

Studies in road transportation have explored visual scanning dynamics during key driving operations, encompassing hazard anticipation, vehicle lane control, and engagement with onboard interfaces. For instance, eye-tracking studies have provided insights into the distraction caused to drivers due to the presence of external stimuli, particularly concerning the effect of billboards (Y. Yang et al., 2024) and in-vehicle display (Ramanathan Parthasarthy et al., 2021) on altering a driver's gaze strategy. Further, eye-tracking technology has been integrated into driver monitoring systems to detect driver drowsiness, distraction and fatigue. The highest potential metrics for fatigue identification is associated with features linked to eye-opening speed and PERCLOS, whereas the average eye closure/open speed exhibits a negative correlation with the amount of human stress (Němcová et al., 2021). Computer vision is also an important technology in AI, and the ID-YOLO system developed by Qin et al. (2022) uses camera-based eye-tracking technology to monitor the presence of significant and critical objects in AOIs or safety relevance to the driver, reducing the interference of irrelevant scene information.

2.2.3. AI coordinates with human: Integrating AI agents for adaptive human-AI teaming

This section focuses on adaptive AI acting as an agent within HAT, rather than as an isolated driver-support or automation module. In this configuration, the AI agent operates within the human cognitive loop, supporting both strategic and tactical decision processes by continuously modelling human states and coordinating actions in a reciprocal manner. Within transportation HMI contexts, such agent-based integration enables shared authority by allowing the AI to dynamically adjust its roles, recommendations, and interventions in response to evolving human states and operational demands. Accordingly, this section reviews how adaptive AI agents operationalise HAT and how dynamically adaptive HMIs mediate this coordination in real-world transportation systems.

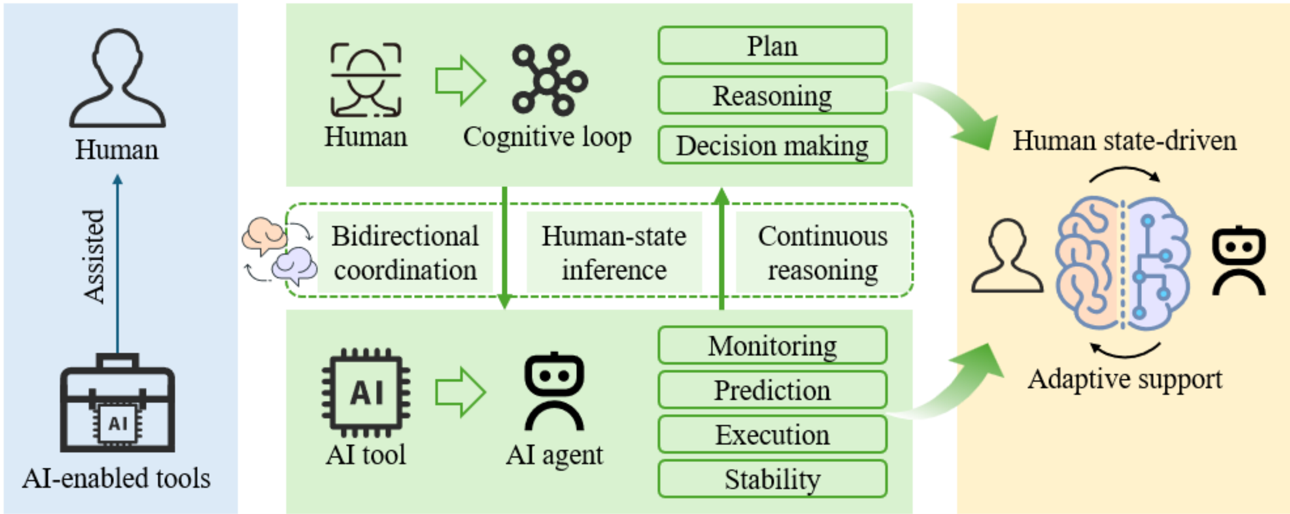


Figure 2-3. AI-enabled human-support system.

2.2.3.1. From autonomous system to adaptive automation assisted by AI agents

Traditional transportation automation has largely been framed through levels of automation (LOA), which specify how control and functions are allocated between humans and machines. Within this paradigm, the primary design objective has been to identify an optimal LOA that maximises system performance while retaining sufficient human involvement. However, empirical evidence consistently shows that increasing LOA can operators' SA (Mica R. Endsley & Kiris, 1995) and task performance (Strand et al., 2014), indicating that fully autonomous solutions do not inherently improve human-automation interaction. These findings expose a structural limitation of autonomy-centred designs: interaction is governed by predefined authority boundaries rather than continuous cognitive coordination. Adaptive automation represents an important step toward HAT by introducing state-aware regulation of automation authority. By monitoring operator workload, attention, or SA, adaptive systems dynamically adjust automation levels to maintain engagement and readiness. Flexible take-over strategies have therefore been proposed to improve human-automation coordination during authority transitions (Calhoun, 2021). In this context, multimodal sensing approaches have been developed to detect preparatory actions for take-over, such as shifts in visual attention, posture, and motor readiness, achieving high classification accuracy and low detection latency (Teshima et al., 2024). Similarly, physiological and behavioural data, including EEG and eye-tracking measures, have been used to predict SA levels and support smooth transitions during take-over requests (Yang et al., 2023).

Table 2-3. Key literatures of the AI-enabled semi-/fully autonomous transport systems.

Application domain	Refs	HF issues	Measurements	Key contributions
Road	(Teshima et al., 2024)	OOTL	Camera, Pressure sensors, Eye tracker	An automated approach was developed to identify the early emergence of preparatory driver actions preceding take-over transitions.
Road	(Yang et al., 2023)	SA, OOTL	EEG, Eye tracker	The use of neural network models allows for better prediction of the driver's SA level at pre- and post-take-over.
Aviation	(Yiu et al., 2025b)	SA, OOTL,	EEG	Proposing two-stage explainable deep learning model based on CNN and LSTM to identify OOTL, so as to support the adaptive automation configuration.
Aviation	(Sauer et al., 2012)	OOTL, Workload	In-cabin sensors	The sensor is applied through noise so as to test the effect of different LOAs on compensating the negative effects caused by noise.
Aviation	(Xu et al., 2025)	SA	Eye tracker	Utilizing a Transformer-based multimodal deep learning model to predict pilot SA levels by fusing high-fidelity eye-tracking (gaze) data with flight control data.

In aviation research, adaptive automation has increasingly been supported by data-driven modelling of pilot cognitive states. Recent advances in deep learning have enabled the identification of latent operator conditions and the dynamic calibration of automation authority. For example, Yiu et al. (2025b) ntroduced a two-stage explainable deep learning architecture that combined convolutional neural networks (CNNs) and LSTM models to analyse EEG signals and detect OOTL states during cruise flight. The resulting framework was designed to inform real-time adjustments of automation configuration. The value of such dynamic regulation is consistent with earlier findings. Sauer et al. (2012) demonstrated, through comparative evaluation, that automation schemes capable of adaptive modulation offer clear advantages over static designs in maintaining human engagement. Beyond engagement monitoring, multimodal approaches have further extended state inference capabilities. Xu et al. (2025) proposed a Transformer-based deep learning model that fused eye-tracking metrics with flight control parameters to estimate pilots' SA levels. In addition to operational application, this SA assessment model was positioned as a training instrument to evaluate novice pilot proficiency.

Despite these advances, adaptive automation remains largely transition-oriented: automation authority is adjusted, but decision-making remains sequential rather than shared. Automation is treated as a tuneable tool, not as an active cognitive participant. This gap motivates a shift toward adaptive AI agents that operate within the cognitive loop, enabling continuous coordination with human operators, an issue addressed in the next section.

2.2.3.2. Adaptive AI agents for human-AI teaming in strategic and tactical decision-

making

HAT extends adaptive automation by conceptualising AI as a cognitive teammate rather than a controllable subsystem. From a socio-technical perspective, the focus shifts from authority transfer to joint sensemaking, coordinated decision-making, and mutual predictability, particularly under abnormal or time-critical conditions. Accident analyses across transportation domains indicate that failures often stem from breakdowns in collaborative decision-making rather than isolated human or system errors, underscoring the need for AI systems that can participate directly in human-centred control processes.

Table 2-4. Key literatures of the HAT.

Application domain	Refs	AI application/function	Key contributions
Road	(Rothfuß et al., 2023)	Decision making	A cooperative automation architecture was formulated to facilitate joint decision making, integrating concepts derived from negotiation and game-theoretic models.
Road	(Srivastava et al., 2024)	Decision-making	An AI-guided trajectory planning approach was introduced to align human–AI situational awareness through calibrated decision support.
Road	(Lai, 2024)	Driving assistance	Proposing a framework of adaptive driving (FAD) to identify four common driving strategies: competitive driving, skill-based behaviour, defensive driving and urgent reactions based on environmental and eye-tracking data.
Aviation	(C. Y. Yiu et al., 2022)	Communication Decision-making	Employing EEG to evaluate the shared SA level between ATCOs under adverse weather condition.
Aviation	(Yin et al., 2023)	AI take-over assistance	Using eye-tracker to sense the pilot's attentional perceptual differences, and when the pilot and the AI agent's attentional characteristics agree, the pilot controls the aircraft, otherwise, the AI-agent

			controls the aircraft.
Road	(S. Zhang et al., 2024)	Decision-making	TrafficGPT was introduced as a multi-LLM fusion framework designed to perceive, interpret, and process complex traffic data streams, thereby supporting human decision-making in traffic management contexts.
Road	(P. Wang et al., 2024)	Decision-making	TransGPT was developed as a domain-specialised generative model capable of synthesising traffic scenarios, explaining traffic dynamics, and responding to traffic-related analytical queries.
Aviation	(F. Li et al., 2024)	Virtual pilots	Virtual pilot serves as the first officer (FO) to assist in automated quick procedures searching according to SOPs.
Aviation	(Lyu & Li, 2025)	Identifying OOTL	An eye-tracking–augmented LLM architecture was introduced to recognise troubleshooting patterns and provide adaptive support when pilots exhibit OOTL conditions.

In road transportation, AI-enabled teaming frameworks have explored how adaptive agents support tactical decisions through intent-aware reasoning and coordinated control. Rather than issuing unilateral commands, these systems align their recommendations with human goals, anticipate operator actions, and adjust behaviour to preserve shared SA and decision continuity (Lai, 2024; Rothfuß et al., 2023; Srivastava et al., 2024). Such approaches position AI agents as partners in tactical reasoning, contributing to conflict resolution, trajectory selection, and manoeuvre planning while retaining human authority.

In aviation contexts, human-AI collaboration has further expanded toward group-level coordination. Rather than evaluating operators individually, shared SA has been assessed across multiple human agents to regulate automation in a manner that supports team decision-making. C. Y. Yiu et al. (2022) implemented an AI agent grounded in

Bayesian neural network (BNN) modelling to align SA between pilots and air traffic controllers during adverse weather operations. By improving state consistency, the system supported coordinated communication and decision processes under elevated workload. It's not just about crew members collaborating with a single AI agent; teaming-oriented AI systems increasingly incorporate explicit models of pilot attention and cognition. Systems such as Air-Guardian (Yin et al., 2023) exemplify this transition by integrating real-time eye-tracking, visual saliency modelling, and liquid neural networks to monitor pilot attention and intervene during attentional lapses. By explicitly modelling both human attentional states and task-relevant environmental cues, such systems enable coordinated human-AI attention management rather than reactive automation, thereby enhancing flight safety without displacing pilot authority.

Recent advances in Generative AI (GAI), particularly Large Language Models (LLMs) serving as sophisticated AI agents, are fundamentally reshaping the strategic scope of HAT in transportation by transitioning AI's role from a passive tool to an active cognitive teammate. The core advantage of these agents lies in their generative reasoning capabilities, which allow for dynamic support of high-level planning, complex problem decomposition, and multimodal information synthesis through natural language interaction. LLM-based agents, such as TrafficGPT (S. Zhang et al., 2024) and TransGPT (P. Wang et al., 2024), demonstrate how generative reasoning can support high-level planning, problem decomposition, and multimodal information synthesis through natural language interaction. By enabling humans to query, explore, and iteratively refine complex transportation scenarios, these agents facilitate shared strategic reasoning rather than static decision support. In aviation, the concept of a virtual co-pilot (V-COP) extends this paradigm to tactical operations by reasoning over standard operating procedures, monitoring cockpit states, and providing context-sensitive guidance during abnormal or high-workload situations (F. Li et al., 2024). Collectively, these systems exemplify how adaptive AI agents function as cognitive teammates within the strategic and tactical layers of HAT. LLM can also be applied to process non-semantic eye-tracking data. Lyu and Li (2025) proposed a framework for detecting and managing troubleshooting behaviours exhibited by ITL pilots. Their approach integrates a LLM with non-semantic eye-tracking data that are transformed into structured Visual Attention Matrices (VAMs). By encoding gaze patterns into a matrix-based representation, the method bridges low-level visual attention signals with higher-level reasoning processes, enabling the model to distinguish between troubleshooting actions and routine monitoring behaviours. VAM representation thus functions as an intermediate abstraction layer, facilitating LLM-based inference while preserving the temporal and spatial characteristics of visual attention.

2.2.3.3. Dynamic human state-driven adaptive HMI for mediating human-AI teaming

While adaptive AI agents enable shared decision-making, effective HAT ultimately depends on how this coordination is enacted through the HMI. In contrast to static interfaces designed for nominal conditions, HAT-oriented HMI must dynamically adapt to human cognitive states, operational context, and task demands. Human-state-driven adaptive HMI therefore constitutes the interaction layer through which HAT becomes operational and sustainable.

Table 2-5. Key literatures of the AI-enabled personalised HMI design.

Application domain	Refs	Functions	Key contributions
Aviation	(Q. Li et al.,	Triggering	Applying fNIRS to monitor ATCO’s cognitive workload,

	2021).	wake-up signals to decrease workload.	so as to develop a user-generate and system-generated data-driven decision support system which could personalise the interface by triggering warnings and wake-up signals.
Railway	(Verstappen et al., 2022)	Providing information and advice.	Driver advisory systems (DAS) including the route context information and coasting advice benefit to decrease the workload.
Road	(Y. Liang et al., 2023)	Enhancing display targets and provision of advice	Introducing a visual reasoning-based AR-HUD mechanism to compensate SA by providing perceptual information, such as identifying road elements, moving elements' intention recognition, and generating driving strategy recommendations.
Road	(J. Dong et al., 2023)	Personalising driving decisions.	Generating personalised driving decision in autonomous driving by modelling the driving scenarios and descriptions based on DL.
Road	(Elmalaki, 2021),	Avoiding potential collision.	Applying RL approach to learn user's behaviour during key decision to realise adaptive and fair HITL in potential collision scenarios.

A growing body of work has demonstrated how physiological and behavioural sensing can drive real-time HMI adaptation. In air traffic control and aviation contexts, fNIRS- and EEG-based workload assessment has been used to personalise information presentation, regulate alert timing, and trigger wake-up interventions to mitigate cognitive fatigue (Q. Li et al., 2021). Eye-tracking technology is one of the most natural form of HMI. Eye-tracking-informed interfaces further adapt visual saliency, information density, and modality selection to support operator perception and comprehension under varying workload conditions (Verstappen et al., 2022). These adaptations do not alter decision authority; instead, they reshape how AI-generated insights are communicated to align with human cognitive capacity.

In road transportation, adaptive HMI increasingly integrates environmental perception with human-state awareness to enhance situational understanding. Vision-based scene interpretation combined with knowledge-graph reasoning has been used to generate context-sensitive driving recommendations displayed via Augmented Reality Head-up Displays (AR-HUDs), supporting perception and comprehension without replacing driver judgement (Y. Liang et al., 2023). Explainable AI techniques further strengthen HAT by translating model outputs into human-interpretable rationales (J. Dong et al., 2023). This model casts the decision-making process as an image captioning task to generate textual descriptions as explanations to enhance trustworthiness. This system innovatively jointly models the correlation between driving scenarios (images) and descriptions (language) based on Transformer structure, which can effectively imitate the learning process of human drivers and generate appropriate driving decisions. Beyond moment-to-moment adaptation, reinforcement learning (RL) and HITL optimisation enable HMI to evolve over longer timescales. By learning individual preferences, risk tolerance, and interaction styles, adaptive interfaces can personalise guidance strategies and information prioritisation across users and scenarios (Elmalaki, 2021). In this sense, HMI functions not as a static display, but as an adaptive mediator that synchronises AI agent behaviour with human cognitive dynamics, closing the loop of HAT.

2.2.4. AI entangles with human cognition: Refining HMI paradigm in human-AI hybrid

2.2.4.1. From human-AI teaming to Hybrid Intelligence: Conceptual Foundations

HAT has been widely promoted as the “next step” beyond classical automation, because it reframes AI as a teammate rather than a tool and emphasizes interdependence, shared intent, transparency, and dynamic role allocation (Mica R. Endsley, 2023). However, in safety-critical transportation settings, the dominant HAT pattern still treats the human and AI as functionally separable cognitive agents who must coordinate through explicit communication (Ali et al., 2025). This framing creates an upper bound: system performance ultimately depends on how efficiently the two agents can maintain cognitive alignment when conditions become dynamic, time-critical, or uncertain.

Why current human-AI teaming is still not enough: the limits of explicit coordination

The limitation of current HAT stems not from inadequate algorithms or interfaces, but from the coordination architecture itself. When human and AI processes remain explicitly separated, the system assumes that alignment can be achieved through messages, cues, prompts, or mode transitions. In transportation operations, this assumption becomes fragile. Distinct cognitive streams create friction: small delays in information exchange can turn into mismatches in situational understanding and shifts in authority can create additional latency or uncertainty.

First, the human and the AI can hold different awareness of the situation. The AI acts on a formal objective and sensed variables. The human uses context, risk cues, and experience. When alignment depends on explicit communication, any gap in what is sensed, represented, or valued becomes a mismatch. **Second,** coordination takes time. Information must be shown, interpreted, and acted on. Under time pressure, the team may not converge to a shared understanding before action is required. Latency then becomes a driver of mismatch. In other words, even "good" explanations can be mis-timed relative to the operator's cognitive control loop. **Third,** explicit coordination also creates authority boundaries. Control handoffs are not just technical events. They change attention, monitoring effort, and intervention readiness (Parasuraman & Manzey, 2010). Empirical evidence from aviation and other safety-critical domains suggests that explicit communication and authority handover themselves introduce cognitive overhead, OOTL, and opportunities for breakdown, particularly under high workload or degraded SA (Ali et al., 2025).

These three factors are tightly coupled. Latency makes mismatches harder to correct. Mismatches make handoffs harder to execute. Handoffs and long periods of automation can further increase latency in re-engagement and deepen mismatch in situation understanding. These issues rarely occur in isolation. They interact, escalate, and reveal the inherent constraint of treating human and AI cognition as parallel tracks that must be continuously bridged rather than jointly organized. Taken together, these limitations suggest a key point: HAT is still largely defined by "who controls what" and "how we communicate," whereas the next transition is defined by how cognition is distributed, synchronized, and stabilized across human and AI over time.

What is Human-AI hybrid intelligence?

Based on the limitation of HAT, decades of cognitive systems research suggest that effective performance in complex environments often emerges from **distributed and extended cognitive structures**, rather than isolated agents. Theories of distributed cognition (Neville A. Stanton, 2016)and the extended mind (Le Guillou et al., 2023) argue that cognitive processes can span individuals, artifacts, and environments, forming integrated cognitive systems rather than discrete components. Building on this paradigm, the concept of HAH intelligence reframes the human-AI relationship from coordination between agents to cognitive coupling within a joint system. Hybrid intelligence does not merely denote higher levels of automation or more adaptive interfaces; instead, it describes systems in which human and artificial cognitive processes become interdependent, continuously co-adapting through shared representations, mutual predictability, and implicit interaction channels (Dellermann et al., 2019; Kamar, 2016)(Kamar, 2016; Dellermann et al., 2019). In such systems, intelligence emerges from the interaction itself, rather than residing solely in either the human or the AI.

Importantly, this shift distinguishes hybrid intelligence from human-state-driven adaptive HMI approaches. Whereas adaptive systems adjust interface features or automation behaviour in response to inferred human states, hybrid intelligence emphasizes long-term co-adaptation, where both human and AI modify their internal models based on ongoing interaction history. The objective is not to optimize momentary assistance, but to establish stable cognitive coupling that supports fluent joint perception, decision-making, and action over time. From this perspective, HMI no longer functions solely as a communication layer between agents, but as a shared cognitive substrate enabling integrated human-AI cognition.

2.2.4.2. How human-AI evolve into hybrid intelligence?

Below is a practical interpretation of *how teaming transitions into hybrid* in transportation HMI—centered on three mechanisms that go beyond human-state driven adaptive HMI and LOA adjustment.

Co-learning and mutual modelling: From separable agent to a coupled cognitive system

In many teaming systems, the AI learns about the human to adjust support, such as workload prediction, takeover readiness, while the human learns about the AI through transparency and explanations (Wohleber et al., 2023). This is often asymmetric. Hybrid intelligence requires reciprocal adaptation. Both sides update their models of the environment, the task strategy, and each other's intent and limits through repeated interaction. Over time, their models co-evolve through repeated interaction, so coordination relies less on explicit messaging and more on learned mutual predictability.

Transportation HMI research already contains early forms of this shift. Shared-control HMI increasingly treats cooperation as a joint system property rather than a handoff problem. Sun et al. (2024) propose a personalized shared control framework that models driving capability and driving style, and uses an intention-aware decision process to allocate authority. Importantly, they evaluate the framework in both HITL simulation and a field test, linking model adaptation to acceptance, comfort, and workload outcomes. This aligns with distributed and joint cognition views because "who does what" is not fixed. It is learned and stabilized through interaction. RL makes the co-learning mechanism explicit at the policy level. Huang et al. (2024) propose a safety-aware RL approach with shared control, where human intervention guides learning to avoid catastrophic actions while improving driving performance. Here, the human is not only a monitored operator. The human becomes a training signal that shapes the AI policy, and the resulting policy changes the human's future interaction strategy. This closes the reciprocal adaptation loop that teaming systems often lack.

Implicit interaction channels: Multimodal interaction coupling beyond visualization

HAT typically relies on explicit channels such as speech, text, alerts, and visualizations. Hybrid intelligence pushes interaction toward multichannel coupling. The goal is not "more information," but better and tighter coordination. When perception and action are distributed across visual, auditory, and haptic channels, the joint system can guide attention, reduce search cost, and keep the operator engaged without relying on slow dialogue or long explanations.

Recent aviation HMI work shows the benefit of cognitive coupling based on multimodal interaction. R. Yu et al. (2025) evaluated different interaction modalities for flight tasks and examined how modality count and task complexity affect pilot performance, workload, and user experience. Their results support a basic design claim for hybrid systems: adding channels can improve interaction, but benefits depend on how modalities are combined and on task demands. This moves the discussion beyond "using multimodal" to "designing hybrid modalities for stable performance". A complementary aviation perspective comes from Gaillet et al. (2025), who interviewed light-aircraft pilots about multimodal feedback. Pilots described the cockpit as already saturated visually and auditorily and expressed interest in tactile and auditory cues to reduce dependence on the visual channel and to support monitoring of critical parameters. This supports the hybrid argument that coupling should not rely on one channel or explicit messaging alone. Automated driving studies point to similar mechanisms in time-critical transitions. Meng et al. (2025) introduced a novel tangible interface for maintaining SA in conditional autonomous driving. Drivers were provided with haptic feedback of operational information about approaching vehicles, aiding drivers in preparing for takeover control.

Embodied & neuro-adaptive interfaces: From showing information to shaping control loops

Unlike the focus on state-triggered adaptation in HAT, for example inferring SA or OOTL and then switching LOA, alerts, or display content, this embodied and neuro-adaptive interactions aim to shape the joint perception–action loop directly. Coordination becomes smoother by default and less dependent on explicit negotiation. The interface is treated as part of the control loop, not only a channel for messages.

The concept of "loop shaping" has been operationalised in recent aviation research. For example, AdaptiveCoPilot, a neuroadaptive cockpit guidance system was proposed that modulates multimodal cues, including visual, auditory, and textual feedback—based on real-time workload estimates derived from fNIRS signals. The system was validated in a virtual reality cockpit environment with licensed pilots, demonstrating its feasibility for state-aware adaptive guidance (Shaoyue. Wen et al., 2025). The contribution is not simply workload detection. It is the attempt to keep the pilot within an effective operating zone by continuously tuning guidance form and load, which changes how pilots sequence actions rather than only informing them. Complementing this neuro-adaptive route, Morcos et al. (2025) develop and test full-body haptic cueing combined with spatial audio to support pilot perception in degraded or denied visual conditions. Their cueing algorithms are designed around compensatory tracking tasks, where tactile and spatial cues directly steer control actions, reducing reliance on visual scanning during critical moments. Evidence from automated driving suggests the same mechanism generalizes across transportation. Nakade et al. (2023) propose a multi-level collaborative steering approach grounded in shared control, where haptic interaction, arbitration rules, and "inclusion" of driver intent allow interventions to be assimilated into trajectory planning instead of being treated as disturbances. This framing explicitly targets continuous cooperation without deactivation, which is a control-loop level coupling strategy rather than a display-level intervention. Together, these studies point to a shared direction: interaction is designed to re-allocate sensing and timing inside the control loop, not merely to present more information.

Overall, embodied and neuro-adaptive interaction work supports the teaming-to-hybrid transition when it moves beyond "detect state, then adapt the interface," and instead stabilizes shared timing, shared intent, and shared action structure through interaction embedded in the loop (Nakade et al., 2023).

2.2.4.3. Implications and emerging opportunities: Refining HMI paradigm in human-AI hybrid systems

The convergence of mutual modelling, implicit channels, and neuro-adaptive coupling (Section 5.2) necessitates a shift from episodic task coordination to three intertwined loop-level challenges: continuous co-adaptation, fluid authority negotiation, and behavioural reliance calibration. These dimensions form a functional closed-loop: co-adaptation redefines the performance boundaries that underpin authority negotiation; the resulting distribution of influence creates the critical friction points where reliance calibration is manifested; and mis-calibrated reliance, in turn, biases the feedback loops essential for further co-adaptation.

From Intermittent Commands to Continuous Co-adaptation

As interaction shifts from intermittent commands to continuous co-adaptation, evaluation and design priorities should move beyond task completion alone toward human-AI in-syncness, namely the degree to which human control intent and AI assistance remain temporally and functionally aligned within the closed loop. In-syncness can be operationalized through interaction-level signatures such as control-loop stability and smoothness, consistency between human inputs and AI compensation, and the corrective conflicts (e.g., counter-steering and oscillatory adjustments). This operationalization is consistent with recent shared-control research that explicitly quantifies human–machine conflicts and intent consistency as measurable interaction phenomena, rather than treating coordination as a purely subjective construct (S. Li et al., 2024). This reframing motivates a control-theoretic agenda for HMI: rather than optimizing what information is shown, future research can treat HMI as a loop-shaping mechanism and systematically tune coupling policies to improve perceived intuitiveness and fluency, while preserving safety margins under disturbances.

A first HMI direction is to design continuous inten-response alignment mechanisms that make assistance legible at interaction speed. Rather than presenting additional guidance content, the interface should externalize the current coupling mode of the system in ways that are perceptually available and easy to regulate, such as gradual changes in haptic gain, steering feel, or guidance strength that operators can sense and tune without leaving the control flow. This suggests a design agenda that connects HFs with control theory: coupling policies should be engineered not only for nominal performance, but also for human-perceivable stability margins, damping, and predictability, so that operators can maintain an accurate "feel" for how the joint loop will respond under disturbances.

Recent GAI-enabled assistance further suggests how continuous co-adaptation may extend beyond low-level control to higher-level operational micro-loops where coordination costs are traditionally high. For example, Yanjie Zhang et al. (2025)

investigated audio-visual interaction for ATCO-pilot readback verification by introducing **generative subtitles** as redundant information alongside voice communication, and examined the associated cognitive effects using fNIRS and eye-tracking measures. Their results highlight a concrete HMI opportunity for hybrid systems: generative AI can support **real-time cross-checking and redundancy management** in communication-intensive tasks, reducing the operator's verification burden without requiring additional explicit dialogue. Building on this direction, AviationCopilot (Z. Zhang et al., 2026), a reliability-oriented LLM copilot framework inspired by human pilot training was proposed. AviationCopilot integrates aviation knowledge and structured information into the training and retrieval pipeline to improve robustness for operational use. Rather than positioning the LLM as a generic conversational layer, this line of work points to domain-grounded copilots that can provide **procedural support and context-conditioned guidance** while maintaining traceable, safety-relevant behaviours.

Context-Aware Negotiation: HMI as a Fluid Authority Mediator

In hybrid transportation HMI, the key question is no longer *who* controls the vehicle or system at a given moment, but **how authority is negotiated and continuously allocated across control dimensions** (e.g., lateral control, longitudinal control, tactical planning, hazard response) under changing context and risk. This calls for an HMI that functions as an **authority mediator**: it makes contribution weights legible, supports smooth re-weighting when conditions change, and prevents abrupt disruptions that break the operator's action flow. A growing body of automated-driving research has already formalized this problem as **driving authority allocation** and **arbitration** in shared control, explicitly targeting conflict resolution between driver intent and automation constraints (Nakade et al., 2023).

An opportunity is to move beyond binary handoffs toward **negotiatory interaction primitives** that resolve intent conflicts without long explanations. When human intent diverges from the system's safety policy, the mediator should support graded, time-bounded convergence rather than "all-or-nothing" takeover. Recent collaborative steering work makes this design space concrete: arbitration rules can be designed to shape interaction types such as cooperation, collaboration, or competition by modulating assistive torque and response characteristics, allowing the operator and automation to express and sense intent through the control channel itself. Complementary approaches frame authority allocation as an optimization or game problem to explicitly resolve driver–automation conflict, suggesting that negotiation can be implemented as continuous re-weighting rather than discrete switching. For HMI, the implication is to treat "conflict arbitration" as a first-class interaction design object, with implementable primitives such as soft constraints (e.g., resistive haptic guidance), reversible commitments (confirm/undo patterns), and short-cycle evidence requests (what condition would change the system's preferred action), all designed to preserve operational continuity.

Emergent coordination via shared affordances: making AI constraints perceptible and actionable

HAH enables interaction in which AI constraints and intent become perceptually actionable, allowing operators to sense what the machine senses and act with minimal cognitive translation. From an ecological HMI perspective, this means designing interaction so that coordination emerges through perception–action coupling rather than through a read–interpret–search cycle. The design target is **shared affordances** (Silva et al., 2013): perceptually available action possibilities that render AI constraints, safety margins, and intent directly usable within the operator's control loop. This framing also aligns with recent embodied-AI arguments that effective coordination is not only the product of internal reasoning, but of tight coupling among perception, action, and the environment, particularly as multimodal foundation models increasingly serve as general-purpose inference cores for embodied agents (M. Chen et al., 2025).

A practical opportunity is to externalize AI constraints and predictions as perceptually available affordances that operators can pick up and use directly. Instead of discrete symbols or verbal instructions, HMI can render boundaries, gradients, or directional "pull" cues that make safe corridors, constraint proximity, and impending violations immediately perceivable. The central question is whether these designs reduce translation cost and representational distance between what the AI knows and what the human can do next, thereby enabling more seamless perception-action coupling.

This perspective strengthens the case for **sensory augmentation** in hybrid HMI, especially when embodied channels can convey constraint information more directly than symbolic displays. Rather than adding more icons or dialogue, future systems can explore pre-reflective cues such as haptic steering feedback that encodes boundary proximity or constraint direction, or spatial audio structures that support orientation and urgency without monopolizing central vision. In automated driving, socially-compliant hierarchical human–vehicle collaboration with **multimodal haptic steering** provides an instructive template: Y. Zhang et al. (2025) propose a hierarchical collaboration framework in which haptic feedback functions as an interaction medium for guidance and authority mediation, enabling drivers to perceive system intent and constraints through the control channel while maintaining continuous involvement in steering actions. Such results reinforce the shared-affordance argument that constraint negotiation can occur through perceptual-motor coupling rather than explicit messaging. Building on related principles in aviation, AR-assisted HUDs can be viewed as perceptual augmentation that places AI-relevant constraints at the point of action. Yuan, Ng, Li, et al. (2024) proposed an AR-assisted HUD design that organizes phase-relevant synthesized information and alerts to support pilot-centred interaction across flight phases and abnormal situations. Read through the hybrid lens, the value of AR-HUD is not simply presenting more data but making system constraints and guidance perceptually available in ways that promote more intuitive coupling between perception and control.

2.2.5. Conclusion

This review contributes a HFs-driven account of how AI-enabled HMI in transportation is shifting from AI-assisted support to human–AI teaming and, ultimately, HAH intelligence, by reframing HMI as a cognitive human–AI interaction loop rather than a set of isolated technical modules.

First, this review contributes a structured synthesis of multimodal sensing for proactive HMI by clarifying how human-state inference can move beyond passive monitoring toward operationally meaningful human state awareness (corresponding to **RQ1**). Specifically, it organizes evidence around what should be inferred (attention, workload, SA-related markers), and why those inferences matter only when they are translated into controllable interface and automation behaviors. Second, this review contributes an integration lens for adaptive AI in strategic and tactical decision layers by articulating teaming as coordinated shared authority rather than isolated interventions (corresponding to **RQ2**). It consolidates recent advances in adaptive automation and agentic/GAI-enabled decision support into a single coordination-centric framing, highlighting the HFs conditions under which authority transitions remain predictable, inspectable, and recoverable. Third, this review contributes a refined interaction-paradigm perspective for hybrid systems by synthesizing interaction modalities that enable tighter cognitive coupling (corresponding to **RQ3**). It positions Section 5 as the forward-looking consolidation of HAH opportunities, emphasizing interaction designs that reduce explicit coordination costs, enhance seamless interaction within multimodal medium, and while preserving controllability and cognitive continuity.

This review intentionally adopts a HFs-driven perspective on AI-enabled HMI, prioritizing the evolution of human-AI relationships and interaction paradigms over detailed algorithmic benchmarking. As a result, it does not provide an in-depth comparison of specific AI models, sensing pipelines, or learning architectures in terms of predictive accuracy or computational efficiency. While this focus enables a clearer synthesis of how AI capabilities are translated into interaction, authority allocation, and cognitive coupling, it may limit the review's utility for readers seeking guidance on algorithm selection or implementation-level optimization. In addition, because many reviewed studies employ heterogeneous sensing setups, labeling strategies, and evaluation metrics, the analysis emphasizes conceptual patterns rather than quantitative cross-study comparison. These limitations reflect a trade-off inherent to theory-oriented reviews, where explanatory coherence and design insight are prioritized over technical exhaustiveness.

Chapter 3. From Unrecognised to Incapacitating: Neuro-ergonomic Differentiation of Spatial Disorientation Typologies Induced by Visual Illusions for Pilot Cognitive Training and Cockpit Design

Within this chapter, an initial empirical study is conducted, representing a crucial precursor to the comprehensive research framework. The central aim of this investigation is to explore the neuro-ergonomic signatures of four defined Spatial Disorientation (SD) classifications: no SD, Type I (recognised), Type II (recognised but manageable), and Type III (incapacitating). To achieve this, Electroencephalography (EEG) and eye-tracking methodologies are utilised to evaluate pilots' SD states through the analysis of brain cortical activity and scanning behaviours, respectively. The outcomes of this study are expected to provide the empirical basis for the future development of offline predictive models for pilot SD types, employing advanced machine learning algorithms.

3.1. Introduction

3.1.1. Research Background of Spatial Disorientation and its Category

3.1.1.1. The Significance of the Issue for Civil Aviation

Spatial disorientation (SD) is known as a pilot's incapacity to accurately understand the attitude, motion, or location of an aircraft in relation to the gravitational vertical (Gaydos et al., 2012), persists as a critical threat to aviation safety across military and civil operations (Stróżak et al., 2018). SD can impair a pilot's ability to interpret instrument data accurately and maintain spatial awareness, leading to incorrect control inputs and poor decision-making (Q. Li et al., 2024). The U.S. Federal Aviation Administration (FAA) attributes 15-22% of general aviation fatalities to SD-related causes (Antunano, 2021), with accident rates soaring to 72% in degraded visual environments (DVEs). This perceptual mismatch between vestibular cues and instrument readings (Newman & FAICD, 2007) induces catastrophic control reversals—for example, 89% of SD accidents involve pilots inadvertently entering a graveyard spiral while believing themselves to be in level flight of visual illusions (Gibb et al., 2011).

3.1.1.2. Types of Spatial Disorientation

During actual flight operations, SD takes on different manifestations. Understanding the different types of SD and their cognitive aspects is vital for developing effective strategies to prevent and address the associated risks. The Type I-III SD taxonomy, first operationalised by the U.S. Air Force (USAF) to analyse 1989-1991 SD accident patterns (Lyons et al., 1994), categorises disorientation states based on pilots' metacognitive awareness and operational capacity: Type I (Unrecognised SD), Type II (Recognised SD), Type III (Incapacitating SD). This framework has been validated across aviation domains and informs modern SD mitigation strategies (Heinle & Ercoline, 2002; Webb et al., 2012):

Despite its high fatality rate, Type I SD remains understudied due to the intrinsic paradox: pilots' unawareness of SD. A study proposed using centre of pressure (CoP) deviations as an objective indicator of Type I SD episodes, demonstrating that pilots exposed to optical flow illusions exhibited 34% greater postural instability compared to baseline (Hao et al., 2022). The effect of Type II SD on flight safety may be understated in comparison to the other two SD kinds. Pilot performance may be impacted by Type II SD, which might include usurping attentional resources to regain or maintain spatial orientation or increasing cognitive effort while recognising being confused. In this context, Landman et al. (2024) explicitly manipulated Type II SD by operating motion stimulus magnitude and measuring its effect on flight performance. The findings showed that cognition is involuntarily disrupted when SD is recognised. Type III SD research is sparse and primarily limited to military aviation case studies. The 1994 USAF report documented that vestibular-ocular reflex saturation during Coriolis illusions caused 89% of Type III pilots to involuntarily release flight controls within 10 seconds (Benson, 2003).

3.1.2. Contributory Factors to Spatial Disorientation

The human perception of spatial orientation relies on information gathered through the visual, vestibular, and proprioception systems. Under normal flight conditions, these systems provide complementary sensory inputs that enable precise spatial perception and motion understanding (L. Huang et al., 2023). Approximately 80% of the flight-related information is obtained through the visual system (Sánchez-Tena et al., 2018). So, visual illusions represent a pivotal mechanism through SD manifestation, particularly in low-visibility environments (Meeks et al., 2023). Among them, lack of visual reference significantly contributes to visual illusion. What the pilot sees outside and inside the cockpit can be divided into adverse weather and malfunctioning instruments. First, adverse weather dramatically limits external visual cues, compelling pilots to rely increasingly on instrumentation (M. Chang et al., 2022). During typhoons or other severe meteorological events, rapid atmospheric changes—including turbulence, wind shear, and sudden visibility reductions—can induce rapid perceptual transitions that overwhelm sensory integration mechanisms (Eusse-Villa et al., 2024). These conditions create a sensory environment where visual-vestibular conflicts become almost inevitable, significantly elevating SD risk. Simultaneously, instrumentation malfunctions further compromise spatial awareness. Discrepancies between instrument readings and sensory perceptions can introduce critical cognitive confusion (van den Hoed et al., 2022). Recent studies have emphasised how delayed or inconsistent instrumentation—such as Attitude Indicators (AI) (Landman et al., 2019; van den Hoed et al., 2022) and weather radar (Hua et al., 2022) —can fundamentally undermine a pilot's spatial perception.

In conclusion, one of the main contributing factors to fatal SD-related accidents has been shown to be the pilot's incapacity to properly monitor the aircraft's flying characteristics and implement corrective measures using visual illusions. A case study is the accident of West Air Sweden flight SWN294 (Authority, 2016). Researchers discovered that improper pitch and roll signals resulted from an incorrect attitude indicator on the left Primary Flight Display (PFD), which was brought on by an issue with the Inertial Reference Unit (IRU). The pilots' SD was made worse by the removal of the pitch and roll comparator indicators from the PFDs when anomalous data were detected. Such incidents underscore the critical need for understanding how visual illusions, including instrumentation failures and adverse weather, interact to compromise pilot perception is essential for developing more effective training protocols and technological interventions to enhance flight safety.

3.1.3. Current Qualitative Evaluations of Spatial Disorientation

Assessing SD necessitates multimodal approaches due to its complex neurocognitive nature. Prevailing methodologies fall into two domains:

3.1.3.1. Performance-driven Metrics: Behavioural and Subjective Data

Operational SD detection traditionally hinges on behavioural degradation and subjective introspection. Behavioural metrics serve as direct indicators of disorientation consequences, typically encompassing task accuracy (Stróżak et al., 2018), operation error (van den Hoed et al., 2022), flight parameter deviations (L. Huang et al., 2023), and postural or physical instability (Hao et al., 2022). For example, in the context of storm movement path awareness, Hua et al. (2022) used measures like location deviation and absolute distance deviation to evaluate spatial awareness. Yet, such metrics conflate SD with skill deficits—a critical validity threat when assessing experienced pilots.

Subjective scales include direct questions regarding the judgment of position, physical state, vertigo, dizziness, etc., to confirm situational or spatial awareness (Yuan, Ng, Li, et al., 2024). Although tools like the Motion sickness questionnaire (MSQ) and simulator Sickness Questionnaire (SSQ) effectively the influential factors leading to SD occurrence (Wojciechowski & Blaszczyk, 2019), they fail to diagnose different SD type cases in a real-time manner. A study also identified the occurrence of SD directly through personal reports based on the Likert scale (Hao et al., 2022). The synergistic integration of behavioural and subjective measures enables triangulation of SD manifestations (Landman et al., 2024). For instance, Bolton and Bass (2009) utilised Situation Awareness-Subjective WORkload Dominance (SASWORD) and Situation

Awareness Global Assessment Technique (SAGAT) alongside objective indicators — such as relative angle, relative distance, relative height, and abeam time from the flight tasks — to comprehensively evaluate the effects of different terrain textures in Synthetic Vision Systems (SVS) on inducing SD. However, while foundational, subjective scales suffer from gaps in self-awareness—pilots cannot recognise and report disorientation in a timely manner.

3.1.3.2. Neuro-ergonomic Probes: Indirect Measures based on Cognitive Process

Emerging cognitive neuroscience tools are bridging this temporal resolution gap. Eye-tracking reveals attentional tunnelling during SD—manifested as dwell time asymmetry on specific Areas of Interest (AOI) to infer SD conflicts (Fudali-Czyz et al., 2022). Other eye-tracking metrics, including fixation and saccades, were applied to capture visual responses to illusions (Bałaj et al., 2019; Kang et al., 2021). Electrophysiological markers show particular promise. Several neuroimaging techniques, such as Electroencephalography (EEG), have also been preliminarily applied to monitor neural activity in SD scenarios. H. Hu et al. (2021) identified sensorimotor cortex theta suppression (C3/C4, 4-7Hz) during vestibular disorientation—a potential neural signature of Type I SD. While the application of neuro-ergonomic techniques to SD detection is still in its early stages, such approaches hold strong potential due to their ability to validate the underlying neural mechanism of SD. These methods could provide insights into how brain activity correlates with spatial perception and cognitive processing changes.

However, extant neuro-ergonomic studies predominantly employ static laboratory paradigms, limiting ecological validity (Bałaj et al., 2019). Few published works have concurrently tracked neural-ocular dynamics across SD subtypes, a gap our study addresses through synchronised 32-channel EEG and gaze pattern analysis.

3.1.4. Research gaps in SD typology assessment

Despite advances in SD characterisation, critical Research Gaps (**RG**) persist in translating laboratory findings to operational countermeasures. First, while sensory illusions (e.g., somatogyral, Coriolis) have been mapped to physiological responses, the USAF separate Type I-III taxonomy remains underexplored in a comprehensive framework. Current studies treat SD as a homogeneous phenomenon, overlooking the cognitive distinctions and dynamic transitions between SD types (**RG-1**). Second, despite single-modality measures of behavioural or neurophysiological metrics, multimodal integration for SD typing lacks ecological validity. So, **RG-2** is exacerbated by the absence of comprehensive neuro-ergonomic patterns linking real-time biomarkers (e.g., prefrontal beta suppression) to gaze anchoring patterns during SD type progression.

3.1.5. The current study

This study introduces a neuro-ergonomic framework combining SD-Probe (a customised subjective scale) with synchronised EEG-eye tracking to address these gaps. We aim to establish multimodal thresholds for SD typing to three typologies induced by visual illusions by correlating EEG biomarkers in different brain regions with gaze dispersion metrics.

Specifically, we designed flight scenarios, including adverse weather conditions (typhoons) and malfunctioning instruments (inconsistent instrument displays) as visual illusions to trigger SD occurrence. A flight scenario without any instrument emergency settings and fare weather was also set as the baseline. EEG and eye tracker were utilised to monitor the occurrence and transition of different SD types in neuro-ergonomics. Next, to label the SD type, a customised SD probe including several questions about flight scenarios was introduced, and the response of SD probe was used to determine specific SD types.

Based on the above study, our hypotheses are as follows:

- **H1:** Non-SD conditions will exhibit the highest and most balanced brain activation patterns. Optimal spatial awareness requires dynamic attention and sensory processing integration, which need multiple brain regions working together (Catherwood et al., 2014). Thus, the non-SD state maintains higher and more balanced levels of brain activity.
- **H2:** Neurophysiological activation will progressively increase with the evolution of Type-I to III and the appearance of visual illusions. Landman et al. (2019) suggested that SD manifests through progressively complex cognitive disruptions. Our hypothesis proposes an incremental increase in neurophysiological activity from Type I to Type III, reflecting escalating perceptual-cognitive challenges.
- **H3:** Visual scanning range will narrow hierarchically from Non-SD to Type III. According to Fudali-Czyz et al. (2022), visual delayed discrimination occurs with the presence of SD because SD disrupts attention distribution. Building on this, Type I-III will be characterised by a narrower visual scanning range, suggesting perceptual constriction.
- **H4:** Type II SD pilots will exhibit neuro-ergonomic profiles closest to non-SD. Although a misperception between sensation and actual state is recognised and

can cause surprise emotion (Landman et al., 2024), appropriate vigilance in Type II SD can help activate a wider range of neuro-ergonomic activity.

3.2. Methodology

3.2.1. Experimental Apparatus

Figure 3-1 shows the setup of the experimental apparatus. To simulate actual flight operations and situations, a fixed-based CockpitSonic Airbus 320 Flight Training Device (FTD) was used. Highly realistic flying circumstances were simulated using the FTD in conjunction with map landscape data from the visual simulation software Prepar3D (version 5.3). Active Sky (version 5) simulated the meteorological environment based on historical weather data. The relevant weather information was updated and presented in real-time on the navigation display (ND), as shown in Figure 3-1 (b)-(d). For monitoring neuro-ergonomic feedback, a wireless EEG headset (EMOTIV Flex) was utilised to record cerebral cortical activity at a sample rate of 128 Hz. Thirty-two saline electrodes were arranged to cover the prefrontal (PFC), parietal (PL), left/right temporal (LTL and RTL), and occipital lobes (OL). Additionally, the pilot's visual engagement with the cockpit was examined using a glasses-based eye tracker (Tobii Pro Glasses 3) with a 256 Hz sampling rate. Besides, pilots were required to use a tablet PC to respond to custom SD probes related to critical spatial awareness information. The researchers evaluated and labelled the participants' SD levels based on multimodal evaluation criteria.

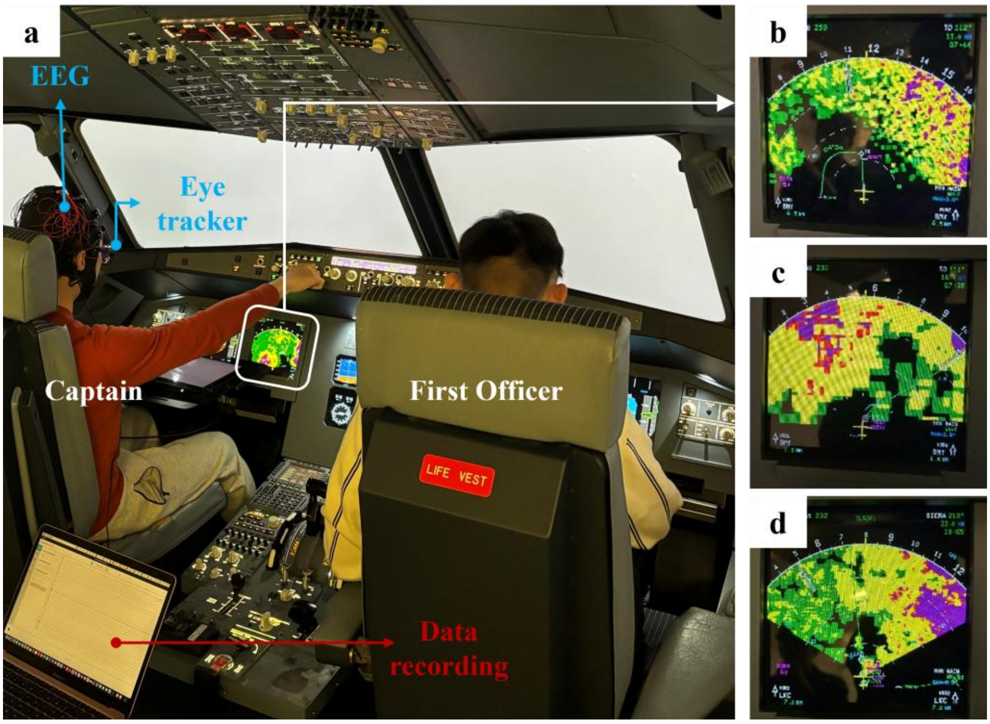


Figure 3-1. Experimental apparatus. (a): Experimental schematic diagram. (b), (c) and (d): Meteorological radar map samples of the typhoon.

3.2.2. Participants

Sixteen cadet pilots (Age: 29.25 ± 6.57; 2 female) from commercial airline companies and flight training organisations were recruited as participants. A priori power analysis (G-Power) with an effect size of f = 0.8 (α = 0.05, power = 0.95) indicated a minimum requirement of 12.26 participants. The sample size (N = 16) exceeds this threshold, ensuring sufficient statistical power to detect neuro-ergonomic effects under visual illusions. All the participants had training experience on the Airbus A320, and the demographics and profile are shown in Table 3-1. Due to the high variability in subjects' training hours, the median (156) was chosen in this study as the categorisation pilot for novice (NP, N=8) or experienced pilot (EP, N=8). Although the participants had relatively limited flight experience, pilots with shorter training histories are more susceptible to experiencing SD under challenging or degraded flight conditions. As the primary objective of this study was to investigate and characterise distinct neuro-ergonomic patterns associated with different SD sub-types, it was essential that the experimental design reliably elicit varying degrees and forms of disorientation. From this perspective, the recruited cadet pilots provided an appropriate participant group for triggering multiple SD manifestations, thereby enabling the identification and comparison of neurophysiological signatures across SD sub-types. Accordingly, the participants' level of flight experience was sufficient to meet the experimental objectives of the study.

Table 3-1 Demographic and profile of the participants.

No.	Age	Gender	Training hours (Ground Theory and simulators)	Pilot category (EP/NP)
1	25	M	600	EP
2	22	M	400	EP
3	22	M	500	EP
4	29	M	400	EP
5	33	M	156	EP
6	28	M	400	EP
7	35	F	30	NP
8	39	M	100	NP
9	46	M	450	EP
10	26	M	150	NP
11	26	M	48	NP
12	28	M	150	NP
13	25	F	50	NP
14	25	M	200	EP
15	25	M	100	NP
16	34	M	100	NP

The participants acted as the captain. This study arranged for a student from an aeronautical engineering major with rich experience in flight simulation to serve as a pseudo-first officer and assist the captain in completing relevant flight tasks. Ethical approval has been granted by the Hong Kong PolyU Institutional Review Board (Reference number: HSEARS20210318002). Pre-experiment informed consent was obtained for all participants in written form. Participants were compensated with monetary incentives in coupons.

3.2.3. Experimental Design

3.2.3.1. Independent variables: visual illusions from internal and external cockpit

Given the limited visual reference caused by both external weather conditions and internal malfunctioning instrument display (i.e., instrument inconsistencies), this study selected two key visual illusions to induce SD: adverse weather conditions and instrument inconsistencies. The following are the details of the two visual illusions settings.

1. Weather conditions: Typhoon Koinu (08/10/2023). The typhoon condition was primarily adopted to simulate degraded visibility and increased crosswind disturbances, which are known to intensify visual illusions, elevate control difficulty, and exacerbate sensory ambiguity, key contributing factors to the onset of SD. Typhoons are characterised by strong winds, heavy precipitation, and low visibility, severely impair pilots' ability to accurately perceive spatial orientation. These conditions heightened the risk of disorientation by affecting the perception of the aircraft's position relative to the ground and horizon. Detailed weather conditions for the arrival and departure airports are listed in Table 3-2. Such conditions allow the experiment to more effectively elicit SD-related cognitive and behavioural responses than nominal weather scenarios. At the same time, considerations of operational realism were incorporated into the

experimental design. Although a typhoon weather system was selected, flight tasks were not conducted during the peak intensity of the storm. Instead, specific time windows were chosen based on the temporal evolution of the typhoon, such that the selected departure time (e.g., UTC 02:00 at Hong Kong International Airport) corresponded to periods when weather severity had not yet reached its maximum. In addition, the planned flight routes avoided regions with the most severe storm activity and non-navigable conditions. This approach balances the need to induce challenging perceptual conditions with the requirement to maintain plausible and operationally consistent flight scenarios. In the fair-weather experiments, meteorological conditions were consistent with visual meteorological conditions defined by the International Civil Aviation Organisation (ICAO), i.e., good visibility (8 kilometres or more) and cloud height sufficient for visual flight.

2. Instrument inconsistencies: Unreliable airspeed and barometric altitude (Triggered by Pitot tube & Static tube (Captain & F/O) blocked; Auto-Pilot & Auto Thrust disengaged). The aircraft cannot measure airspeed, altitude and vertical speed correctly, and the relevant instrument indications may be seriously deviated. Discrepancies may occur in the Captain, F/O and standby instruments. As a result of the discrepancy, the pilot may have a false spatial perception because the instrument displays do not correspond to the actual state. What's more, manual control of the aircraft increases cognitive workload, contributing significantly to the risk of SD.

Table 3-2 Detailed adverse weather conditions for arrival and departure at the airport.

Airport	Time (UTC)	Weather condition	Interpretation

Departure airport (ZGGG)	08/10/2023 01:45	METAR ZGGG 080145Z 09002MPS 360V130 CAVOK 19/09 Q1009 NOSIG=	Routine weather report for ZGGG at 01:45 UTC: wind speed 2m/s, direction 090 and varying between 360 and 130. Visibility over 10km, no clouds, no significant weather. Temperature 19°C, dew point 9°C. Atmospheric pressure 1009 hectopascals. No significant changes expected in the next 2 hours.
Arrival airport (VHHH)	08/10/2023 02:30	METAR VHHH 080230Z 01021KT 9999 +SHRA FEW006 SCT018 23/20 Q1011 TEMPO 36025G35KT 3500 SHRA FEW010CB SCT020=	Routine weather report for VHHH at 02:30 UTC: Wind 210 degrees at 21 knots. Visibility greater than 10km, heavy rain showers. Few clouds at 600 ft, scattered clouds at 1800 ft. Temperature 23°C, dew point 20°C, pressure 1011 hPa. Temporary winds are 360 degrees at 25-35 knots, visibility is 3500m, there are rain showers, a few cumulonimbuses at 1000 ft, and scattered clouds at 2000 ft.
Arrival airport (VHHH)	08/10/2023 02:26	SPECI VHHH 080226Z 01021KT 9999 4300S +SHRA FEW006 SCT018 22/20 Q1011 WS R07R TEMPO 36025G35KT 3500 SHRA FEW010CB SCT020=	Special weather report for VHHH at 02:26 UTC: Wind 210 degrees at 21 knots. Visibility greater than 10km, 4300m in heavy rain showers. Few clouds at 600 ft, scattered clouds at 1800 ft. Temperature 22°C, dew point 20°C, pressure 1011 hPa. Wind shear on Runway 07R. Temporary winds are 360 degrees at 25-35 knots, visibility is 3500m, there are rain showers, a few cumulonimbuses at 1000 ft, and scattered clouds at 2000 ft.
Arrival airport (VHHH)	08/10/2023 02:20	SPECI VHHH 080220Z 01021KT 9999 -SHRA FEW008 SCT020 22/20 Q1010 RERA WS R07R TEMPO 36025G35KT 3500 SHRA FEW010CB SCT020=	Special weather report for VHHH at 02:20 UTC: Wind 210 degrees at 21 knots. Visibility greater than 10km, light rain showers. Few clouds at 800 ft, scattered clouds at 2000 ft. Temperature 22°C, dew point 20°C, pressure 1010 hPa. Recent rain, wind shear on Runway 07R. Temporary winds are 360 degrees at 25-35 knots, visibility is 3500m, there are rain showers, a few cumulonimbuses at 1000 ft, and scattered clouds at 2000 ft.
Arrival airport (VHHH)	08/10/2023 02:00	METAR VHHH 080200Z 01022KT 9999 SHRA FEW008 SCT020 23/20 Q1010 TEMPO 36025G35KT 3500 SHRA FEW010CB SCT020=	Routine weather report for VHHH at 02:00 UTC: Wind 220 degrees at 22 knots. Visibility greater than 10km, rain showers. Few clouds at 800 ft, scattered clouds at 2000 ft. Temperature 23°C, dew point 20°C, pressure 1010 hPa. Temporary winds are 360 degrees at 25-35 knots, visibility is 3500m, there are rain showers, a few cumulonimbuses at 1000 ft, and scattered clouds at 2000 ft.

3.2.3.2. Dependent variables: Three SD types classification.

The dependent variables in this study were the typologies of SD types (Type I, II, and III) and their neurobehavioral correlations. Dynamic SD types need to be evaluated and labelled. First, the SD-Probe was customised in this study to assess the pilot's perception of the flight situation. Regarding the definition of SD, this study designed thirteen multi-choice questions (Appendix II) related to the spatial information of the flight tasks as the customised SD probes to monitor and evaluate the dynamic SD type changes. Thirteen questions were evenly distributed throughout the experiment session, with an interval of about 2-3 minutes. The SD probe was deliberately embedded at moments that did not interfere with primary flight control or critical task execution. Specifically, SD probes typically evaluate the pilot's awareness of the aircraft's position, velocity, orientation, flight phase, and other safety-critical information. The probe questions were designed to be of simple to moderate difficulty and were not associated with strict response time constraints, thereby minimising additional time pressure. All these questions were popped out on the tablet PC appropriately. When participants saw the probe, they were required to view relevant instrument information or manuals for quick answers quickly. As a result, the SD probe functioned as a low-intrusiveness assessment of the pilot’s spatial awareness state rather than as a competing secondary task. Consequently, it is unlikely to have introduced a significant increase in overall cognitive or psychological load beyond that induced by the flight scenario itself.

Second, referring to the Quick Reference Handbook (QRH) of A320 and consulting with captions with rich flying experiences, this study sets the objective Standard operating procedures (SoPs) when suffering with the pre-designed emergency situations. Table A1 in Appendix lists the critical SoPs when met with unreliable airspeed and barometric altitude. In brief, the criteria for determining the pilots’ SD type are based on whether the pilot’s operational behaviour conforms to the SoPs and whether the SD probes can be answered correctly. Table 3-3 shows the criteria used to label the SD types based on the UASF definitions (Lyons et al., 1994).

Table 3-3 The criteria for the evaluation of three SD types.

Visual Illusion Setting	Answer of SD-probe	Whether comply with the SoP	Label	Remark
Without visual illusions	Right	Yes/No *	Non-SD	Without SD
	Wrong	Yes/No *	Type I SD (unrecognised)	Lack of awareness of the current situation, although there is no emergency.
With visual illusions	Right	Yes	Non-SD	Without SD
		No	Type II SD (recognised)	Recognised SD, but capable to react properly
	Wrong	Yes	Type I SD (unrecognised)	Unrecognised SD
		No	Type III SD (Incapacitating)	Aware of SD but unable to handle it.

* Since there is no visual illusion set in session 1, pilots do not need to execute the relevant SoPs.

The no visual illusion conditions without flight emergencies serve as a baseline for comparison. There is no SoP designed for emergencies and there should be no severe SD occurring in this condition. So, the execution of SoP in this session is not necessary. If the subject answered the SD probe correctly, then the subject was labelled as non-SD. If the subject answered the question incorrectly no matter whether complied with the SoP, he/she was marked as Type I SD, which means that the subject was not aware of SD.

Under visual illusion conditions, the determination of SD type needs to be discussed on a case-by-case basis:

- If the subject answered correctly and can execute the SoP step by step, it was marked as non-SD. However, assuming that the flight operation was not in compliance with the SoP, although the pilot was able to respond to the abnormalities, in this case he or she would not be able to determine the cause of the visual illusion. Even if the answer is correct, it only means that the subject recognises SD. In this case, it is marked as Type II SD.
- If the subject answered incorrectly but could comply with the overall SoP, he or she was labelled as Type I SD as the pilot was not aware of the issues. If the subject was aware of the SD but cannot execute the SoP with the clear steps, indicating that he or she was incapacitated, he or she was labelled as Type III SD.

3.2.3.3. Flight Route Design.

A round-trip route between Hong Kong International Airport (VHHH) and Guangzhou Baiyun International Airport (ZGGG), as illustrated in Figure 3-2 (a), was included in the experimental scenario. Specifically, during adverse weather conditions, the chosen flight route was disturbed by the edge of Typhoon Koinu's moving path. The weather conditions near the Hong Kong area deteriorated gradually due to the influence of the typhoon's path, with increasing wind speeds and rainfall intensity. In contrast, the impact of the weather on the Guangzhou area was relatively smaller than that of the Hong Kong area. Figure 3-2 (b) and (c) compare the visual horizon in fair and adverse weather, respectively.

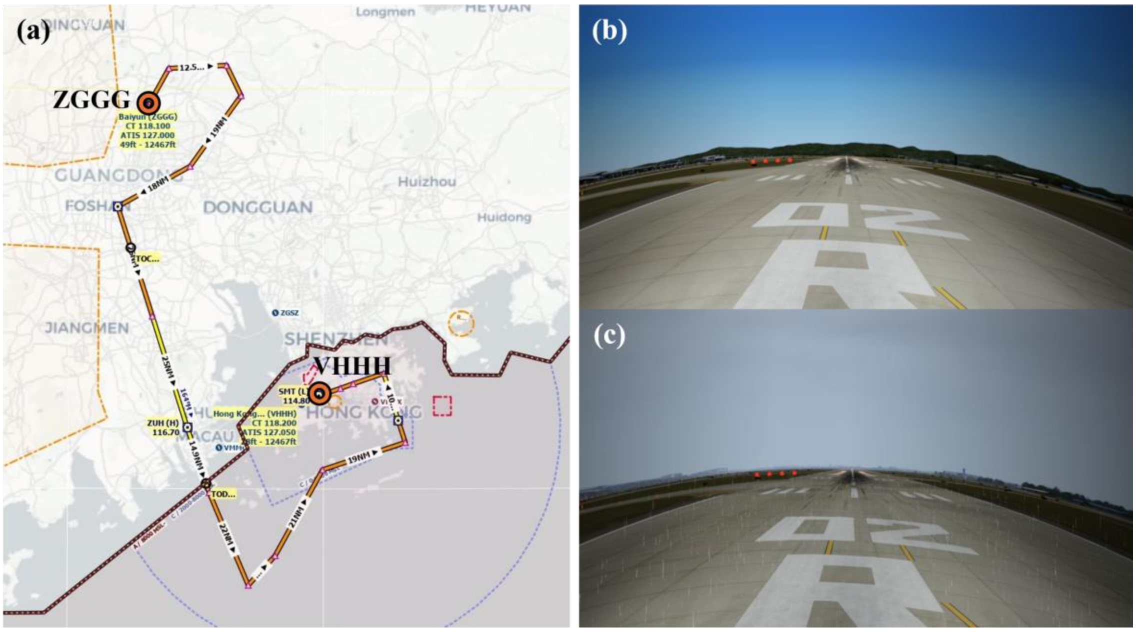


Figure 3-2. Flight route design and weather condition setting. (a) Flight route between Guangzhou Baiyun International Airport (ZGGG) and Hong Kong International Airport (VHHH), (b) Fair weather condition, (c) Adverse weather conditions (Typhoon Koinu).

3.2.4. Experiment Procedure

The experiment consists of two sessions: no visual illusions and the visual illusions that trigger SD. Table 3-4 presents the details of the experimental setting of each session. Each experimental session started with taxiing out at the departure airport until the aircraft safely landed in the destination airport. For the visual illusion session, a total of five instrument inconsistencies were set for emergencies in climbing (N=1), cruising (N=3), and approaching (N=1). Specifically, when the aircraft approached the pre-set trigger point, we started to simulate the "unreliable airspeed and barometric altitude". Auto Pilot & Auto Thrust disengaged were also triggered simultaneously.

Table 3-4 The details of the experimental setting of two sessions.

Session	Flight route	Internal visual illusion: malfunctioning instrument display	External visual illusions: adverse weather conditions with DVEs
1	VHHH-ZGGG	Without malfunctioning instrument display	Fair weather
2	ZGGG-VHHH	Instrument inconsistencies: (1) Pitot tube & Static tube (Captain & F/O blocked), (2) Auto-Pilot & Auto Thrust disengaged.	Typhoon Koinu

The participants were asked to read the information sheet and flight plan, write down their demographics, and sign a consent form. Then, the EEG headset and eye tracker were mounted on the participant's head. Before starting the first round of studies, participants got fifteen minutes to get familiar with the simulator. The fair and adverse weather experiment sessions were presented to the subjects randomly to avoid any learning effect. The experiment in fair weather served as a baseline. After a fifteen-minute rest, the participants were required to read the flight plan for the second round of experiments and then start the second flight experiment. Each round of experiments takes 40 minutes, and the whole experiment lasted about two hours.

3.2.5. Data Analysis

3.2.5.1. Neuro-ergonomic Data Analysis

To figure out the neuro-ergonomic patterns of different SD types and the impact of visual illusions on SD, EEG and eye tracking data were analysed accordingly. Figure 3-3 presents the general procedure of data analysis.

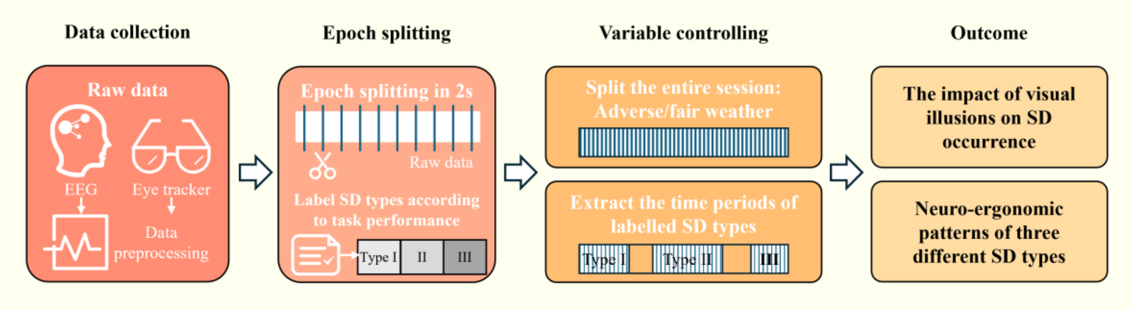


Figure 3-3. Data-analysed framework.

MATLAB (R2024b) and EEGLAB (v2025.0.0) toolbox (Delorme & Makeig, 2004) were utilised for EEG pre-processing, including channel localisation, bandpass filtering (1-45 Hz), independent component analysis (ICA), artefact removal, and re-referencing. The detailed description of each step is shown below:

1) Channel Localisation: Electrode coordinates were aligned to the standard 10-20 system using EEGLAB's BESA file for 4-shell dipfit spherical model.
2) Filtering: A bandpass FIR filter (1-45 Hz) was applied to remove low-frequency drifts and high-frequency noise.
3) Independent Component Analysis (ICA): Extended Infomax ICA decomposed data into 32 components. Components corresponding to ocular, cardiac, and muscle artifacts were identified through visual inspection and characteristic scalp topographies and were subsequently removed from the dataset.
4) Manual Verification: Poor-quality epochs and bad channels were manually eliminated by interpolation.
5) Re-referencing: To lessen reference bias in the next studies, the data were re-referenced to the common average following the elimination of artifacts.

After the pre-processing, the individual EEG segments that corresponded to the four labelled SD types (including Type I to III, and non-SD) were extracted, respectively. Each SD types of segments were then divided into 2-second EEG epochs to compensate for artefacts. Power Spectral Density (PSD) inside theta (4-8 Hz), alpha (8-12 Hz), beta (12-30 Hz), and gamma (30-45 Hz) bands waves were then calculated using the Fast Fourier Transform (FFT). The individual neurophysiological response data for each 2-second EEG epoch were evaluated using the PSD, which is the most commonly used EEG measurement for detecting brain activity (Fernandez-Blanco et al., 2020; Q. Li, K. K. H. Ng, et al., 2023). The FFT was employed to derive the PSD in $\frac{\mu V^2}{Hz}$ by converting the EEG signals from the time domain to the frequency domain for each EEG frequency band (M. Zhang et al., 2022). Due to the experimental environment, EEG data will inevitably contain outliers, which can seriously affect the results and lead to undesirable statistical outcomes. In this regard, we used the interquartile range (IQR) statistical method to detect outliers and replace any outliers at the upper and lower limits of each frequency wave. Lastly, to validate data quality, Cohen's d effect sizes were computed via equation (1) to evaluate systematic differences between processed (IQR-corrected) and original data (data automatically processed by EEGLAB) across two experimental sessions. Cohen's d was calculated per channel-frequency pair using pooled standard deviation and the results were presented as Table 3-5:

$$d = \frac{\bar{X}_{processed} - \bar{X}_{original}}{\sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}} \quad (1)$$

where $\bar{X}$ = mean power, s^2 = variance, n = trial counts.

Table 3-5 Data validation via sensitivity analysis (Cohen's d effect sizes).

Session	Mean_d (Theta, Alpha, Beta, Gamma)	Std_d (Theta, Alpha, Beta, Gamma)
1	[-0.0415, -0.0324, -0.0509, -0.1036]	[0.0639, 0.0728, 0.0792, 0.1240]

2	[-0.035, -0.0357, -0.0484, -0.1008]	[0.0613, 0.0651, 0.0688, 0.1125]

The small Cohen's d values in both sessions suggest that there are negligible differences between the manual calibration and the automated toolbox processing. This implies that automated preprocessing pipelines are robust and produces results comparable to manual corrections. The minor discrepancies observed are likely not meaningful in practical terms and do not significantly affect the overall data quality or subsequent analyses.

For eye tracker analysis, Time of Interest (TOI), which is associated with four SD types' metrics, was extracted via Tobii Pro Lab. Importantly, AOIs were also shown in Figure 3-4, to investigate the attention distribution over the various instruments in the cockpit. The chosen AOI includes the main panel in the cockpit, which allows for a more detailed exploration of the pilot's attention to various instrument information under different situations. The main panel includes a Primary Fight Display (PFD), Navigation Display (ND), Electronic Centralised Aircraft Monitoring (ECAM), and Integrated Standby Instrument System (ISIS). The flight controlling displays include Flight Control Units (FCU) and ECAM control. In addition, the AOI includes the overhead panel, the external view, and the area where the flight plan is placed.

Figure 3-4. The distribution of area of interests (AOIs) in the cockpit.

To quantitatively compare pilot visual behaviours across the two experimental sessions, a set of eye-tracking metrics was computed based on the gaze data associated with the defined AOIs. For each session, entropy was calculated as equation (2) to evaluate the distributional complexity of visual attention across different AOIs—higher entropy indicates a more dispersed attention pattern, while lower entropy suggests focused attention on fewer instruments.

$$H = -\sum_{i=1}^{N} p_i \log_2 p_i \tag{2}$$

Additionally, the number of AOI transitions was computed to quantify the frequency with which pilots shifted their gaze between different cockpit regions, providing insight into scanning behaviour and information search strategies. The fixation count per AOI was also analysed to assess which instruments or display elements received the most visual attention in each session. Moreover, the total number of fixations was used as a general measure of visual workload, and the scanpath length—defined as the cumulative number of fixations throughout the session—was employed to represent the extent of visual exploration. These metrics collectively enabled a comprehensive comparison of pilot attention allocation and scanning dynamics between the two sessions.

3.2.5.2. Statistical Analysis

Using a within-subjects experimental approach, the current study examined how visual illusions affect the onset and progression of various SD kinds from a neuro-ergonomic standpoint. This study did not take gender differences or flight experiences into account. Outliers and equal variances were eliminated, and the normal distribution was ascertained using the Shapiro-Wilk test. To evaluate the impact of visual illusions on the occurrence of SD. Paired T-test was first computed to summarise the key neurophysiological metrics with significant differences under the two experimental conditions (i.e., no visual illusion and visual illusion). The Wilcoxon Signed-Rank Test was selected if the data did not conform to the normal distribution. A linear mixed-effects model (LMM) was then used to detect the significant differences in neuro-ergonomics metrics between the four SD types (including Type I-III and non-SD), as well as the training hours. The Friedman test was used for the anomalous values' differential statistics. The significance level was adjusted using the Bonferroni adjustment. A statistically significant value was $P \leq 0.05$.

3.3. Result

3.3.1. Neurophysiological Behaviours: Cerebral Cortical Activity

Topographic heat maps of PSD results (Figure 3-5) were visually inspected to gauge alpha, beta, theta and gamma band activity changes across SD-type changes. Generally speaking, non-SD states exhibited higher average PSD and more balanced distribution compared to the states that suffered from SD. Type III SD (incapacity) had the most heterogeneous distribution of cortical activity but the second-highest PSD intensity. Type I SD (unrecognised) had the lowest PSD values in alpha (PSD=2.1), beta (PSD=1.1) and gamma (PSD=0.95), while Type II (recognised) was lowest in theta (PSD=6.1). Thus, increased neurophysiological activation from Type I-III SD was confirmed with H2. Notably, cerebral cortical activity was elevated in PFC, TL and OL regions when suffering from SD.

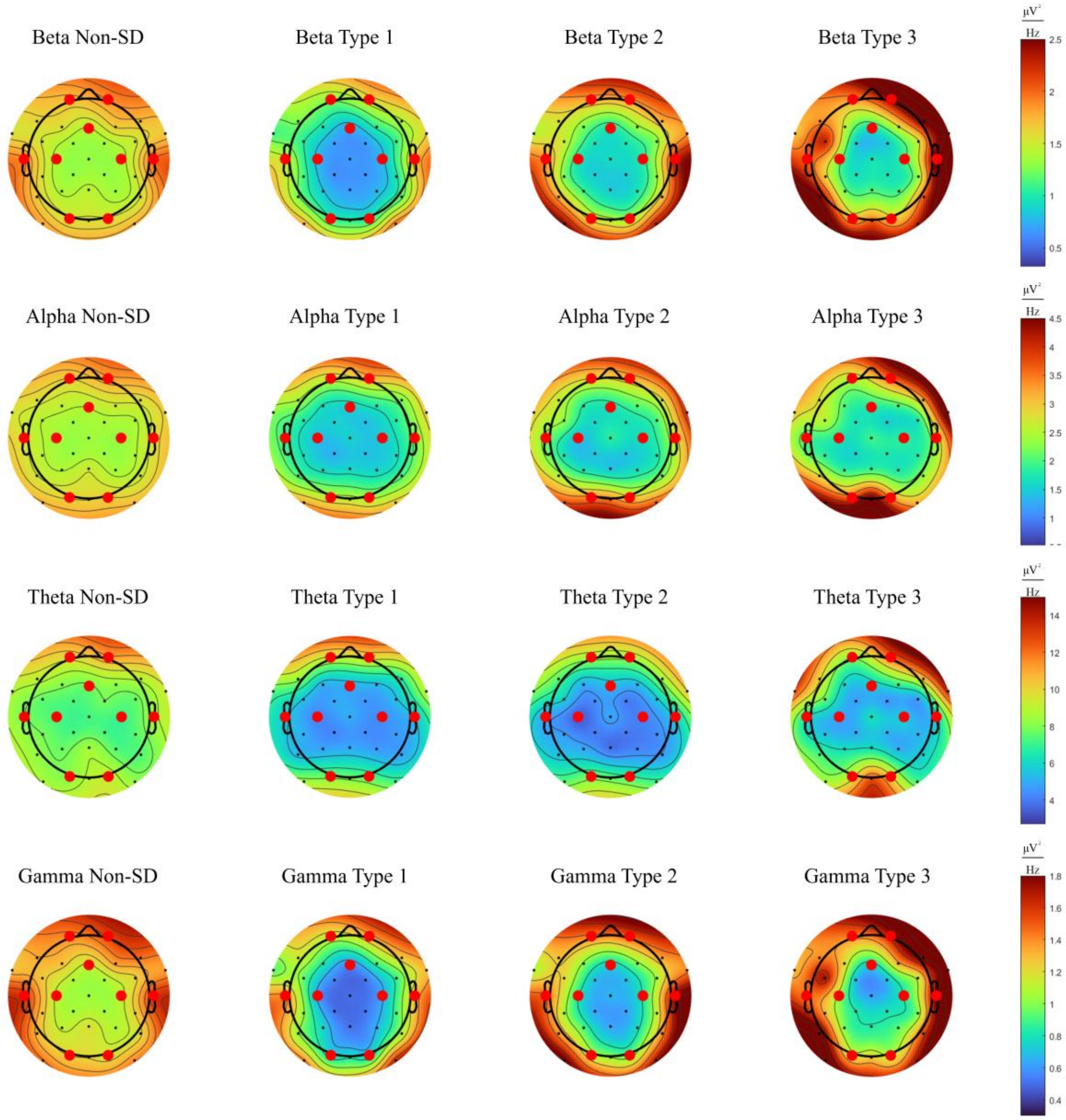


Figure 3-5. The heat map of the PSD in the beta, alpha, theta and gamma band of four SD types.

Table 3-6 shows the result of the mixed linear model of the EEG data. The model contained 10,240 observations across 26 groups, with an average of 393.8 of group size. First, concerning the SD types effect, PSDs were significantly reduced in the Type I to III SD groups compared to the baseline group (Non-SD) ($p \leq 0.05$), which confirms with H1. The relative reduction was most pronounced in the Type I SD group ($\beta =$

$-0.717, p < 0.001$). The Type II and III SD groups similarly showed a significant decrease ($p \leq 0.05$). Tukey post-hoc tests (Table 3-7) showed significant differences between non-SD and all three other SD types ($p \leq 0.05$), between the Type I and III SD, and between the Type II and III SD. This significant difference between non-SD and Type II SD rejects H4. No significant difference was shown between Type I and II SD. Second, with respect to electrode locations, electrode locations (e.g., CP5, F3, F4, F7, F8, FC5, FC6, FT9, FT10, Fp1, Fp2, O1, O2, Oz, P7, P8, PO9, PO10, T7, T8) showed significant differences in EEG signals ($p \leq 0.05$). Finally, taking into account the interaction effect between SD types and the EEG band, the Type I SD showed a significant positive interaction effect in the gamma band ($\beta = 0.336, p = 0.042$). All SD types showed a significant negative interaction effect in the theta band ($p \leq 0.001$). This demonstrates a mechanism for differences in band activity across SD types.

Table 3-6 Mixed linear model regression results.

Variables	Coef.	Std.Err.	z	P>|z|
Intercept	2.067	0.217	9.522	0.000
SD_Type[T.Type-I]	-0.717	0.117	-6.135	0.000
SD_Type[T.Type-II]	-0.265	0.132	-2.009	0.045
SD_Type[T.Type-III]	-0.409	0.148	-2.759	0.006
EEG_Band[T.Beta]	-1.175	0.116	-10.154	0.000
EEG_Band[T.Gamma]	-1.482	0.116	-12.811	0.000
EEG_Band[T.Theta]	5.768	0.116	49.854	0.000
SD_Type[T. Type-I]:EEG_Band[T.Beta]	0.216	0.165	1.307	0.191
SD_Type[T. Type-II]:EEG_Band[T.Beta]	0.230	0.184	1.250	0.211
SD_Type[T. Type-III]:EEG_Band[T.Beta]	0.353	0.206	1.715	0.086
SD_Type[T. Type-I]:EEG_Band[T.Gamma]	0.336	0.165	2.034	0.042
SD_Type[T. Type-II]:EEG_Band[T.Gamma]	0.209	0.184	1.135	0.257
SD_Type[T. Type-III]:EEG_Band[T.Gamma]	0.231	0.206	1.124	0.261
SD_Type[T. Type-I]:EEG_Band[T.Theta]	-1.549	0.165	-9.373	0.000
SD_Type[T. Type-II:EEG_Band[T.Theta]	-2.060	0.184	-11.194	0.000

SD_Type[T. Type-III]:EEG_Band[T.Theta]	-0.944	0.206	-4.583	0.000

Table 3-7 Tukey post-hoc tests results.

Group1	Group2	Meandiff	p-adj	Lower	Upper	Reject
Non-SD	Type I	-0.9514	0.0000	-1.1760	-0.7269	TRUE
Non-SD	Type II	-0.7756	0.0000	-1.0257	-0.5256	TRUE
Non-SD	Type III	-0.3691	0.0039	-0.6488	-0.0893	TRUE
Type I	Type II	0.1758	0.2769	-0.0762	0.4278	FALSE
Type I	Type III	0.5824	0.0000	0.3008	0.8639	TRUE
Type II	Type III	0.4066	0.0031	0.1043	0.7088	TRUE

Figure 3-6 compares the cerebral cortical activity with or without visual illusions. * represents the neurophysiological behaviours of brain regions with significant differences in two visual illusions after Paired T-test. Overall, the average PSD of no visual illusion condition in the four frequency bands was lower than with visual illusions, which confirms H2. Among these, visual illusions had the greatest impact on PL, showing significant differences in all four frequency bands (Alpha: $p = 0.012$, Beta: $p = 0.023$, Theta: $p = 0.027$, Gamma: $p = 0.035$). In addition to gamma, PFC had significant differences in the three frequency bands (Alpha: $p = 0.045$, Beta: $p = 0.045$, Theta: $p = 0.040$). In the entire TL region, only RTL was significantly different for visual illusions in the alpha band ($p = 0.046$). Alpha ($p = 0.022$) and theta ($p = 0.030$) were observed significant differences in OL.

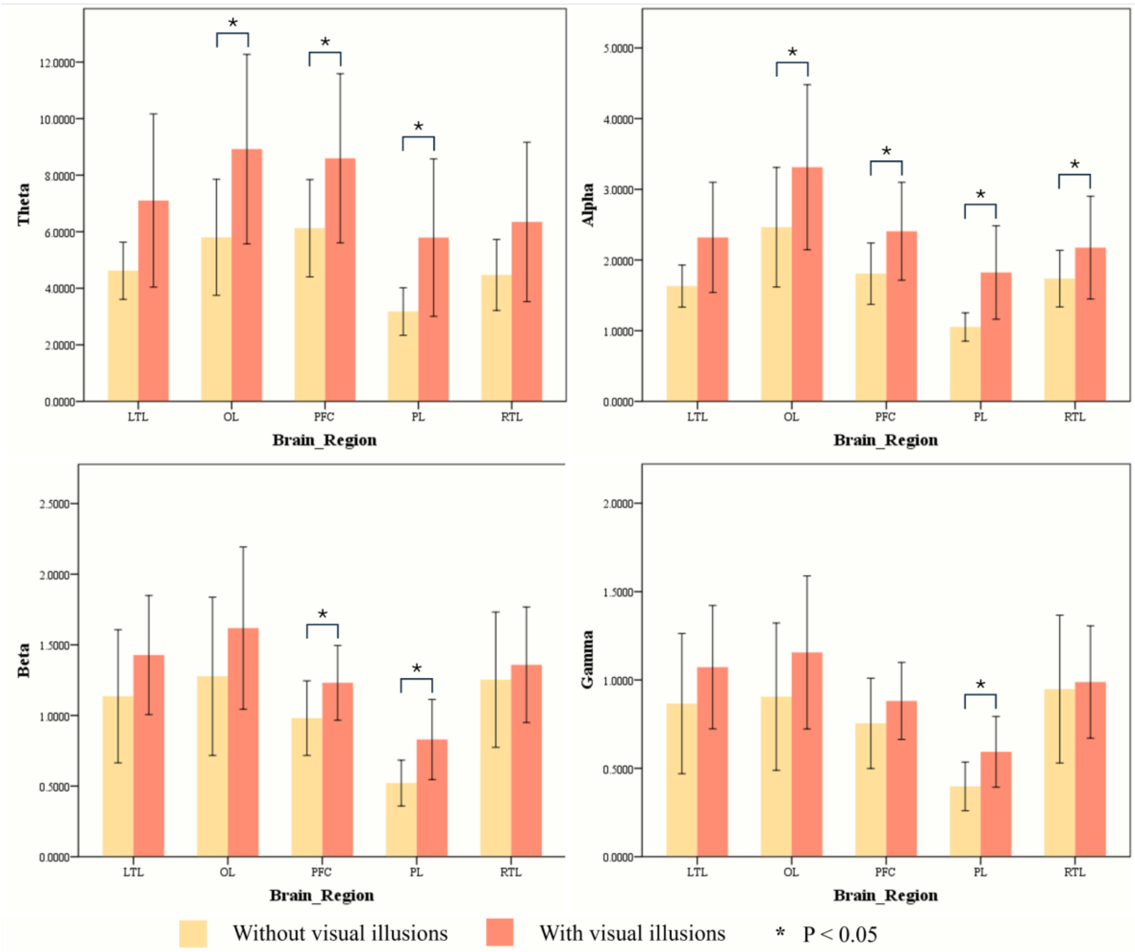


Figure 3-6. Paired T test result of the comparison of four wave bands in brain regions between visual illusions.

3.3.2. Visual Scanning Behaviour Analysis

The heat map illustrates visual scanning patterns of all participants in different visual illusion settings (Figure 3-7 (a)-(b)), and different SD types (Figure 3-7 (c)-(f)). Results indicated that pilots did not glance at the external view much, regardless of external weather conditions. It is evident that pilots paid more attention to the ND, ISIS, and ECAM to eliminate the effects of visual illusions. Moreover, visual illusions also caused pilots to pay more attention to data such as the aircraft's flight attitude, which was reflected in the more of the fixation falls on the PFD's pitch and roll scale, underscoring the pivotal role of attitude information in maintaining spatial orientation.

The three SD types also clearly exhibit different visual scanning patterns. In the non-SD state, pilots allocated more visual attention to the flight plan, FCU, and ECAM control area. Pilots reduced their fixation on the flight plan in the Type I SD state. Since the inconsistent instrument display was triggered, the pilot's view of ISIS and ECAM increased. Pilots' visual scanning was the most narrowly focused on the Type II SD state (recognised disorientation), consistent with H3. In the severe Type III SD state, pilots became overly concerned with the ISIS, ECAM, and ECAM control area while neglecting crucial information on the PFD and ND. Meanwhile, compared with Type II SD, Type III SD also increased the control of FCU. Entropy values, which indicate the dispersion of gaze across multiple AOIs, were generally higher in session-2 (S2, with visual illusion) than in session-1 (S1, without visual illusion), suggesting more distributed attention during the visual illusions. As Table 3-8 shows, the entropy of total fixation duration increased from 2.35 (S1) to 2.61 (S2), while average fixation duration entropy rose from 3.25 to 3.42. Similarly, the entropy associated with the number of fixations and total glance duration also exhibited slight increases in S2 (2.66 and 2.60, respectively) compared to S1 (2.52 and 2.36), implying a more varied scanning strategy in the later session. In terms of visual transition behaviour, the number of AOI transitions increased from 7,991 in S1 to 8,735 in S2, indicating a higher frequency of gaze shifts between cockpit instruments. Fixation count per AOI also varied substantially. For instance, S2 demonstrated notably higher fixation counts on key displays such as the Primary Flight Display (PFD_PF: 132,638 in S2 vs. 63,316 in S1), Navigation Display (ND_PF: 107,070 vs. 35,025), and External View (35,730 vs. 18,021), suggesting an increased demand for visual monitoring or information processing during that session. Conversely, fixations on secondary panels such as

ECAM Control were only moderately increased or remained comparable across sessions. Finally, the overall scanpath length—used as a cumulative indicator of visual exploration—was significantly greater in S2 (395,883 fixations) than in S1 (158,868 fixations), reinforcing the interpretation that pilots exhibited more extensive gaze activity and a broader visual scan with visual illusions.

Table 3-8 Comparison of entropy between with/without visual illusions.

	Total duration of fixations entropy	Average duration of fixations entropy	Number of fixations entropy	Total duration of Glances entropy	Average duration of Glances entropy
S1	2.3516	3.2472	2.5191	2.3615	3.0658
S2	2.6077	3.4159	2.6551	2.6026	3.31296

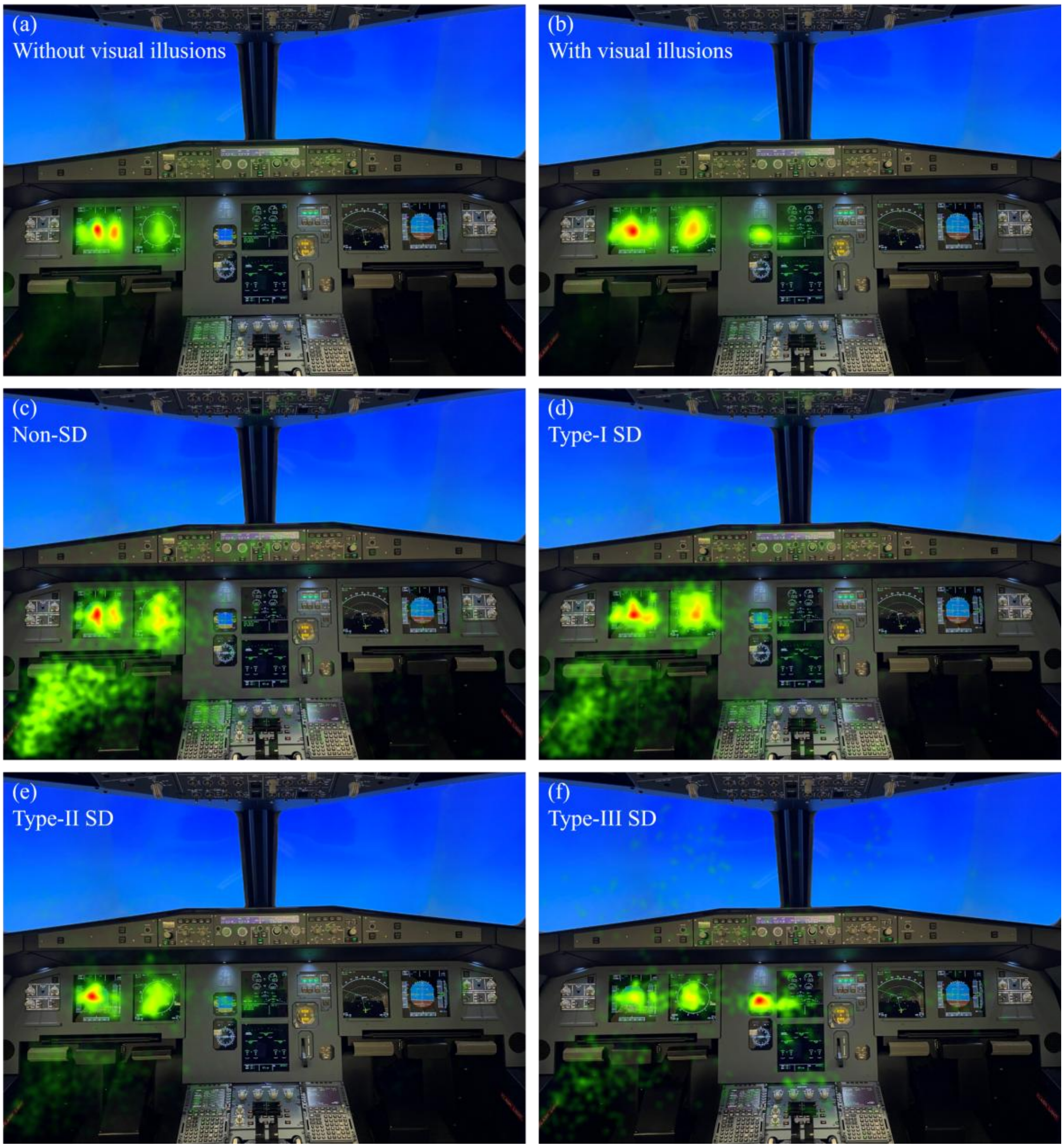


Figure 3-7. Heat map of the impact of visual illusions on SD type changes during operating flight tasks. (a) The heat map without visual illusions, (b) The heat map with visual illusions, (c) Non-SD, (d) Type-I SD, (e) Type-II SD, and (f) Type-III SD.

Lastly, this study systematically analysed the significant differences in eye movement behaviour among different SD types through the Kruskal-Wallis nonparametric test (Table 3-9). Overall, 22 indicators were significantly different among the four SD types. In particular, number of glances, total duration of fixations, and total duration of glances belong to the dynamic characteristics of attention. Pupil diameter, number of whole fixations, and whole fixation pupil diameter represent the intensity and complexity of cognitive processing. The above findings support that different SD types significantly affect eye movement behaviour and cognitive processing processes.

Table 3-9 Eye-tracking metrics of statistically significant differences of Kruskal-Wallis nonparametric test.

Metric	H-Statistic	p-value
Total duration of fixations	84.474	0.000
Average duration of fixations	14.342	0.002
Min duration of fixations	7.898	0.048
Max duration of fixations	17.443	0.001
Number of fixations	77.669	0.000
Average pupil diameter	68.184	0.000
Total duration of whole fixations	70.517	0.000
Average duration of whole fixations	17.160	0.001
Min duration of whole fixations	8.311	0.040
Max duration of whole fixations	15.288	0.002
Number of whole fixations	64.573	0.000
Time_to_1st_whole_fixations	19.331	0.000
Duration of first whole fixations	19.504	0.000
Average whole fixation pupil diameter	62.934	0.000
Total duration of Glances	82.637	0.000
Min duration of Glances	14.985	0.002
Number of Glances	91.027	0.000
Time to first Glance	19.676	0.000
Number of saccades in AOI	38.417	0.000
Time to entry saccade	18.184	0.000
Time to exit saccade	17.559	0.001
Peak velocity of entry saccade	7.993	0.046

3.3.3. Effects of Training Hours on SD Typology

To further investigate the generalizability of the findings to pilots with various levels of experience, we examined the distribution of SD types across all participants. As categorised in Table 3-10, novice pilots (NP, <156 training hours, n = 8) exhibited a higher incidence of Type I and III SD episodes compared to experienced pilots (EP, ≥156 hours, n = 8).

Table 3-10 Distribution of SD types across all pilot categories.

	No-SD	Type-I SD	Type-II SD	Type-III SD
Novice Pilot, NP	62	64	54	24
Experienced Pilot, EP	81	50	55	18

A LMM analysis (Table 3-11) was utilised to examine the impact of pilot training hours on the distribution of SD types. The Beta and Gamma frequency bands showed significant group differences ($p \leq 0.05$). Overall, the amplitude of neural activity was lower in EP than in NP. During non-SD and Type I SD state, EP exhibited significant lower beta and gamma synchronisation in PFC-TL regions. No significant differences were found in the Type II SD condition. Type III SD showed the greatest impact of training hours. Type III SD exhibited PFC decoupling compared to other SD types for beta, while PL, TL and OL remained prolonged activation in NP. OL presented the most significant difference between the two pilot categories.

Table 3-11. A linear mixed effects model result for investigating the influence of pilot training hours.

	Frequency Bands (Beta/Gamma)	PFC (Fz, Fp1, Fp2, FC1, FC2, F3, F4, F7, F8)	PL (Cz, C3, C4, Pz, P3, P4, CP1, CP2)	TL (FC5, FC6, CP5, CP6, T7, T8, P7, P8, FT9, FT10)	OL (PO9, PO10, Oz, O1, O2)
Non-SD	Beta	F3, F7, F8	P3	CP6, T8	PO9, O2
	Gamma	Fp1, F7, F3, F4,	Cz, CP2	All: FC5, FC6,	PO9, PO10

		F8, Fp2		CP5, CP6, T7, T8, P7, P8, FT9, FT10	
Type I SD	Beta	Fp1, F7, F3, F8, Fp2	C3	P8, CP6, T8	PO10
	Gamma	Fp1, F7, F3, F4, F8, Fp2	C3, CP1, Pz	FC5, FT9, CP5, P7, T8, FC6	Oz
Type II SD	Beta	-	-	-	-
	Gamma	-	-	-	-
Type III SD	Beta	F7, F3, F4, F8	C3, CP1, P3, Pz, P4, CP2, C4	All: FC5, FC6, CP5, CP6, T7, T8, P7, P8, FT9, FT10	All: PO9, PO10, Oz, O1, O2
	Gamma	F7, F3, F4, F8	Pz, P4, CP2, C4	FC5, FC6, FT9, T7, CP5, P8, CP6, T8	PO9, Oz, O2

3.4. Discussions

3.4.1. Theoretical Insights

3.4.1.1. Identifying Neuro-ergonomic Patterns of Different SD Types.

The findings from the neurophysiological analysis revealed that the PSD in non-SD involved higher levels of cognitive integration with balanced theta, alpha, and gamma activity, suggesting optimal attention allocation and alertness-workload tradeoff (Q. Li et al., 2024). In contrast, Type I SD (unrecognised SD) showed reduced alpha and gamma activity in OL, indicating an impaired visual illusion to maintain standard cognitive integration and information processing (P. Li et al., 2023). Type II SD (recognised) and Type III SD (incapacitating) were associated with heightened beta and gamma activity, particularly in the PFC and TL, which points to increased arousal and attempts to cope with heightened demands (T. Lu et al., 2023). These patterns suggest that Type I SD represents a failure in recognition and cognitive adjustment. In contrast, Type II SD and Type III SD reflect a state of heightened cognitive alertness and emotional response.

The LMM result (Table 3-6) indicated that the staggering theta coefficient ($\beta=5.768$, $z=49.854$, $p<0.001$) confirms its role as the primary neural stabiliser. For time-related cognitive tasks like information collection and working memory, theta waves consistently reflect mental workload (Q. Li, K. K. H. Ng, et al., 2023). Increased theta power in non-SD indicated effective attentional engagement, while the spike in theta activity in PFC and OL during Type III SD reflected the brain's effort to manage overwhelming disorientation, especially for visual information management. The reduction in alpha activity with increasing SD levels suggested a progressively demanding cognitive state, highlighting an increasing workload level since the pilot cannot quickly identify the cause factor of inconsistent instrument display and how to deal with it (Q. Li, H. C. Leung, et al., 2023). While Type I SD showed the lowest PSD in alpha, implying lack of alertness. The increased beta and gamma bands' power are associated with higher task demands and active states. Especially in Type II SD and Type III SD, the beta and gamma activity levels have significantly increased in PFC and TL. In Type I SD, occipital gamma enhancement suggests initial visual cortex efforts to resolve sensory conflicts, compensating for theta depletion. Conversely, Type III SD exhibited pathological prefrontal gamma hyperactivation that exceeded neural integration capacity, reflecting failed top-down regulation. This biphasic pattern supports the EEG oscillations (Herrmann & Demiralp, 2005): Early-stage SD recruits occipital gamma for error correction, while advanced SD transitions to maladaptive frontal gamma surges, marking cognitive overload and spatial awareness collapse.

While this study currently utilised conventional spectral metrics (e.g., PSD) to differentiate SD subtypes, recent advances in divergence-based EEG analysis—such as belief f-divergence (J. Huang et al., 2023; Xiao et al., 2025) and fractal Rényi divergence (Y. Huang et al., 2023)—offer promising avenues to capture nonlinear neural dynamics during SD. For instance, approaches such as belief f-divergence and generalised f-divergence can offer novel insights into EEG signal complexity by quantifying subtle differences in the underlying probability distributions of neural activity (J. Huang et al., 2023). Belief f-divergence could also enhance detection of such transitions by identifying abrupt changes in prefrontal theta-gamma coupling—a marker of cognitive conflict. Additionally, the integration of higher-order fractal belief Rényi divergence offers a robust framework to capture the multi-scale and self-similar characteristics of EEG signals, which are critical for distinguishing between varied SD conditions. By incorporating these advanced divergence measures into our analytical pipeline, future studies may improve the sensitivity and specificity of machine learning classifiers, enabling more accurate detection and differentiation of SD types. This integration could facilitate the development of adaptive and predictive systems in cockpit environments, ultimately advancing human-machine collaboration and enhancing flight safety.

As for the attention allocation perspective, under visual illusion conditions, pilots focused more on key instruments related to aircraft control, such as the ND, ISIS, and ECAM, to manage emergencies and reallocate visual resources to maintain spatial awareness. Bad weather conditions also make pilots more alert and enhance scans and cross-checking of multiple AOIs. Differences between SD types were also evident in visual scanning behaviour. During non-SD, the flight plan receives the most attention, and FCU has the highest number of fixation points among all SD types, indicating better spatial awareness and control of aircraft (Yuan, Ng, Li, et al., 2024). Surprisingly, Type II SD exhibited the narrowest focusing scanning mode, and the number of fixations of AOIs required to address inconsistent instrument displays such as ISIS and ECAM-control did not significantly increase. According to the Multiple-Resource Theory (MRT) (C. D. Wickens, 2002, 2008), as perceived cognitive workload shifts away from optimal levels, pilots need to integrate conflicting spatial information (such as cross-checking instrumental data) in a short period of time, which suppresses visual information breadth processing in the OL, forming a cognitive tunnelling effect in the visual attention allocation (Van der Burg et al., 2024). Type III SD showed over-fixation on specific instruments, neglecting critical displays like the PFD and ND, thereby reflecting impaired information prioritisation.

3.4.1.2. Dynamic Transitions between SD Topologies

The neuro-ergonomic differentiation of SD subtypes provides a static snapshot of disorientation states, the temporal progression between subtypes during prolonged disorientation remains a critical frontier. Integrating findings with conflict monitoring theory (Dignath et al., 2020) and alertness-workload trade-off theory (Yan et al., 2024), this study further explored the dynamic transitions between SD types. The mixed linear model (Table 3-6) and post-hoc analyses (Table 3-7) revealed critical nonlinear interactions between SD types, electrode locations, and frequency bands, providing a mechanistic basis for understanding transition pathways.

Non-SD to Type I SD: Attentional Decoupling: Prolonged monotonous flight missions make pilots less alert and underload (Kamzanova et al., 2014), which leads to failure to detect visual illusions in a timely manner. The difference in neurophysiological patterns from non-SD to Type I SD was the largest between other SD typologies transitions. The transition from non-SD to Type I SD exhibited the most widespread spectral disruptions, with significant power differences ($p \leq 0.05$) across all frequency bands (alpha, beta, gamma, theta) in PFC, PL, and TL. OL, however, showed selective beta ($O1, p = 0.01$, $Oz, p = 0.00$, $O2, p = 0.00$) and gamma ($O1, p = 0.01$, $Oz, p = 0.01$, $O2, p = 0.01$) alterations, suggesting early-stage visual processing deficits despite preserved low-frequency oscillations. This decoupling aligns with “perceptual insensitivity", where pilots under low cognitive engagement fail to detect mismatches. Behaviourally, this manifests as undetected deviations from flight parameters, with fixation patterns resembling non-SD states but lacking corrective actions.

Type I to Type II SD: Cognitive Surprise and Conflict Ignition: Prolonged Type I episodes shift to Type II SD, characterised by beta rebounds in OL ($PO9, p = 0.02$, $O1, p = 0.01$, $Oz, p = 0.02$) and PFC ($Fp1, p = 0.04$). This transition reflects conscious conflict detection, while pilots were still able to reconcile sensory. What more, theta suppression coupled with beta surges, reflecting compensatory top-down control (Haslbeck & Zhang, 2017; C. S. Yu et al., 2016). The lack of significant theta differences between Type I and II suggests that conflict recognition relies more on beta/gamma reallocation than theta-mediated workload modulation. The anterior cingulate cortex (ACC), a part of the brain involved in cognitive regulation, is not directly monitored because to the restricted monitoring capabilities of EEG equipment, but it also seems to react when conflict arises (Botvinick et al., 2001). ACC is usually associated with Fz, FC1, and FC2. These sites showed further increases in alpha, beta, and gamma after the discovery of conflict, but theta showed a decrease. Notably, PL showed no spectral differences, implying disengagement from spatial integration networks during conflict resolution. Paradoxically, this hyperactivation narrows visual scanning ranges, prioritising instrument fixation over environmental monitoring—a hallmark of adaptive attention (Engström, 2011).

Type II to Type III SD: Cognitive Overload and Compensatory Hyperactivation: Sustained hyperactivity in Type II SD depletes reserves in the PFC, precipitating a cascade toward Type III incapacitation. Although there were no significant differences in the EEG bands at individual electrode locations in the transition from Type II to Type III. In PFC, alpha, beta and gamma all showed further decreases. However, theta continued to be enhanced, highlighting the role of theta as a stabiliser whose depletion accelerates the transition to Type III.

3.4.1.3. Cognitive Mechanisms Behind Different SD Types.

The neuro-ergonomic differentiation of SD subtypes aligns with the MRT (C. D. Wickens, 2008), which posits that cognitive resources are distributed across distinct neural subsystems. Constrained by a limited cognitive processing pool, MRT frames cognition as a competition for domain-specific resources (e.g., visual vs. executive). Our findings suggest that transitions between SD types reflect dynamic reallocations of these resources under dynamic cognitive demands.

Alertness-Workload Tradeoff Theory: According to the alertness-workload trade-off theory (Yan et al., 2024), the balance where increased workload may enhance alertness up to a point, beyond which excessive demand degrades performance due to stress or cognitive overload. The relationship between workload and alertness posits an inverted

U-shaped relationship (Mélan & Cascino, 2022). Peak performance (non-SD) occurs at a balance where workload sustains alertness without overwhelming cognitive capacity. Therefore, the human states in non-SD could be treated as the optimal zone for alertness-workload. Conversely, low workload can reduce alertness (Type I). The observed beta/gamma decreased in Type I but surged during Type II signifies a paradoxical interplay between heightened vigilance and cognitive overload. The H4 hypothesis expected Type II SD to return to a neural state similar to non-SD due to proper vigilance. However, the results showed that beta and gamma power remained significantly higher in the non-SD state than in Type II SD across multiple electrode regions (beta in Fz, CP1, PO9, P4, and FC2, gamma in CP1, Pz, CP2, and FC2). Such results rejected H4. Conscious conflict recognition (Type II) triggers compensatory neural activation to sustain task performance, but this effort depletes metabolic reserves, adaptively narrowing attentional bandwidth. Consciously conflicts recognition triggers compensatory neural activation due to the increment of alertness and workload, which continues until Type III triggers hyperactivation.

Conflict Monitoring Theory: The conflict-monitoring theory (Dignath et al., 2020) suggests that a specialised monitoring system detects conflicts between incompatible reactions, and that these conflicts cause attentional filters to shift and control mechanisms to adjust to the task demand. According to Type II SD's characteristic, cognitive abilities are recruited when conflict circumstances are detected (Pires et al., 2018). Enhanced Beta and Gamma activity reflects a surge in conflict monitoring and working memory load but also results in a narrowing of the visual scanning range (overcrowding of visual resources). This would explain why H4 was rejected, i.e., it was found that the cognitive compensation effect after conflict makes the neuro-ergonomic pattern different from the optimal state of non-SD. Data pertaining to the ACC, a region of the brain involved in cognitive regulation that also seems to react when conflict arises, corroborate this conclusion (Botvinick et al., 2001). The ACC activation signals (beta and gamma) signals unresolved sensory conflicts, prompting conscious error correction. Reduced visual scan range in Type II SD reflects strategic focus on critical instruments, not passive tunnelling. Further incapacitation leads to Type III SD. The progressive narrowing of visual scanning of main panel of cockpit exemplifies cognitive tunnelling in Type III, where hyperfocus on the conflicting instrument.

3.4.2. Practical Implications

This study revealed the neurophysiological and visual scanning patterns exhibited by pilots experiencing the three distinct types of SD compared to those not suffering from SD in order to develop a comprehensive, multimodal framework for SD detection. Targeted countermeasures are necessary to prevent and mitigate the effects of different SD types.

3.4.2.1. Development of Pilot-Support System based on SD Detection.

The study's conclusions demonstrate the possibility of incorporating real-time neuro-ergonomic monitoring methods into the cockpit's pilot-support system. By tracking alterations in cerebral cortex activity, especially in the PFC, OL, and TL, such systems might identify early indicators of SD and activate automated support systems to help pilots cope with SD. The results also have implications for improving pilots' decision-making capabilities under visual illusions. Strategies should also be developed to manage fixation biases by the human-machine interface (HMI) design, particularly during unrecognised disorientation (Type I SD), to prevent over-fixation on non-critical displays. These interventions could help redirect visual focus towards critical instruments, thereby improving spatial awareness and reducing the risk of SD.

3.4.2.2. Improvement of Pilot Training Programmes.

The observed elevation in beta and gamma synchronisation among NP during non-SD and Type I SD states underscores a critical neuroergonomic paradox: heightened neural activity does not equate to superior performance. EP exhibited significantly lower beta and gamma power in prefrontal-temporal (PFC-TL) regions during non-SD and Type I SD ($p \leq 0.05$), suggesting streamlined neural processing honed through repetitive scenario exposure. In contrast, NP's hyperactivation likely reflects compensatory overengagement—inefficient recruitment of frontal executive networks to sustain baseline performance, akin to novice pilot overactivating motor cortices during simple manoeuvres. This difference in varied training hours could serve as one of the evaluations of the pilot training. By providing the neurophysiological state of the pilot during training, it can also be used to improve more targeted pilot training programs.

More training and practice should be arranged to simulate SD situations and train how to resolve and prevent SD correctly. Training should reinforce decision-making strategies for managing visual illusions and inconsistent instrument readings, emphasising the importance of effective information processing and prioritisation. Incorporating the multimodal SD detection framework into flight simulators could help pilots better understand their responses to SD and improve their ability to recognise and mitigate disorientation (Friedrich et al., 2021). This approach would enable pilots to become more familiar with different SD states' cognitive and visual demands, ultimately enhancing their ability to cope with challenging flight conditions.

3.5. Limitation and Future research

First, the study was limited by the controlled simulation conditions, which only involved a limited type of visual illusion through inconsistent flight parameter displays and weather conditions. Next, the fixed-based simulator does not replicate gravitational acceleration which serves as an important component contributing to SD. While the simulator provides a controlled and repeatable environment for isolating cognitive and perceptual mechanisms underlying SD, it cannot fully reproduce certain contributing factors present in real flight, particularly vestibular cues, sustained motion feedback, and complex aircraft–environment dynamics. Nevertheless, this fixed-based setup restricts the ability to fully capture the dynamic interactions present in real flight scenarios, particularly those involving six degrees of freedom (6-dof). As a result, some forms of multisensory conflict and long-term adaptation effects may be attenuated compared to in-flight conditions. These limitations may influence the intensity and temporal characteristics of SD-related responses but are unlikely to alter the relative patterns of cognitive and neurophysiological differences observed across experimental conditions. Accordingly, the findings should be interpreted as reflecting relative, mechanism-driven effects rather than absolute representations of in-flight disorientation.

Additionally, exploring varied environmental factors and emergency scenarios could enhance our understanding of the interplay between cognitive load and flight performance. Future research should incorporate more comprehensive simulation environments that include gravitational feedback to better understand the complex cognitive and physiological responses to spatial disorientation. Then, limited by the number of subjects, the current research is less involved with subjects' training hours and flying experiences for triggering different types of SD. We acknowledge that future studies with larger cohorts and a greater representation of low-to-mid experience pilots are needed to further reinforce and validate the generalisability of these findings.

As for the neuro-ergonomic pattern classification, future data analysis can incorporate advanced divergence-based methods, such as belief f-divergence, generalised f-divergence, and fractal Renyi divergence. These methods can extract high-order statistical features and fractal properties from EEG signals to distinguish complex signal patterns. Especially when processing physiological signals with self-similarity and multi-scale characteristics, they can provide richer information and more robust feature descriptions for pattern classification. Developing adaptive automation systems based on neurophysiological and visual metrics could also provide dynamic support to pilots in challenging situations. Future cockpit systems should focus on enhancing human-machine collaboration to reduce the risk of SD, rather than merely relying on automation to replace human judgement.

3.6. Concluding remarks

This study investigated the impact of visual illusions caused by adverse weather conditions and malfunctioning instrument emergencies on neurophysiological patterns and visual scanning behaviours across different SD types. Compared to the no SD state, different SD types showed varied neuro-ergonomic patterns related to cognitive demands and processing abilities. The results confirm significant differences in internal neural processing in contexts with high cognitive conflict and attention allocation demands, even when surface performance is stable. Non-SD involved higher cognitive integration, with more balanced and elevated cerebral activity. Type I SD (unrecognised SD) showed the lowest PSD values, indicating a failure in information perception. Type II SD (recognised SD) involved a lower cognitive workload than Type III SD, which showed the greatest differences in brain activity, reflecting heightened alertness, especially in the PFC and OL. Visual scanning also showed differences with different SD types. The widest scan area was observed when vigilance was highest in non-SD. With visual illusion, scanning area gradually began to focus on the main instrument panel, showing adaptive attention. But as the pilot became incapacitated, the scanning behaviour gradually changed from controlled adaptive to cognitive tunnelling (Type III SD). The practical implications of these findings include integrating real-time neuro-ergonomic monitoring systems in the cockpit to detect early signs of SD and guide the pilot support systems' design.

Chapter 4. Towards pilot's spatial awareness enhancement: An Adaptive Attention-Awareness Pilot-Aircraft Interaction System in future cockpit

This chapter introduces the framework of Adaptive Attention-Awareness Pilot-Aircraft Interaction System (3A-PAI). The primary objective of the 3A-PAI system is to enhance pilot spatial awareness and attention stability through multimodal sensing, robust recognition of cognitive states, and timely, adaptive feedback. Building upon the neuro-ergonomic signatures of four Spatial Disorientation (SD) types defined in the preceding chapter, this chapter solves two main works, SD and LoAC recognition and prediction, respectively. First, this work extends the research by applying advanced machine-learning models to realise automatically SD types of classification and abnormal scanning series detection. Second, the proposed system is then designed to detect early signs of SD and Loss of Attention Control (LoAC), predict their occurrence, and deliver context-relevant interventions to support safe and effective pilot performance.

For SD prediction, this research applies LightGBM, achieving a remarkable 96.8% accuracy in four-class SD classification. For LoAC prediction, this chapter develops a novel joint model that combines probabilistic models, optimisation algorithms, and classification algorithms. This joint modelling approach leverages the strengths of both Hidden Markov Models (HMMs) and Conditional Random Fields (CRFs): HMMs provide a generative understanding of latent states and temporal dependencies, while CRFs offer robust discriminative prediction capabilities, particularly effective in integrating complex, overlapping features. This synergy enables a comprehensive and accurate approach to LoAC prediction.

4.1. Introduction

4.1.1. Background and Motivation

Significance of Spatial Disorientation (SD) in aviation safety: Spatial Disorientation (SD) has long been recognised as a critical factor in aviation safety, contributing to a significant proportion of flight incidents and accidents, particularly under demanding environmental conditions (Fudali-Czyz et al., 2022). When a pilot's impression of location, velocity, or attitude deviates from reality, it's known as SD, and it frequently results in poor judgments or delayed reactions when flying (van den Hoed et al., 2022). This is especially true in conditions of low visibility or monotonous environments, where visual cues are degraded. Vision is typically the dominant sense, providing approximately 80% of the sensory input required for orientation (Newman & FAICD, 2007). However, non-visual senses can be tricked by low visibility, such as at night, in fog, or on uneven ground, resulting in confusion and illusions (Sánchez-Tena et al., 2018).

Relationship between Loss of Attention Control (LoAC) and SD: A growing number of evidence indicates that loss of attention control (LoAC) is a major antecedent of SD. LoAC may manifest as channelled attention (overfocusing on a task), distraction (diverting attention away from critical tasks), or a startle and surprise response (Ghaderi & Saghafi). Cognitive overload, a common occurrence in complex, data-intensive cockpits, can disrupt cognitive processing, deplete neural resources, and impair situational awareness (Alreshidi et al., 2023). When pilots experience attentional lapses or fail to maintain effective visual scanning, their susceptibility to SD rises, particularly in phases of flight characterised by dynamic task demands and information overload.

The high lethality of unrecognised SD (Type I) reveals a critical gap in current pilot assistance systems: the inability of pilots to correct problems they do not perceive (Meeks et al., 2023). This highlights the urgent need for an objective, real-time detection system that can alert pilots before they become aware of (or incapacitated by) SD, thereby directly addressing the most lethal forms of the phenomenon. This means that traditional self-correction or instrument cross-checks (which rely on conscious recognition) will be ineffective if the pilot is not aware of the disorientation.

Importance of adaptive human-machine interaction (HMI) in aviation: While traditional pilot training and procedural safeguards remain valuable, they are increasingly insufficient to entirely mitigate the risks posed by SD (Mumaw et al., 2020; Newman & FAICD, 2007). As modern cockpits become progressively more information-rich and operationally complex, pilots face higher cognitive loads and are more susceptible to human error, making SD a persistent safety concern (Yamin et al., 2024). Despite significant advances in avionics and human-machine interaction (HMI) design, existing cockpit systems are largely reactive rather than proactive. They primarily rely on the pilot's ability to actively monitor and interpret information, offering limited real-time support for maintaining attention and situational awareness, especially during high-workload or cognitively challenging phases of flight (Zuo et al., 2025).

Consequently, any effective cockpit system designed to combine SD detection and mitigation must provide an external, objective alert or intervention mechanism that can operate independently of the pilot's subjective state or misperceptions. Such a system would be crucial for overcoming the inherent limitations of human self-monitoring (Yiu et al., 2025a). In this context, adaptive HMI has emerged as a promising approach. By continuously monitoring the pilot’s cognitive and physiological state, adaptive systems can offer timely, individualised interventions, potentially reduce SD risk and enhancing both safety and performance. However, realising such a system requires novel sensing and inference capabilities that extend beyond current practice to accurately and unobtrusively capture pilot state in real-time. Thus, this research addresses this critical need by integrating multi-modal sensing to build a proactive pilot-aircraft interaction system.

4.1.2. Research objectives, and research questions of this study

The primary objective of this study is to develop and empirically validate a **multimodal, adaptive pilot support framework** capable of real-time SD and LoAC recognition, and prediction. Specifically, this research aims to:

- Develop robust models for SD type recognition and prediction based on EEG-derived features.
- Construct phase-specific normative models of visual scanning using eye-tracking, enabling detection of attention-control anomalies.
- Integrate neurophysiological and behavioural data streams within an adaptive HMI framework, culminating in real-time, context-sensitive HMI interventions.

This study is guided by the following research questions (RQ):

- RQ1: How can multi-modal physiological and behavioural data, specifically EEG and eye-tracking signals, be effectively fused to accurately detect and classify different types of SD and LoAC in pilots?

- RQ2: What predictive models can be developed to forecast the onset of SD and LoAC based on real-time EEG and eye-tracking indicators, enabling proactive intervention before critical performance degradation occurs?
- RQ3: How can the detection and prediction outcomes be integrated into an adaptive interaction system that dynamically adjusts cockpit interface designs to mitigate SD and LoAC in varying operational and environmental conditions?

4.1.3. Overview and Contributions of the Study

To address these objectives, This study proposes the **Adaptive Attention-Awareness Pilot-Aircraft Interaction System (3A-PAI)**, a comprehensive framework that synthesises neurophysiological and eye-tracking data for real-time pilot state monitoring and adaptive HMI feedback. The ability of the proposed 3A-PAI system to objectively detect and predict SD, particularly Type I SD, goes beyond merely assisting pilots who recognise disorientation. It is able to intervene in situations where the pilot's own sensory systems have failed completely, providing a potentially life-saving 'source of truth' that directly addresses the most critical failure modes in SD-related accidents. The main methodological and conceptual innovations include:

- Development of a new conceptual and technical framework for the 3A-PAI system, demonstrating how real-time pilot state can be used to drive adaptive HMI interventions.
- Introduction of hybrid model combining machine learning (ML) algorithm, HMM and CRF for scanning anomaly detection, leveraging phase-specific features,

- Introduction of hybrid labelling strategies combining expert and weak labels for robust model training,
- Implementation of a closed-loop, context-adaptive HMI interface that dynamically intervenes based on pilot cognitive state.

The remainder of This study is organised as follows: Section 2 reviews relevant literature; Section 3 details the proposed 3A-PAI framework; Section 4 describes the experimental methodology and dataset; Sections 5 and 6 present methods and results for SD recognition and scanning behaviour supervision; Section 7 conclude with findings, implications, and future research directions.

4.2. Literature review

4.2.1. Spatial Awareness and Loss of Attention Control in Aviation

SD represents a breakdown in SA, often precipitated by LoAC. When visual scanning deteriorates—either through fixation tunnelling or neglect of critical displays—the integration of spatial cues is impaired, increasing susceptibility to perceptual illusions and orientation errors. Empirical evidence indicates that LoAC not only correlates with degraded SA but also serves as an early warning sign for impending SD events. This relationship underscores the importance of jointly monitoring attention control and spatial orientation to prevent accidents.

4.2.1.1. Theoretical concepts of Spatial Awareness (SA)

In aviation, the term "spatial awareness" (SA) describes the pilot's precise observation and comprehension of their position, orientation, and velocity in relation to the surface of the Earth and other objects in the operating environment. In aviation research, Endsley's three-level model of situation awareness—perception, understanding, and projection—has been extensively used to characterise the mental processes that underlie SA (Mica R Endsley, 1999). In dynamic flight contexts, SA enables pilots to integrate sensory inputs with mental models, thereby guiding effective decision-making and control actions.

Vision, the vestibular system (inner ear), and the proprioceptive system (sensory receptors in the skin, muscles, tendons, and joints) are the three primary sensory inputs that make up SA (Newman & FAICD, 2007). When these visual, vestibular, and proprioceptive inputs differ or are inconsistent in intensity, direction, and frequency, sensory conflicts arise, leading to illusions and triggering SD. Vision is thought to be the largest contributor to orientation (Gil-Cabrera et al., 2021; Ledegang & Groen, 2018).

Significant progress has been made in predicting pilot human factors using ML models (Q. Li, K. K. H. Ng, et al., 2023). The integration of EEG and eye tracking data is crucial for identifying and predicting SD types. Some studies have investigated the identification of a single SD type. Geva et al. (2025) showed the left frontal area decreased significantly (34%) in theta power (4–7.5 Hz) during the visual illusions. These physiological markers can be captured by ML models, enabling them to distinguish different types of SD. The Random Forest model has demonstrated optimal performance in detecting SD types in helicopter pilots (Sehoon et al., 2024).

Despite progress in SD detection and prediction, challenges remain. Rather than simply detecting a general condition called "SD," the system must be able to distinguish between different types of SD. This fine-grained distinction is crucial, as different types of SD arise from varying sensory conflicts and may require different corrective information or display strategies. This precise intervention can more effectively address pilots' specific dilemmas, significantly improving flight safety.

4.2.1.2. Theoretical Concepts of Loss of Attention Control (LoAC):

LoAC occurs when pilots fail to maintain effective allocation of visual and cognitive resources across relevant cockpit and environmental cues. In the cockpit, LoAC often manifests as suboptimal visual scanning, excessive fixation on irrelevant instruments, or neglect of critical areas, thereby undermining both SA and task performance (Mengtao et al., 2023). Integrating these eye-tracking metrics with ML has enabled more sophisticated methods for recognising and predicting LoAC. For explainable OOTL identification, Lyu, Li, Lee, et al. (2024) developed a linear temporal logic approach based on visual attention. Furthermore, including Area of Interest (AOI) characteristics into eye tracking features can greatly enhance machine learning performance (Lyu, Li, Qu, et al., 2024). In an Air traffic controller (ATC) study, Random Forest achieved the highest performance with the accuracy of 79% among three classifications of attention type classification (Z. Li et al., 2024). More advanced approaches, are now being used to analyse the temporal dynamics of neurophysiological data with highly accurate time-series characteristics, allowing for the prediction of an impending LoAC, rather than just post-hoc detection. Yiu et al. (2025a) examined the pilots attention level based on Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) with the accuracy beyond 99%.

While extensive research has explored both SD and LoAC, they are often studied in isolation. The existing literature typically focuses on either the physiological and perceptual causes of SD or the cognitive markers of attention failure. A significant gap exists in research that systematically integrates both streams of information. This fusion is essential for building a truly comprehensive and proactive system that addresses both the causes and effects of human error in high-stakes environments.

4.2.2. Adaptive Human-Machine Interaction (HMI)

Pilots are increasingly adopting supervisory roles, intervening only as needed (Yiu et al., 2024), which demands an evolution in HMI. This need is further underscored by the United States Air Force's (USAF) Technology Horizons project, which warned that by 2030, natural human capacities and advanced technologies would become increasingly misaligned, making humans the weakest link in generalised processes and systems (Dahm, 2010). Thus, a dynamic distribution of tasks between human pilots and aircraft interaction systems is imperative for superior overall performance.

4.2.2.1. Adaptive automation system in transportation

The evolution of aviation HMI has progressed from static, pre-defined interfaces to more sophisticated, adaptive systems that can dynamically adjust to the pilot's state, task demands, and environmental conditions (Feigh et al., 2012). In aviation, adaptive HMI seeks to optimise information delivery, task allocation, and alerting strategies to match the pilot’s current needs and capabilities (J. Liu et al., 2016). This paradigm shift, often termed adaptive automation (Kaber et al., 2001), moves beyond the traditional fixed allocation of functions between human and machine, instead advocating for a dynamic and context-sensitive distribution of control. The core principle is to maintain an optimal level of pilot engagement and workload, preventing both cognitive overload and vigilance decrements that can arise from under-stimulation. Pilots mostly depended on autopilot and auto-thrust systems in general aviation. The aircraft profile was not adequately monitored as a result of the decreased alertness. Pilots' lengthy idle time makes them OOTL (Mica R. Endsley & Kiris, 1995). With this background, A significant body of research has explored various levels and forms of adaptive automation. Yiu et al. (2025b) and Xu et al. (2025) developed LSTM and Transformer to realise the adaptive automation level according to pilot's EEG engagement, respectively. Specifically, the former study identified the high delta and theta waves with low beta and gamma waves contributed positively to the OOTL state. The later study further combined eye tracking data to interpret the SA level to customise the level of automation. This research has provided a theoretical foundation for developing systems that can intelligently adjust the level of automation, and the function between the pilot and the aircraft, such as shifting from manual to semi-automated flight control during high-workload periods or providing predictive alerts for potential system failures.

4.2.2.2. Adaptive human-machine interface and interaction in transportation

Adaptive HMI systems, by dynamically adjusting to real-time pilot cognitive states, are fundamentally designed to maintain optimal workload levels. A new idea for a Cognitive Pilot-Aircraft Interface (CPAI) was presented in a groundbreaking study by J. Liu et al. (2016). The purpose of CPAI, an adaptive knowledge-based supporting system, is to ensure aviation safety in SPO. Human errors from pilots can be prevented and improved human-avionics system synergy can be supported with the suggested CPAI system implementation, which is based on real-time sensing of the pilot's physiological and cognitive states. Estimates of the pilot's cognitive states drive the expert decision logics in the CPAI system, which provide adaptivity to the avionics HMI. This makes the system unique. Reallocating the task-load with different degrees of automation moderates the pilot's workload and fatigue levels, which determine the range of optimal performance. W. Yu et al. (2023) further built on this by using eye trackers to mine implicit expert knowledge and predict workload levels. They designed a SA improvement interaction system that uses the pilot's online perceived information and expert knowledge to infer tip information via the Head-up Display (HUD) and convey missing SA information. Understanding the development and recent developments in adaptive HMI requires an awareness of this fundamental background.

With the evolution of Artificial Intelligence (AI), pilot support system generates targeted tips information with the benefit of the Large Language Model (LLM). According to an empirical study, pilots who used the AdaptiveCoPilot system (Shaoyue Wen et al., 2025) showed shorter task completion times and greater frequencies of optimum cognitive load states, particularly in working memory and perception. This neuroadaptive guiding system, which was developed for pilot virtual reality preflight checklist training, is a state-of-the-art example of adaptive HMI. AdaptiveCoPilot uses functional Near-Infrared Spectroscopy (fNIRS) neuroimaging classification algorithms to evaluate pilot cognitive workload in real-time. It creates a context-aware LLM by combining behavioural data, these categorised cognitive burden levels, and well-crafted adaptive rules. With the ultimate objective of preserving the pilot's ideal workload, the LLM then cleverly interprets this input to dynamically influence adjustments in feedback modality (visual, aural, or textual) and information load. This system demonstrates the profound potential of integrating advanced AI (specifically LLMs) with neurophysiological sensing for highly personalised and context-aware adaptive guidance.

Despite the significant technological advancements, fundamental HFs challenges persist. These include effectively managing the risks of pilot over-reliance on automation, preventing disengagement, and mitigating the potential for diminished situational awareness that can arise from highly automated systems. Ensuring robust transparency in automation interfaces remains an absolutely crucial requirement, particularly during dynamic task allocation and critical automation handoffs, to prevent catastrophic outcomes.

4.2.3. Literature Summary and Identified Research Gap

The reviewed literature establishes that **SA and attention control are deeply interdependent** in aviation safety. LoAC can both precede and exacerbate SD, and technological systems that fail to address both concurrently are inherently limited in their preventive capabilities. While adaptive HMI, including HUD and LLM-generated systems, shows promise in supporting SA, its effectiveness is constrained by reliance on single-modality inputs and reactive design philosophies.

The key research gaps are therefore:

1. Future research needs to move beyond crude perceptual bias detection to fine-grained identification of different types of perceptual biases, thereby providing more targeted and effective corrective information.
2. Existing research often examines SD and LoAC as isolated phenomena, lacking a systematic approach that integrates physiological and cognitive information from both.

3. Limited implementation of adaptive AR-HUD interfaces that dynamically tailor information presentation to both flight phase and pilot cognitive state.

Originality and Novelty of the Present Study:

This study addresses these gaps by introducing the **Adaptive Attention-Awareness Pilot-Aircraft Interaction System (3A-PAI)**, which fuses EEG-based SD type recognition with eye-tracking-based LoAC detection to create a closed-loop, adaptive AR-HUD. By combining proactive prediction, phase-specific supervision, and context-sensitive intervention, the 3A-PAI framework advances both the methodological and applied frontiers of adaptive HMI in aviation.

4.3. Proposed Framework: Adaptive Attention-Awareness Pilot-Aircraft Interaction System (3A-PAI)

4.3.1. Overview of the proposed 3A-PAI system

The primary objective of the 3A-PAI system is to enhance pilot spatial awareness and attention stability through multimodal sensing, robust recognition of cognitive states, and timely, adaptive feedback. By continuously integrating neurophysiological (EEG) and behavioural (eye-tracking) data, the system aims to detect early signs of SD and LoAC, predict their occurrence, and deliver context-relevant interventions to support safe and effective pilot performance.

The Adaptive Attention-Awareness Pilot-Aircraft Interaction System (3A-PAI) is designed to proactively monitor, recognise, and mitigate pilot SD and LoAC in real time. This framework addresses the critical need for dynamic and context-sensitive support in high-risk, information-dense aviation environments, where lapses in attention or disruptions in spatial orientation can have severe safety consequences. Figure 4-1 illustrates the architecture and interaction between critical modules.

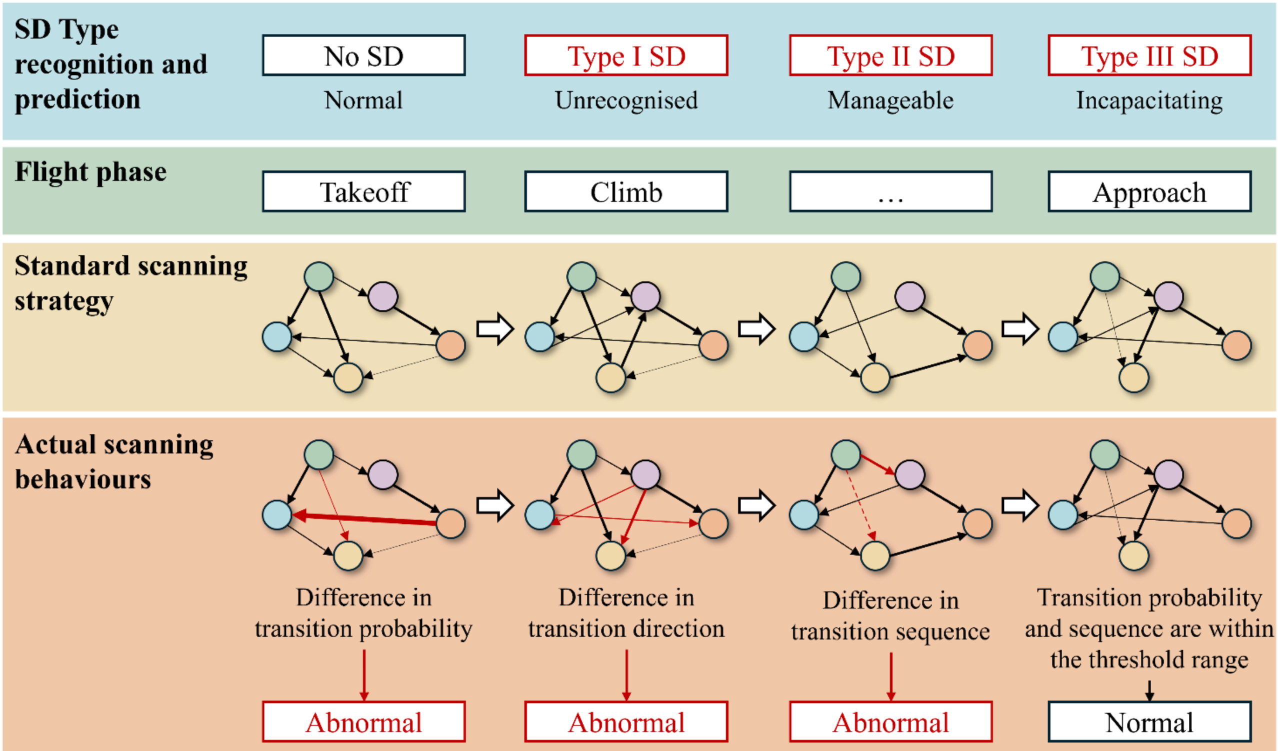


Figure 4-1. The crucial architecture of Adaptive Attention-Awareness Pilot-Aircraft Interaction System (3A-PAI) of the detection for SD and LoAC.

The system is underpinned by two central design principles:

- Adaptivity: The system dynamically tailors its monitoring, recognition, and feedback mechanisms according to both flight phase and individual pilot state. This includes real-time adaptation of detection thresholds, pattern recognition models, and AR-HUD interface cues based on current operational context and pilot condition.
- Attention-Awareness Integration: By jointly modelling attention control (via eye-tracking) and spatial awareness (via EEG), the system provides a holistic

understanding of pilot cognitive status, enabling integrated supervision and intervention for both overt scanning behaviour and covert neurocognitive shifts.

4.3.2. Architecture of the 3A-PAI system

4.3.2.1. Multimodal neuro-physiological data acquisition

This module collects high-resolution, synchronised neurophysiological and behavioural data in real time. EEG signals are continuously sampled to capture spatial awareness-relevant brain activity, while eye-tracking devices monitor gaze position, fixation patterns, and visual scanning dynamics. Both data streams are precisely time-aligned and segmented according to flight phase, establishing a comprehensive foundation for downstream analysis.

4.3.2.2. SD typology recognition and prediction module

Leveraging advanced ML algorithms, this module processes EEG-derived features—such as multi-band power spectral densities and frequency ratios—across the sensor array to recognise and classify SD episodes into predefined typologies (e.g., No SD, Type I/II/III). Both recognition and prediction functions are supported, enabling the system to both identify current SD status and anticipate high-risk intervals before onset. Classification models (including LightGBM, XGBoost, and others) are continually validated and updated based on accumulating data, ensuring robust generalisation to diverse pilot profiles and operational conditions.

4.3.2.3. Phase-specific scanning supervision module

This module is responsible for the continuous assessment of pilot visual attention control within the context of dynamically shifting flight demands. Eye-tracking data are mapped to predefined Areas of Interests (AOIs) and grouped into cognitively meaningful clusters (e.g., primary instruments, controls, external view). Hidden Markov Models (HMMs), augmented by fixation duration encoding, are used to establish normative, phase-specific scanning sequences. Deviations from these standards are detected in real time using joint HMM frameworks, with both event-based and timestamp-based anomaly detection, drawing on both weak and expert-provided labels. Dynamic thresholds for OOB (Out-of-Bounds) and UNK (Unknown) gaze states further support context-sensitive supervision.

4.3.2.4. Adaptive HMI intervention module

This final module constitutes the system's real-time adaptive interface layer. Based on the continuous inference of SD and LoAC states from the previous modules, the AR-HUD dynamically presents intervention cues—such as spatial orientation reminders, scanning guidance, and critical area highlighting—to assist pilots in regaining or maintaining optimal awareness and attention. The interface content, modality, and urgency are contextually modulated according to both the phase of flight and the assessed risk level, ensuring timely and minimally intrusive support.

In summary, the 3A-PAI system operationalises a closed-loop, adaptive, and multimodal approach to pilot state monitoring and support. Its architecture bridges cutting-edge neuro-ergonomics, ML, and HMI design, providing a scalable foundation for next-generation intelligent cockpit environments.

4.4. Experiment Methodology and Dataset Construction

4.4.1. Participants

A total of twenty-four pilots (First officer = 5, cadet pilot = 19) participated in this study. They were recruited based on their prior flight training experience, specifically with Airbus A320 simulators. Their average flying hours were 318.083 ± 237.109. Participants were informed of experimental details and provided written consent prior to involvement. Ethical approval was obtained from the Institutional Review Board of Hong Kong Polytechnic University (Ref: HSEARS20210318002).

4.4.2. Apparatus and Experimental Settings

This study utilised a fixed-base CockpitSonic Airbus A320 Flight Training Device (FTD), designed to replicate authentic flight operations. Realistic environmental visuals were rendered by Prepar3D software (version 5.3), augmented with real-time and historical weather data simulated by Active Sky software (version 5).

Neurophysiological responses were monitored using an EMOTIV Flex wireless EEG headset with 32 saline-based electrodes covering key cortical regions (prefrontal, temporal, parietal, occipital lobes), sampling at 128 Hz. Figure 4-2 (a) illustrated the configuration of electrodes and corresponding cortical regions. Visual engagement and attentional dynamics were tracked via Tobii Pro Glasses 3 eye-tracking glasses, sampling at 256 Hz. AOIs were defined around critical cockpit instruments including Primary Flight Display (PFD), Navigation Display (ND), Electronic Centralised Aircraft Monitoring (ECAM), Integrated Standby Instrument System (ISIS), and external views. Two additional categories were defined: "Out of Bounds (OOB)" for fixation points outside the cockpit's forward view (indicative of head turns) and

"Unknown (UNK)" for fixations within the forward view but not falling within specified AOIs. In addition, these separate AOIs were consolidated into meaningful groups to reflect typical cockpit scanning patterns. The detailed AOI groups configuration was shown as Figure 4-2 (b).

- **PF (Pilot Flying):** PFD_PF, ND_PF
- **PM (Pilot Monitoring):** PFD_PM, ND_PM
- **Central Panel:** ECAM, ISIS, LDG
- **Control:** ECAM_Control, FCU, Overhead_Panel, Flight Plan
- **External View:** External_View
- **UNK:** Unknown (areas within cockpit forward view but outside predefined AOIs)
- **OOB:** Areas out of bounds, such as left or right windshields.

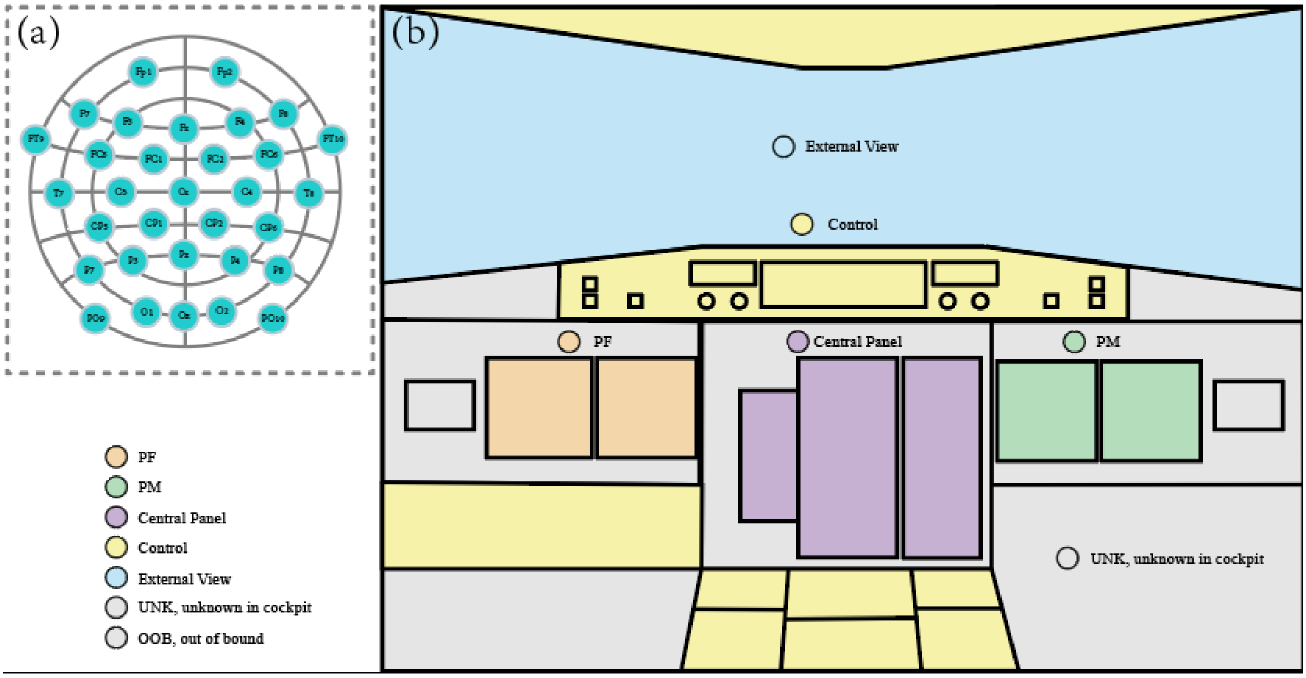


Figure 4-2. Experimental apparatus. (a) EEG configuration, (b) AOI groups configuration.

4.4.3. Experimental Procedure

Two main conditions were defined to induce SD: adverse weather scenarios and cockpit instrument inconsistencies. Adverse weather involved simulated conditions reflecting the real-world impact of Typhoon Koinu (08/10/2023), characterised by reduced visibility, strong crosswinds, and precipitation, creating significant spatial perception challenges for pilots (Gibb et al., 2011). Instrument inconsistency scenarios included simulated malfunctioning of airspeed indicators and altimeters due to pitot-static blockage, coupled with autopilot disengagement. These instrument failures critically disrupted the pilot's accurate spatial orientation and increased cognitive load significantly (Meeks et al., 2023; Newman & FAICD, 2007).

Figure 4-3 illustrates the overall experiment procedure. The flight routes were the round trip between the Hong Kong International Airport (ICAO code: VHHH) to Guangzhou Baiyun Airport (ICAO code: ZGGG), which took roughly 40 minutes for each trip. Experimental scenarios were structured across standard flight phases (take-off, climb, cruise, descend, and approach), enabling precise monitoring of SD episodes and attention management across distinct operational contexts. The weather conditions throughout the experiment simulated the real historical weather conditions during the period of Typhoon Koinu (08/10/2023) UTC 02:00 - 03:00. A total of four instrument failures were triggered in climb, cruise, and descend phases.

Throughout flight tasks, custom SD probes were administered at intervals of approximately 3 minutes via a tablet interface. Figure 4-3 shows the example of the probes. Each probe involved a multi-choice question about the current and future flight situation. Correctness to these probes served as critical reference points for annotating instances of SD. Subjects are tagged with a time marker at the beginning and at the end of answering the probes. After analysing the correctness of the subject's answers and the associated actions, the time markers were replaced by manual labels such as SD-0, SD-I, SD-II, and SD-III.

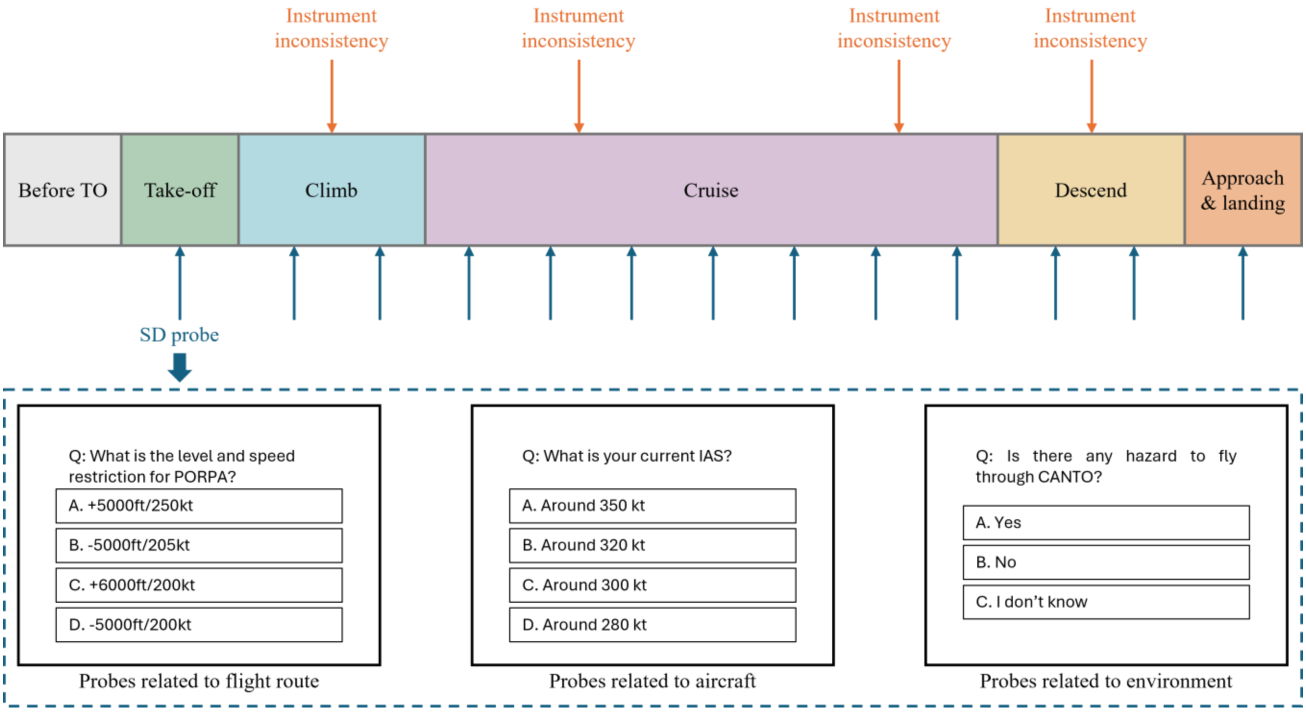


Figure 4-3. Experiment procedure and samples of SD probes.

4.4.4. Dataset Construction and Annotation

Two types of neurophysiological-based data, EEG and eye tracking data were extracted using the synchronous markers with SD probes. Table 4-1 presented the key metrics extracted from neuro-physiological data.

Table 4-1. Key metrics extracted from neuro-physiological data.

Features	Description
Power spectrum density (PSD)	The distribution of average power of a signal in a frequency in a brain region.

Average * / Max. / Min / Std. / Sum. duration of fixation	Describes statistical measures (average, maximum, minimum, standard deviation, or sum) of the time spent fixating on a single AOI.
Duration of fixation slope	The rate of change in fixation duration over time, indicating trends in visual attention.
Fixation count	The total number of fixations recorded, reflecting the intensity of visual exploration.
Gaze point X/Y *	The horizontal (X) and vertical (Y) coordinates of the gaze point on the visual display, capturing spatial attention.
Gaze path length *	The cumulative distance travelled by the gaze point, representing visual scanning behaviour.
AOI Hit	The count of times the gaze enters specific AOI, indicating visual focus on predefined regions.
AOI entropy	A measure of randomness or variability in gaze distribution across different AOIs, reflecting attentional dispersion.
Probability of AOI	The likelihood of gaze being directed to specific AOIs, quantifying attentional preference.

* represents the calculated metrics including t-1 and t-2 of the adjacent sliding window, delta, average, and range of the three adjacent sliding window)

4.4.4.1. Data Preprocessing Workflow

The EEG signals were processed using EEGLAB (v2025.0.0) in MATLAB (R2024b). Preprocessing included bandpass filtering (1-45 Hz), independent component analysis (ICA) for artifact removal, channel interpolation, and referencing to common averages. EEG data segments corresponding to annotated SD types and normal operations were segmented into epochs of 2 seconds to standardise temporal analysis intervals. Subsequently, spectral features (alpha, beta, theta, gamma power) were extracted using Fast Fourier Transform (FFT). The function of SD probe is to help manual labelling the neuro-physiological data as the ground truth. Merat et al. (2014) and Q. Li et al. (2024) suggested that the lead time could be up to around 20 s to achieve the better performance of recognition and prediction of SA. Similarly, in order to extract data from subjects under different SD types, EEG data from the first 20 s of subjects' responses to the SD-

Probe were extracted for subsequent studies.

Eye-tracking data underwent rigorous preprocessing through Tobii Pro Lab (version 24.21.435) software. In the extraction of phase-specific scanning sequences, only fixation events were included, while saccades were excluded from further analysis. Because saccades, as rapid eye movements between fixation points, are typically associated with transitions rather than information acquisition. The inclusion of saccade events may introduce noise and temporal discontinuity into the modelling of sequential gaze patterns, particularly when using HMMs to characterise standard scanning behaviours. Fixation and gaze data were segmented by AOI, gaze path length, fixation durations, AOI-based entropy, and probability of AOI calculated by HMM model were computed to characterise attentional dynamics comprehensively. Unlike the EEG, this study wanted to investigate the scanning behaviour of the pilot throughout the entire flight phase, therefore, the eye tracking data from the time of doing the "Before Take-off" checklist to the time of successful landing were kept intact and segmented according to the different flight phases.

4.4.4.2. Manual Annotation Procedure

Manual annotation involved expert analysis of flight performance, probe response accuracy, and observed cockpit scanning behaviour. Experts classified SD instances into three categories based on pilot responses and behavioural performance: Type I (unrecognised SD), Type II (recognised but managed SD), and Type III (recognised but incapacitating SD), consistent with prior definitions (Figure 4-4). Scanning anomalies indicative of LoAC were similarly annotated based on deviations from phase-specific scanning sequences. SD probes also served as the ground truth of the scanning behaviour. The eye movement data from 20 seconds before the SD probe appeared to the end of the answer were extracted. The criterion for labelling scanning series (SS) is determined by Eq 3. If the pilot answered correctly and can execute the SoP step by step, it was marked as normal scanning behaviour. Otherwise, corresponding eye movement data segments were marked as abnormal scanning series. Figure 4-4 also illustrates the process of manual annotation of SD types.

$$SS_s^n = \begin{cases} 1, & abnormal \\ 0, & normal \end{cases} \tag{3}$$

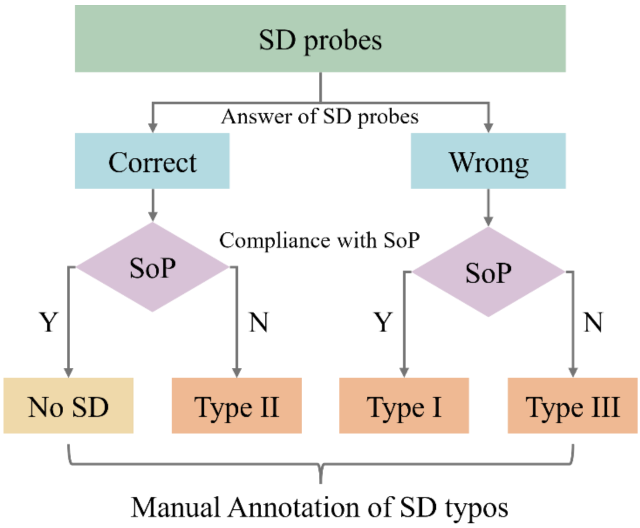


Figure 4-4.The criteria of manual annotation of four SD typologies

4.5. Recognition and Prediction of Spatial Disorientation

The overall workflow of SD types recognition and prediction module is shown in Figure 4-5.

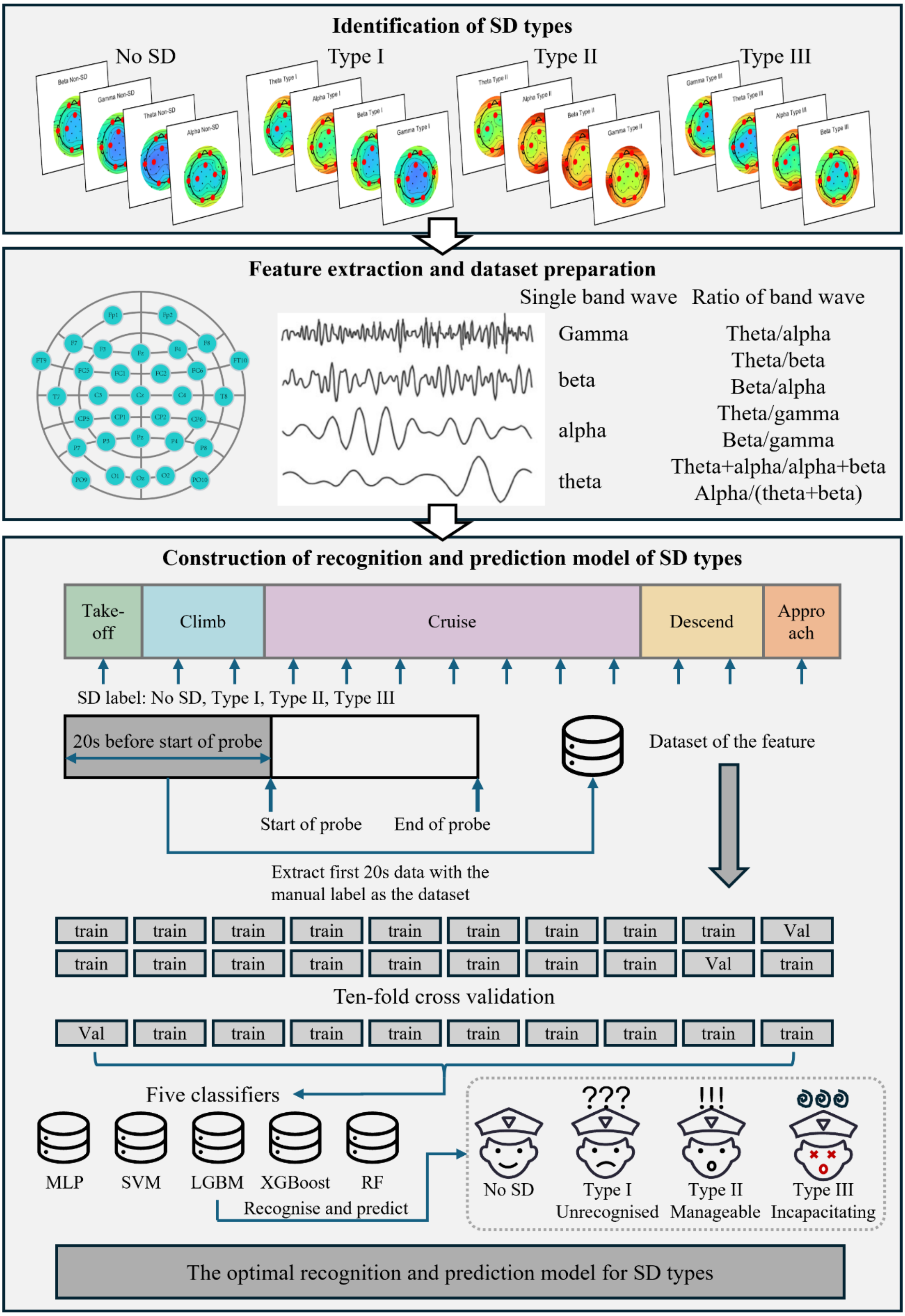


Figure 4-5. The overall workflow of SD types recognition and prediction module

4.5.1. Feature Extraction and Dataset Preparation

Neurophysiological data derived from EEG signals were processed to extract features relevant for recognising and predicting SD. As described in section 4.4.4, we extracted data from the 20 seconds before the subjects answered the SD probe and manually labelled the data based on the feedback. Specifically, the EEG features included power spectral densities (PSDs) from alpha, beta, theta, and gamma frequency bands across 32 EEG channels. Additionally, frequency band ratios were calculated to enhance discriminative capability, including theta/alpha, theta/beta, beta/alpha, theta/gamma, beta/gamma, (theta+alpha)/(alpha+beta), and alpha/(theta+beta). The description of the four frequency bands and their ratio was shown in Table 4-2.

Table 4-2. Description of the four frequency bands and their ratio.

No.	Band waves or ratio between the band waves	Corresponding human factor issues	Application scenario
1	Alpha	Relaxation, drowsiness, reduced cognitive workload	Monitoring relaxation states during breaks, detecting drowsiness in operators (e.g., pilots, drivers).
2	Beta	High alertness, increased cognitive workload, stress.	Evaluating mental workload in complex tasks, monitoring stress levels in high-demand environments (e.g., air traffic control).
3	Theta	Fatigue, vigilance, mind-wandering	Fatigue detection in driving/aviation, vigilance monitoring in monotonous tasks.
4	Gamma	High-level cognitive processing, focus, memory encoding.	Measuring focus and cognitive engagement in learning or problem-solving scenarios.
5	Theta/alpha	Fatigue, vigilance	Fatigue detection in driving/aviation
6	Theta/beta	Fatigue, loss of attention	Continuous task attention

			monitoring
7	Beta/alpha	Vigilance, cognitive activation	Task complexity and workload assessment
8	Theta/gamma	Working memory, cognitive process	Working memory load in complex tasks
9	Beta/gamma	Executive function, higher level of cognition	Complex operation and task switching
10	Theta+alpha/alpha+beta	Fatigue, vigilance index	Fatigue and vigilance assessment
11	Alpha/(theta+beta)	Cognitive efficiency, stress, workload	Cognitive mastery and stress state adaptation

The combined dataset from all participants consisted of labelled epochs classified into four SD categories: SD 0 (No SD), Type I (unrecognised SD), Type II (recognised but manageable SD), and Type III (recognised but incapacitating SD). Labels were assigned based on participants' performance on spatial orientation probes administered 20 seconds after EEG data acquisition, evaluated against standard operating procedures (SOPs). The class distribution within the final dataset was: SD0 (2937 samples), SD1 (1583 samples), SD2 (1204 samples), and SD3 (404 samples).

4.5.2. Classification Model Development

Five distinct classifiers—Random Forest, Support Vector Machine (SVM), Multilayer Perceptron (MLP), XGBoost, and LightGBM—were compared in their effectiveness to classify these SD types. The dataset was stratified and split into training (70%) and testing (30%) subsets. Ten-fold cross-validation was performed on the training set to ensure robust performance estimation and hyperparameter tuning. Performance metrics, as shown in Table 4-3, included accuracy, precision, recall, and F1-score. Among these classifiers, LightGBM and XGBoost outperformed others, achieving the highest test accuracy (both over 90%). Specifically, the superior performance of LightGBM can be attributed to its ability to handle large-scale datasets efficiently and its gradient boosting framework, which optimally combines weak learners. Additionally, its built-in feature importance mechanism enhances interpretability and ensures effective utilisation of relevant features in the classification task (Ke et al., 2017). Given these results, LightGBM was selected as the primary classification model for subsequent analysis.

Table 4-3. Performance of the compared models for classifying four distinct SD typologies.

	LightGBM	XGBoost	Random Forest	MLP	SVM
Accuracy	0.968	0.933	0.883	0.841	0.704
Precision*	0.972	0.964	0.914	0.868	0.692
Recall*	0.914	0.910	0.874	0.846	0.738
F1 score*	0.970	0.934	0.883	0.856	0.704

* represents the average value of the four categories (SD-0, Type I, Type II, and Type III).

4.5.3. Application and Further Prediction with LightGBM

The trained LightGBM model was applied to a new dataset derived from entire experimental sessions (not limited to epochs preceding probes), simulating real-time SD recognition scenarios. The model predictions were utilised to generate continuous SD-type classifications across experimental epochs. The distribution of predicted SD types and the probability statistics for each category over the entire experimental dataset was quantified as Table 4-4. The Type I SD accounts for the largest proportion of all SD types, which is correlated with the previous study that unrealised SD is the most common of disorientation. Additionally, the prediction confidence (maximum predicted probability) for each epoch was assessed. High confidence levels (mean confidence exceeding 0.9 across epochs) indicated reliable and stable model predictions. Epochs with lower prediction confidence (<0.6) were identified as 0.267, representing potential uncertainty intervals where further human verification might be required.

Table 4-4. Counts and probability distribution statistics of four SD typologies.

	Type I	Type II	Type III
Count	27883	4716	334
	Prob. of Type I	Prob. of Type II	Prob. of Type III
Mean	0.6981	0.1487	0.0295
Std	0.3260	0.2038	0.0638
Min	0.0126	0.0000	0.0000
Max	0.9999	0.9785	0.8658

The predictive results indicated that the LightGBM model consistently achieved high accuracy and reliability. Detailed numerical evaluations of prediction probabilities and confidence metrics were prepared for use in adaptive HMI interfaces, facilitating proactive intervention strategies in aviation operational contexts.

4.6. Phase-Specific Scanning Series Mining

To model pilot scanning behaviour effectively, a method that captures the sequential dependencies and spatial relationships between eye movements is essential. Traditional ML models, such as SVMs or Random Forests, typically operate on a fixed-size feature vector, treating each data point independently. This approach fails to account for the crucial temporal and transitional information inherent in a pilot's scanning path. A pilot's gaze is not a series of isolated events but a continuous, structured sequence moving between different instruments. Graph-based models are therefore the natural choice, as they represent the cockpit displays as nodes and the transitions between them as edges (W. Yu et al., 2022). This representation allows for the direct modelling of the inter-instrument transition probabilities and the overall scanning structure, which are key indicators of a pilot's cognitive state and SA.

Since the classifiers based on ML are ill-suited for the intricate task of sequence-level anomaly detection in a time series, HMM and CRFs are specifically designed for probabilistic modelling of sequential data. The combination of HMM for generating the standard model and CRF for precise, context-aware anomaly detection provides a robust and powerful framework that traditional classifiers cannot match.

4.6.1. Eye Tracking Data Processing and Dataset Construction

4.6.1.1. Dataset preparation

The raw data were obtained from the Tobii Pro Glasses 3 and initially filtered to retain only fixation events, with saccades and instances labelled as "Eyes Not Found" or "Unclassified" excluded. This selection ensured a clear focus on deliberate attention processes. Each fixation was labelled with an AOI identifier (AOI_ID) and grouped by AOI categories (AOI_GROUP), as presented in Figure 4-2 (b). OOB and UNK are also included in AOI statistics because OOB and UNK within a reasonable range are also normal scanning behaviours.

In eye-tracking data from real-world or simulated flight, a non-negligible proportion of samples may be labelled as “UNK” (AOI group ID=-1), representing gaze points within the main field of view but not falling into any defined AOI. These UNK samples, if left untreated, may fragment event segmentation and bias downstream statistical analysis or sequence modelling (e.g., HMM). In order to enhance sequence clarity for subsequent analyses, we designed an interpolation strategy to reassign plausible AOI identities to certain UNK events, based on temporal and contextual continuity.

Fixations in the same AOI with adjacent timestamps are considered as a single event. Temporal continuity was established by creating sequences segmented into runs based on continuous fixations within the same AOI and event type. The gaze sample sequence is segmented into events using the joint constancy of multiple attributes:

$$event\ id = cumsum\left[\Delta(AOI\ Group) \vee \Delta(Type) \vee \Delta(dur) \vee \Delta(AOI\ ID)\right] \qquad (4)$$

For each event with $AOI\ Group\ ID = -1$, we check its temporal context:

- Case 1: Dwell duration matching with neighbouring events. If the preceding or following event belonged to valid AOI group, and has exactly the same dwell duration (dur_ms), then all UNK samples in the current event are reassigned to that AOI group:

$$If \begin{cases} AOI\ Group_{curr} = -1 \\ AOI\ Group_{prev/next} \in Valid\ AOI\ Group \Rightarrow assign\ AOI\ Group_{curr} := AOI\ Group_{prev/next} \\ dur_{curr} = dur_{prev/next} \end{cases}$$

(5)

- Case 2: Short-duration UNK fixations (under 500 milliseconds) that occurred between two identical AOI events were interpolated to the nearest valid AOI group(preceding or following, if available).
- Otherwise, the UNK event remains unassigned.

This interpolation strategy improves sequence continuity and reduces artificial fragmentation due to brief or contextually resolvable UNK events, ensuring more accurate statistical and probabilistic modelling in subsequent visual scanning analysis

4.6.1.2. Determination of OOB and UNK Thresholds

Following interpolation, statistical thresholds for OOB and UNK fixation ratios were established across different flight phases (e.g., climb, cruise, approach). The proportions of dwell time attributed to OOB and UNK fixations were calculated per participant and per flight phase. Thresholds were derived statistically using mean plus two standard deviations (mean+2SD) and the 95th percentile method, serving as criteria to detect anomalous attentional behaviours in later analyses. This study also invited two experts served as the caption position to fine-tune the OOB and UNK ratio according to their actual flight experience. These thresholds provided a robust quantitative basis for the identification and subsequent warning of potential attentional deviations during flight operations.

Table. Tuned threshold setting of OOB and UNK in different flight phases.

Phase	Metric	Thr_mean2sd	Thr_p95	Thr_tuned
Takeoff	OOB_ratio	0	0	0
	UNK_ratio	0.08	0.07	0.075
Climb	OOB_ratio	0.13	0.14	0.15
	UNK_ratio	0.16	0.14	0.15
Cruise	OOB_ratio	0.12	0.11	0.15
	UNK_ratio	0.14	0.13	0.15
Descend	OOB_ratio	0.07	0.06	0.06
	UNK_ratio	0.09	0.09	0.10
Approach	OOB_ratio	0.11	0.10	0.10
	UNK_ratio	0.06	0.06	0.06

Collectively, these preprocessing procedures generated a clean and well-structured eye-tracking dataset, ready for phase-specific standard scanning sequence modelling and anomaly detection in subsequent analytical steps.

4.6.2. Construction of Standard Scanning Strategies based on Hidden Markov Model

In this phase, standard scanning sequences were constructed using time-expanded HMM based on the AOI groups.

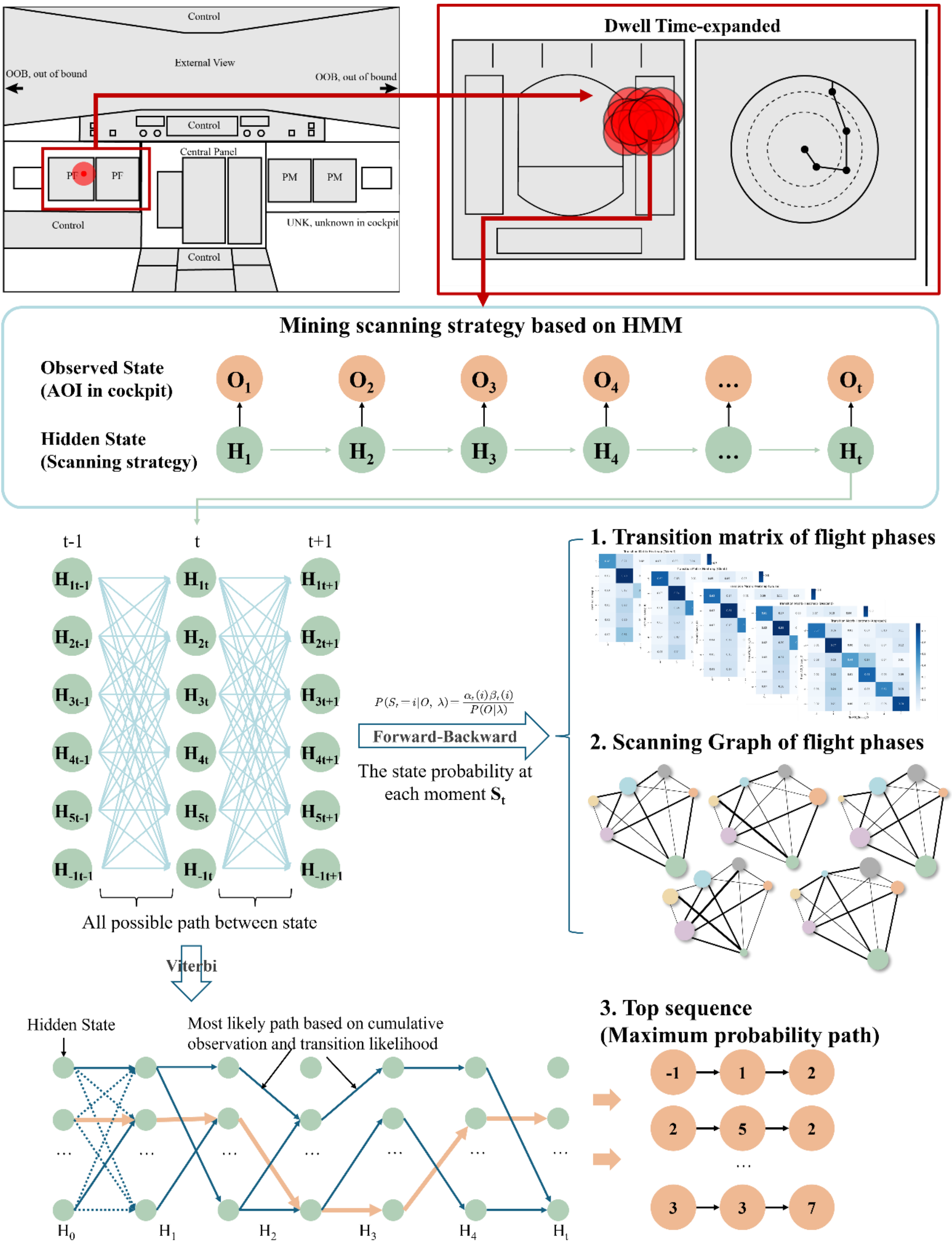


Figure 4-6. The construction of the time expanded HMM for mining standard scanning

patterns.

4.6.2.1. Development of Time-expanded HMM Model

Recognising that fixation duration conveys critical attentional information, fixation duration times were explicitly integrated into the HMM framework. Traditional HMM methods have limitations to capture the true dynamic structure of visual scanning; hence, an innovative "time-expanded" method was implemented, wherein each observation not only denotes a spatial AOI but is temporally expanded to reflect the actual dwell duration, thereby preserving the time-resolved nature of pilot scanning behaviour. Each fixation was discretised into observation frames of fixed duration (100 ms intervals), allowing fixation duration to be implicitly encoded within the expanded observation sequences. We define a set of hidden states corresponding to each AOI group, with emission distributions strongly favouring their corresponding AOI group observations (via Dirichlet or Laplace smoothing for numerical stability):

$$P(O_t = a_j | S_t = s_i) = \begin{cases} 1 - \varepsilon \cdot (N-1), & i = j \\ \varepsilon, & i \neq j \end{cases} \tag{6}$$

where N is the number of AOI groups and ε is a small smoothing constant.

The initial AOI-to-AOI group transition matrix T is derived from the previously computed standard scanning transition probability matrix P, after removing OOB. Transition probabilities between states were smoothed using a Dirichlet prior to stabilise the model parameters and avoid sparsity. For AOI groups $i,j \in A$:

$$T_{ij} = \frac{Count(i \to j) + \varepsilon}{\sum_k Count(i \to n) + N_\varepsilon} \tag{7}$$

The HMM was then trained using the time-expanded AOI observation sequences via the Viterbi (maximum likelihood) algorithm, with a prior (edge inertia) to balance between data-driven transition updates and the prior matrix:

$$Objective = \lambda \cdot \log P_{prior}(S_{1:T})\,(1-\lambda) \cdot \log P_{data}(S_{1:T}|O_{1:T}) \tag{8}$$

where λ is controlled by the edge inertia parameter ($\lambda = 0.5$ in this study). Each AOI group served as discrete states.

4.6.2.2. Transition Probability Matrices and Standard Scanning Strategies

Initially, AOI group transitions during fixation events were analysed to determine the standard scanning patterns for each distinct flight phase. For each flight phase, we aggregated all available eye movement event as pilot runs. Only fixation events on meaningful AOI groups were retained, excluding OOB, so as to focus on purposeful visual scanning relevant to cockpit operation. For each run, the AOI group ID sequence was extracted as

$$\begin{gathered}[(a_1,d_1),\ (a_2,d_2),\ \ldots,\ (a_n,d_n)],\ a_i \in A \\ S = [s_1, s_2, \ldots,\ s_n],\ s_i \in\ A\end{gathered} \tag{9}$$

Where A denotes the set of AOI groups in the cockpit for that phase.

For each phase, we built an empirical transition count matrix C, where element C_{ij} represents the count of transitions from AOI group i to AOI group j. To avoid zero probabilities and ensure stability, Dirichlet smoothing with parameter α was applied:

$$C_{ij} = Count(s_t = i,\ s_{t+1} = j) + \alpha \tag{10}$$

The transition probability matrix P is then given by row-wise normalisation:

$$P_{ij} = \frac{Count(i \rightarrow j) + \alpha}{\sum_k [Count(i \rightarrow k) + \alpha]} \tag{11}$$

Where P_{ij} is the estimated probability of transitioning from AOI group i to AOI group j.

For each AOI group, we computed both the number of visits and the total dwell time, yielding the following metrics:

Visit ratio:

$$Visit\_pct_i = \frac{Number\ of\ visits\ to\ i}{Total\ visits} \tag{12}$$

Dwell time ratio:

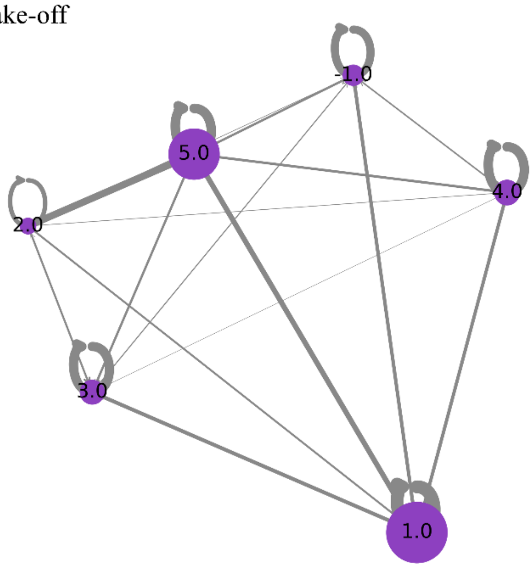
$$Dwell\_pct_i = \frac{Total\ dwell\ time\ on\ i}{Total\ dwell\ time} \tag{13}$$

Using the PrefixSpan algorithm, we extracted frequent scanning subsequences of length three that appeared above a minimum support threshold.

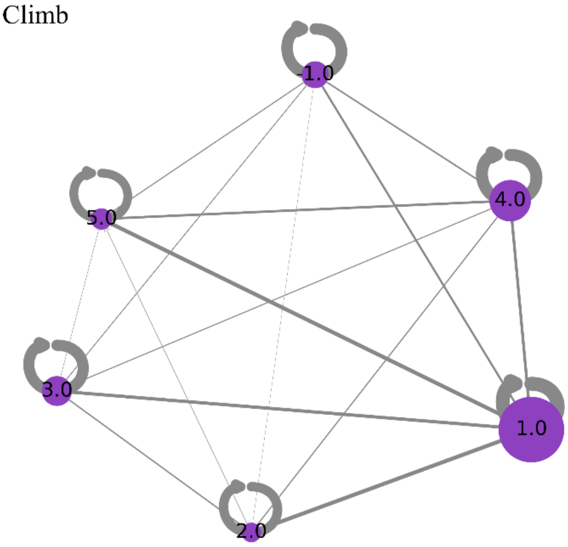
$$S_{freq} = \left\{ s \middle| support(s) \geq \min\_sup,\ s = 3 \right\} \tag{14}$$

These represent **standard scan series** reflecting typical visual routines in each phase. These provided quantitative descriptions of the visual scanning behaviour, identifying the most common (top-sequence) scanning patterns. Visualisations of these transitions were generated as scan graphs (Figure 4-7), where nodes represent AOIs and edge thickness encodes transition probability.

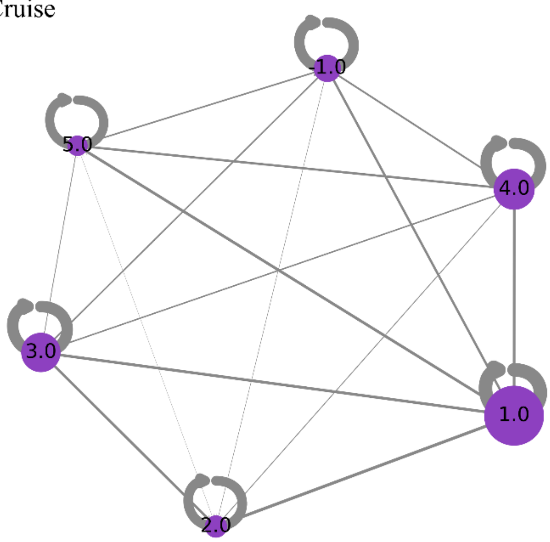
Take-off



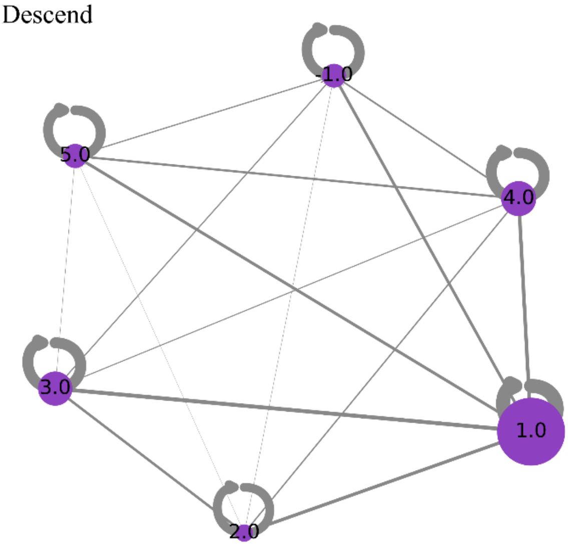
Climb



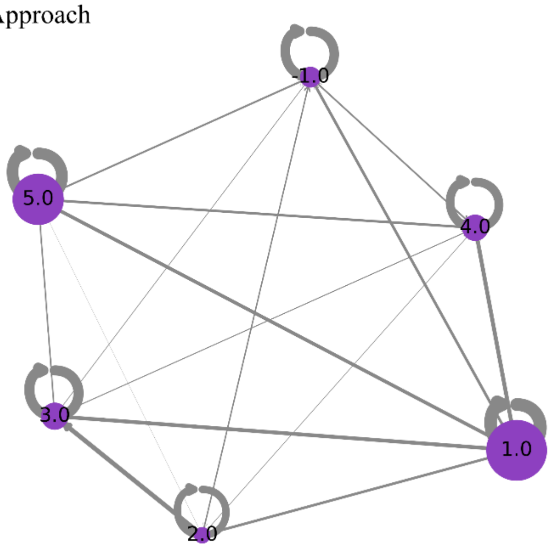
Cruise



Descend



Approach



-1.0: UNK

1.0: PF Area (Pilot Flying): PFD_PF, ND_PF

2.0: PM Area (Pilot Monitoring): PFD_PM, ND_PM

3.0: Central Panel: ECAM, ISIS, LDG

4.0: Control: ECAM Control, FCU, Overhead Panel, Flight Plan

5.0: External View: External View

Figure 4-7. Scanning graph of the transition matrix in different flight phases between AOI groups.

4.6.2.3. Abnormal Scanning Sequence Detection via HMM Log-Likelihood

The Log-Likelihood per frame (LLH/frame) for each sequence was computed via time-expanded HMM, providing an objective metric of sequence conformity to the established scanning patterns. A baseline LLH distribution was created using sequences labelled as "normal" or representative. Because an HMM marginalises over all possible hidden state sequences, direct computation of the observation likelihood $P(\mathbf{O}|\lambda)$ is intractable for long sequences. Instead, the forward algorithm provides an efficient dynamic programming solution.

Let:

- $O=(o_1,o_2,...,o_T)$: observation sequence (time-expanded AOI group series)
- N: number of hidden states (AOI groups)
- $\pi(i)$: initial probability of state i
- a_{ij}: transition probability from state i to j
- $b_j(o_t)$: emission probability of observing o_t in state j
- $\lambda=(\pi,\ A,\ B)$: the set of HMM parameters

The forward algorithm provides an efficient dynamic programming solution. Following is the steps for calculating LLH threshold.

1) Initialisation $(t=1)$:

$$\alpha_1(i)=\pi(i)\cdot b_i(o_1),\ \forall i\in\{1,...,\ N\} \tag{15}$$

2) Recursion $(t=2,...,\ T)$:

$$\alpha_t(j) = \left[\sum_{i=1}^{N} \alpha_{t-1}(i) \cdot a_{ij} \right] \cdot b_j(o_t), \ \forall j \in \{1, ..., N\} \tag{16}$$

3) Termination:

$$P(\mathbf{O}|\lambda) = \sum_{j=1}^{N} \alpha_T(j) \tag{17}$$

4) Log-likelihood:

$$\log - likelihood = \log(P(\mathbf{O}|\lambda)) \tag{18}$$

5) Normalised Log-Likelihood: Given that the sequence length T may vary, this study used normalised log-likelihood:

$$LLH_{norm}(\mathbf{O}) = \frac{1}{T} \log P(\mathbf{O}|\lambda) \tag{19}$$

For each phase, a threshold τ_{LLH} is determined using the 5th percentile of LLH_{norm} among standard runs. Any sequence with:

$$LLH_{norm}(\mathbf{O}) < \tau_{LLH} \tag{20}$$

is marked as abnormal.

Finally, the computed LLH thresholds were combined with previously established thresholds for OOB and UNK fixation ratios. The final abnormality flag is:

$$Abnormal(\mathbf{O}) = \begin{cases} True, \ if \ r_{OOB} > \tau_{OOB} \\ True, \ if \ r_{UNK} > \tau_{UNK} \\ True, \ if \ LLH_{norm}(\mathbf{O}) < \tau_{LLH} \\ False, \ otherwise \end{cases} \tag{21}$$

By leveraging the forward algorithm to compute the sequence log-likelihood under a phase-specific HMM, we can sensitively identify scanning sequences that, while not necessarily OOB or UNK-dominated, deviate from standard temporal–spatial attention dynamics. This combination created a comprehensive, three-fold rule-based alert criterion. Scanning behaviour that deviated significantly from standard patterns or exceeded established LLH, OOB, or UNK thresholds was marked as anomalous, triggering attention-control alerts during real-time monitoring.

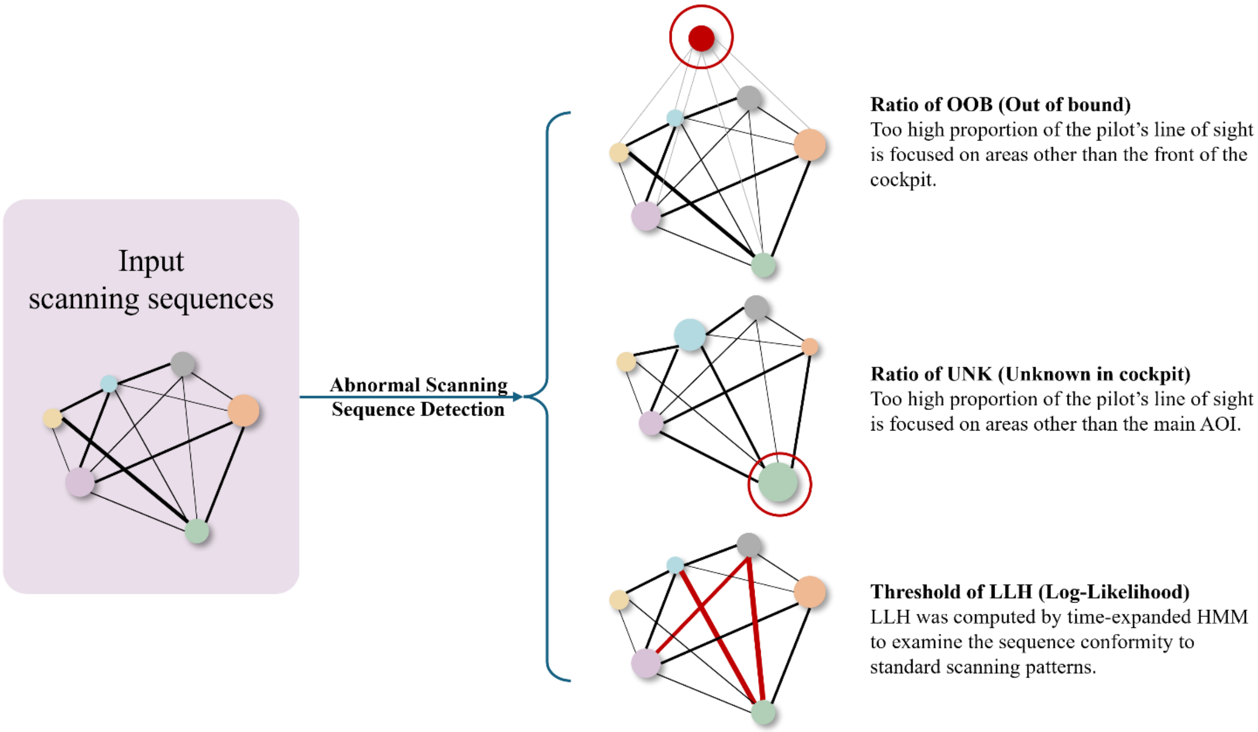


Figure 4-8. Three rules of detecting the abnormal scanning sequence.

To evaluate the robustness and reliability of the trained time-expanded HMMs, differences between transition matrices derived from observed sequences and those predicted by time-expanded HMMs were analysed quantitatively. Discrepancies were visualised using difference heatmaps, clearly illustrating areas of divergence between observed scanning behaviours and modelled standard patterns. This allowed for precise identification of scanning deviations during specific phases.

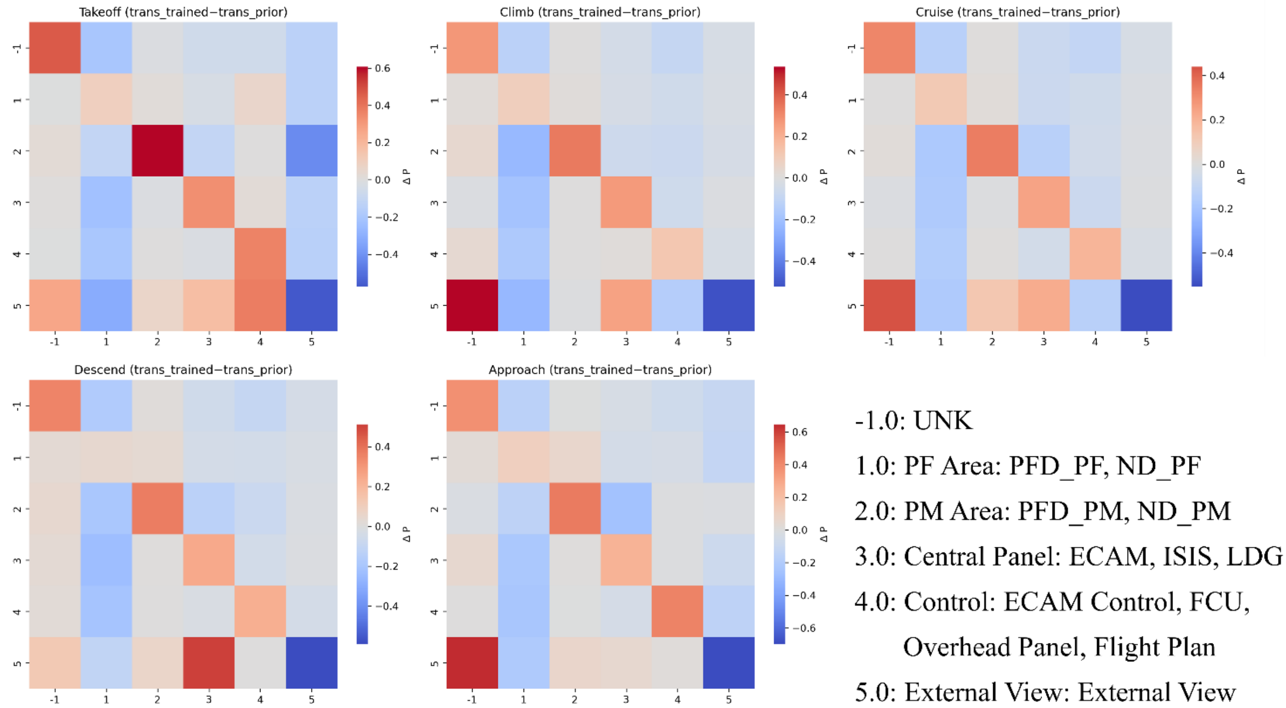


Figure 4-9. The difference of the transition matrix between observed sequences and those predicted by HMMs.

This time-expansion HMM modelling scheme allows the learned model to reflect both where and how long the pilot allocates visual attention, enabling finer-grained supervision, benchmarking, and adaptive assistance for safety-critical cockpit scenarios. It also provides a statistically robust foundation for phase-specific scanning pattern recognition and abnormality detection in subsequent stages.

4.6.3. Timestamp-based HMM-CRF Joint Training for Scanning Anomaly Prediction

HMMs possess an inherent advantage in modelling hidden, latent states and their transitions within sequential data. By first employing an HMM to model the underlying cognitive or operational states of pilot eye-movement behaviour, we can obtain probability distributions or sequences pertaining to these states. The outputs of the

HMM can then serve as powerful, temporally contextualised features, which are fed into the CRF model. The CRF is subsequently able to effectively integrate this higher-order sequential information provided by the HMM with our meticulously designed eye-movement features, which encompass rich context detail and interdependencies. In light of the respective strengths and limitations of HMM and CRF, this study adopts a strategy of HMM-CRF joint training. Figure 4-10 illustrates the overall workflow of the HMM-CRF joint training framework.

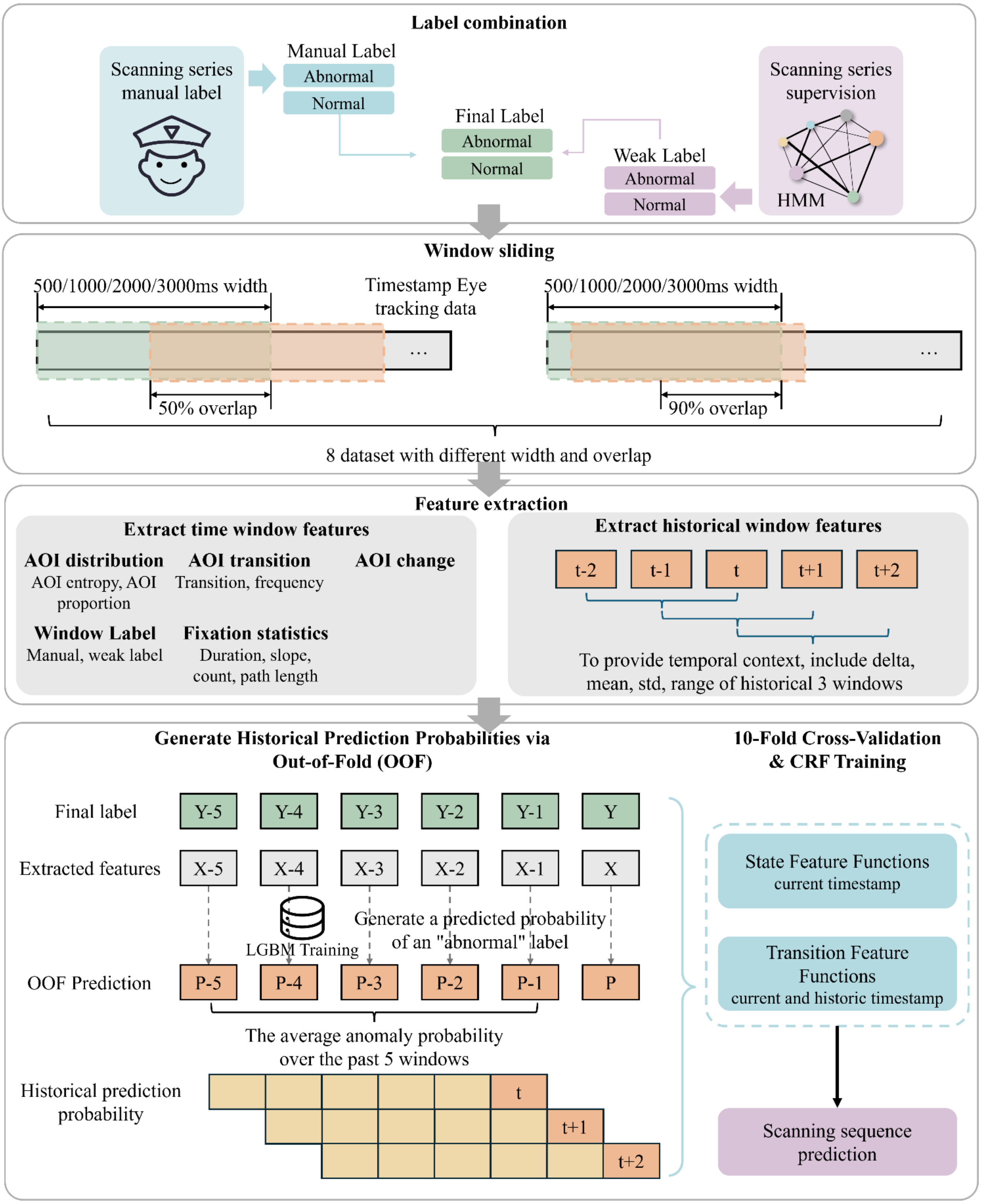


Figure 4-10. Workflow of timestamp-based HMM-CRF joint training framework.

4.6.3.1. Dataset preparation for CRF training

Weak and manual labelling integration strategy. By generating weak labels through HMM, this study can achieve the goal of using fewer manual labels to fine-tune the dataset that requires a fully supervised CRF model. First, weak labels were assigned to all fixation events using the previously trained HMMs, which defines tri-level abnormality detection for each three scanning sequences. Meanwhile, manual expert labels were transferred from the SD probe annotations, where available, to the corresponding events. A unified event-based dataset was then constructed with both weak and manual labels for each run. For time intervals where manual labels were missing, the weak label was adopted, yielding a comprehensive, high-coverage ground truth dataset.

Feature Extraction from Timestamped Data. Raw timestamp-level fixation samples, each labelled with either manual or weak labels, were segmented into overlapping sliding windows. This ablation study systematically varies two key window parameters: window width and overlap percentage, providing a comprehensive analysis of their individual and combined effects on model performance. Five different window widths were evaluated: 500ms, 1000ms, 2000ms, 3000ms and 5000ms. This range was chosen to explore the impact of short-range and gradually long-range contextual information on the performance of the HMM-CRF joint model. For each selected window width, two overlap percentages were rigorously tested: 50% and 90%. Overlap is a critical parameter for ensuring data continuity and preventing information loss at the window edges, particularly when applying various windowing functions during signal processing. It ensures that temporal data in the overlapping portion contribute to multiple calculations of adjacent windows, effectively mitigating the "zeroing" effect that can be introduced by windowing functions. A 50% overlap is generally considered the optimal minimum value for certain windowing functions, such as the Hanning window, as it ensures complete data capture without significant gaps. Conversely, a 90% overlap, while increasing the computational burden due to redundant processing, provides a higher degree of data redundancy and may produce a "smoothing effect" that can help identify subtle trends or improve overall signal representation. Overall, this design resulted in a total of ten unique experimental configurations (four window widths × two overlap percentages):

- 500ms width with 50% and 90% overlap,
- 1000ms width with 50% and 90% overlap,
- 2000ms width with 50% and 90% overlap,
- 3000ms width with 50% and 90% overlap.
- 5000ms width with 50% and 90% overlap

For each window, a rich set of statistical and trajectory features was computed as Table 4-5. Detailed features include AOI distribution, entropy, fixation duration statistics, gaze path length, number of unique AOIs, transition counts, and AOI dominance changes.

Table 4-5. Summary of key features after sliding windows.

Feature	Description	Formula / Method
AOI entropy	Diversity of AOI distribution	$-\sum p_i \log_2 p_i$
Mean dwell	Mean dwell duration (ms)	$\bar{d} = \frac{1}{n}\sum d_j$
Dwell slope	Linear trend in dwell duration	Slope of least-squares fit

Gaze path length	Total path traversed by gaze	$L=\sum_{j=2}^{n}\sqrt{(g_{x,j}-g_{x,-1})^2+(g_{y,j}-g_{y,j-1})^2}$
AOI transitions	Number of AOI switches	$\sum_{j=2}^{n}\mathbb{I}(a_j\neq a_{j-1})$
Main AOI change	Flag if dominant AOI changes across windows	$\mathbb{I}(AOI_{main,\,t}\neq AOI_{main,t-1})$
Rolling stats	Mean, std, range over previous 3 windows	Rolling calculations

Due to the inherent imbalance in the distribution of "normal" and "abnormal" labels within the dataset, conventional assignment of the majority label to each time window would result in a highly skewed training set and potential model bias. To mitigate this, we implemented a phase-specific thresholding strategy for the assignment of abnormal window labels, tailored to the empirical distribution of abnormal samples in each flight phase. For each window, let R_{abn} denote the proportion of "abnormal" frames within the window. For each phase, a threshold τ_{phase} is empirically determined based on the distribution of abnormal samples in the training data. The window label is then assigned as:

$$win\_label=\begin{cases}\text{“abnormal”},\ if\ R_{abn}>\tau_{phase}\\ \text{“normal”},\ otherwise\end{cases}\tag{22}$$

4.6.3.2. Intermediate LightGBM-based Sequence Probability Modelling:

To provide an intermediate, data-driven assessment of abnormal scanning probability at the window level, we employed the LightGBM algorithm to model the relationship between windowed eye-tracking features and the abnormality label. The resulting out-of-fold probability predictions are further utilised to enhance the downstream CRF sequence modelling. To address this, a two-stage approach was developed.

First, a LightGBM classifier was trained using out-of-fold (OOF) group cross-validation to predict the probability of abnormality for each window. The LightGBM classifier is trained on $(X_{train},\ y_{train})$, where X are the window features and y is the binary abnormality flag (0=normal, 1=abnormal). OOF prediction probabilities $\widehat{p_i}$ are computed for test windows as:

$$\widehat{p_i} = P(y_i = 1\,|X_i;\theta_{LGBM}) \tag{23}$$

This process produces an unbiased, OOF-estimated probability for each window.

The resulting historical probability was used to generate high-safety sliding window features (e.g., the rolling mean of the past several windows' probabilities), ensuring strict prevention of data leakage and maintaining causal structure for subsequent CRF modelling. For each window w_t within a subject-phase sequence, the mean abnormality probability over the preceding five windows (excluding the current window) is computed as:

$$abn\_in\_win5_t = \frac{1}{n}\sum_{j=1}^{n} \widehat{p_t} - j \tag{24}$$

where $n = \min(5, t-1)$, and $\widehat{p}_{t-j}$ are the historical predicted probabilities, ensuring strict causality (no future data leakage).

4.6.3.3. CRF Model Training and Evaluation:

Features for each window include: 1) Statistical features from the timestamp window (Section 4.6.3.1), 2) Historical and rolling window features (section 4.6.3.2), and 3) LightGBM-derived probability feature (Section 4.6.3.2). To leverage both local and sequential dependencies in windowed eye-tracking data, we adopted the linear-chain CRF as the primary sequence model for abnormality detection. A CRF model was subsequently trained in a grouped ten-fold cross-validation manner, with extensive hyperparameter search to optimise sequence labelling accuracy.

The conditional probability of a label sequence given the observation series is modelled by the linear-chain CRF:

$$P(y|X) = \frac{1}{Z(X)} \exp\left(\sum_{t=1}^{T} \sum_{k} \lambda_k f_k(y_{t-1}, y_t,\ x_t)\right) \tag{25}$$

where f_k are feature functions (over current and previous labels, and window features), λ_k are learned weights, $Z(X)$ is the normalisation partition function. This model thus captures both state-observation compatibility (how well window features support a given label) and label transition dynamics (temporal label smoothness/structure).

For each combination of regularisation parameters$(c_1,\ c_2)$, a CRF is trained via the LBFGS algorithm. The parameters control the strength of L1 and L2 regularisation, respectively:

$$\mathcal{L}_{CRF} = -\log P(y|X) + c_1 \|\lambda\|_1 + c_2 \|\lambda\|_2^2 \tag{26}$$

Performance was evaluated in terms of accuracy, F1 (abnormal class), and ROC-AUC. The inclusion of LightGBM probability features markedly improved the CRF's sequence labelling performance, particularly for abnormal event detection. On the test set, the trained CRF predicts the most probable label sequence $\hat{y}$ using the Viterbi algorithm:

$$\hat{y} = arg\max_{y} P(y|X) \tag{27}$$

Marginal probabilities for the “abnormal” class at each window t are also computed for further ROC/AUC analysis. Table 4-6 compared the four sliding window’s performance, and larger window size showed the best performance.

Table 4-6. The performance of the ablation experiment with different width and overlap proportion.

	Accuracy	F1(normal)	F1(abnormal)	ROC-AUC
500ms width with 50% overlap	0.83	0.87	0.74	0.89
500ms width with 90% overlap	0.83	0.87	0.73	0.88
1000ms width with 50% overlap	0.83	0.87	0.74	0.90
1000ms width with 90% overlap	0.82	0.87	0.73	0.86
2000ms width with 50% overlap	0.86	0.88	0.83	0.94
2000ms width with 90% overlap	0.85	0.87	0.83	0.93
3000ms width with 50% overlap	0.86	0.88	0.83	0.94
3000ms width with 90% overlap	0.86	0.88	0.83	0.94
5000ms width with 50% overlap	0.78	0.83	0.68	0.89
5000ms width with 90% overlap	0.78	0.83	0.67	0.83

Across the eight different datasets used in this ablation study, the combined HMM-CRF joint model consistently achieved top performance when configured with a 3000ms window width. This key finding is robust, regardless of the overlap percentage (50% or 90%), demonstrating the model's strong and dominant dependence on the temporal extent of the provided input context. This result directly points to the critical importance of the HMM-CRF model for broad context understanding for the specific tasks represented by these datasets. It strongly suggests that the patterns or dependencies required for accurate annotation extend beyond shorter temporal ranges, necessitating a wider observation window.

The empirical observation that the 3000ms window (the maximum width tested) consistently outperforms the HMM-CRF model strongly suggests that the underlying phenomena or patterns in the dataset require a relatively long temporal context to be accurately modelled by the HMM-CRF. This is highly consistent with the ability of CRF to capture high-order dependencies and model global relationships within a sequence. This characteristic implies that larger context windows better capture global patterns and long-range dependencies, which improves the model's capacity to produce predictions that are globally consistent and well-informed, eventually leading to an improvement in overall model performance.

While this design strengthens internal validity, the generalizability of the findings to pilots with substantially different total flight hours, aircraft backgrounds, or operational cultures warrants further investigation. It is important to note that the proposed framework is grounded in relative neurophysiological modulation patterns and phase-specific scanning dynamics rather than aircraft-specific control parameters or fixed performance thresholds. Such mechanism-oriented features are expected to exhibit structural consistency across experience levels, although their magnitude and expression may vary. Future research will therefore expand the dataset to include pilots with broader experience ranges and diverse aircraft types, enabling cross-cohort validation and robustness testing. In addition, adaptive modelling strategies—such as transfer learning, domain adaptation, and incremental fine-tuning—can be incorporated to facilitate model recalibration when applied to new pilot populations. Through these extensions, the framework can evolve from a controlled experimental system toward a more generalisable and operationally scalable human-AI collaboration architecture.

In operational cockpit environments, physiological and behavioural sensors are inevitably subject to noise, artefacts, or temporary signal loss. Ensuring robustness under such conditions is therefore critical for the practical deployment of the 3A-PAI framework. The system architecture mitigates single-point failure through multimodal redundancy, integrating neurophysiological features, visual scanning dynamics, and flight-phase context rather than relying on any single signal source. This design enables partial state estimation when one modality degrades, while preprocessing and quality-control procedures reduce the impact of transient noise. From a safety perspective, fail-safe mechanisms are incorporated. When signal reliability or model confidence falls below predefined thresholds, adaptive interventions are suspended and the AR-HUD reverts to a baseline, non-adaptive display mode. This degradation-to-safe strategy prevents misleading guidance and preserves human authority in safety-critical situations.

Overall, the final model delivered robust, temporally precise discrimination of attention-control anomalies across time, with high reliability as demonstrate ed in cross-validated results and confusion matrix analyses. This timestamp-based HMM-CRF hybrid modelling approach, with the LightGBM-CRF bridge, enables fine-grained, real-time detection of scanning anomalies, providing a critical foundation for adaptive pilot support and feedback in the 3A-PAI system.

4.7. Adaptive HMI to Mitigate SD through AR-HUD

To bridge the gap between theoretical prediction models and practical flight safety applications, we developed a real-time closed-loop prototype system. This system integrates the predictive algorithms established in the previous sections into a Perception-Decision-Action loop, designed to dynamically mitigate SD and LoAC through a personalised head-up display (HUD).

By presenting data without requiring users to divert their gaze from their normal viewpoint, a HUD simplifies the integration of instrument references and external visual cues through its transparent display technology (Nicholl, 2014). Augmented reality (AR) technology further expands the capabilities of HUDs by overlaying digital information onto the physical world. Augmented reality head-up displays (AR-HUDs) integrate data from various sources, such as navigation aids, waypoints, terrain awareness, and traffic data, and overlay it in real time onto the pilot's field of view (Yuan, Ng, Li, et al., 2024).

AR-HUDs have significant potential for mitigating SD by providing critical information and helping pilots correct perceived errors. Research suggests that task-specific eye scanning strategies are required for SD mitigation (Hua et al., 2022). Visual illusions require increased attention to instruments, while vestibular illusions benefit from HUD engagement. This suggests that AR-HUDs can adaptively adjust their display content and guidance based on the type of SD detected.

Based on the standard scanning strategies in different flight phases summarised via HMM, this study summarised the key information and transition strategies to guide the adaptive interface design. Table shows the detail of the design guidelines, which can also serve as the training strategies of scanning distribution in critical flight phases.

Table 4-7. Adaptive HUD design guidelines for four SD types in different flight phases.

	No SD	SD Type I (Unrecognised)	SD Type II (Manageable)	SD Type III (Incapacities)
Take-off	Minimalist display. Only highlight key information related to V-speeds and	When the pilot's gaze fixates on areas other than PF area or external view for an	When a parameter in PF area (e.g., airspeed) or external view is abnormal but	AR-HUD should overlay a full-screen attitude indicator on the external view,

	runway centreline alignment.	extended period, the AR-HUD should use a subtle sparkle or dynamic line to guide the gaze back to the core instruments (PF area and external view).	within a manageable range, the AR-HUD should highlight the parameter and use a green border or arrow to indicate its target value or acceptable range.	augment key parameters from PF and external view, and use concise text to warn of the danger.
Climb	Only display the most essential flight parameters, such as the Flight Path Vector, target heading, and altitude.	When the pilot's scanning pattern deviates from the PF to control panel or central panel cross-check, the AR-HUD should subtly fade in and out to highlight the next required AOI to redirect the scanning path.	When PF area (e.g., vertical speed) or control panel (e.g., ECAM screen) deviates from the preset value, the AR-HUD should project the correct value as a reference and use a stable cursor to indicate the target.	Providing an integrated Flight Path Vector that shows the correct climb attitude and heading in real time on the AR-HUD, while also locking the display of key parameters from PF area.
Cruise	Maintain a minimalist and non-intrusive display. Key parameters (e.g., altitude, heading) should only be shown on the AR-HUD as a secondary reference.	The core AOI is control panel (e.g., FCU, ECAM screen). If the scanning pattern deviates from the core control panel for too long, the AR-HUD's projection of control panel may briefly flash to gently prompt the pilot.	When control panel (e.g., FCU) shows a slight deviation, the AR-HUD should use augmented text to display the operation which should be taken.	In the event of a significant deviation in altitude or heading, the AR-HUD should dynamically lock the key data from control panel and PF area and prompt the pilot for correction.
Descend	Highlight critical information related to the approach path. The AR-HUD should project the planned descent path, while displaying target	When the pilot fails to perform the high-frequency PF area to central panel/control panel cross-check, the AR-HUD should highlight PF area (ND) or use an arrow	When PF area (e.g., descent rate) or control panel (e.g., FCU) is exceeded, the AR-HUD should highlight the corresponding instrument area,	When the pilot cannot effectively control the descent, the AR-HUD should provide mandatory path and attitude guidance. Key data from PF area and

	altitude and speed, enabling the pilot to seamlessly integrate external and internal data.	to guide the gaze from PF area to central panel or control panel.	provide corrective information (e.g., "increase thrust").	control panel are displayed in a locked manner with an emergency alert.
Approach	Dynamically overlay the runway centerline, glideslope, and localiser deviation on the external view. The AR-HUD should display the key information from PF and external view in a concise manner, ensuring effective integration of internal and external information.	When the pilot's gaze remains fixed on besides PF area or external view for an extended period without switching, the AR-HUD should use dynamic cues (e.g., an arrow tracking the pilot's gaze) to guide them to cross-check between the dual-core AOIs.	AR-HUD should overlay the correct status information on the corresponding location of external view, use a red highlight to indicate the incorrect configuration, and provide a short instruction (e.g., "Gear Down" according to the checklist of QRH manual).	AR-HUD should immediately enter a mandatory guidance mode to show the checklist item by item. It should provide full landing path guidance in the external view, integrate key data from PF area and external view.

This table is crucial for demonstrating the practical application of the 3A-PAI system. It directly links the detected SD types and indicators (from the previous sections) with the proposed AR-HUD interventions. This demonstrates the closed-loop nature of the system and how it provides a practical solution for SD mitigation, thus reinforcing the core value proposition of the user paper.

To evaluate the proposed system’s feasibility, we further developed this mitigation system prototype, as Figure 4-11 shows.

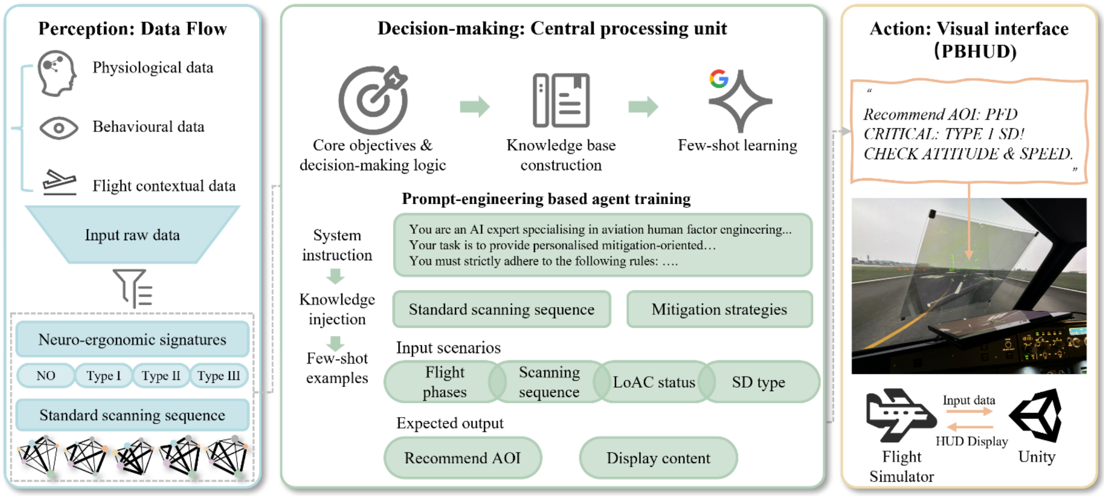


Figure 4-11. LLM-assisted mitigation system construction.

1) Perception: Data flow and acquisition

The perception module aggregates three distinct data streams to form a comprehensive understanding of the pilot-aircraft system:

- Physiological data: Real-time EEG signals captured to monitor cognitive states.
- Behavioural data: Eye-tracking coordinates used to assess scanning patterns.
- Contextual data: Flight phases, aircraft attitude parameters, and telemetry data.

These synchronised data streams serve as the input for the previously trained ML models (including the HMM-CRF-LightGBM framework). The processing core utilises these inputs to identify neuro-ergonomic signatures and compare real-time scanning behaviour against the standard scanning sequences derived for each flight phase.

2) Decision-making: Central processing unit

The decision module serves as the system's brain. Hosted on a central control PC, this processing core runs the computational models and the reasoning engine. It functions as an onboard "reasoning agent", continuously evaluating the fused neuro-ergonomic and contextual data to determine whether an intervention is warranted. We explored conventional LLM fine-tuning strategies to further enhance model robustness in aviation-specific contexts. Specifically, the model was trained on a larger set of high-quality expert-labelled input-output pairs using supervised fine-tuning to better capture the mapping from quantitative pilot-state signals to adaptive interface recommendations. We also incorporated iterative refinement based on expert evaluations of model outputs, promoting safer and more interpretable guidance. Together, the integrated Few-Shot-Learning (FSL) and fine-tuning workflow enables the generation of context-sensitive, real-time visual cues aligned with detected SD and LoAC states. Upon detecting a risk (e.g., SD Type I – III or high LoAC), the PC generates specific graphical or textual commands based on the predefined mitigation strategies.

3) Action: Visual interface

The action module executes the mitigation strategies through a custom Projection-Based HUD (PBHUD). While HUDs integrate navigation and traffic data to aid situational awareness (Yuan, Ng, Li, et al., 2024), research suggests that task-specific scanning guidance is essential for effective SD mitigation (Hua et al., 2022). The prototype included three parts: 1) PBHUD interface, 2) central control screen, and 3) central control PC. The visual interface was developed using Unity (version 2022.3.2f1). It connects to the FTD software (Prepar3D v5) via the SimConnect API to receive real-time telemetry (Yuan, Ng, Yiu, et al., 2025). The central PC acts as the hub for data transmission and reasoning. It processes the input streams and renders the Unity-based interface. The rendered interface is mirrored from the central control screen and projected onto the ARHUD reflective film.

Translating the 3A-PAI framework from an experimental prototype into certified aviation technology entails significant technical and regulatory considerations. On the technical side, integrating multimodal neuro-ergonomic sensing within operational cockpits requires aviation-grade hardware reliability, stable long-duration signal acquisition, deterministic real-time computation, and system redundancy. Regulatory hurdles are equally substantial. Adaptive AI-driven systems must demonstrate validation transparency, performance consistency, and explainability across diverse operational conditions. Certification authorities require traceable decision logic, well-defined human–automation authority allocation, and conservative fail-safe behaviour, particularly for systems influencing pilot cognition.

A realistic implementation pathway is therefore incremental. Initial application in simulator-based training or advisory-only cockpit modes would allow progressive validation and HFs assessment. With accumulated operational evidence, structured safety cases and confidence monitoring mechanisms could support gradual integration into certified environments. Such a staged approach aligns technological innovation with the conservative evolution required in aviation regulation.

4.8. Concluding remarks

This study aimed to address the critical challenges of SD and LoAC in aviation by developing and validating a multimodal, adaptive pilot support framework. The proposed Adaptive Attention-Awareness Pilot-Aircraft Interaction System (3A-PAI) successfully integrates real-time neurophysiological and behavioral data to monitor, recognise, and intervene against these pervasive threats to flight safety.

To address our research questions, this work delivered several key findings. First, we developed robust predictive models based on EEG signals to accurately classify and predict four distinct types of SD. By comparing five different classifiers—Random Forest, SVM, MLP, XGBoost, and LightGBM—we found that LightGBM achieved the optimal performance, with a remarkable 96.8% accuracy in four-class SD classification. This demonstrates the feasibility of using neurophysiological data for real-time, fine-grained SD recognition.

Second, to address the challenge of LoAC, we employed HMMs to mine and establish phase-specific normative visual scanning patterns. These models serve as a coarse-grained baseline for monitoring a pilot's overall scanning behaviour, enabling the detection of initial deviations from a standard attention-control strategy.

Finally, to achieve a more precise level of supervision, we developed a more advanced framework by combining a HMM, CRF with LightGBM. This hybrid approach, utilising a sliding window technique, enabled fine-grained, real-time prediction of scanning anomalies. This joint training methodology allows the model to simultaneously benefit from the HMM's generative understanding of latent states and temporal dependencies, as well as the CRF's robust capability in handling complex, overlapping features and performing precise discriminative prediction. Therefore, the joint model allows our system to provide a timely and accurate alert for LoAC by identifying subtle deviations in the pilot’s scanning behaviour that would otherwise be missed.

The contributions of the 3A-PAI system extend beyond methodological innovation to significant practical and conceptual implications for aviation safety and human-machine interaction. The development of a hybrid model for scanning anomaly detection and a hybrid labelling strategy represents a novel approach to training robust pilot-state models. The successful integration of these models into a closed-loop HMI provides a concrete demonstration of how adaptive HMI can be realised in a safety-critical domain.

Ultimately, this research offers a new paradigm for pilot support, shifting the focus from reactive assistance to proactive, real-time intervention. The 3A-PAI system's ability to forecast cognitive degradation and sensory failures could dramatically reduce the number of SD-related accidents and enhance overall flight safety, particularly in single-pilot operations where the cognitive burden is high.

While this study represents a significant step forward, several avenues for future research exist. Further work is needed to validate the 3A-PAI system in a wider range of flight scenarios, including extreme weather and high-stress emergency situations, to test the robustness of the predictive models. A particularly promising direction is the integration of Large Language Models (LLMs) into the framework. LLMs could act as a cognitive co-pilot, not only interpreting the system's SD and LoAC predictions but also synthesising this information with flight parameters, weather data, and operational procedures. This would enable the system to generate natural language explanations or context-aware guidance, shifting the HMI from simple visual cues to a more intelligent, conversational support system. This could also help in exploring the optimal timing, modality, and level of automation for HMI interventions, preventing issues of over-reliance or automation complacency. Specifically, 3A-PAI goes beyond conventional human-AI teaming by not only supporting task-level coordination and decision assistance, but also by continuously modelling the pilot's cognitive state and embedding this information into adaptive visual guidance. However, the system does not fully constitute a human-AI hybrid, as final authority, intent formation, and control remain explicitly human-centred, and the AI does not autonomously substitute or merge with human cognitive processes. Accordingly, 3A-PAI represents an intermediate, transitional stage in which AI begins to interact with, and adapt to, human cognitive states in real time, laying the groundwork for future human–AI hybrid systems while preserving clear human oversight in safety-critical aviation contexts. Finally, future research could investigate the long-term effects of using such adaptive systems on pilot training and skill retention, ensuring that these technological advancements do not diminish fundamental piloting skills. How to embedded 3A-PAI with the implicit interaction approach in the cognitive loop and operations in the cockpit is also a crucial research for the application of human-AI hybrid intelligence.

Chapter 5. Exploring the Human-Centric Interaction Paradigm: Augmented Reality-Assisted Head-Up Display Design for Collaborative Human-Machine Interface in Cockpit

Building upon the foundational work of the 3A-PAI system, this chapter presents the design and implementation of an adaptive Augmented Reality Head-Up Display (AR-HUD). A crucial first step involved verifying and evaluating the usability of AR glasses as a viable platform for AR-HUD development. The AR-HUD interface was subsequently designed by applying the strategies outlined in the previous chapter, with a specific investigation into its potential for assisting pilots during subtle emergency situations. To accommodate diverse operational contexts, specific HUD designs were developed for various scenarios, including different flight phases and emergency conditions. The effectiveness of the AR-HUD was assessed through two primary methods: eye-tracking data from the AR glasses to evaluate scanning behaviour modifications, and the situation awareness present assessment method (SPAM) to quantify improvements human states in maintain optimal situational awareness and workload level.

5.1. Introduction

Human-centric solutions aim to achieve better collaboration human-machine-interface (CHMI) by utilising automation and real-time data exchange to further enhance operational efficiency and reduce human errors (C. Zhang et al., 2023). By enabling a new paradigm of human-cybernetic interfaces (HCI), cybernetic systems can help create a balance between automation and human interaction with right amount of workload and certain level of situational awareness (SA) (Pilla & Moreira, 2022). With a focus on human-centric design, HCI aims to create interfaces that enhance user experience, productivity, and safety (Krippendorff, 2019). By considering human factors (HFs), cognitive abilities, and ergonomics, cybernetic systems are designed to enhance user experience, by providing clear and meaningful feedback to users, and ensuring that machines can understand and respond to human input in a natural and user-friendly manner (Q. Li et al., 2021).

In commercial jet, the Pilot Flying (PF) assumes responsibility for aircraft control, management, and task delegation. Concurrently, the Pilot Monitoring (PM) is tasked with instrument monitoring and providing assistance to the PF. For cost concerns and pilot shortage in the market, airlines are initiating a reduction in cockpit crew members, from dual-pilot operations (DPO) to single-pilot operations (SPO), in flight operations (Tran et al., 2018). However, various pilot association and unions concern the SPO adoption and such advocation may pose an unacceptable risk in flight operations. One could expect that pilot in SPO may face human performance degradation due to the increased workload and loss of situational awareness (SA) by overwhelmed tasks. This degradation often results in human errors, and it may possibly induce accidents or incidents (Maurino et al., 2017). As the number of crew diminishes, the absence of a PM in case of a single pilot's failure could compromise flight safety. Transitioning to SPO requires a substantial automation support both on the flight deck and on the ground, and the pilot can stay focus on ground communication and flight controls (Cummings et al., 2016). Therefore, ensuring aviation safety requires vigilant monitoring and maintenance of optimal human performance across all potential scenarios, particularly in SPO contexts. The evolution of automation and monitoring technology, including reliable automated operating systems, remote pilot-support systems, and human-centric cockpit display designs, can mitigate pilot performance degradation and enhance flight safety to a certain extent (J. Liu et al., 2016).

Pilots need to simultaneously process flight information from multiple instruments' CHMI while performing flight operations and checklist. This necessitates frequent shifts in their gaze points, such as the external environment, primary flight display (PFD), electronic centralised aircraft monitor (ECAM), and others. Intuitively, the mental and physical demands imposed on a pilot during flight operation can be overwhelming (Q. Li, K. K. H. Ng, et al., 2023). Numerous instruments can easily increase workload and cause pilots to lose attention of hazard. In addition, the intricate flight environment can precipitate a loss of SA. SA, the ability to comprehend and perceive the surrounding environment, is crucial for pilots to appropriately manage unexpected situations (Cho Yin Yiu et al., 2022). SA encompasses the identification of pertinent information (Level-1), understanding its implications (Level-2), and the capacity to anticipate or respond effectively to changes or events within that environment (Level-3) (M. Endsley et al., 2000). The assessment of pilots' SA and workload serves as a preliminary step towards designing adaptive interventions to support pilots with degraded performance (Nguyen et al., 2019), especially in SPO scenario. Existing monitoring measures encompass both subjective and objective methodologies. Traditional SA evaluations often rely on expert experiential judgment, which is known for its low accuracy and susceptibility to personal bias. In order to mitigate the influence of subjective evaluations, contemporary SA evaluation methods commonly employ assessment scales. These scales encompass techniques such as the situation awareness global assessment technique (SAGAT), the situation awareness rating technique (SART), and the situation awareness present assessment method (SPAM). These evaluation methods gauge operators' SA levels using uniform standards by responding to questions based on the current environment while pausing or executing the flight task in a cockpit (Weiland et al., 2013). Furthermore, SART and SPAM can capture real-time SA changes due to their immediate evaluation during the ongoing task. Other than subjective measurement, objective evaluations rely on wearable bio-sensing technology to conduct behavioural and neurophysiological assessments (Ziv, 2017). Eye tracking technology has emerged as a valuable tool in HF studies within the field of aviation. This technology enables researchers to precisely monitor and analyse the eye movements and gaze patterns of pilots during various aviation tasks and scenarios (Liang et al., 2021; White & O'Hare, 2022). In detail, eye tracking technology utilises cameras and infrared sensors to track the movements of the eyes. These eye trackers can accurately identify and record the point of gaze, pupil dilation, and fixations (periods of focused visual attention) with high temporal and spatial resolution (Mengtao et al., 2023). To maintain appropriate SA and workload levels during rapidly changing flight scenarios, pilots must allocate their mental resources and visual attention to assimilate environmental and system information during different flight phases. Therefore, a combination of both subjective and objective evaluation methods can accurately and comprehensively monitor and measure a pilot's SA (Q. Wu et al., 2021). For example, the attentional distribution, mental fatigue, and SA in the face of flight hazard situations were identified using a self-report approach and eye tracker (Z. Lu et al., 2020). The timely detection of a pilot's state degradation in the SPO system, particularly neuro-psychological changes, has significantly aided the development of SPO and enhanced flight safety (Millot & Pacaux-Lemoine, 2013).

However, pilot performance, such as changes in SA and behavioural responses, may vary between SPO and DPO, which needs to be initially identified to support SPO development.

Considering that pilots need to process amounts of information during flight tasks, Head-Up-Display (HUD), as an integrated information transparent display, can serve as an assisted HMI to support pilots. The HUD is a transparent screen positioned in front of the pilot's field of vision. It provides an instantaneous overview of the information from various instrument interfaces (Tran et al., 2018). It is appropriate for applications in HUD for the head-mounted Augmented Reality (AR) to show rich textual and graphical superimposed information in the operating environment (Gardony et al., 2021). Nowadays, AR technology have been applied in the military fighter' HUD design as the AR assisted HUD to augment pilot's spatial awareness. With the implement of AR, AR-assisted HUD (AR-HUD) helps pilot reduce the reaction time and task completion time (W.-C. Li et al., 2022). Human-centric design positively affects users' experience and performance during flight tasks. The wearable AR-HUD's increased visual experience of the actual environment benefits the cognitive process and enhances SA and task performance (J. Liu et al., 2016). Airbus, Boeing, Commercial Aircraft Corporation of China (COMAC) and other aircraft manufacturers are now actively investigating the possible of AR-HUD in a cockpit and further improvement in pilot performance (W.-C. Li et al., 2022). Thus, the combination of CHMI design with AR technologies is highly innovative research, as it can present personalised and adaptive CHMI designs to pilots in the form of AR based on pilot's performance data, enabling them to maintain optimal task performance.

CHMI will be one of the key elements in future cockpit design, and it helps study the pilot interaction with audio and visual representation from the HCI (Neville A. Stanton et al., 2019). This study aims to design an AR-immersive HUD and explore the effect of AR-HUD on a single pilot's SA and workload level under different flight operated scenarios. The proposed research applies subjective and objective methods to collect multidimensional data to monitor and evaluate SA and workload changes to identify the difference in psychological and physiological feedback in both DPO and SPO scenarios. The experiment, including SPO and DPO situations, will be conducted by recruiting professional pilots wearing Microsoft HoloLens as the AR-HUD in the A320 cockpit simulator to simulate a flight from Hong Kong International Airport (VHHH) to Macau International Airport (VMMC). During the experiment, this study tracks pilots' visual attention allocation data by HoloLens and collected perceived SA level by the designed SPAM probes. Lastly, implementing AR in the CHMI of cockpit design introduces a human-centric pilot-support system that will contribute to aviation safety and allows the implementation of SPO to provide a feasible solution.

5.2. The interaction paradigm of AR-assisted HUD design

An experimental prototype was designed, also as an experimental platform, consisting of 1) a flight simulator, 2) an AR-HUD based on HoloLens, and 3) SPAM-based probes for assessing pilot SA. Figure 5-1 shows the experimental design setting. To simulate the real flight situation to the greatest extent, this study also set up a pseudo-Air Traffic Control officers (ATCOs) to simulate the approach and departure procedures from the air traffic control (ATC) tower. ATC issues instructions and communicates to pilots, such as the taxiing instruction ‘HX123, Taxi Hotel 7, Hotel 9, Juliet, holding point Juliet 11, RWY 25L’.

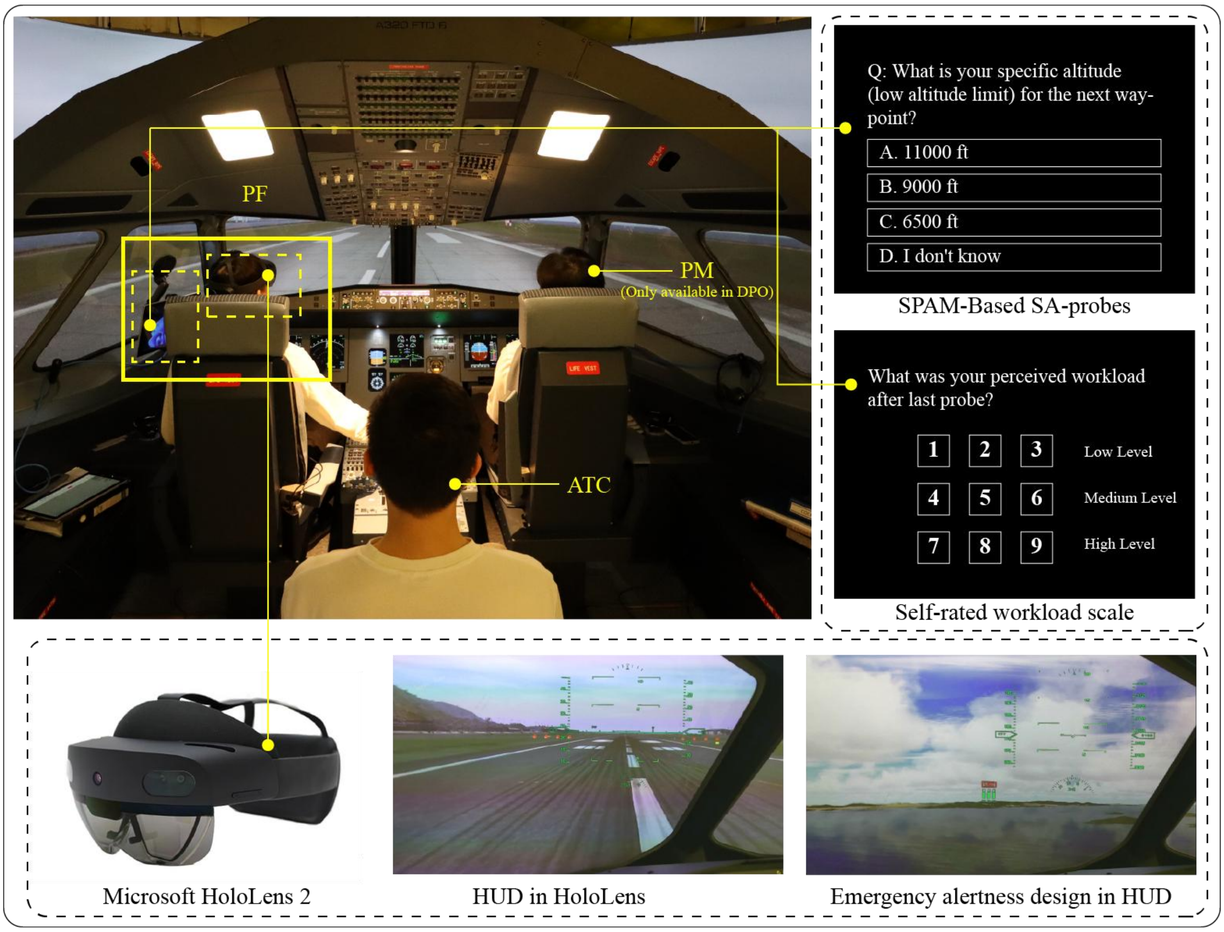


Figure 5-1. The experimental platform.

5.2.1. Experimental flight training simulator

A non-full motion CockpitSonic Airbus 320 Flight Training Device (FTD) was used to simulate the real flight operation and scenario. The FTD includes a 220° surrounded projection visual system and all the flight deck hardware. The FTD, together with the map terrain data from Prepar3D version 5.3, a visual simulation platform, allow for maximum simulation of realistic flight scenarios. The instructor operating station based on Jeehell flight management guidance system (FMGS) Pro can trigger various types of events, which can be adopted for simulating various inflight emergency scenarios for flight training and assessment.

5.2.2. AR-HUD design and development

For the AR-HUD development, the Microsoft HoloLens 2 was used in this study as the display terminal. Additionally, by utilising the Microsoft Mixed Reality Toolkit (MRTK) (Kevin Semple, 2022) in conjunction with Unity, this study was able to deploy and modify the HoloLens application directly within Unity through the Holographic Remoting feature and relevant MRTK assets. Given the similarity between the experimental platform, the CockpitSonic A320 Flight Training Device (FTD), and the Airbus A350 HUD design, the AR-HUD opted to reference the latter as a basis for the HMI design. After referring to the publicly released Unity asset by Maloke Games (Games, 2021) and Polyester Studio (Studio, 2016), 4 different HUD display modes were designed as shown in Figure 5-2, corresponding to different flight stages: taxiing and take-off, cruise, landing/rollout, and emergency scenarios.

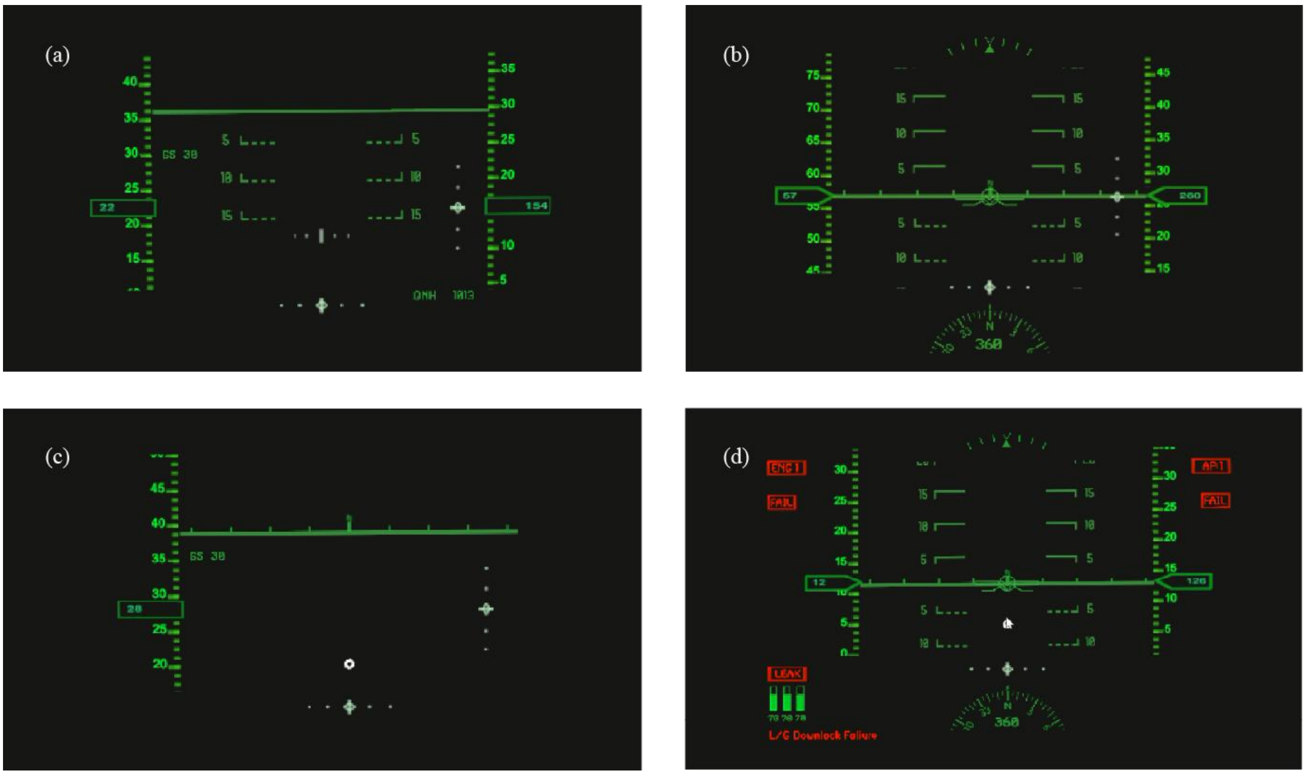


Figure **5-2**. HUD interface design for three flight phases and emergency scenarios. (a).

Taxi and take-off; (b) Cruise; (c) Landing/Rollout; (d) Emergency alertness design.

It is crucial to set up a data transmission channel between FTD and HUD display. The process of extracting data from the FTD involved establishing a connection between the FTD and the Unity software through the SimConnect software. A wrapper, functioning as a dynamic library, was developed to connect the Unity and SimConnect and facilitate the retrieval of variables and flight data. Then, the HoloLens device is able to effectively reflect the real-time status of the airplane via continuous data streaming from FTD.

5.2.3. Probes for assessing pilot SA and workload level.

The SPAM-based probes and workload self-rated scale were employed for measuring SA and workload, respectively. SPAM is a real-time approach to measuring SA and is not required to pause the flying, which was used to assess SA according to the responding time and accuracy. Specifically, this study designed different SPAM-based probes to measure subject' SA level in different experimental scenarios and employed 9-likert self-rated scale to monitor the workload level. Subjects were required to place the highest priority on responding to probes.

On the one hand, in order to assess SA, 2 sets of 24 multi-choice questions (Self-rated probes based on SPAM) were designed and pop-out at a specific moment one by one via a tablet PC placed on the left-hand side of the captain. The difficult level of SPAM-based probes is same. The participants are required to search the situational and operational information to respond to the pop-outed SA question by clicking one of the choices. The SA probe involves assessing the subject's short-term memory, perception of flight phases and performance, familiarity with the flight plan and so on. For example:

Q: What is the name of the previous waypoint that you just have flown through?

A: PECAN

B. CHALI

C. RUMSY

D. I don't know.

For the assessment of subject's SA level, the experimenter labelled the subject's SA level based on the response time and correctness when the subject answered the SA probes. If the response time was more than 10 seconds and the SA probe answered incorrectly, it was labelled as SA-Low level. If the response time was more than 10 seconds and the SA probe answered correctly, or if the response time was less than 10 seconds but the SA probe answered incorrectly, it was labelled as SA-Medium level. If the reaction time was less than 10 seconds and the SA probe answered correctly, it was labelled as SA-High level.

On the other hand, this study also set up a 9-likert workload self-assessment scale to measure the perceived workload level during each of the two self-assessment periods, as Figure 5-1 shows. The workload self-assessment scale was popped out just before the SA probe. That is, subjects were first required to self-rate the workload level, and then the SA probe was immediately pop up automatically. Therefore, this study could analyse the relationship between workload, SA, and different flight operation modes. Furthermore, the participants were asked to complete a designed post-experiment questionnaire to collect their subjective feedback about workload and operation perception.

5.3. Methodology

5.3.1. Participants

This study evaluated 21 subjects in total, and the demographics and profile of the participants were summarised in Table 5-1. Demographics and profile of the participants. Fifteen cadet pilots (3 female) were recruited from airline company's Cadet Pilot training programme, their age ranging from 22 to 37 (average age = 29.40 ± 4.29). Therefore, the experiment counted the hours they received training, ranging from 80 - 720 hours, with an average of 438.267 ± 249.68 hours. 6 experienced pilots (all male) on Second Officer (SO) from airlines were also recruited, ranging in age from 25 to 34 (average age = 29.83 ± 3.19). Different from the cadet pilot, the experiment counted the training time of SO on the simulator, which ranged from 200 - 250 hours, with an average of 208.33 ± 20.41 hours. All the subjects had professional training experience on the Airbus A320. All subjects were well informed the experimental data are confidential and the experiment is not related to any pilot assessments of their affiliated company. The subjects received HKD $500 supermarket coupons as remuneration of their efforts and time for the experiment. This experiment was conducted in the *Human Factors and Ergonomics Laborator*y, Department of Aeronautical and Aviation Engineering, *The Hong Kong Polytechnic University, HKSAR.* This study was conducted following the procedure under the ethical standards of the 1975 Helsinki Declaration. It was also approved by the *Hong Kong Polytechnic University Institution Review Board* regarding clinical human subjects in advance (No. HSEARS20210318002).

Table 5-1. Demographics and profile of the participants.

Experienced pilots (SO)			
No.	Age	Gender	Flight hours (On simulators)
1	25	Male	200
2	28	Male	200
3	29	Male	250
4	31	Male	200
5	32	Male	200
6	34	Male	200
Cadet student pilots			
No.	Age	Gender	Training hours (Ground Theory)
1	22	M	80
2	26	M	84
3	33	M	600
4	27	M	600
5	30	M	200
6	28	M	650
7	34	F	660
8	28	F	200
9	35	M	600
10	25	M	210
11	31	F	570
12	29	M	720
13	37	M	750
14	32	M	150
15	24	M	500

5.3.2. Experimental design

An experimental scenario with a flight route started from Hong Kong International Airport (VHHH) to Macau International Airport (VMMC), as shown in Figure 5-3. The callsign is ‘HX123’, and the entire flight, including taxiing, take-off, departure, climb, cruise, descend, approach, and landing phases, takes around 35 minutes for each session. Cirrocumulus clouds with a visibility of 25,000ft were set at an altitude of 2000-5000ft.

Meanwhile, 3 different level of emergency events that can be fixed were inserted to complicate flight operation. The easier emergency situations include 1) LDG not Up/Down Locked, and 2) ADR and AP failure. More difficult cases include 1) ENG-1 failure, and 2) ENG-1 fire. Whatmore, the experiment also included a monitoring situation: whether the Inner Tank was leaking. In order to avoid the learning effect of the subject, each round of the experiment chose one of the above 3 situations of emergency as the task. So, each round of experiment includes a high- and low-difficulty emergency, and a monitoring task. By simulating a controlled number of emergencies, this study is able to assess the impact of different cockpit configurations and AR-HUDs on SA and workload under stressful conditions. This design enables the separation and comparison of pilot performance and cognitive load in different scenarios, providing insight into the potential benefits of the AR-HUD in emergency situations.

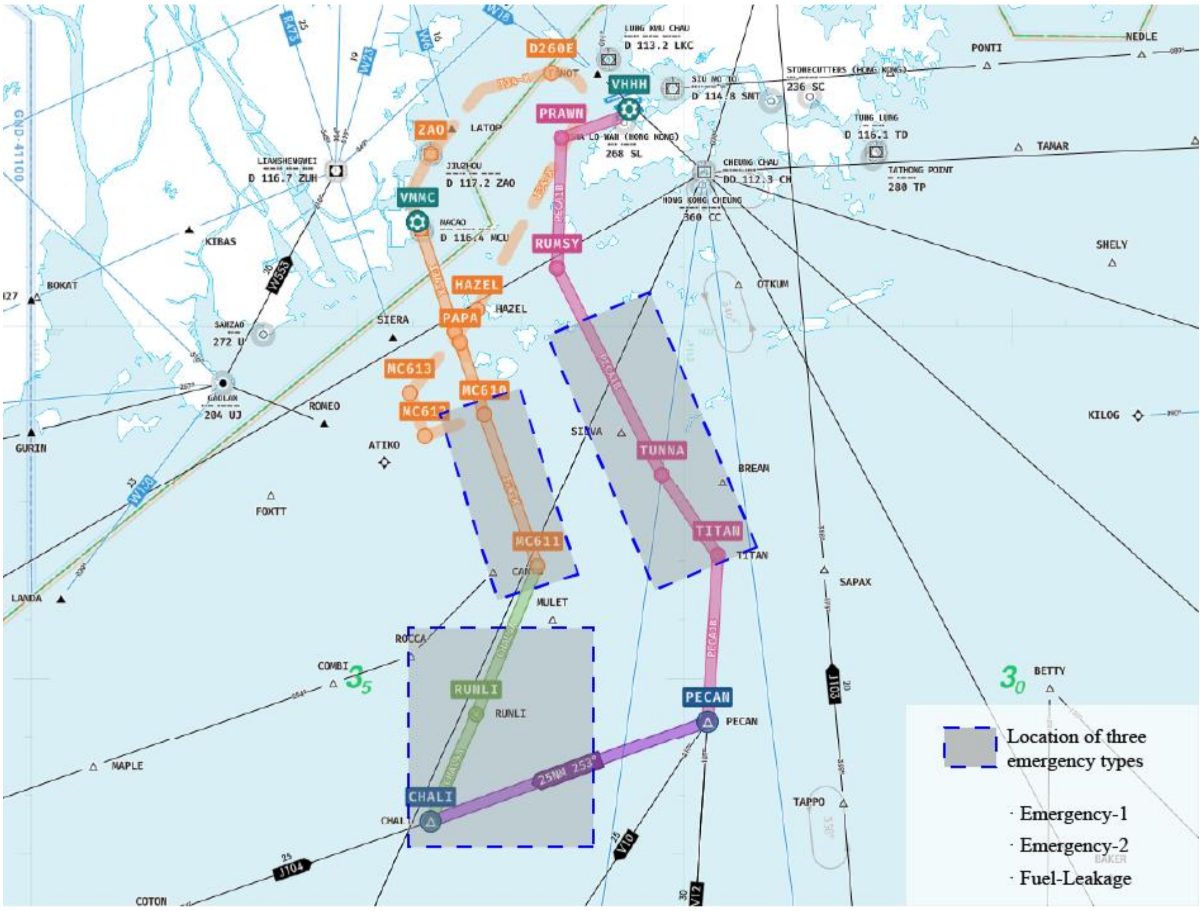


Figure 5-3. The flight route from Hong Kong international airport (ICAO:VHHH; IATA: HKG) to Macau international airport (ICAO: VMMC; IATA: MFM)

5.3.3. Experimental procedure

Prior to the experiment, subjects were informed about the overview and procedure of this experiment, signed the consent form, and completed the demographic survey. All information related to the flight simulation was given, such as airport and aircraft data, flight plan, and flight route, etc. In order to examine the effect of AR-HUD on different operation scenarios, this study considers two independent variables: 1) the availability of AR-HUD; 2) the number of pilots. Therefore, the entire experiment contains 3 sessions in total, and a one-way flight route (VHHH-VMMC) was regarded as one session. The 3 sessions are 1) DPO, 2) SPO without HUD (referred to as SPO) and 3)

SPO with HUD (referred to as SPO-HUD), as shown in **Table 5-2**. Following that, subjects were equipped with HoloLens to record their eye-tracking data in all the 3 sessions and present the HUD information only in the HUD scenario.

During the experiment, all subjects were required to sit and wear the HoloLens comfortably. The integrated eye tracker was calibrated using the built-in calibration routine. Each subject was given 20 minutes to familiarise themselves with the operation of the HoloLens and FTD. Afterwards, the subjects completed the 3 sessions of flight simulation experiments in sequence. In order to avoid a learning effect on the subjects, the experimental sequence of DPO, SPO and SPO-HUD was assigned randomly so as not to affect the results of the experiment. The flight task started from request push back (when the aircraft engine was turned on), and ended with roll out. The checklist was executed from 'BEFORE TAKEOFF' to 'LANDING'. If landing was not possible, the go-around procedure was carried out until reaching the position where a successful landing would have taken place. The duration of each session was about 35 minutes, and the total duration of the experiment was about 2 hours. All the instructions from the tower, departure, approach, etc., as well as the processing of the emergencies, were given by a pseudo-ATC on site. In the DPO scenario, the experiment also arranged a flight training instructor to mimic PM functions and assist the PF to complete the flight tasks. It should be note that the PM position does not exist in SPO scenario with and without HUDs.

During the flight tasks, the experiment randomly triggers 3 emergencies, as described in the experiment design, and the subject needs to solve the emergencies as best as possible according to the standard operating procedures (SOPs), and at the same time, the subject needs to report the incident to the ATC. Considering the possibility that the cadet pilot may not be able to completely solve the more difficult emergencies, ATC will provide guidance to the subject when appropriate. However, in this case, once ATC provides guidance and intervention, the subject's task of handling the emergency is considered a failure.

Table 5-2. Session allocation

Session	HUD available	Position	Scenario
1	N	PF	DPO
2	N	PF	SPO
3	Y	PF	SPO-HUD

Measuring SA level and self-rating of workload were required within every 1-2 minutes throughout the flight task. The SPAM Probe intervals depend specifically on the experimental design, ranging mainly from 2-7 min (Feng et al., 2021; Gregoriades & Sutcliffe, 2018; Naderpour et al., 2016). A shorter interval allowed us to capture potential fluctuations in SA and workload within the available time frame during the flight task, providing a more detailed understanding of their dynamics over time. This granularity could help identify critical moments or transitions where the AR-HUD's assistance might have a more pronounced effect. Meanwhile, the 24 probes can be reasonably distributed in different flight phases. Taxi, take-off, departure, and approach all include 2 probes, cruise phase is the longest with twelve probes, followed by the more critical landing phase with 4 probes.

During the insertion of SPAM-based probes in the experiment, to avoid interfering with the pilots' normal operations, the insertions were made when the pilots were not currently engaged in any operational action. When a question popped up on the tablet, it was accompanied by an audio beep that required the subject to place the highest priority on answering the question. The SPAM was presented to the subject in the form of a multiple-choice question on a tablet PC, and the subject simply clicked on the answer. Immediately after answering the SPAM questions, the 9-likert self-rated workload scale popped up. the pilot directly clicked on the value that matched his or her situation. One SPAM and one workload self-assessment are one unit, and there are 24 units in total for each session.

The eye-tracking data recorded by HoloLens, performance in SA probes, workload feedback, and performance in dealing with emergencies were the dependent variables within the entire flight phases. Furthermore, the participants were required to complete a designed post-experiment questionnaire to collect their subjective feedback form about workload and operation perception for different flight scenarios and AR-HUD.

5.3.4. Data collection

5.3.4.1. Measurement based on eye-tracking data.

Apart from the prototype developing, the eye-tracking data is also one of the important factors to assess the pilot performance. Kapp discovered that there is not any direct access to eye-tracking data extraction by implementing a Unity 3D package in HoloLens 2 (Kapp et al., 2021). Therefore, their team created an open-source toolkit, Augmented-Reality-Eye-Tracking-Toolkit (ARETT), for researchers to obtain the desired data by inserting the package into the project directory in Unity. This toolkit adds a virtual camera to the 3D scene that matches the location, projection, and resolution of the integrated front-facing camera. This enables the projection of the 3D gaze position to the virtual 2D camera image and, hence, to the webcam image. The virtual camera is preconfigured to match the integrated, front-facing webcam of the HoloLens2. Nevertheless, ARETT required a virtual cockpit environment to capture the position of the labelled panels and displays. Therefore, this study constructed a virtual cockpit environment based on the structure of the FTD and the flight tasks and divided the entire cabin into twenty Aera of Interests (AOIs). Twenty AOIs includes Autopilot, ECAM, ECAM Control Panel, External View 1, External View 2, External View 3, FMS Left, FMS Right, Flap Parking Brake, Ground, HUD_1, HUD_2, HUD_3, Head Up Panel, Joystick Right, Joystick Left, Left PFD, Right PFD, Table, and Throttle Lever. The virtual cockpit environment and the locations of the twenty AOIs are shown in Figure 5-4. ARETT can record the AOI, timestamp and velocity of the subject's gaze point as eye-tracking data.

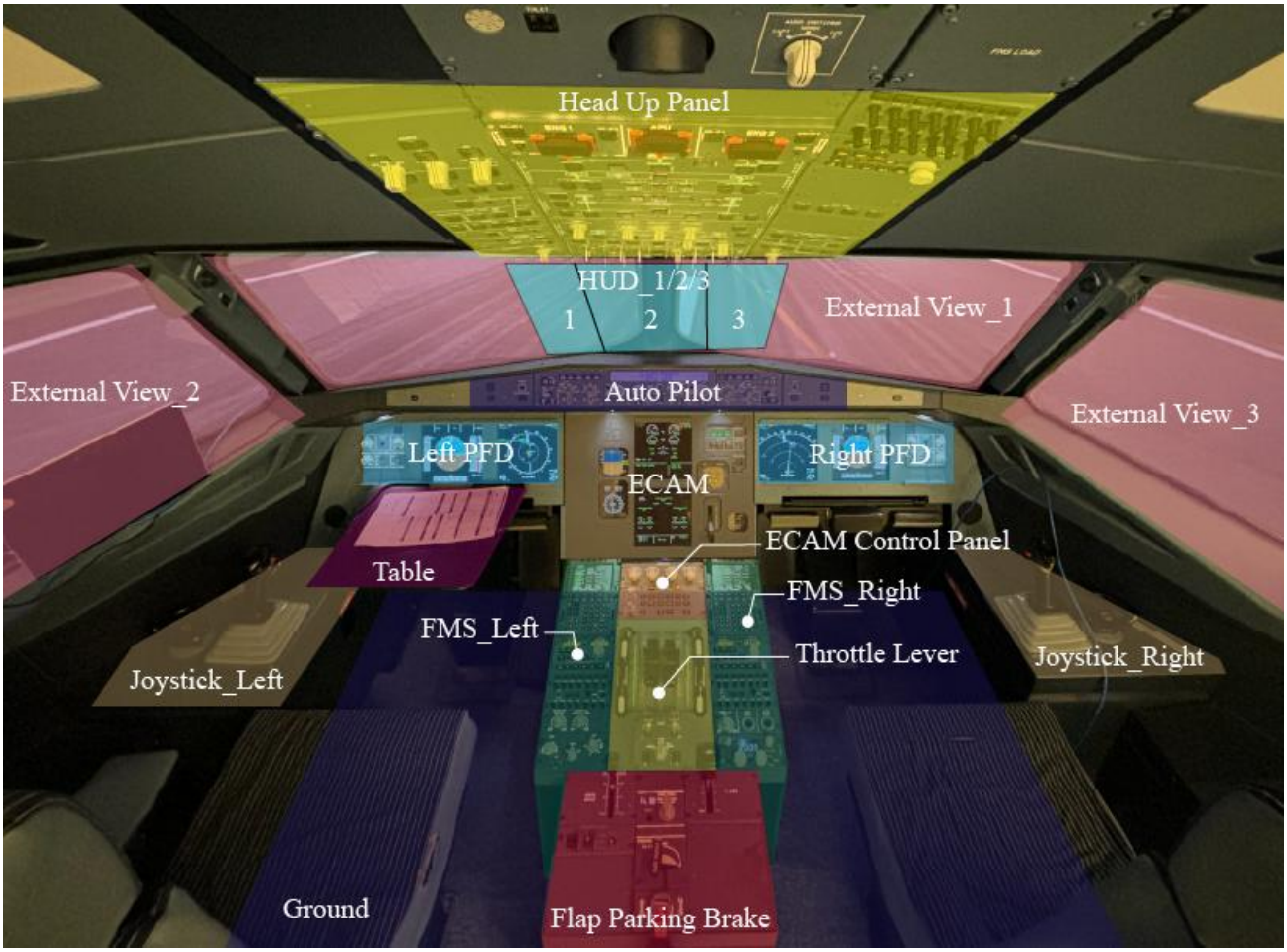


Figure 5-4**.** The virtual cockpit environment and the locations of the 20 AOIs

5.3.4.2. Measurement based on subjective feedback.

The self-rated probes for measuring SA and collecting workload feedback were designed to assess SA based on the psychology software tools, *E-prime 3.0.* The touch-screen PC with the E-prime installed was placed on the left side of the PF so that it does not affect the normal flight task of the PF. For measuring SA, self-rated probes recorded the answer chosen by the subject, reaction time and accuracy of each SA probe. These metrics were used to label the SA-Level of each subject.

5.3.4.3. Measurement based on task performance data.

Thirdly, this study also recorded the task performance data, that is, the complete time for dealing with the emergencies. The researcher first recorded the emergency trigger time. When the pending emergency instructions in ECAM screen are completely eliminated, that is, ECAM memo without blue items, the emergencies processing is deemed to be completed. The researcher immediately recorded the completion time.

5.3.5. Data analysis

5.3.5.1. Eye-tracking data processing

Gaze data can also be accessed within Unity 3D via the Mixed Reality Toolkit (MRTK) through ARETT (Kapp et al., 2021). ARETT records raw gaze data robustly with a data rate of 30 Hz, with high spatial accuracy and precision. The recorded data from ARETT mainly includes 3 types of eye-tracking data, namely, 1) Time data: timestamp; 2) Gaze data: a gaze origin and a direction vector; 3) AOI data: position of the gaze point on the AOIs. The specific data recorded, and description are shown in Table 5-3.

Table 5-3**.** The key indicators of the eye tracker data of ARETT.

Key indicators	Description
Gaze Origin(x/y/z)	Gaze origin in the global reference frame
Gaze Point AOI (x/y/z)	Position of the gaze point on the AOI in global coordinates
Gaze Point AOI Target Name	Name of game object representing the AOI
Number of Fixations	The number of fixations occurring during TOIs
Number of Saccades	The number of saccades occurring during TOIs

The eye-tracking data was processed according to an open-source toolkit using R-Language (Kapp et al., 2021). Data processing can be divided into 3 parts, namely, 1) Pre-processing, 2) fixation identification, and 3) post-processing. The processing steps are shown in Figure 5-5.

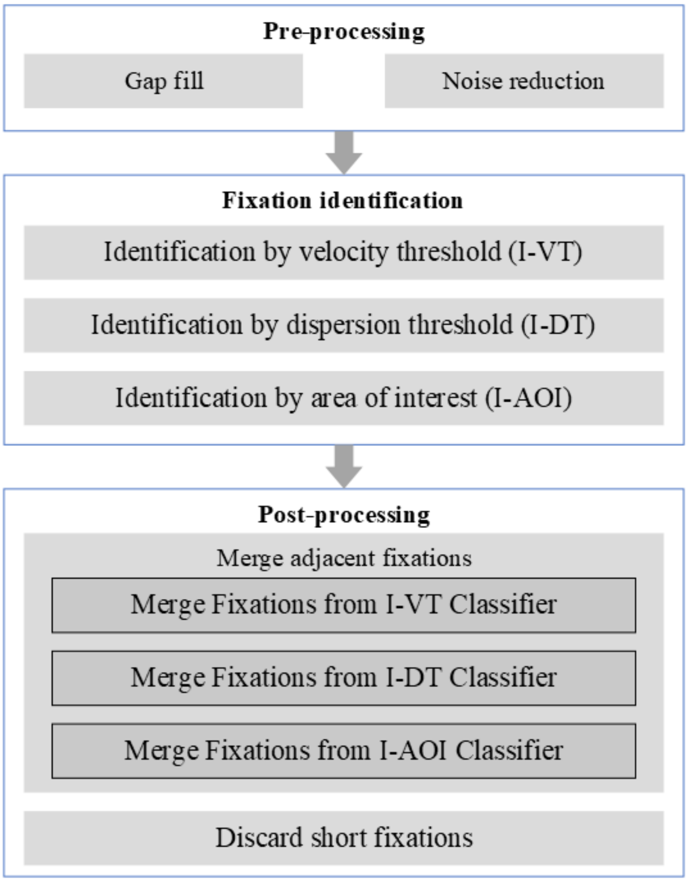


Figure 5-5. Eye-tracking data processing

(1) Pre-processing:

Gap fill is designed to fill in any gaps in the data frame by linear interpolating the missing values. It then fills in the missing values by linearly interpolating between the indeed rows and the valid rows.

Noise Reduction reduces noise in eye-tracking data by applying a median value in a specified window. It iterates over all gaze points and replaces each data point with the result of processing all data points with the median value in a window centric on the currently replaced data point. This is consistent with the method processed by Tobbi.

Since HoloLens 2 has an eye tracking rate of 30 Hz, if noise reduction is applied to 3 samples, the window size is 1/30Hz = 100ms.

(2) Fixation identification:

Identification (Classify Fixations) by Velocity Threshold (I-VT): Classifies each gaze point according to the gaze point velocity calculated by ARETT. The general idea behind an I-VT filter is to classify eye movements based on the velocity of the directional shifts of the eye. The velocity is most commonly given in visual degrees per second (°/s) (Salvucci & Goldberg, 2000). If it is above a certain threshold the sample for which the velocity is calculated is classified as a saccade sample and below it is seen as part of a fixation (Kosel et al., 2023). The default velocity threshold is 100°/s, as described by Tobii for their mobile eye trackers (AB, 2024).

Identification (Classify Fixations) by Dispersion threshold (I-DT): Each gaze point is classified according to the dispersion of the gaze point. It uses distance-dispersion threshold to identify fixation points with a specified threshold and minimum duration. According to Llanes-Jurado et al. the acceptable parameters for the HTC Vive Pro Eye HMD range from 1-1.6 degrees and 0.25-0.4 seconds, with 1 degree and 0.25 seconds being the optimal values (Llanes-Jurado et al., 2020).

Identification (Classify Fixations) by area of interest (I-AOI): Every gaze point is categorised based on whether it lands on the AOI and the minimum fixation time. Initially, all gaze points that hit the AOI are classified as fixations, and all other gaze points are labeled as saccades. Consecutive gaze points on the same AOI are grouped as a fixation if the total duration exceeds the minimum fixation time. If the minimum fixation duration is not met, these gaze points are reclassified as saccades.

(3) Post-processing:

Merge adjacent fixations: 1) Merge Fixations from I-VT Classifier is to merge fixations in the I-VT classifier checks whether the gap between each fixation is shorter than the specified maximum time between fixations, and whether the angle between the fixations before and after the gap is less than the specified angle. If this is the case, the two fixations and the gap are combined into a new combined fixation point. According to Tobii's definition, the default values for Maximum Time and Maximum Angle are 75 ms and 0.5 degrees respectively. 2) Merge Fixations from I-DT Classifier is to merge the fixations that the distance between these fixations is less than the maximum time and that the specified maximum dispersion is not exceeded. This is used to merge fixations on the same AOI that are separated by less than the specified time. 3) Merge Fixations from I-AOI Classifier is to combine fixations when they are on the same AOI and closer to each other than the max time specified. This is a custom implementation for merging fixations on the same AOI with a gap smaller than the specified time between them.

Discard Short Fixations: discards fixations with durations shorter than the specified minimum duration and reclassifies them as gaps. All parameters including a minimum fixation duration are consistent with I-DT and I-AOI.

5.3.5.2. Statistical data processing

IBM SPSS (v.23.0) was used for the statistical analysis to discuss the benefits of AR-HUD and the difference across flight operation mode. Gender differences were not considered in this study. The Shapiro-Wilk test was used to determine the normal distribution. One-way repeated ANOVA was used to test the significance of the three-source data, including eye-tracking data, self-rated probes feedback, and emergency handling time, between 3 experimental scenarios when the data were all normally distributed, with no outliers and equal variances. The Bonferroni correction was used to correct the significance level. Friedman test was selected for difference statistics for the abnormal values. $P \leq 0.05$ was statistically significant.

5.4. Results

5.4.1. Self-rate probe result

This study compared the self-rated workload, response time, and accuracy of the SA probe under 3 different operating scenarios. Figure 5-6 (a) illustrates the average workload for each probe. The average workload levels for DPO, SPO, and SPO-HUD scenarios are 3.68, 4.40, and 3.53, respectively. Among the 24 self-rated workload feedback evenly distributed throughout the entire flight process, as shown in the Table 5-4, the probes with significant differences were 03, 04, 05, 06, 07, 08, 09, 13, 14, 17 ($P \leq 0.05$). In other words, there are statistical differences in the 3 operation modes during the take off and cruise stages.

Table 5-4. Statistical differences in self-rated workload of 24 probes under three scenarios.

Probes	DPO		SPO		SPO-HUD		P
	Average	SD	Average	SD	Average	SD	
03	3.5238	1.7210	5.5714	1.8048	4.7143	1.9785	0.000
04	3.1429	1.6818	5.4286	2.1348	4.0476	2.0119	0.000
05	2.9048	1.6095	5.1429	1.9821	3.3333	1.5916	0.000
06	2.4762	1.2498	3.8571	1.8785	2.8571	1.6518	0.007

07	2.6667	1.2780	4.2381	1.9724	3.7619	1.9976	0.003
08	2.7619	1.4459	3.9048	1.8949	2.8095	1.6917	0.019
09	3.6667	2.2657	3.5714	2.03891	2.3810	1.2836	0.041
13	4.0000	2.0494	3.9524	2.3974	2.3333	1.3166	0.000
14	4.0000	1.8974	4.2381	2.6628	2.6190	1.9615	0.029
17	3.7143	1.9011	5.1429	2.6322	4.3810	2.3974	0.048

For the SA assessment, this study firstly labelled all the subjects in each probe according to the SA assessment criteria introduced in section 2.3. Then, the number of SA levels labelled as high, medium, and low was counted for each probe. Table 5-5 showed the distribution of the three SA-Level labelled in different scenarios. Pilots under SPO-HUD scenario got the higher level of SA than other 2 operated mode. However, in the statistical analysis, there was no significant difference in the distribution of the number of the three SA-Levels under the 3 scenarios. Secondly, this study conducted Friedman test on the SA level of each subject labelled under each probe, and the result was also shown in Table 5-5. The labelled SA-level in Probes 03, 05, 06, 09, 11, 12, 16, 17, 19, 20, and 22 were significantly different under the 3 scenarios. The flight phases corresponding to the above probes are departure, cruise, approach, and landing. The distribution of SA-levels appears minimal between DPO and SPO scenario. Because when the SPAM-based probe is inserted, the pilot is not currently engaged in a heavy operation, and regardless of the scenario, the pilot is able to have more time to search for information to answer the question. In addition, pilots need to search for information independently, regardless of the presence or absence of a PM. One of the functions of the HUD is to integrate and display key information during flight, aiding pilots in faster information retrieval. Therefore, the data show minimal differences between DPO and SPO scenarios, while SPO-HUD shows a significant improvement.

Table 5-5. Distribution of the three SA-Levels labelled in the three scenarios.

	DPO	SPO	SPO-HUD
High-level	245	231	326
Medium-level	195	198	145
Low-level	64	75	33

Probes with significant difference ($P \leq 0.05$): 03，05，06，09，11，12，16，17，19，20，22

The average reaction times for answering SA probes in DPO, SPO and SPO-HUD were $10889.9147 \pm 13009.6094$ ms, $11643.8849 \pm 15301.2183$ ms, and 6332.6567 ± 5668.2635 ms respectively. Figure 5-6 (b) shows the average reaction time for each probe. Among the 24 probes, 01, 03, 07, 08, 12, 16, 19, 20, 21, 22, and 24 showed a significant difference in SA reaction time under 3 flight scenarios as shown in Table 5-6. Lastly, the correctness to SA probes showed significant differences among the 3 operational scenarios ($P = 0.000$). The percentage of correctness of SA probe were 65.08%, 62.30%, and 72.02% under DPO, SPO, and SPO-HUD, respectively.

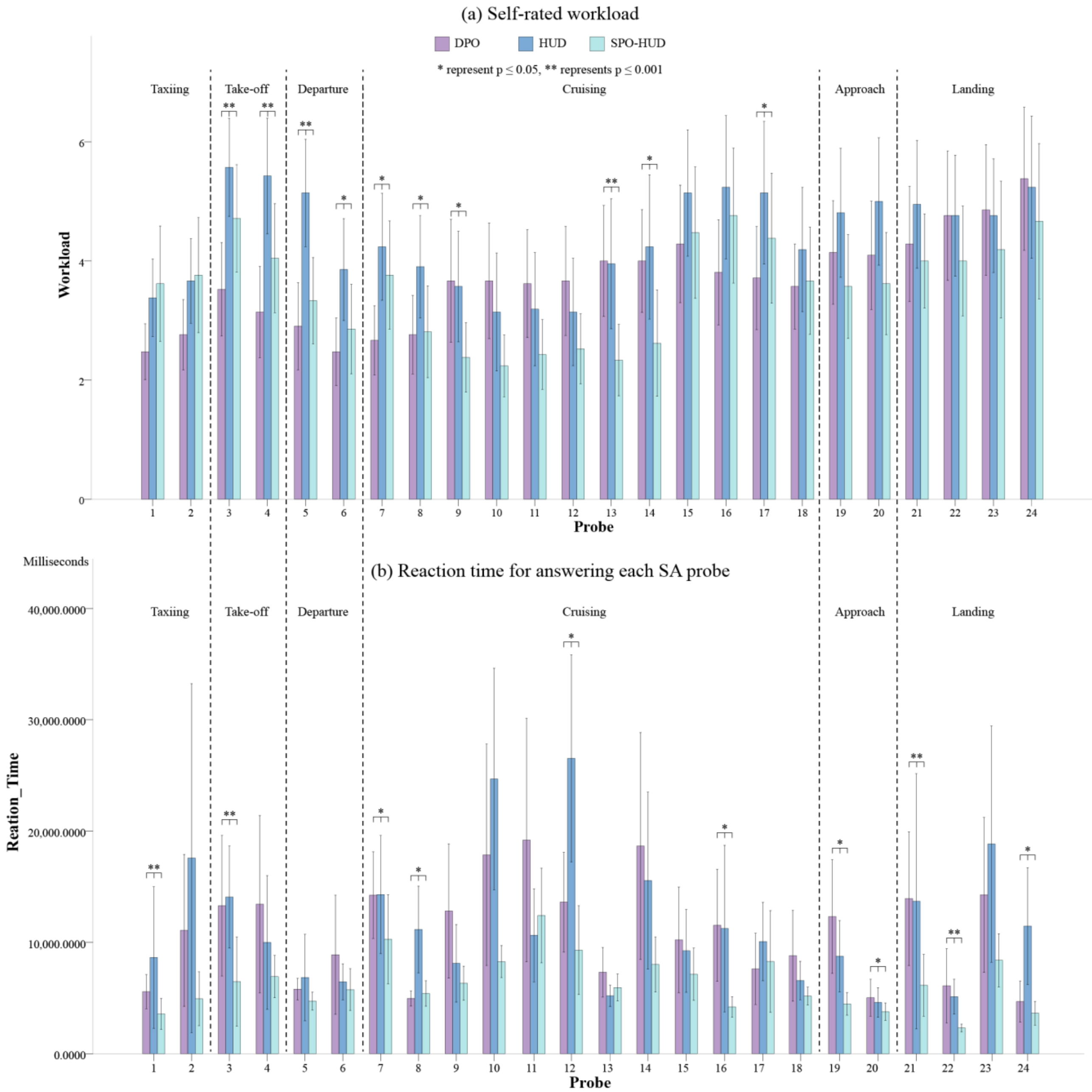


Figure 5-6. Change in the average of 24 self-rated workloads and reaction time of SA probes for the three scenarios. (a) Self-rated workload; (b) Reaction time for answering each SA probe.

Table 5-6. Significant differences in SA probe's reaction time under three scenarios.

Probe	DPO		SPO		SPO-HUD		P
	Average / ms	SD / ms	Average / ms	SD / ms	Average / ms	SD / ms	
01	5580.5238	3358.4824	8643.3333	13964.8397	3580.1905	3050.9094	0.001
03	13299.1905	13844.2914	14073.1429	10056.1763	6479.8095	8797.8897	0.001
07	14250.8095	8540.4910	14299.3333	11637.1840	10285.0476	8807.2361	0.041
08	4966.0000	1470.9961	11166.0952	8564.3277	5406.8095	2511.6792	0.006
12	13622.7143	9832.1594	26528.2857	20439.4149	9296.4762	8766.1294	0.002
16	11542.1429	11056.5774	11248.3333	16431.7780	4207.0000	2010.5980	0.002
19	12320.6190	11233.8090	8761.3333	7031.4009	4471.0476	2232.3739	0.019
20	5031.3333	3627.9345	4611.2381	2936.2539	3781.2381	1706.7725	0.041
21	13935.3810	13206.2267	13698.3810	25176.0650	6150.9048	6109.9938	0.001
22	6099.6667	7345.1484	5121.7143	3435.1703	2314.9524	761.3548	0.000
24	4684.9524	4048.4733	11464.2857	11538.6776	3641.5238	2352.8169	0.005

Overall, in the SPO scenario equipped with HUD, the pilot had the lowest workload level, the highest SA response accuracy, and the lowest reaction time. The sorting of the 3 scenarios is: SPO-HUD, DPO, and SPO. Therefore, in terms of SA probe and self-rate workload, HUD can play a role in compensating PM, reducing PF workload, and improving SA level.

5.4.2. Pilot performance data for emergencies

In the 3 flight scenarios, 3 emergencies were set separately. In order to avoid the learning effect of the subject, the emergency levels in each round of experiments are different, but the difficulty settings are basically the same: divided into high and low difficulty emergency processing, and a continuous monitoring fuel tank task. In DPO, SPO, and SPO-HUD scenarios, the average completion time for difficult emergencies is: 0:03:47, 0:03:33, and 0:02:51; The average completion time for low difficulty emergencies is: 0:00:19, 0:00:27, and 0:00:09; The average time for monitoring fuel leakage and reporting it to ATC is: 0:05:00, 0:03:55, and 0:00:44, as shown in Table 5-7. It should be noted that when calculating the reaction time of fuel leaks, subjects that failed to successfully monitor fuel leaks were not included. In the DPO scenario, 4 pilots were unable to successfully detect fuel leaks, while in the SPO without HUD scenario, 6 pilots were unable to successfully detect fuel leaks. However, in the SPO-HUD scenario, all pilots quickly detected fuel leaks at the HUD prompt.

Table 5-7 shows the statistical analysis of the performance data of 3 levels of emergencies in different scenarios. There were significant differences for the completion time among the 3 types of emergencies in DPO, SPO, and SPO-HUD scenarios.

Table 5-7. Completion time for handling emergencies in three different flight scenarios

	DPO	SPO	SPO-HUD
Emergency 1	0:03:47**	0:03:33**	0:02:51**
Emergency 2	0:00:19**	0:00:27**	0:00:09**
Fuel Leakage	0:05:00**	0:03:55**	0:00:44**

** represents $P \leq 0.01$

5.4.3. Eye tracking data

This study summarised the gaze point of 20 AOIs using eye tracking data collected by HoloLens, as shown in Table 5-8. Among the 3 scenarios, 7 AOIs in total, namely, HUD_ 1. HUD_ 2. External View_ 1. External View_ 2. Left PFD, Auto Pilot, and ECAM are all basically higher than the average number of gazes. AutoPilot, ECAM, ECAM Control Panel, External View 2, FMS_Left, Ground, HUD_1, HUD_2, HUD_3, Joystick_Right, Left PFD, Right PFD, and Table 5-8 counts the percentage of 20 AOIs where the gaze point is located and significant differences among the 3 operation scenarios ($P \leq 0.05$). In the DPO scenario, the 5 AOIs with the highest number of gazes are: Left PFD (19.02%), ECAM (17.87%), and HUD_ 2 (14.67%), External View1 (12.46%), HUD_ 1 (9.55%). In the SPO scene, the 5 AOIs with the highest number of gazes are ECAM (19.20%), Left PFD (18.72%), and HUD, respectively_ 2 (14.14%), External View1 (12.26%), and External View2 (8.80%). In SPO-HUD scenario, the 5 AOIs with the highest number of fixations are HUD_ 2 (28.13%), External View1 (14.87%), ECAM (13.46%), HUD_ 1 (12.33%), Left PFD (8.64%).

Table 5-8. The percentage of AOI distribution in DPO, SPO, and SPO-HUD scenarios.

No.	DPO		SPO		SPO-HUD	
	Name of AOIs	Percentage	Name of AOIs	Percentage	Name of AOIs	Percentage
1	LeftPFD	0.196947647	ECAM	0.188201891	HUD_2	0.279185861
2	ECAM	0.182273359	LeftPFD	0.180389224	ExternalView1	0.146349983
3	HUD_2	0.150449485	HUD_2	0.13532219	ECAM	0.134183133
4	ExternalView1	0.132084789	ExternalView1	0.120829936	HUD_1	0.125336333
5	HUD_1	0.097068524	HUD_1	0.089246099	LeftPFD	0.085150256
6	ExternalView2	0.088644474	ExternalView2	0.086862524	ExternalView2	0.074124457
7	AutoPilot	0.080680405	AutoPilot	0.082914361	AutoPilot	0.057794271
8	HUD_3 *	0.030650326	HUD_3 *	0.029962008	Ground	0.014522047
9	Table	0.02475652	Ground	0.020284605	HeadUpPanel *	0.006846306
10	Ground	0.020713202	Table	0.014701404	Table	0.006540076

* represents represent AOIs that did not rank in the top 10 in all three scenarios.

This study calculated the heat map of gaze point distribution. Yellow colour block represents more fixation times, while black colour block represents less annotation times. Figure 5-7 shows the heat map of the gaze point distribution in DPO (left), SPO (middle), and SPO-HUD (right) scenarios, respectively. For the X and Y axes, the number of gaze points in the AOI of Auto Pilot, ECAM, Left PFD, and Table under DPO and SPO scenarios is higher than that in SPO-HUD scenarios. While SPO-HUD showed a significant reduction in the number of fixations on both PFD and ECAM but increment on the AOI of HUD_1. Under the X-Z coordinate axis, there is no significant difference in heat maps among the 3 scenarios. Under the Y-Z coordinate axis, the position of the gaze point under DPO and SPO scenarios is slightly lower than that under SPO-HUD scenarios. It is evident that when wearing the HUD, the pilot places more focus on the HUD rather than looking down for other dashboard information. Also, with the HUD installed, the pilot's eyesight is more evenly distributed.

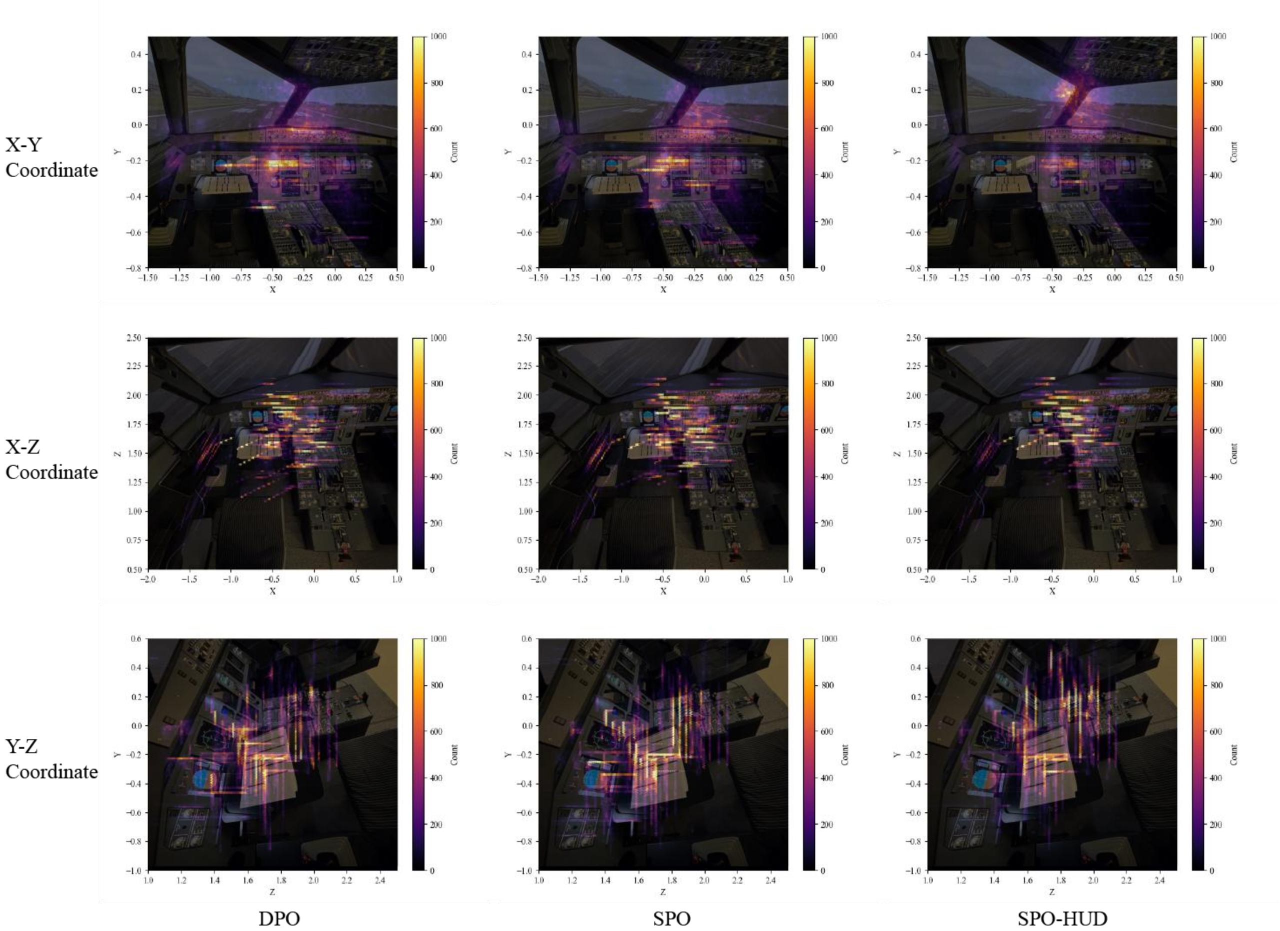


Figure 5-7. The heat map of gaze distribution of all subjects during the three different scenarios from all participants.

5.4.4. Post-experimental feedback

This experiment also designed a user feedback questionnaire for user experience. In the question 'Which flight mode do you prefer?', 11 subjects chose DPO mode, 6 subjects chose SPO-HUD mode, and the remaining 4 subjects chose SPO mode. In the evaluation of the degree of assistance of AR-HUD in completing flight missions, more than half of the subjects believed that AR-HUD was helpful to some extent (14/21), while a small number of subjects believed that the assistance was low or not helpful (7/21). However, wearing HoloLens generates some level of heat over time, which can affect user comfort. Eleven participants revealed that wearing AR-HUD would cause some discomfort, 3 subjects were neutral on the comfort level, and 6 participants felt more comfortable. Finally, due to the incomplete development functionality of AR-HUD, in the post experimental feedback, 10 subjects believed that AR-HUD was relatively satisfied with the overall performance, 8 subjects felt somewhat dissatisfied, and 6 subjects were in a neutral state.

5.5. Discussion

The study is to contribute to the technology, to the next generations of such support-devices and CHMI in cockpit design and to ideas about how to optimise and improve aviation safety and pilot experience. Specifically, this work uses SA-probe reaction time, correctness, eye-tracking data in HoloLens, handling time for emergency, and post-experiment questionnaire to analyse the effect of AR-HUD on pilot's SA and workload between SPO and DPO scenarios based on flight/neuro-physiological performance difference.

The HMI design in aviation is of paramount importance, as it directly influences pilot performance, HFs, and flight safety. With the advanced automated systems, the primary duties of pilots are progressively shifting towards supervisory roles, intervening only when necessary (Post, 2021). These new roles require a corresponding evolution in the HMI. Firstly, current HMI design in flight cockpit is static and do not take into consideration the dynamic variations in HFs, such as mental workload, SA, and so on. Secondly, Industry 5.0 enables the integration of human intelligence and capabilities with advanced technologies. Since cybernetics emphasising human centredness and HCI serving as a medium for human-machine communication, HCI enables real-time feedback mechanisms, adaptive control, and intelligent decision-making, empowering individuals to actively participate and contribute to the intelligent engineering processes (Mehrali et al., 2018). By reconfiguring the HMI according to the operator's tasks and behaviours has been shown to improve performance (C. A. Miller & Hannen, 1999). The advancements in display technologies, particularly HUD, have significantly enhanced the presentation and integration of crucial flight information. Assisted technologies have proven highly beneficial in mitigating information overload and effectively supporting pilots in recovering from degraded performance, particularly during critical flight phases like landing (Y. S. Liang et al., 2023). These technologies provide valuable assistance that aids pilots in managing complex tasks and making crucial decisions, ultimately enhancing safety and performance in challenging situations. J. Liu et al. (2016) also explored the cognitive pilot-aircraft interface (CPAI) for SPO. CPAI implemented new sensing technologies to monitor pilots' cognitive capacity and adapted individuals to support decision-making. Moreover, AR technology allows for improved usability and intuitive interaction with the HMI in the cockpit, ensuring efficient and error-free interaction between the pilot and the cockpit.

These multi-dimensional results show that the pilot suffered the least SA-loss to some extent under SPO-HUD scenario. In this study, the SA level in the 3 operated modes was rated by combining reaction time and correctness, and SPO-HUD scenario had the highest number of SA levels labelled as high. In the three-level SA distribution, DPO and SPO had similar distributions. Results indicate that access to the AR-HUD led to significant improvements to pilots' SA level. The presentation of information in a way that does not require the pilot to divert visual attention and cognitive resources into the cockpit away from external events and primary flight references such as possible with a HUD has been documented to reduce workload and increase SA level (Neville A. Stanton et al., 2020). It has been demonstrated that the display of synthetic environmental information in HUD, as opposed to Head-Down Display (HDD), enhances pilot flight performance (Prinzel Iii et al., 2004). Whilst previous research has demonstrated that HUD can be detrimental to SA, particularly when task-irrelevant information is presented in demanding situations (Richards & Lamb). Therefore, such assistance methods should only offer help when necessary, as providing unwarranted assistance can also lead to human errors (Ververs & Wickens, 1998). Too much information presented in the view will degrade the detection of unexpected events due to attentional tunnelling (Ernst et al., 2021). This study has simplified the layout of the HUD design during cruising phases to deduce the distraction to some extent.

Precise evaluation of mental workload can aid in preventing operator errors and enable timely intervention by forecasting performance deterioration resulting from either excessive workload or under-stimulation (Aricò et al., 2015). The researchers are interested in adjusting the difficulty of the task and the allocation of responsibilities accordingly by dynamically monitoring the operator's mental abilities (Pang et al., 2023). The result of self-rated workload also indicated a reduction in pilot workload. In addition, as expected, the perceived workload of the PF increases in the SPO scenario due to the absence of the PM. For example, the communication between PM and ATC, or some other monitoring tasks that would otherwise be performed by PM, are all done by the PF. Therefore, the perceived workload of the PF is higher especially during take-off and before landing. In other words, the absence of PM might induce a decrease in captains' SA and operational performance, and increase their workload, leading to a higher possibility of encountering flying mistakes.

For the emergencies handling, all 3 operated modes show significant differences at different levels of emergencies. When a fuel tank leak occurs, the HUD pops up a leak warning, so the pilot can detect the relevant event faster. In the DPO environment, the 2 pilots check the status information in ECAM at irregular intervals, so the time to monitor a tank leak is lower than in the SPO scenario. The SPO scenario shows the worst performance in monitoring the events without an alert. When dealing with emergencies, the HUD converts the instruction to an oral guidance that follows the pilot processing steps to guide the completion of the emergency. This allows the pilot to spend less time continually looking at ECAM and be able to respond more quickly and correctly(S. H. Kim & Kaber, 2014).

Besides the results reported above, the cross-correlations of flight operations and eye-movement behavioural pattern shifts were revealed towards each scenario. In the flight cockpits, pilots encounter high mental demands as they are required to monitor, comprehend, and intervene in numerous complex automatic operating modes (Sarter & Woods, 1995). In general, pilots appear to spend more time fixating on instruments compared to what they see outside the cockpit (about 60–70% on instruments and about 30–40% on the outside world (Colvin et al., 2005; Helleberg & Wickens, 2003;

Christopher D Wickens et al., 2017). To effectively fulfil these tasks, pilots may need to adapt specific visual scanning patterns that vary based on the aircraft's state and mode of operation. One of the purposes of this study is to analyse visual scan patterns in different scenarios, in order to explore the ability of HUD for pilots to monitor external events. Eye-tracking results indicated there existed different scan-patterns between 3 operated scenarios. By the visualisation of the gaze point, it is found that 3 operated modes differ from each other in 3 dimensional coordinates. DPO and SPO show similar gaze patterns with more attention put on ECAM and PFD, while SPO-HUD scenario focuses more on HMI in HoloLens. It can also be observed in the heat map of Z-Y coordinates. The heat map also revealed that SPO-HUD scenario showed a more equitable distribution of gaze between instruments and external view. It is worth noting that the position of the gaze point is lower in the DPO and SPO scenarios than in the SPO-HUD scenario. With the use of a HUD, the pilot's eye does not need to switch back and forth between the external view and the inside instrument, reducing the visual burden on the pilot (Fadden et al., 2001; X. Hu & Lodewijks, 2020). Similar studies have also shown that pilots' flight performance is higher when using a HUD than when using other types of displays (W. C. Li et al., 2020; Neville A. Stanton et al., 2019).

In addition, a percentage of subjects in the post-experimental feedback questionnaire reported that AR-HUD assistance to be minimal or unhelpful. One possible contributing factor is the current limitations in the functionality of the AR-HUD. This study primarily focused on displaying essential aircraft status information such as altitude, speed, pitch, and roll angles, while certain features such as the Flight Mode Annunciator (FMA) and heading/track guidance were not fully integrated into the AR-HUD. These additional features have been shown to play a significant role in enhancing SA and decision-making in HUD systems. Thus, the absence of these functionalities in the current study may have impacted the perceived effectiveness of the AR-HUD. Another potential factor is the participants' varying levels of familiarity and experience with HUD usage. The composition of our participant group, which primarily consisted of cadet pilots, may have contributed to their varying levels of familiarity and experience with HUD usage. While the use of HUD is part of their training curriculum, it is possible that some participants had not received extensive or systematic training in HUD usage, leading to a lack of familiarity and limited understanding of its full functionality. This lack of familiarity with HUD technology and its functionalities could have influenced their perception of the AR-HUD assistance. To better address these challenges, future research should focus on improving the functionality of the AR-HUD to include essential features such as FMA and flight direction indicators. Additionally, a more diverse participant pool, including pilots with varying levels of HUD experience, should be considered to ensure a comprehensive understanding of the benefits and limitations of the AR-HUD in real-world flight scenarios.

5.6. Limitations

The development of the AR-HUD functionality is still in the preliminary stages and the extracted flight data is limited. However, the function of AR is not only limited to the presentation of data; AR-HUD can also be used for movement guidance. For example, highlighting the panel that should be operated with arrows and highlighting instead of just the current voice alerts. There was one subject reported that using the AR-HUD may produce a distraction where attention is distributed in the icon of the HUD. The distraction is called cognitive tunnelling. As mentioned in the discussion section, some irrelevant information presented on the HUD could capture pilots’ attention and cause misaccommodation. Excessive information of low relevance to this phase can reduce the pilot's ability to monitor external events. Follow-up research could address the issue of cognitive tunnelling to some extent by further analysing the HMI design and the key information required for different flight phases to be redesigned.

Whatmore, understanding pilots' comfort and acceptability is crucial when evaluating the feasibility of implementing AR-HUD in SPO scenarios. Participants reported that the discomfort and heat dissipation when wearing HoloLens 2. Indeed, the development of AR-HUD wearable devices is an ongoing process that requires advancements in technology to address various factors contributing to pilot comfort and acceptance. Currently, many AR glasses available on the market tend to be relatively heavy, which can cause discomfort and fatigue when worn for extended periods. Technological progress is vital to reduce the weight of these devices, making them more comfortable for pilots to wear during flight operations. On the other hand, future ergonomic research could involve studying factors such as the design of the AR-HUD system, ergonomics of the wearable device, and the integration of advanced technologies to minimise weight, improve comfort, and provide a user-friendly experience. Pilots will then be willing to use and wear the AR-HUD for extended periods of time, making it a real application. In summary, with the rising human-interaction technologies, these limitations will be improved and discussed in the future development of the AR-HUD.

5.7. Concluding remarks

This exploratory preliminary AR-HUD development was validated in different operated mode. First of all, an AR-HUD, which could be used as an interaction medium for HCI, was designed based on Microsoft HoloLens 2. Then, this study conducted an experimental study (n = 21) to investigate the effect and feasibility of AR-HUD on SA and workload of pilots between DPO and SPO scenarios. SPAM-based SA probe and self-rated workload scale were designed to monitor and evaluate the SA and workload level during the entire flight phases. HoloLens captures the location and velocity of the gaze point to further analyse behavioural performance. In addition, this study also triggered three-level of emergencies during the flight task and collected the completion time as one of the factors for evaluating AR-HUD. Therefore, this study combined both subjective and objective methods to explore the feasibility of AR-HUD. The main conclusions and contributions are summarised as follows:

- The perceived mental workload of pilots was lower with the AR-HUD. Especially during the take-off and cruise phases, the perceived workload during SPO-HUD scenario is significantly different from both DPO and SPO scenarios.
- Wearing the HUD improved pilot performance in answering SPAM-based questions (higher correctness and shorter reaction times) and the labelled SA levels were higher than those of DPO and SPO, especially for departure, cruise, approach, and landing.
- When dealing with three levels of accidents, wearing a HUD has a positive effect on correctly detecting and dealing with accidents, compensating to some extent for missing PM.
- SPO-HUD, DPO and SPO presented different visual scan patterns: in the DPO and SPO scenarios, the pilot focused more on the PFD and ECAM, but in the SPO-HUD scenario, the pilot focused more on the HUD regions and external view in the AR glasses. In addition, the subject's gaze point had higher coordinates and more focussed vision in the SPO-HUD scenario.

In conclusion, the integration of HCI with new interaction technologies, specifically the AR-HUD in this case, places greater emphasis on the human factors, aiding in maintaining the pilot in an optimal state. This study serves as an exploratory investigation, laying the foundation for the development of an adaptable pilot-support system with the AR-HUD serving as the human-centric interactive terminal.

Chapter 6. Optimising Pilot-Aircraft Interaction: A Low-Cost Projection-Enhanced Head-Up Display (PE-HUD) with Neuro-Ergonomic Validation

Building upon the evaluation of the adaptive AR-HUD utilising AR Glasses in the preceding chapter, this chapter pursues two primary objectives. First, it introduces and validates an alternative Augmented Reality (AR) presentation format: the Projection-Enhanced HUD (PE-HUD). This approach aims to circumvent the adoption barriers associated with traditional AR glasses solutions, such as visual conflicts in head-mounted displays and the prohibitive costs of waveguide Head-Up Displays (HUDs). Second, this chapter seeks to further verify the effectiveness of this new HUD system specifically during critical flight phases.

The proposed low-cost PE-HUD achieves collimated imagery through retroreflective film-screen symbiosis, eliminating the need for complex optics. Based on subject feedback from the previous AR-HUD experiment, the PE-HUD interface was refined to provide greater emphasis on the flight control director. Distinct from the methodology employed in the last chapter, this study adopted functional Near-Infrared Spectroscopy (fNIRS) as a novel approach to evaluate the PE-HUD's impact during specific flight phases, namely takeoff and landing. These phases were chosen due to their demonstrated higher cognitive workload and greater potential to benefit from HUD assistance. This research establishes neuro-ergonomic evidence for phase-sensitive HUD validation, thereby advancing cost-effective HMI interface design in aviation.

6.1. Introduction

6.1.1. Pilot-aircraft interaction in modern cockpit

Modern aviation operations demand pilots continuous monitoring of multi-source cockpit instrumentations, creating complex human-computer interaction (HCI) challenges. Pilots must simultaneously process primary flight parameters (e.g., pitch, altitude), navigation data, and system status indicators across spatially distributed displays (Christopher D Wickens et al., 2017). Such operational demands impose dual cognitive costs: degraded situational awareness (SA) from excessive visual scanning and elevated mental workload (Dorneich et al., 2016). Specifically, this fragmented information architecture imposes high visual scanning demands, leading to cognitive tunnelling effects where critical signals may be overlooked during high-workload phases (C. D. Wickens, 2008; Christopher D. Wickens & Alexander, 2009). While conventional head-up displays (HUDs) partially mitigate these issues by projecting critical data within pilots' forward line of sight (Aromaa et al., 2020), their static symbology and prohibitive retrofitting costs limit effectiveness in modern glass cockpits.

6.1.2. Projection-Enhanced HUD: A cost-effective enhanced display technology

Augmented Reality (AR) technologies demonstrate potential for HCI enhancement through context-aware visualisation (Gardony et al., 2021; Spicer et al., 2017). The dynamic cues offered by augmented visualisation in AR devices have the potential to elevate SA levels by directing attention more effectively across the flight deck (W.-C. Li et al., 2020). Therefore, integrating pilot aids, such as augmented visualisation like AR technique, with HUD system, should be a feasible consideration in the early stages of flight desk design to enhance interaction efficiency and aviation safety (W.-C. Li et al., 2022; N. A. Stanton et al., 2016). Experimental AR-HUD implementations, such as terrain outline projections onto runway axes during low-visibility approaches, have decreased deviation errors in the tracking task (S. H. Kim & Kaber, 2014). However, two fundamental barriers hinder aviation adoption of AR systems. First, optical see-through head-mounted displays (OST-HMDs) induce visual accommodation conflicts between virtual content and external scenes, causing simulator sickness (Yuan, Ng, Li, et al., 2024). Meanwhile, waveguide-based HUDs require precision optical systems costing over $500,000, making fleet-wide retrofits economically impractical.

To address these constraints, a projection-enhanced HUD (PE-HUD) leverages retroreflective film-screen symbiosis. This configuration achieves collimated virtual image projection without complex optical assemblies by projecting flight symbology onto a windshield-mounted retroreflective surface via a transparent reflective film.

6.1.3. Validation with neuro-ergonomic approaches

Traditional HMI interface evaluation paradigms in aviation predominantly rely on performance metrics (e.g., reaction time, error rate) and subjective feedback such as NASA-TLX (Biernacki & Lewkowicz, 2024), which lack exploration in decoding cognitive adaptation patterns during complex HCI. Neuro-ergonomic approaches address this limitation by providing direct biomarkers of cerebral hemodynamic responses and visual scanning behaviour (Mehta & Parasuraman, 2013). For the neural cortical activity, fNIRS enables the quantification of phase-specific executive and activated levels through oxygenated/deoxygenated haemoglobin dynamics. Eye-tracking captures scan path entropy, Areas of Interest (AOI) transition frequencies, and dwell time ratios between synthetic HUD elements and physical instruments (Li et al., 2015). This integrated approach facilitates three-tier validation: cognitive mechanism analysis of information processing, correlation of neural efficiency with flight phase transitions, and groundwork for adaptive interfaces through neurofeedback integration.

6.1.4. The current study

In order to explore a more natural interaction paradigm, this study developed a PE-HUD system that displays and alerts critical flight information by projecting it onto the front windshield. This study applies neuro-ergonomic techniques to assess the cognitive demands and SA of pilots assisted by the PE-HUD in low visual conditions during critical flight phases such as take-off and landing. This cost-adaptive approach redefines the economics of cockpit upgrades while maintaining critical flight path awareness.

6.2. Methodology

6.2.1. Participants

This study involved the evaluation of ten cadet pilots from commercial airlines. Participants' ages ranged from 22 to 38 years (mean age = 30.20 ± 7.00 years). The study also recorded the number of training hours completed by each participant, which varied from 80 to 500 hours, with an average of 136.2 ± 172.69 hours. All participants had undergone professional training on the Airbus A320. Each participant was awarded HKD $100 in supermarket vouchers as an incentive. This experiment was conducted in the Human Factors and Ergonomics Laboratory, Department of Aeronautical and Aviation Engineering, The Hong Kong Polytechnic University, HKSAR. This study was conducted following the procedure under the ethical standards of the 1975 Helsinki Declaration. It was also approved by the Hong Kong Polytechnic University Institution Review Board regarding clinical human subjects in advance (No.

HSEARS20210318002).

6.2.2. Apparatus

The equipments required for the experiment are mainly: 1) a flight simulator, 2) the developed PE-HUD, 3) brain physiological collection equipment (Functional Near-Infrared Spectroscopy, fNIRS), and 4) scanning behaviour collection equipment (eye tracker). Figure 6-1 shows the developed PE-HUD prototype.

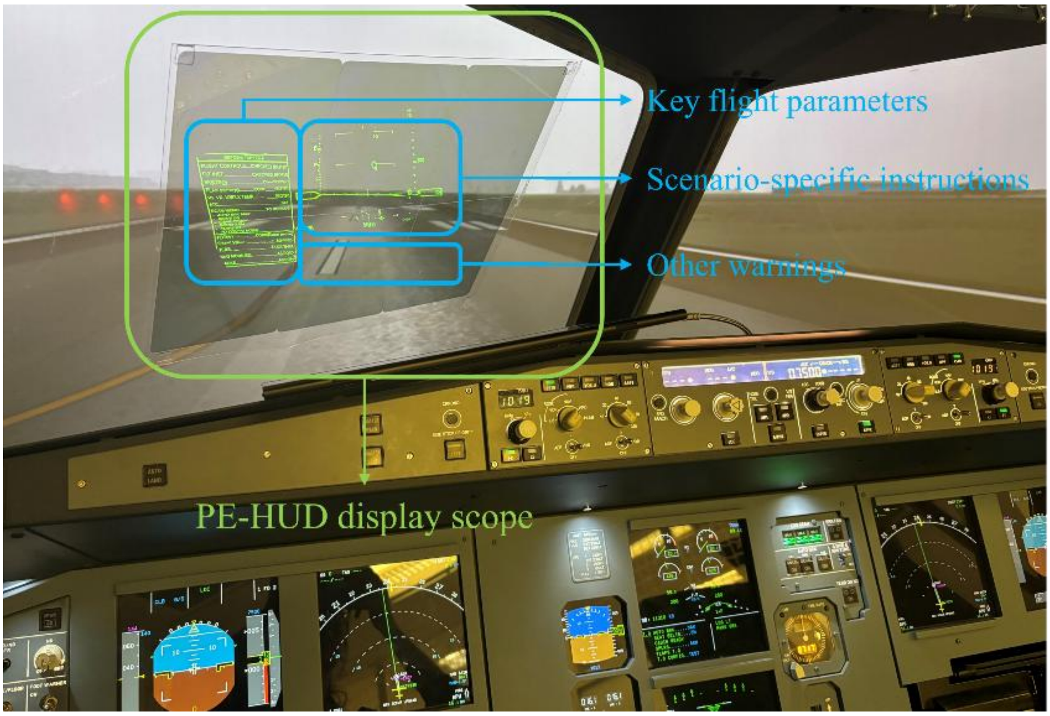


Figure 6-1. Projection-Enhanced Head-Up Display (PE-HUD) prototype.

6.2.3. Projection-enhanced HUD development

Figure 6-2 shows the framework for developing PE-HUD systems. The development of the PE-HUD system began with the interface design using the Unity game engine. According to the requirements of each critical flight phase, this study designed different interfaces, including the format, alert, and guidance design. To enable the PE-HUD system to display real-time flight data, Simconnect API was utilised to establish a connection with the Prepar3D V5 flight simulator. This allowed the Unity-based HUD interface to receive a continuous stream of live aircraft data, including parameters such as airspeed, altitude, heading, and engine performance.

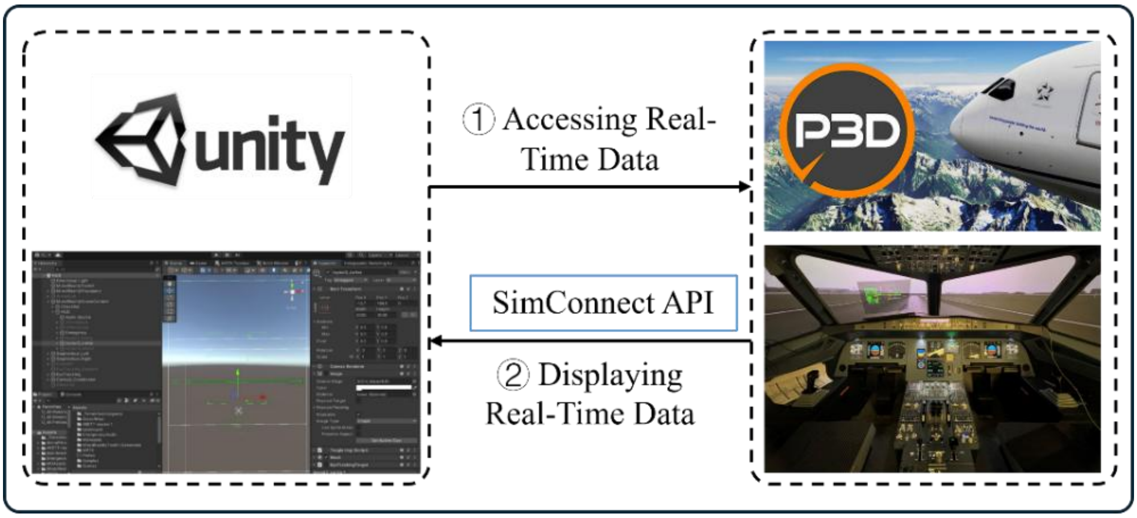
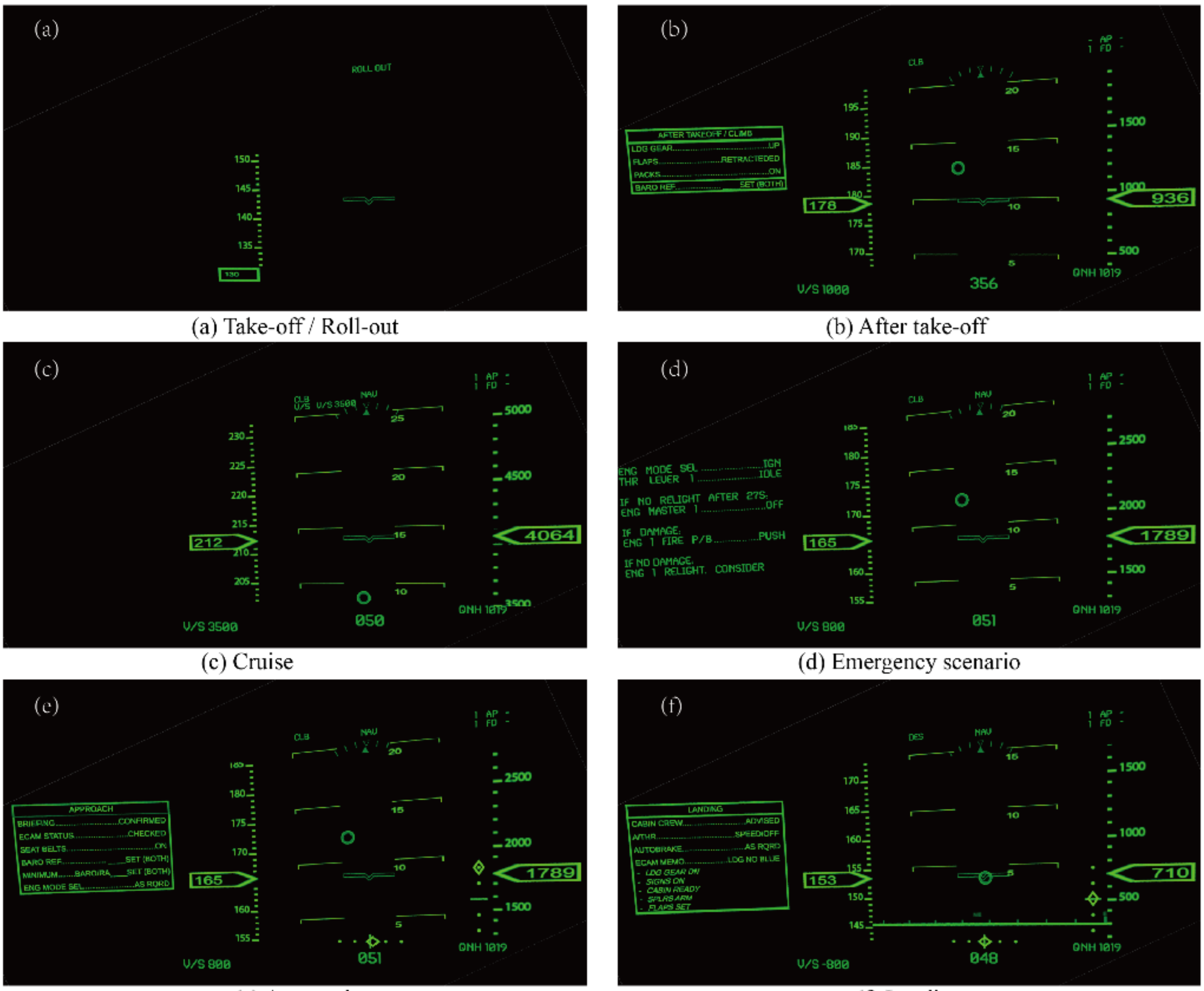


Figure 6-2. The framework of developing PE-HUD systems.



(a) Take-off / Roll-out (b) After take-off (c) Cruise (d) Emergency scenario (e) Approach (f) Landing

Figure 6-3. Adaptive HUD interface design

6.2.4. Experimental procedures

The experiment comprised two counterbalanced sessions: 1) Baseline SPO: Single-pilot operation (SPO) without HUD; 2) HUD Intervention: SPO with HUD. A custom flight scenario replicated a round-trip RCTP-RCMQ route (Figure 6-4). Each 15–20 minute session included all flight phases: takeoff, departure, climb, cruise, descent, approach, and landing. Weather conditions featured cirrocumulus clouds with 5 km visibility between 1,000-5,000 ft. Additionally, this design evaluated pilot emergency response capabilities to sudden engine failures under controlled conditions. So each session incorporated one randomised remediable emergency: ENG-1 or 2 flameout.

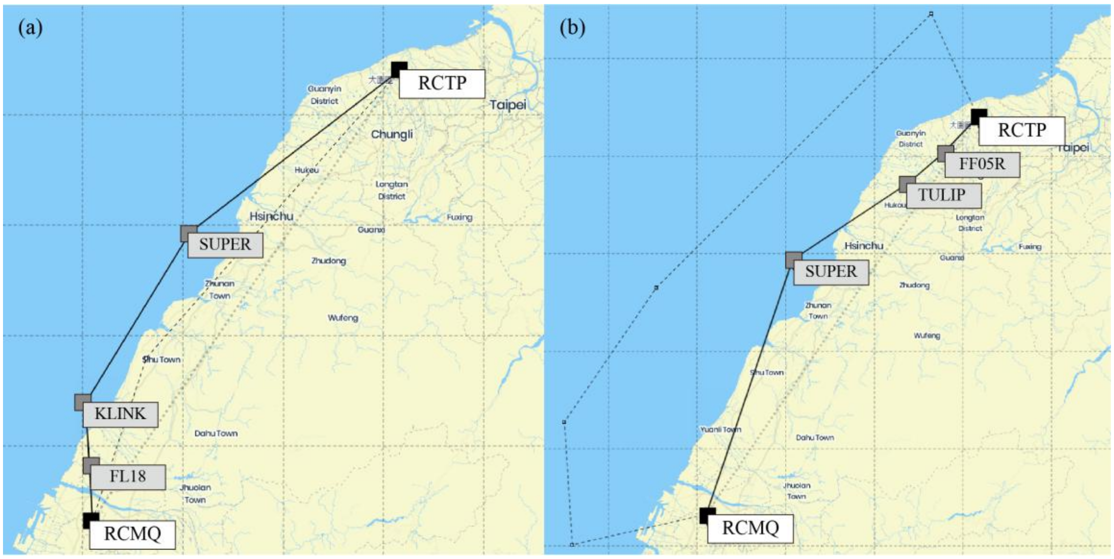


Figure 6-4. The flight route between RCTP and RCMQ: (a) the route from RCTP to RCMQ, (b) the route from RCMQ to RCTP.

In order to avoid a learning effect on the subjects, the experimental sequence of SPO and HUD was assigned randomly so as not to affect the results of the experiment. The experiment was paused at specific points during the pilot's execution of takeoff and landing flight tasks to insert the SAGAT probe. Specifically, the experiment was paused after the pilot successfully raised the landing gear during takeoff and when the altitude was lower than 2500 ft and the landing gear was down during landing phase. Subjects were required to answer six consecutive questions about the current flight phases' critical information.

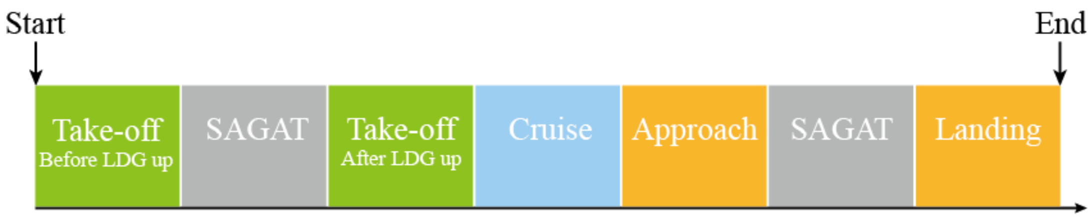


Figure 6-5. Experimental procedure of each session.

6.2.5. Neuro-ergonomic data acquisition

Neurophysiological data. In this study, the Nirsmart (3000A), developed by Jiangsu Danyang Huichuang Medical Equipment Co., Ltd., was employed as the blood haemoglobin data acquisition device. The device operated with a sampling frequency of 10 Hz, a filtering parameter of 2.0 Hz, an analogue-to-digital conversion rate of 24 bits, and wavelengths of 760 nm and 850 nm, respectively. A total of 35 Source-Detector (SD) probes—comprising 19 sources and 16 detectors—were positioned across several brain regions: the prefrontal cortex (PFC), left motor cortex (LMC), right motor cortex (RMC), left occipital lobe (LOL), right occipital lobe (ROL), left temporal lobe (LTL), and right temporal lobe (RTL).

Eye-tracking data acquisition. This study used the Tobii Glasses 3 eye-tracking system as the data acquisition device. The Tobii Glasses 3 is a wearable, binocular eye-tracking system with a sampling frequency of 100 Hz. The Tobii Glasses 3 records eye movement data, including fixations, saccades, and blink information, which are captured and stored for subsequent analysis.

6.2.6. Experimental procedures

The experiment consisted of two sessions in total. These sessions include 1) single-pilot operation (SPO) session: SPO without a HUD, and 2) HUD session: SPO with HUD, as summarised in Table 6-1. An experimental flight scenario featuring a round-trip route between Taiwan Taoyuan International Airport (RCTP) and Taichung International Airport (RCMQ) was developed. Each session of the flight, encompassing take-off, departure, climb, cruise, descent, approach, and landing, lasted approximately 15 - 20 minutes. Weather conditions included cirrocumulus clouds with visibility set at 5 km, spread between altitudes of 1000 to 5000 feet. Additionally, two remediable emergency events were introduced to two sessions to add complexity to the flight operations: an “ENG-1 flameout” and an “ENG-2 flameout”. Each flight session featured one of these emergencies.

Table 6-1. Session allocation.

Session	HUD available	Scenarios	Flight route	Emergency
1	N	SPO	RCTP-RCMQ	ENG-1 Flameout
2	Y	HUD	RCMQ-RCTP	ENG-2 Flameout

To ensure all participants had an equal level of familiarity with the interface and interactive modes of the PE-HUD, a tutorial briefing and practice session on the PE-HUD device were conducted prior to the trials. Comprehensive information regarding the flight simulation—such as flight plan and route—was provided. Following the briefing, participants were equipped with Tobii Glasses 3 to capture eye-tracking data and fNIRS to monitor brain activity throughout both sessions.

6.2.7. Data processing

6.2.7.1. Physiological data processing

NirSpark analysis software (v1.8.1) was used to process the raw data obtained from the fNIRS acquisition, operating within the MATLAB environment (MathWorks, USA). The data processing was carried out in the following steps.

- Step 1: Converting the raw data (Light intensity) into optical density data.
- Step 2: Denoising and filtering optical density data.
- Step 3: Calculating the relative change of haemoglobin concentration by modified Beer-Lambert law.

- Step 4: Superimposing and averaging the blood oxygen concentrations collected during the three flight phases using the Hemodynamic Response Function (HRF). A generalised linear model (GLM) was applied to fit the HRF to the analysis.

6.2.7.2. Eye-tracking data processing

For visual scanning behaviour analysis, Times of Interests (TOIs) were established and associated with two critical flight phases, namely Take-off and Landing, via Tobii Pro Lab (v 1.181.37603). The Aera of Interests (AOIs) were shown in Figure 6-6 to explore the scanning behaviours in flight phases and with or without HUD.

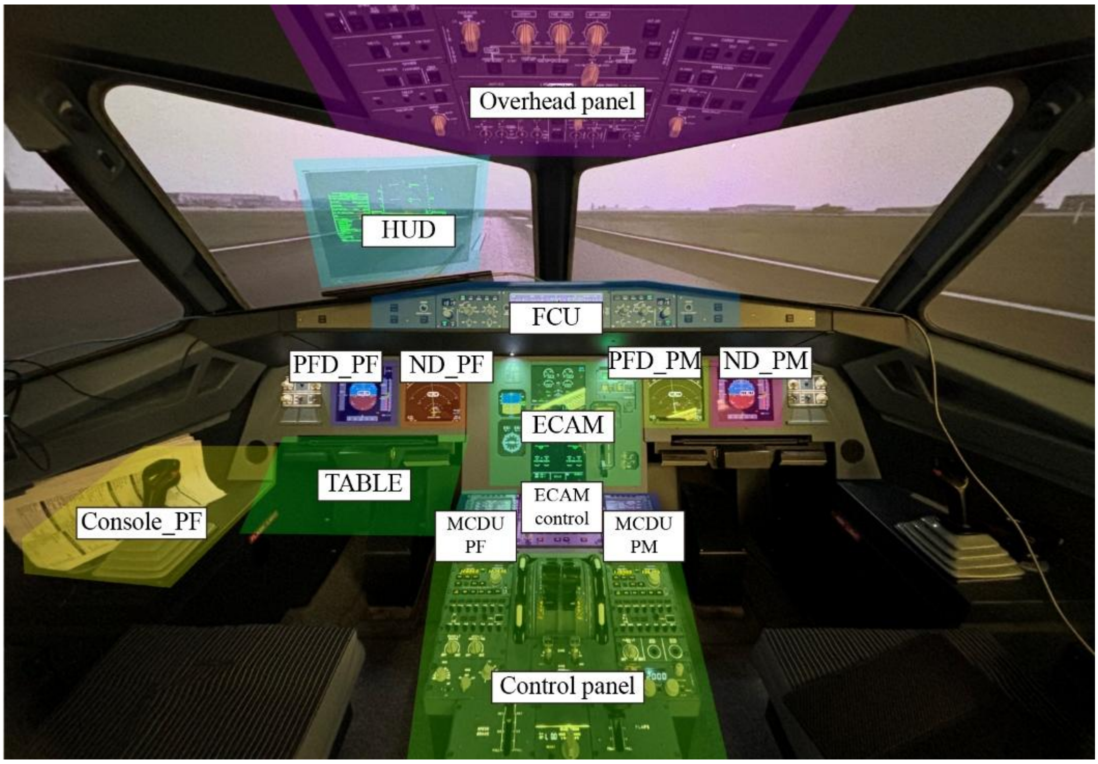


Figure 6-6. AOI configuration.

6.3. Result

6.3.1. Physiological result

Figure 6-7 shows the difference between the take-off and landing phases in SPO and HUD scenarios. Pilots exhibited higher beta values, indicative of increased neural activity, when using the PE-HUD system compared to the baseline condition, particularly in the occipital lobe (OL). This effect was more pronounced during the landing task compared to the take-off task. Specifically, the contrast between the PE-HUD and baseline conditions showed stronger activation differences in the landing phase across several brain regions, including the PFC, OL, and motor cortex (MC).

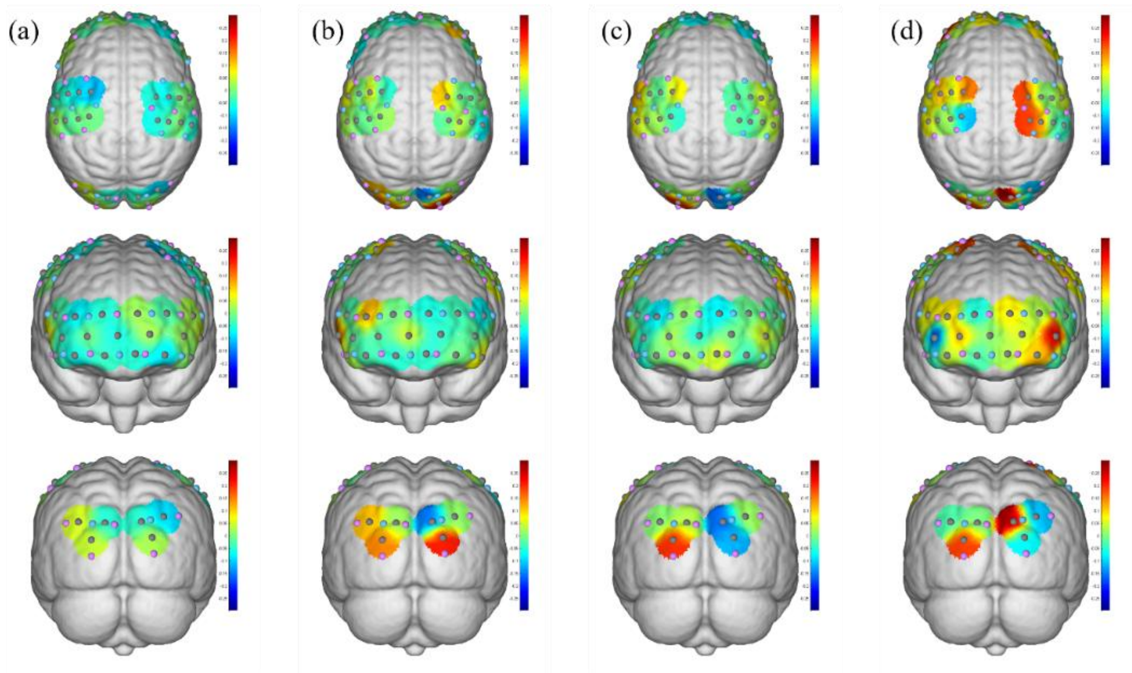


Figure 6-7. Beta value of different critical flight phases in two scenarios. (a) T.O in SPO scenario, (b) T.O in HUD scenario, (c) Landing in SPO scenario, (d) Landing in HUD scenario.

6.3.2. Visual scanning behaviour

The heat map (Figure 6-8) compares fixation distributions between the two scenarios over the main panel. Regarding the entire flight phase, the results indicated that the PE-HUD system increased the frequency of pilots' head-up glances, resulting in a rise in the proportion of time spent observing the external visual environment. Furthermore, the PE-HUD condition also resulted in an increased visual emphasis on the flight control unit (FCU). Conversely, the PE-HUD condition was associated with reduced fixation time on the Primary Flight Display (PFD) and Navigation Display (ND). Table 6-2 examines the differences in the eye-tracking metrics across the main instruments using paired t-test. Conversely, the AR-HUD condition was associated with reduced fixation time on the MCDU, PFD and ND, and the increment on the HUD, indicating that the augmented visual information provided by the HUD helped to offload some of the pilots' instrument scanning requirements.

Table 6-2. The summary of eye-tracking metrics with significant difference between SPO and AR-HUD scenarios.

Name of Metrics	Name of AOIs	P-value
Average duration of fixations	FCU	0.04
Total duration of fixations	Console-PF	0.03
	HUD	0
	MCDU-PF	0.01
	ND-PF	0
	PFD-PF	0.03
	Table	0.03
Number of fixations	Console-PF	0.03
	HUD	0
	MCDU-PF	0.01
	ND-PF	0
Number of saccades in AOI	ECAM-Control	0.03

	HUD	0
	MCDU-PF	0
	ND-PF	0
Number of glances	HUD	0
	MCDU-PF	0.01
	ND-PF	0
	Table	0.02
Total duration of glances	Console-PF	0.03
	HUD	0
	MCDU-PF	0.01
	ND-PF	0
	PFD-PF	0.03
	Table	0.03

When comparing the critical flight phases (take-off and landing), the take-off phase increased the focus on altitude and airspeed information, aligning with the heightened sensitivity to these parameters in the initial climb-out. In addition, the greater emphasis on pitch angle and altitude cues during landing highlights the system's ability to provide salient and relevant information in the final approach and roll-out phases.

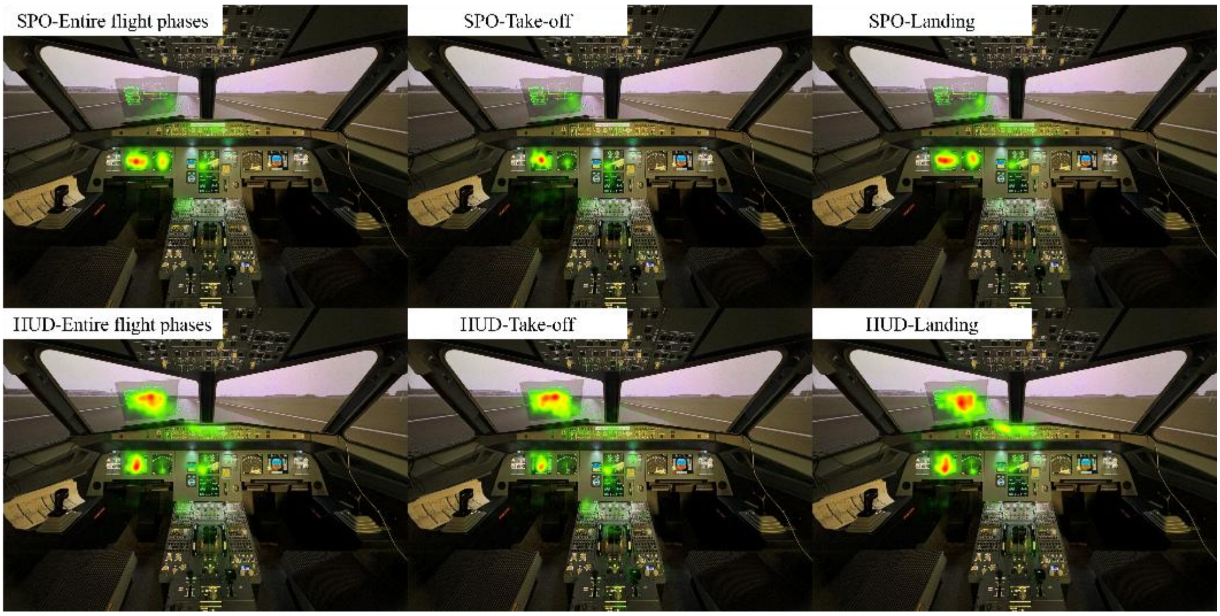


Figure 6-8. The comparison of fixation distribution between take-off, cruise, and

landing phases with or without PE-HUD.

6.4. Discussion

The higher beta values recorded in the fNIRS data indicate increased neural activity and arousal in the underlying cortical regions. Beta oscillations (13-30 Hz) are typically associated with focused attention, information processing, and heightened cognitive engagement (Q. Li et al., 2021). The observed increase in cortical activation, especially in the PFC, OL and MC, suggests that the PE-HUD system enabled pilots to engage more cognitive resources and maintain higher SA during the challenging landing phase of flight. The elevated beta responses observed in the PE-HUD condition, particularly in the OL responsible for visual processing, suggest that pilots were able to more effectively allocate attentional resources and engage in more active information extraction from the augmented visual environment. Similarly, the increased prefrontal and parietal activations point to the PE-HUD's ability to support higher-level cognitive functions, such as decision-making, spatial awareness, and task planning, which are essential for safe and precise landing manoeuvres (Ahlstrom, 2015). The motor cortex activation further implies that the PE-HUD facilitated more coordinated and responsive control inputs from the pilots, potentially improving their ability to maintain aircraft control and trajectory during the final approach and touchdown (R. Li et al., 2021). In these low visual conditions, pilots must rely more heavily on the information provided by the aircraft's instrumentation and displays to maintain SA and execute a successful landing (Ahlstrom, 2015). The results indicate that the PE-HUD system was particularly effective in supporting pilots' cognitive processes and resource allocation during this critical phase of flight.

The increased head-up glances and proportion of time spent observing the external visual environment in the PE-HUD condition suggest that the enhanced display enabled pilots to maintain better SA and a more holistic understanding of the flight environment. The PE-HUD system appears to have facilitated a more integrated and seamless information processing strategy by reducing the need for frequent transitions between the HUD and other cockpit displays. The heightened visual attention on the FCU in the HUD session indicates that pilots were more actively engaged in monitoring and controlling the aircraft's state, potentially leading to improved flight path management and responsiveness. The reduced fixation time on the PFD and ND, in turn, implies that the PE-HUD was able to offload some of the pilots' instrument scanning requirements, allowing them to dedicate more cognitive resources to other essential tasks.

These findings provide neuro-ergonomic evidence that the PE-HUD technology can enhance pilots' SA and cognitive engagement, particularly during high-workload scenarios such as instrument landings. The increased cortical activation, especially in the occipital, prefrontal, and parietal regions, suggests that the PE-HUD system enabled pilots to perceive, comprehend, and project the evolving flight environment, potentially contributing to improved operational performance and safety more efficiently.

Specifically, based on the above two feasible AR-HUD realisation, we can further develop the pilot & system-centred training scheme:

First, immersive SD-exposure training can be implemented by leveraging the projection-based and wearable AR-HUD systems to simulate degraded visual environments, crosswind disturbances, and ambiguous motion cues under controlled simulator conditions. Instead of merely briefing pilots about SD, the system can deliberately induce graded levels of sensory ambiguity, allowing trainees to directly experience illusion-prone situations while maintaining operational safety. By integrating real-time neuro-ergonomic monitoring (e.g., EEG-derived patterns and visual scanning deviations), the training environment can dynamically adjust task difficulty to ensure that different SD sub-types are sufficiently elicited. This approach transforms SD training from passive awareness instruction into an adaptive, mechanism-driven exposure paradigm.

Second, personalised vulnerability profiling can be established through the joint modelling framework developed in Chapters 4. By analysing each trainee's neurophysiological responses and scanning-sequence characteristics across flight phases, the system can identify individual susceptibility patterns, such as a tendency toward attentional narrowing, delayed instrument cross-checking, or increased cognitive instability under workload escalation. Based on these profiles, targeted training sessions can be designed to reinforce phase-specific scanning strategies and strengthen instrument reliance under degraded cues. In this manner, the AR-HUD does not merely display information, but functions as a training mediator that translates state-recognition outputs into individualised corrective guidance.

Third, guided correction and progressive cue fading can be embedded within training sessions to prevent over-reliance on augmented cues. When early signs of disorientation or loss of attentional control are detected, the AR-HUD can provide minimal yet salient visual prompts to redirect attention toward critical instruments or restore balanced scanning patterns. As trainees demonstrate improved stability and scanning consistency, these prompts can be gradually reduced, encouraging the internalisation of effective cognitive strategies. This fade-out mechanism ensures that AR assistance serves as a scaffold rather than a substitute for pilot cognition.

Finally, scenario-based resilience training can incorporate decision-gate reinforcement under time pressure and environmental uncertainty. By simulating weather degradation or operational complexity within plausible boundaries, trainees can practise recognising early cognitive destabilisation and executing predefined recovery scripts, such as stabilising attitude, prioritising instrument cross-check, or adjusting flight strategy. Through repeated exposure within a closed-loop sensing and feedback framework, pilots can develop not only technical proficiency but also cognitive resilience against SD-triggering conditions.

Lastly, the technology-centred HMI system will inherently introduce the risk of automation surprise and overreliance. Sudden, opaque, or overly directive interventions may disrupt SA or encourage passive dependence on system guidance. This adaptive HMI design driven by 3A-PAI system (proposed in Chapter 4) mitigates these risks through conservative intervention principles. First, AR-HUD cues are minimal and state-triggered rather than continuously prescriptive, reducing cognitive intrusion. Second, adaptive guidance remains transparently linked to observable flight parameters and scanning behaviour, limiting unexpected system actions. Third, the system maintains advisory authority only; pilots retain full control and decision responsibility, and adaptive assistance is suspended when model confidence is insufficient.

6.5. Concluding remarks

This study effort focused on developing and evaluating a novel low-cost projection-enhanced head-up display (PE-HUD) system for commercial aviation applications. Leveraging principles of reflection and projection, the proposed PE-HUD was designed to provide a more natural and comfortable interface for pilots, eliminating the need for cumbersome AR glasses or other transparent screens.

The findings revealed the positive effects of the PE-HUD on improving SA capabilities and engagement in brain activities. During the critical landing phase, the PE-HUD provided the most significant support for pilots' cognitive capabilities, followed by the take-off phase. However, the benefits were more limited during the cruise segment. The eye-tracking data further demonstrated that the PE-HUD was able to enhance the integration of critical information from the traditional cockpit displays, such as the PFD and ND, as well as certain functions of the ECAM. Notably, the PE-HUD also increased pilots' engagement with the FCU, suggesting enhanced SA and control of the aircraft. Importantly, the study revealed that pilots' attention patterns on the PE-HUD varied across the different flight phases, highlighting the need for adaptive and phase-specific HUD designs. Since the current PE-HUD is an incomplete prototype, the data can give value to the design of an adaptive PE-HUD, which can present the most suitable HUD for pilots in the future.

Chapter 7. Conclusion

7.1. Summary of progress

This research set out to address a critical challenge in aviation safety: the mitigation of Spatial Disorientation (SD) through an innovative, pilot-centric approach. By bridging the fields of neuro-ergonomics and adaptive Human-Machine Interaction (HMI), this research successfully established a comprehensive framework to enhance pilot cognitive resilience, thereby providing a proactive solution to a persistent aviation threat.

The central objective of this study was to transform the understanding of SD from a purely phenomenological perspective to a quantifiable, neurophysiological and behavioural model, and to use this understanding to create a practical, adaptive mitigation system.

1) **From Phenomenon to Quantifiable Data**: Through the integration of EEG and eye-tracking data, this research successfully established a clear neurophysiological and behavioural signature for different types of SD. We moved beyond traditional behavioural observation to quantitatively characterise the subtle changes in a pilot's cognitive state—such as shifts in mental workload and attentional control—that precede or accompany SD onset. This finding provides the foundational evidence needed for real-time monitoring of a pilot's cognitive state.
2) **From Theoretical Concept to Functional Framework**: The research constructed an Adaptive Attention-Awareness Pilot-Aircraft Interaction System (3A-PAI) driven by neurophysiological and behavioural data. By leveraging

advanced ML models, the framework was shown to possess the predictive capability to assess and classify different SD types. This represents a significant leap from traditional, static cockpit designs, which lack the ability to actively respond to a pilot's internal state. In addition, phase-specific scanning patterns were mined through probabilistic modelling of sequential data. This joint training methodology allows the model to simultaneously benefit from the HMM's generative understanding of latent states and temporal dependencies, as well as the CRF's robust capability in handling complex, overlapping features and performing precise discriminative prediction. Therefore, the 3A-PAI system could provide alert and tips information to enhance cognitive resilience based on the prediction of SD and LoAC.

3) **From Design Principles to a Validated Paradigm:** Within the established framework, we designed and validated adaptive interaction strategies. The use of an Augmented Reality Head-Up Display (AR-HUD) served as a key medium for these strategies, demonstrating its potential to deliver timely, context-aware information that can guide a pilot back to a state of correct spatial orientation. Two different AR interaction technologies have been implemented: AR glasses that enable more diverse interactions and projection AR that is more user-adaptable. The validation of these paradigms in a simulated environment provides compelling evidence of their effectiveness in enhancing pilot cognitive resilience and mitigating the effects of disorientation.

7.2. Contributions of the research

This research makes three key contributions that form a complete, closed-loop research cycle, advancing the state-of-the-art in both academic theory and practical application.

1) **A Novel Neuroergonomic Understanding of SD:** This research provides a systematic, neuro-ergonomics driven human factors analysis of the cognitive mechanisms underlying SD. By quantifying the neurophysiological and behavioural markers of disorientation, it fills a critical gap in the existing literature, offering a new, evidence-based foundation for understanding and predicting this complex phenomenon.
2) **A Complete Adaptive HMI Solution:** We have not only proposed a concept but have built a complete, data-driven, and validated solution for SD mitigation. This work demonstrates a concrete, applied pathway for how real-time cognitive monitoring can inform intelligent, adaptive systems, marking a significant advancement over the limitations of current static flight deck interfaces and manual mitigation techniques.
3) **A Blueprint for Future Cockpit Interaction Paradigms:** Beyond SD, this research provides a blueprint for future cockpit interaction paradigms. By exploring how an AR-HUD can serve as a medium for an intelligent, adaptive system, we have laid the groundwork for a new generation of pilot-centric flight decks. This work highlights the potential for future aviation systems to proactively collaborate with the pilot, not just react to their inputs, thereby contributing to the next wave of innovation in aviation safety standards and design philosophy.

7.3. Limitation and Future works

7.3.1. Limitations

This research, while providing a closed-loop solution for the adaptive HMI to mitigate

SD to enhance cognitive resilience, is not without its limitations. Future efforts aimed at overcoming these constraints are necessary to mature the proposed solution from an experimental prototype into an operationally viable system.

1) **Experimental Scenario Fidelity:** The study was carried out within a simulated flight setting. Although high-fidelity simulators reproduce numerous operational characteristics, they do not entirely capture the physiological and psychological demands experienced during real-world flight, such as exposure to G-forces, mechanical vibrations, and genuine risk perception. The lack of these stressors may alter pilots' neurophysiological responses and potentially constrain the ecological validity of the induced SD conditions.
2) **Subject Population:** The study was conducted with a relatively small cohort of participants, which may limit the extent to which the findings can be extrapolated to the broader pilot population. Although individual variability was partially controlled, validation across a larger and more heterogeneous sample—including pilots with varying levels of experience and age ranges—would strengthen confidence in the system's general applicability.
3) **Predictive Model Fidelity:** The classification of SD subtypes relies on a defined set of neurophysiological and behavioural features. Consequently, model performance, particularly in terms of generalisation, depends on the representativeness and diversity of the training data. The robustness of these models under unfamiliar or highly dynamic operational conditions remains to be empirically verified.

7.3.2. Future works

Building upon the robust framework established in this thesis, future research can expand in two primary directions to advance both the technical maturity of the system and the depth of human-machine interaction in the cockpit.

7.3.2.1. Enhancing System Robustness and Generalisation

The adaptive HMI strategy used in this study was first explored and then pre-designed for the system. Future work could incorporate reinforcement learning (RL), treating the adaptive system as an intelligent agent. This agent autonomously learns and optimises its adaptive policy through interaction with the environment (the flight scenario in the simulator and the pilot's real-time cognitive state). For example, when a pilot experiences high cognitive load, the RL agent could explore different ways of presenting information (such as reducing the amount of information, changing colors, and adding sound cues) and determine which policy is most effective by evaluating the pilot's subsequent cognitive state changes (reward function). This approach enables the system to find the optimal adaptive policy, even one that exceeds that preset by human experts.

A critical area of inquiry is the system's long-term impact on pilot behaviour. Future studies should investigate the potential for automation dependency and skill degradation over extended use. Research should be designed to explore how the adaptive HMI system can be calibrated to maintain a pilot's cognitive alertness and independent decision-making skills, rather than simply displaying information tips.

7.3.2.2. Developing an LLM-Assisted Virtual Co-pilot

This thesis has demonstrated a powerful model for proactive SD mitigation. Future work can extend this concept to a new paradigm of human-machine collaboration by developing an LLM-assisted virtual co-pilot.

The emergence of LLM provides a new avenue for this collaboration. A significant area of future work is to develop an LLM-assisted visual co-pilot. This system could be trained to interpret cockpit instrument data and procedural information, then verbally or visually assist the Pilot Flying (PF) or Pilot Monitoring (PM) with routine or complex tasks. This would offload a substantial portion of the pilots' mental and visual workload related to instrument scanning and monitoring, allowing them to focus more on higher-level decision-making and manual flight control, ultimately enhancing overall flight safety and efficiency.

Appendix I - SoP and QRH required when processing "UNRELIABLE SPEED INDICATION".

UNRELIABLE SPEED INDICATION/ADR CHECK PROC		
Condition: Doubt that airspeed or Mach number indication is unreliable. **Aim:** If possible, determine a reliable charge or continue the flight using data from the Unreliable Airspeed Flight Table in the Air Performance section.		
●	**If the safe conduct of the flight is impacted**	
<u>MEMORY ITEMS</u>		
-	AP/FD	OFF
-	A/THR	OFF
-	PITCH /THRUST:	
•	Below THRUST RED ALT	15°/TOGA
•	Above THRUST RED ALT and Below FL100	10°/CLB
•	Above THRUST RED ALT and Above FL100	5°/CLB
-	FLAPS	Maintain current CONFIG
-	SPEEDBRAKES	CHECK retracted
-	L/G	UP
●	**When at, or above MSA or Circuit Altitude:**	
-	Level off for troubleshooting	
-	**GPS ALTITUDE**	Display on MCDU

Appendix II - The customised SD Probes in two experimental sessions.

No.	Questions for VHHH - ZGGG	A	B	C	D	Answer
1	01Q: What is the level and speed restriction for PORPA?	+5000ft/250kts	-5000ft/205kts	+6000ft/200kts	-5000/200kts	B
2	02Q: What is the transition altitude for this departure SID?	5000ft	7000ft	9000ft	FL110	C
3	03Q: What is your current IAS?	Around 250 kt	Around 280 kt	Around 300 kt	Around 320 kt	C
4	04Q: What is your current heading?	Around 30°	Around 100°	Around 180°	Around 300°	D
5	05Q: What is your current altitude?	Above FL180	Above FL190	Above FL200	Above FL210	C
6	06Q: How far is the distance to IDUMA?	0-5 NM	5-10 NM	10-15 NM	None of the above	B
7	07Q: What is your current IAS?	Around 300 kt	Around 310 kt	Around 320 kt	Around 330 kt	A
8	08Q: What is your current IAS?	Around 250 kt	Around 280 kt	Around 300 kt	Around 330 kt	D
9	09Q: What is the initial altitude that you need to descend?	About FL150	About FL120	About FL100	About FL70	B
10	10Q: What is your current IAS?	Around 250 kt	Around 280 kt	Around 300 kt	Around 330 kt	B
11	11Q: What is your current altitude?	Above FL90	Above FL100	Above FL110	None of the above	A
12	12Q: What is the decision altitude (in Baro) for ILS02R?	DA:243ft	DA:244ft	DA:245ft	DA:248ft	B
13	13Q: What is the LOC frequency and ident for ILS02R?	108.5/IDM	110.35/IBB	111.5/IPP	111.9/IXL	A
No.	**Questions for ZGGG - VHHH**	**A**	**B**	**C**	**D**	**Answer**
1	01Q: What is the speed limit of GG412?	230kts	205kts	220kts	250kts	A
2	02Q: What runway did you T/O for departure?	02R	02L	20R	20L	A
3	03Q: What is your current IAS?	Above 260 kt	Above 300 kt	Above 330 kt	None of the above	C
4	04Q: What is your current heading to POU?	Around 180	Around 200	Around 220	Around 240	D
5	05Q: What is your current altitude?	Above FL180	Above FL190	Above FL200	None of the above	C
6	06Q: What is your current IAS?	Above 300 kt	Above 330 kt	Above 350 kt	None of the above	B
7	07Q: Is there any Hazard to fly through waypoint BORDA and ROCCA?	Yes	No	I don't know		A
8	08Q: What is your current altitude?	Above FL190	Above FL200	Above FL210	None of the above	A
9	09Q: What is your heading to BORDA?	Around 156	Around 188	Around 208	Around 220	C

10	10Q: Do you think the storm will go through ROCCA?	Yes	No	I don't know		B
11	11Q: What is your current IAS?	Around 180 kt	Around 230 kt	Around 250 kt	Around 280 kt	C
12	12Q: What is your current heading to VHHH?	Around 042	Around 090	Around 166	Around 338	A
13	13Q: Which runway will you land on?	07L	07R	25L	25R	A

Appendix III – Post-experiment questionnaire about HUD design

Question	Answer
What information do you usually get on the HUD?	1. Speed. 2. FD, ILS, altitude, airspeed. 3. Heading, altitude, VSI, glide slop, localiser, pitch, altitude. 4. Altitude. 5. FMA, speed, ILS, altitude. 6. Pitch, speed, altitude, flight director. 7. Airspeed and altitude. 8. Airspeed altitude checklist. 9. Info from PFD. 10. Flight director.
What functions do you hope to develop for HUD?	1. Flight director. 2. Flaps indicator. 3. Night enhance vision. 4. Augmented environment. 5. No. 6. Flight path marker, waypoint azimuth indicator. 7. No need to reset when you start to fly. 8. Flaring clue of runway. 9. FMA, ILS ident/DME, VFE, etc. 10. The flight director (+) sign can keep it in HUD instead of the circle dot.
What do you think are the advantages of using a HUD?	1. Reduce workload. 2. More situational awareness, lesser workload. 3. Align vision with flight display. No need to switch vision between instruments and outside environment. 4. Easy acquisition of information in one place, situational awareness. 5. When runway in-sight in approach, no need to head down to spot important information. 6. Increase situation awareness. 7. No need to look down. 8. Look out. 9. During LVO. 10. Don't need to move the head to check the instrument, and can do the checklist with outside situational awareness.
Are there any discomforts or	1. No. 2. No.

inconveniences when using the HUD?	3. Need more practice before using it in real flight. 4. No. 5. No. 6. No. 7. None. 8. Yes, need to sit in a particular position in order to align with the visual. 9. No. 10. During the landing, the HUD image likes a wiper, I need ignore it.
Do you have any other suggestions for the development of AR-HUD?	1. No. 2. -A way to make the process more automatic for display. - Display font can be bigger in general. 3. No. 4. More accurate information presentation. 5. No. 6. Data link integration to indicate close traffic. 7. None. 8. Colour display. 9. No. 10. No.

References

AB, T. (2024). Tobii Pro Lab User Manual. Tobii Connect: Tobii Connect.

Agrawal, S., & Peeta, S. (2021). Evaluating the impacts of situational awareness and mental stress on takeover performance under conditional automation. *Transportation Research Part F-Traffic Psychology and Behaviour, 83*, 210-225. doi:10.1016/j.trf.2021.10.002

Ahlstrom, U. (2015). Weather display symbology affects pilot behavior and decision-making. *International Journal of Industrial Ergonomics, 50*, 73-96. doi:10.1016/j.ergon.2015.09.008

Ali, H., Pham, D.-T., Alam, S., Schultz, M., Li, M. Z., Wang, Y., . . . Duong, V. N. (2025). Human-AI Hybrids in Safety-Critical Systems: Concept, definition and perspectives from Air Traffic Management. *Advanced Engineering Informatics, 65*, 103256. doi:https://doi.org/10.1016/j.aei.2025.103256

Alreshidi, I., Moulitsas, I., & Jenkins, K. W. (2023). Multimodal Approach for Pilot Mental State Detection Based on EEG. *Sensors, 23*(17), 7350. Retrieved from https://www.mdpi.com/1424-8220/23/17/7350

Amditis, A., Pagle, K., Joshi, S., & Bekiaris, E. (2010). Driver-Vehicle-Environment monitoring for on-board driver support systems: Lessons learned from design and implementation. *Applied Ergonomics, 41*(2), 225-235. doi:10.1016/j.apergo.2009.03.002

Antunano, M. J. (2021). *Medical Facts for Pilots: Spatial Disorientation, safety brochure AM-400-03/1*. Retrieved from Federal Aviation Administration: https://www.faa.gov/pilots/safety/pilotsafetybrochures/media/spatiald.pdf

Aricò, P., Borghini, G., Graziani, I., Imbert, J.-P., Granger, G., Benhacene, R., . . . Colosimo, A. (2015). Air-Traffic-Controllers (ATCO)-Neurophysiological analysis of training and workload ATCO. *Ital. J. Aerosp. Med, 12*, 18-35.

Aromaa, S., Väätänen, A., Aaltonen, I., Goriachev, V., Helin, K., & Karjalainen, J. (2020). Awareness of the real-world environment when using augmented reality head-mounted display. *Applied Ergonomics, 88*, 103145. doi:https://doi.org/10.1016/j.apergo.2020.103145

Authority, S. A. I. (2016). *Final report RL 2016:11e Accident in Oajevágge, Norrbotten County, Sweden on 8 January 2016 involving the aeroplane SE-DUX of the model CL-600-2B19, operated by West Atlantic Sweden AB.* . Retrieved from AVIATION ACCIDENT DATABASE: https://www.aviation-accidents.net/west-air-sweden-bombardier-cl600-2b19-se-dux-flight-swn294/

Ayas, S., Donmez, B., & Tang, X. (2023). Drowsiness Mitigation Through Driver State Monitoring Systems: A Scoping Review. *Human Factors*. doi:10.1177/00187208231208523

Bałaj, B., Lewkowicz, R., Francuz, P., Augustynowicz, P., Fudali-Czyż, A., Stróżak, P., & Truszczyński, O. (2019). Spatial disorientation cue effects on gaze behaviour in pilots and non-pilots. *Cognition, Technology & Work, 21*(3), 473-486. doi:10.1007/s10111-018-0534-7

Benson, A. J. (2003). *Spatial Disorientation - A Perspective*.

Berberian, B., Somon, B., Sahaï, A., & Gouraud, J. (2017). The out-of-the-loop Brain: A neuroergonomic approach of the human automation interaction. *Annual Reviews in Control, 44*, 303-315. doi:https://doi.org/10.1016/j.arcontrol.2017.09.010

Bergasa, L. M., Almería, D., Almazán, J., Yebes, J. J., Arroyo, R., & Ieee. (2014, Jun 08-11). *DriveSafe: an App for Alerting Inattentive Drivers and Scoring Driving Behaviors.* Paper presented at the IEEE Intelligent Vehicles Symposium (IV), Dearborn, MI.

Biernacki, M. P., & Lewkowicz, R. (2024). The role of visual conditions and aircraft type on different aspects of pilot workload. *Applied Ergonomics, 118*, 104268. doi:https://doi.org/10.1016/j.apergo.2024.104268

Bolton, M. L., & Bass, E. J. (2009). Comparing perceptual judgment and subjective measures of spatial awareness. *Applied Ergonomics, 40*(4), 597-607. doi:https://doi.org/10.1016/j.apergo.2008.04.020

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review, 108*(3), 624-652. doi:10.1037//0033-295x.108.3.624

Calhoun, G. (2021). Adaptable (Not Adaptive) Automation: Forefront of Human–Automation Teaming. *Human Factors, 64*(2), 269-277. doi:10.1177/00187208211037457

Catherwood, D., Edgar, G. K., Nikolla, D., Alford, C., Brookes, D., Baker, S., & White, S. (2014). Mapping brain activity during loss of situation awareness: an EEG investigation of a basis for top-down influence on perception. *Hum Factors, 56*(8), 1428-1452. doi:10.1177/0018720814537070

Causse, M., Mercier, M., Lefrançois, O., & Matton, N. (2025). Impact of automation level on airline pilots' flying performance and visual scanning strategies: A full flight simulator study. *Applied Ergonomics, 125*, 104456. doi:https://doi.org/10.1016/j.apergo.2024.104456

Ceken, S. (2025). The role of cognitive flexibility and psychological resilience in enhancing operational performance among air traffic controllers. *International Journal of Aviation, Aeronautics, and Aerospace, 12*(1), 3.

Çevik, İ., & Dahanayake, A. (2022, 2022//). *Fatigue Risk Indicator Model for Aviation Cabin Crew's Fatigue Assessment.* Paper presented at the Reliability and Statistics in Transportation and Communication, Cham.

Chang, M., Huang, L., You, X., Wang, P., Francis, G., & Proctor, R. W. (2022). The black hole illusion: A neglected source of aviation accidents. *International Journal of Industrial Ergonomics, 87*, 103235. doi:https://doi.org/10.1016/j.ergon.2021.103235

Chang, W. J., Chen, L. B., & Chiou, Y. Z. (2018). Design and Implementation of a Drowsiness-Fatigue-Detection System Based on Wearable Smart Glasses to Increase Road Safety. *IEEE Transactions on Consumer Electronics, 64*(4), 461-469. doi:10.1109/TCE.2018.2872162

Chen, J., Xue, L., Rong, J., & Gao, X. (2022). Real-time evaluation method of flight mission load based on sensitivity analysis of physiological factors. *Chinese Journal of Aeronautics, 35*(3), 450-463. doi:https://doi.org/10.1016/j.cja.2021.11.010

Chen, M., Wang, C., He, X., Zhu, F., Wang, L., & Vasilakos, A. V. (2025). Embodied Artificial Intelligence-Enabled Internet of Vehicles: Challenges and Solutions. *IEEE Vehicular Technology Magazine, 20*(2), 63-70. doi:10.1109/MVT.2025.3544675

Chui, K. T. (2022). Driver stress recognition for smart transportation: Applying multiobjective genetic algorithm for improving fuzzy c-means clustering with reduced time and model complexity. *Sustainable Computing: Informatics and Systems, 35*, 100668. doi:https://doi.org/10.1016/j.suscom.2022.100668

Cifrek, M., Medved, V., Tonkovic, S., & Ostojic, S. (2009). Surface EMG based muscle fatigue evaluation in biomechanics. *Clinical Biomechanics, 24*(4), 327-340. doi:10.1016/j.clinbiomech.2009.01.010

Colvin, K., Dodhia, R., & Dismukes, R. (2005). *Is pilots' visual scanning adequate to avoid mid-air collisions?* Paper presented at the 2005 International Symposium on Aviation Psychology.

Craik, A., He, Y. T., & Contreras-Vidal, J. L. (2019). Deep learning for electroencephalogram (EEG) classification tasks: a review. *Journal of Neural Engineering, 16*(3). doi:10.1088/1741-2552/ab0ab5

Crestelo Moreno, F., Roca Gonzalez, J., Suardíaz Muro, J., & García Maza, J. A. (2022). Relationship between human factors and a safe performance of vessel traffic service operators: A systematic qualitative-based review in maritime safety. *Safety Science, 155*, 105892. doi:https://doi.org/10.1016/j.ssci.2022.105892

Cummings, M. L., Stimpson, A., & Clamann, M. (2016). Functional requirements for onboard intelligent automation in single pilot operations. In *AIAA Infotech@ Aerospace* (pp. 1652).

Dahm, W. (2010). *Technology Horizons: A Vision for Air Force Science & Technology During 2010-2030*.

Debener, S., Ullsperger, M., Siegel, M., Fiehler, K., von Cramon, D. Y., & Engel, A. K. (2005). Trial-by-trial coupling of concurrent electroencephalogram and functional magnetic resonance imaging identifies the dynamics of performance monitoring. *Journal of Neuroscience, 25*(50), 11730-11737. doi:10.1523/jneurosci.3286-05.2005

Dekker, S., & Hollnagel, E. (2004). Human factors and folk models. *Cognition, Technology & Work, 6*(2), 79-86. doi:10.1007/s10111-003-0136-9

Dellermann, D., Ebel, P., Söllner, M., & Leimeister, J. M. (2019). Hybrid Intelligence. *Business & Information Systems Engineering, 61*(5), 637-643. doi:10.1007/s12599-019-00595-2

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J Neurosci Methods, 134*(1), 9-21. doi:10.1016/j.jneumeth.2003.10.009

Di Flumeri, G., De Crescenzio, F., Berberian, B., Ohneiser, O., Kramer, J., Aricò, P., . . . Piastra, S. (2019). Brain–Computer Interface-Based Adaptive Automation to Prevent Out-Of-The-Loop Phenomenon in Air Traffic Controllers Dealing With Highly Automated Systems. *Frontiers in Human Neuroscience, 13*. doi:10.3389/fnhum.2019.00296

Dignath, D., Eder, A. B., Steinhauser, M., & Kiesel, A. (2020). Conflict monitoring and the affective-signaling hypothesis—An integrative review. *Psychonomic Bulletin & Review, 27*(2), 193-216. doi:10.3758/s13423-019-01668-9

Dong, J., Chen, S., Miralinaghi, M., Chen, T., Li, P., & Labi, S. (2023). Why did the AI make that decision? Towards an explainable artificial intelligence (XAI) for autonomous driving systems. *Transportation Research Part C: Emerging Technologies, 156*, 104358. doi:https://doi.org/10.1016/j.trc.2023.104358

Dong, M. M. (2023). Investigating the Users' Preferences of Heart Rate Data Types and Visualizations on a Smartwatch. *International Journal of Human-Computer Interaction*. doi:10.1080/10447318.2023.2233130

Dorneich, M. C., Rogers, W., Whitlow, S. D., & DeMers, R. (2016). Human Performance Risks and Benefits of Adaptive Systems on the Flight Deck. *International Journal of Aviation Psychology, 26*(1-2), 15-35. doi:10.1080/10508414.2016.1226834

Dunbar, J., Gilbert, J. E., & Lewis, B. (2020). Exploring differences between self-report and electrophysiological indices of drowsy driving: A usability examination of a personal brain-computer interface device. *J Safety Res, 74*, 27-34. doi:10.1016/j.jsr.2020.04.006

Eisma, Y. B., Borst, C., van Paassen, R., & de Winter, J. (2021). Augmented Visual Feedback: Cure or Distraction? *Human Factors, 63*(7), 1156-1168. doi:10.1177/0018720820924602

Elmalaki, S. (2021). *FaiR-IoT: Fairness-aware Human-in-the-Loop Reinforcement Learning for Harnessing Human Variability in Personalized IoT*. Paper presented at the Proceedings of the International Conference on Internet-of-Things Design and Implementation, Charlottesvle, VA, USA. https://doi.org/10.1145/3450268.3453525

Endsley, M., Sollenberger, R., Nakata, A., & Stein, E. S. (2000). *Situation awareness in air traffic control: Enhanced displays for advanced operations*. Retrieved from

Endsley, M. R. (1999). Situation awareness in aviation systems. *Handbook of aviation human factors, 11*, 257-276.

Endsley, M. R. (2023). Supporting Human-AI Teams:Transparency, explainability, and situation awareness. *Computers in Human Behavior, 140*, 107574. doi:https://doi.org/10.1016/j.chb.2022.107574

Endsley, M. R., & Kiris, E. O. (1995). The Out-of-the-Loop Performance Problem and Level of Control in Automation. *Human Factors, 37*(2), 381-394. doi:10.1518/001872095779064555

Engström, J. (2011). *Understanding Attention Selection in Driving: From Limited Capacity to Adaptive Behaviour.* (Ph.D.). Chalmers Tekniska Hogskola (Sweden), Sweden. ProQuest Dissertations & Theses Global: The Sciences and Engineering Collection database. (27755776)

Ernst, J. M., Ebrecht, L., & Korn, B. (2021). Virtual Cockpit Instruments-How Head-Worn Displays Can Enhance the Obstacle Awareness of Helicopter Pilots. *IEEE Aerospace and Electronic Systems Magazine, 36*(4), 18-34. doi:10.1109/Maes.2020.3040511

Eusse-Villa, L., Pellizzari, C. B., Franceschinis, C., Thiene, M., Borga, M., & Scolobig, A. (2024). Identification of maladaptive behavioural patterns in response to extreme weather events. *Scientific Reports, 14*(1). doi:10.1038/s41598-024-60632-3

Fadden, S., Ververs, P. M., & Wickens, C. D. (2001). Pathway HUDs: are they viable? *Human Factors, 43*(2), 173-193.

Fan, Y. Z., Gu, F., Wang, J., Wang, J. P., Lu, K. J., & Niu, J. W. (2021). SafeDriving: An Effective Abnormal Driving Behavior Detection System Based on EMG Signals. *Ieee Internet of Things Journal, 9*(14), 12338-12350. doi:10.1109/jiot.2021.3135512

Feigh, K. M., Dorneich, M. C., & Hayes, C. C. (2012). Toward a Characterization of Adaptive Systems: A Framework for Researchers and System Designers. *Human Factors, 54*(6), 1008-1024. doi:10.1177/0018720812443983

Feng, C., Wanyan, X., Liu, S., Chen, H., & Zhuang, D. (2021). An approach to situation awareness (SA) assessment in flight simulation: SA dynamic circulation (SADC) model. *Human Factors and Ergonomics in Manufacturing & Service Industries, 31*(5), 559-569. doi:https://doi.org/10.1002/hfm.20903

Fernandez-Blanco, E., Rivero, D., & Pazos, A. (2020). EEG signal processing with separable convolutional neural network for automatic scoring of sleeping stage. *Neurocomputing, 410*, 220-228. doi:10.1016/j.neucom.2020.05.085

Friedrich, M., Lee, S. Y., Bates, P., Martin, W., & Faulhaber, A. K. (2021). The influence of training level on manual flight in connection to performance, scan pattern, and task load. *Cognition, Technology & Work, 23*(4), 715-730. doi:10.1007/s10111-020-00663-8

Fudali-Czyz, A., Lewkowicz, R., Francuz, P., Strózak, P., Augustynowicz, P., Truszczynski, O., & Balaj, B. (2022). An Attentive Blank Stare Under Simulator-induced Spatial Disorientation Events. *Human Factors*. doi:10.1177/00187208221093827

Gaillet, A. G., Scannella, S., & Garcia, J. (2025). *Exploring Multimodal Feedback for Light Aircraft*

Pilots: An Interview Study. Paper presented at the Proceedings of the 36th Conference on l'Interaction Humain-Machine. https://doi.org/10.1145/3765712.3765713
Games, M. (2021). Advanced Aircraft HUD - GUI Pack & Code. Retrieved from https://assetstore.unity.com/packages/tools/gui/advanced-aircraft-hud-gui-pack-code-151686
Gardony, A. L., Martis, S. B., Taylor, H. A., & Brunyé, T. T. (2021). Interaction strategies for effective augmented reality geo-visualization: Insights from spatial cognition. *Human–Computer Interaction, 36*(2), 107-149.
Gaydos, S. J., Harrigan, M. J., & Bushby, A. J. (2012). Ten years of spatial disorientation in US Army rotary-wing operations. *Aviation, space, and environmental medicine, 83*(8), 739-745.
Gegoff, I., Tatasciore, M., Bowden, V., McCarley, J., & Loft, S. (2023). Transparent Automated Advice to Mitigate the Impact of Variation in Automation Reliability. *Human Factors*. doi:10.1177/00187208231196738
Geva, G., Getter, N., Blecher, B., Ben-Ari, O., Gordon, B., Nakdimon, I., & Shriki, O. (2025). Reduction in EEG theta power as a potential marker for spatial disorientation during flight. *Scientific Reports, 15*(1), 1682. doi:10.1038/s41598-025-85219-4
Ghaderi, A., & Saghafi, F. Relationships Between Pilot Gaze Patterns and Control Lapses in Challenging Instrument Approach Flight. *The International Journal of Aerospace Psychology*, 1-18. doi:10.1080/24721840.2025.2531740
Gibb, R., Ercoline, B., & Scharff, L. (2011). Spatial disorientation: decades of pilot fatalities. *Aviation, space, and environmental medicine, 82*(7), 717-724. doi:10.3357/asem.3048.2011
Gil-Cabrera, J., Tornero Aguilera, J. F., Sanchez-Tena, M. A., Alvarez-Peregrina, C., Valbuena-Iglesias, C., & Clemente-Suárez, V. J. (2021). Aviation-Associated Spatial Disorientation and Incidence of Visual Illusions Survey in Military Pilots. *The International Journal of Aerospace Psychology, 31*(1), 17-24. doi:10.1080/24721840.2020.1841562
Gregoriades, A., & Sutcliffe, A. (2018). Simulation-based evaluation of an in-vehicle smart situation awareness enhancement system. *Ergonomics, 61*(7), 947-965. doi:10.1080/00140139.2018.1427803
Hao, C., Cheng, L., Guo, L., Zhao, R., Wu, Y., Li, X., . . . Li, J. (2022). Detection of unrecognized spatial disorientation: A theoretical perspective. *Technology and Health Care, 30*(1_suppl), 469-480. doi:10.3233/thc-thc228043
Haslbeck, A., & Zhang, B. (2017). I spy with my little eye: Analysis of airline pilots' gaze patterns in a manual instrument flight scenario. *Appl Ergon, 63*, 62-71. doi:10.1016/j.apergo.2017.03.015
Heinle, T. E., & Ercoline, W. R. (2002). *Spatial disorientation: causes, consequences and countermeasures for the USAF.* Paper presented at the The RTO HFM Symposium, AIR FORCE RESEARCH LAB WRIGHT-PATTERSON AFB OH HUMAN EFFECTIVENESS DIRECTORATE, La Coruña, Spain.
Helleberg, J. R., & Wickens, C. D. (2003). Effects of data-link modality and display redundancy on pilot performance: An attentional perspective. *The International Journal of Aviation Psychology, 13*(3), 189-210.
Hernandez-Sabate, A., Yauri, J., Folch, P., Piera, M. A., & Gil, D. (2022). Recognition of the Mental Workloads of Pilots in the Cockpit Using EEG Signals. *Applied Sciences-Basel, 12*(5). doi:10.3390/app12052298
Herrmann, C. S., & Demiralp, T. (2005). Human EEG gamma oscillations in neuropsychiatric disorders. *Clinical Neurophysiology, 116*(12), 2719-2733. doi:https://doi.org/10.1016/j.clinph.2005.07.007
Hu, H., Fang, Z., Qian, Z., Yao, L., Tao, L., & Qin, B. (2021). Stress Assessment of Vestibular Endurance Training for Civil Aviation Flight Students Based on EEG. *Frontiers in Human Neuroscience, 15*. doi:10.3389/fnhum.2021.582636
Hu, X., & Lodewijks, G. (2020). Detecting fatigue in car drivers and aircraft pilots by using non-invasive measures: The value of differentiation of sleepiness and mental fatigue. *J Safety Res, 72*, 173-187. doi:10.1016/j.jsr.2019.12.015
Hua, L. S., Ling, C., & Thomas, R. (2022). Effects of delayed weather radar images on pilots' spatial awareness. *Applied Ergonomics, 98*. doi:ARTN 103598
10.1016/j.apergo.2021.103598
Huang, J., Song, X., Xiao, F., Cao, Z., & Lin, C.-T. (2023). Belief f-divergence for EEG complexity evaluation. *Information Sciences, 643*, 119189. doi:https://doi.org/10.1016/j.ins.2023.119189
Huang, L., Hou, Y., Chen, Y., You, X., Proctor, R. W., Francis, G., & Chang, M. (2023). How the black hole illusion environment affects operational performance at different flight phases in aviation. *Applied Ergonomics, 113*, 104048. doi:https://doi.org/10.1016/j.apergo.2023.104048

Huang, W. H., Liu, H. C., Huang, Z. Y., & Lv, C. (2024). Safety-Aware Human-in-the-Loop Reinforcement Learning With Shared Control for Autonomous Driving. *IEEE Transactions on Intelligent Transportation Systems, 25*(11), 16181-16192. doi:10.1109/tits.2024.3420959

Huang, Y., Xiao, F., Cao, Z., & Lin, C. T. (2023). Higher Order Fractal Belief Rényi Divergence With Its Applications in Pattern Classification. *Ieee Transactions on Pattern Analysis and Machine Intelligence, 45*(12), 14709-14726. doi:10.1109/TPAMI.2023.3310594

Hussain, M. I., Rafique, M. A., Kim, J., Jeon, M., & Pedrycz, W. (2023). Artificial Proprioceptive Reflex Warning Using EMG in Advanced Driving Assistance System. *Ieee Transactions on Neural Systems and Rehabilitation Engineering, 31*, 1635-1644. doi:10.1109/TNSRE.2023.3254151

Janssen, C. P., Donker, S. F., Brumby, D. P., & Kun, A. L. (2019). History and future of human-automation interaction. *International Journal of Human-Computer Studies, 131*, 99-107. doi:10.1016/j.ijhcs.2019.05.006

Jiang, G. Y., Chen, H., Wang, C. Y., & Xue, P. X. (2022). Mental Workload Artificial Intelligence Assessment of Pilots' EEG Based on Multi-Dimensional Data Fusion and LSTM with Attention Mechanism Model. *International Journal of Pattern Recognition and Artificial Intelligence, 36*(11). doi:10.1142/s0218001422590352

John, A. R., Singh, A. K., Do, T. T. N., Eidels, A., Nalivaiko, E., Gavgani, A. M., . . . Lin, C. T. (2022). Unraveling the Physiological Correlates of Mental Workload Variations in Tracking and Collision Prediction Tasks. *Ieee Transactions on Neural Systems and Rehabilitation Engineering, 30*, 770-781. doi:10.1109/tnsre.2022.3157446

Joo, T., & Shin, D. (2019). Formalizing Human–Machine Interactions for Adaptive Automation in Smart Manufacturing. *IEEE Transactions on Human-Machine Systems, 49*(6), 529-539. doi:10.1109/THMS.2019.2903402 , ISSN= 2168-2305

Kaber, D. B., Riley, J. M., Tan, K.-W., & Endsley, M. R. (2001). On the Design of Adaptive Automation for Complex Systems. *International Journal of Cognitive Ergonomics, 5*(1), 37-57. doi:10.1207/S15327566IJCE0501_3

Kamar, E. (2016). *Directions in hybrid intelligence: complementing AI systems with human intelligence*. Paper presented at the Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence, New York, New York, USA.

Kamzanova, A. T., Kustubayeva, A. M., & Matthews, G. (2014). Use of EEG Workload Indices for Diagnostic Monitoring of Vigilance Decrement. *Human Factors, 56*(6), 1136-1149. doi:10.1177/0018720814526617

Kanaan, D., Powell, M., Lu, M., & Donmez, B. (2024). In-vehicle displays for driving automation: a scoping review of display design and evaluation using driver gaze measures. *Transport Reviews*. doi:10.1080/01441647.2024.2336921

Kang, Y., Lazaro, M. J., & Kim, S. (2021). Crosschecking through verbal reports under spatial disorientation scenarios: Evidence from eye tracking metrics. *International Journal of Industrial Ergonomics, 86*. doi:10.1016/j.ergon.2021.103202

Kant, V., Nejeeb, I., Alya, S., Jain, P., & Singh, R. (2025). Supporting cognitive intelligence for smart manufacturing systems using HMI design: challenges and fundamental issues. *Journal of Engineering Design, 36*(11), 1961-1981. doi:10.1080/09544828.2024.2343746

Kapp, S., Barz, M., Mukhametov, S., Sonntag, D., & Kuhn, J. (2021). ARETT: Augmented Reality Eye Tracking Toolkit for Head Mounted Displays. *Sensors, 21*(6), 2234. Retrieved from https://www.mdpi.com/1424-8220/21/6/2234

Ke, G. L., Meng, Q., Finley, T., Wang, T. F., Chen, W., Ma, W. D., . . . Liu, T. Y. (2017, Dec 04-09). *LightGBM: A Highly Efficient Gradient Boosting Decision Tree.* Paper presented at the 31st Annual Conference on Neural Information Processing Systems (NIPS), Long Beach, CA.

Kevin Semple, M., Annebelle Olminkhof, Mohit Patel, Vinnie Tieto, Christopher McClister, Qian Wen. (2022). What is Mixed Reality Toolkit 2? Retrieved from https://learn.microsoft.com/en-us/windows/mixed-reality/mrtk-unity/mrtk2/?view=mrtkunity-2022-05

Kim, H., Lee, H., Park, J., Paillat, L., & Kim, S. C. (2023). Vehicle Control on an Uninstrumented Surface With an Off-the-Shelf Smartwatch. *IEEE Transactions on Intelligent Vehicles, 8*(5), 3366-3374. doi:10.1109/tiv.2023.3236458

Kim, S., Kang, Y., & Lazaro, M. J. (2025). Development and validation of spatial disorientation scenarios using virtual reality and motion simulator. *Applied Ergonomics, 125*, 104457. doi:https://doi.org/10.1016/j.apergo.2024.104457

Kim, S. H., & Kaber, D. B. (2014). Examining the Effects of Conformal Terrain Features in Advanced Head-Up Displays on Flight Performance and Pilot Situation Awareness. *Human Factors and Ergonomics in Manufacturing & Service Industries, 24*(4), 386-402. doi:10.1002/hfm.20501

Kirwan, B., Venditti, R., Giampaolo, N., & Sánchez, M. V. (2024). A Human Centric Design Approach for Future Human-AI Teams in Aviation. *Human Interaction and Emerging Technologies (IHIET 2024), 1*(1).

Kosel, C., Mooseder, A., Seidl, T., & Pfeffer, J. (2023). Measuring Teachers' Visual Expertise Using the Gaze Relational Index Based on Real-world Eye-tracking Data and Varying Velocity Thresholds. *arXiv preprint arXiv:2304.05143*.

Krippendorff, K. (2019). The cybernetics of design and the design of cybernetics. *Design Cybernetics: Navigating the New*, 119-136.

Lai, H.-Y. (2024). Advancements in intelligent driving assistance: A machine learning approach to identify real-time driving strategies using environmental, eye movement, control-related, and kinetic-related data. *Advanced Engineering Informatics, 62*, 102745. doi:https://doi.org/10.1016/j.aei.2024.102745

Lampe, D., & Deml, B. (2022). Reducing passive driver fatigue through a suitable secondary motor task by means of an interactive seating system. *Applied Ergonomics, 103*. doi:ARTN 103773
10.1016/j.apergo.2022.103773

Landman, A., Davies, S., Groen, E. L., van Paassen, M. M., Lawson, N. J., Bronkhorst, A. W., & Mulder, M. (2019). In-flight spatial disorientation induces roll reversal errors when using the attitude indicator. *Applied Ergonomics, 81*, 102905. doi:https://doi.org/10.1016/j.apergo.2019.102905

Landman, A., Kalogeras, D., Houben, M., & Groen, E. L. (2024). Orientation Comes First: Becoming Aware of Spatial Disorientation Interferes with Cognitive Performance. *Human Factors, 66*(2), 377-388. doi:10.1177/00187208221103931

Le Coze, J. C. (2022). The 'new view' of human error. Origins, ambiguities, successes and critiques. *Safety Science, 154*, 105853. doi:https://doi.org/10.1016/j.ssci.2022.105853

Le Guillou, M., Prevot, L., & Berberian, B. (2023). Bringing Together Ergonomic Concepts and Cognitive Mechanisms for Human-AI Agents Cooperation. *International Journal of Human-Computer Interaction, 39*(9), 1827-1840. doi:10.1080/10447318.2022.2129741

Ledegang, W. D., & Groen, E. L. (2018). Spatial Disorientation Influences on Pilots' Visual Scanning and Flight Performance. *Aerospace Medicine and Human Performance, 89*(10), 873-882. doi:10.3357/amhp.5109.2018

Li, F., Chen, C.-H., Lee, C.-H., & Feng, S. (2022). Artificial intelligence-enabled non-intrusive vigilance assessment approach to reducing traffic controller's human errors. *Knowledge-Based Systems, 239*, 108047. doi:https://doi.org/10.1016/j.knosys.2021.108047

Li, F., Feng, S., Yan, Y., Lee, C. H., & Ong, Y. S. (2024, 25-27 June 2024). *Virtual Co-Pilot: Multimodal Large Language Model-enabled Quick-access Procedures for Single Pilot Operations.* Paper presented at the 2024 IEEE Conference on Artificial Intelligence (CAI).

Li, P., Zhang, Y., Liu, S., Lin, L., Zhang, H., Tang, T., & Gao, D. (2023). An EEG-based Brain Cognitive Dynamic Recognition Network for representations of brain fatigue. *Applied Soft Computing, 146*, 110613. doi:https://doi.org/10.1016/j.asoc.2023.110613

Li, Q., Chen, C.-H., Ng, K. K. H., Yuan, X., & Yiu, C. Y. (2024). Single-pilot operations in commercial flight: Effects on neural activity and visual behaviour under abnormalities and emergencies. *Chinese Journal of Aeronautics*. doi:https://doi.org/10.1016/j.cja.2024.04.007

Li, Q., Leung, H. C., Ho, M. H., Leung, K. L., Ng, K. K. H., & Yiu, C. Y. (2023). The Effects of Aeronautical Decision-Making Models on Student Pilots' Situational Awareness and Cognitive Workload in Simulated Non-Normal Flight Deck Environment. *The International Journal of Aerospace Psychology, 33*(3), 197-213. doi:10.1080/24721840.2023.2231506

Li, Q., Ng, K. K. H., Fan, Z., Yuan, X., Liu, H., & Bu, L. (2021). A human-centred approach based on functional near-infrared spectroscopy for adaptive decision-making in the air traffic control environment: A case study. *Advanced Engineering Informatics, 49*, 101325. doi:https://doi.org/10.1016/j.aei.2021.101325

Li, Q., Ng, K. K. H., Yiu, C. Y., Yuan, X., So, C. K., & Ho, C. C. (2023). Securing air transportation safety through identifying pilot's risky VFR flying behaviours: An EEG-based neurophysiological modelling using machine learning algorithms. *Reliability Engineering & System Safety, 238*, 109449. doi:https://doi.org/10.1016/j.ress.2023.109449

Li, R., Mayseless, N., Balters, S., & Reiss, A. L. (2021). Dynamic inter-brain synchrony in real-life inter-personal cooperation: A functional near-infrared spectroscopy hyperscanning study. *Neuroimage, 238*, 118263. doi:10.1016/j.neuroimage.2021.118263

Li, S., Deng, L., Hu, J., Kang, S., Qiu, J., & Li, Q. (2024). A comprehensive approach to evaluate human–machine conflicts in shared steering systems. *Accident Analysis & Prevention, 207*, 107758. doi:https://doi.org/10.1016/j.aap.2024.107758

Li, W.-C., Yan, Z., Zhang, J., Braithwaite, G., Court, S., Lone, M., & Thapa, B. (2020, 2020//). *Evaluating Pilot's Perceived Workload on Interacting with Augmented Reality Device in Flight Operations.* Paper presented at the Engineering Psychology and Cognitive Ergonomics. Cognition and Design, Cham.

Li, W.-C., Zhang, J., Court, S., Kearney, P., & Braithwaite, G. (2022). The influence of augmented reality interaction design on Pilot's perceived workload and situation awareness. *International Journal of Industrial Ergonomics, 92*, 103382. doi:https://doi.org/10.1016/j.ergon.2022.103382

Li, W. C., Horn, A., Sun, Z., Zhang, J. Y., & Braithwaite, G. (2020). Augmented visualization cues on primary flight display facilitating pilot's monitoring performance. *International Journal of Human-Computer Studies, 135*. doi:ARTN 102377

10.1016/j.ijhcs.2019.102377

Li, W. C., Yu, C. S., Greaves, M., & Braithwaite, G. (2015). How Cockpit Design Impacts Pilots' Attention Distribution and Perceived Workload during Aiming a Stationary Target. *6th International Conference on Applied Human Factors and Ergonomics (Ahfe 2015) and the Affiliated Conferences, Ahfe 2015, 3*, 5663-5669. doi:10.1016/j.promfg.2015.07.781

Li, Z., Li, F., & Lyu, M. (2024). Tracking the Unseen and Unaware: Deciphering Controllers' Detection Failures to Warnings Through Eye-Tracking Metrics. *International Journal of Human–Computer Interaction*, 1-20. doi:10.1080/10447318.2024.2448877

Liang, C. R., Liu, S., Wanyan, X., Liu, C. P., Xiao, X., & Min, Y. C. (2023). Effects of input method and display mode of situation map on early warning aircraft reconnaissance task performance with different information complexities. *Chinese Journal of Aeronautics, 36*(1), 105-114. doi:10.1016/j.cja.2022.06.011

Liang, N., Yang, J., Yu, D., Prakah-Asante, K. O., Curry, R., Blommer, M., . . . Pitts, B. J. (2021). Using eye-tracking to investigate the effects of pre-takeover visual engagement on situation awareness during automated driving. *Accid Anal Prev, 157*, 106143. doi:10.1016/j.aap.2021.106143

Liang, Y., Zheng, P., & Xia, L. (2023). A visual reasoning-based approach for driving experience improvement in the AR-assisted head-up displays. *Advanced Engineering Informatics, 55*, 101888. doi:https://doi.org/10.1016/j.aei.2023.101888

Liang, Y. S., Zheng, P., & Xia, L. Q. (2023). A visual reasoning-based approach for driving experience improvement in the AR-assisted head-up displays. *Advanced Engineering Informatics, 55*. doi:ARTN 101888

10.1016/j.aei.2023.101888

Liao, M., Wang, H., Liu, Z., & Hu, E. (2025). Rumination and psychological resilience in Chinese civil aviation flight students: the chain mediating role of proactive coping and generalized anxiety disorder. *Frontiers in Psychology, Volume 16 - 2025*. doi:10.3389/fpsyg.2025.1606045

Liu, J., Gardi, A., Ramasamy, S., Lim, Y., & Sabatini, R. (2016). Cognitive pilot-aircraft interface for single-pilot operations. *Knowledge-Based Systems, 112*, 37-53. doi:10.1016/j.knosys.2016.08.031

Liu, Q., Li, F., Ng, K. K. H., Han, J., & Feng, S. (2025). Accident investigation via LLMs reasoning: HFACS-guided Chain-of-Thoughts enhance general aviation safety. *Expert Systems with Applications, 269*, 126422. doi:https://doi.org/10.1016/j.eswa.2025.126422

Liu, Q., Liu, Y., Leng, X., Han, J., Xia, F., & Chen, H. (2020). Impact of chronic stress on attention control: Evidence from behavioral and event-related potential analyses. *Neuroscience bulletin, 36*(11), 1395-1410.

Llanes-Jurado, J., Marín-Morales, J., Guixeres, J., & Alcañiz, M. (2020). Development and Calibration of an Eye-Tracking Fixation Identification Algorithm for Immersive Virtual Reality. *Sensors, 20*(17), 4956. Retrieved from https://www.mdpi.com/1424-8220/20/17/4956

Lohani, M., Payne, B. R., & Strayer, D. L. (2019). A Review of Psychophysiological Measures to Assess Cognitive States in Real-World Driving. *Frontiers in Human Neuroscience, 13*. Retrieved from <Go to ISI>://WOS:000461777800001

Lu, K., Sjörs Dahlman, A., Karlsson, J., & Candefjord, S. (2022). Detecting driver fatigue using heart rate variability: A systematic review. *Accident Analysis & Prevention, 178*, 106830. doi:https://doi.org/10.1016/j.aap.2022.106830

Lu, T., Li, Y., Zhou, C., Tang, M., & You, X. (2023). The Influence of Emotion Induced by Accidents and Incidents on Pilots' Situation Awareness. *Behavioral Sciences, 13*(3), 231. Retrieved from https://www.mdpi.com/2076-328X/13/3/231

Lu, Z., Happee, R., & de Winter, J. C. (2020). Take over! A video-clip study measuring attention, situation awareness, and decision-making in the face of an impending hazard. *Transportation Research Part F: Traffic Psychology and Behaviour, 72*, 211-225.

Lyons, T. J., Ercoline, W. R., Freeman, J. E., & Gillingham, K. K. (1994). Classification problems of U.S. Air Force spatial disorientation accidents, 1989-91. *Aviation, space, and environmental medicine, 65*(2), 147-152. Retrieved from http://europepmc.org/abstract/MED/8161326

Lyu, M., & Li, F. (2025). Do you need help? Identifying and responding to pilots' troubleshooting through eye-tracking and Large Language Model. *International Journal of Human-Computer Studies*, 103617. doi:https://doi.org/10.1016/j.ijhcs.2025.103617

Lyu, M., Li, F., Lee, C.-H., & Chen, C.-H. (2024). VALIO: Visual attention-based linear temporal logic method for explainable out-of-the-loop identification. *Knowledge-Based Systems, 299*, 112086. doi:https://doi.org/10.1016/j.knosys.2024.112086

Lyu, M., Li, F., Qu, X., & Li, Q. (2024). Flashlight model: Integrating attention distribution and attention resources for pilots' visual behaviour analysis and performance prediction. *International Journal of Industrial Ergonomics, 103*, 103630. doi:https://doi.org/10.1016/j.ergon.2024.103630

Martins, N. A. R., Annaheim, S., Spengler, C. M., & Rossi, R. M. (2021). Fatigue Monitoring Through Wearables: A State-of-the-Art Review. *Frontiers in Physiology, 12*. doi:ARTN 790292
10.3389/fphys.2021.790292

Matthews, G., Neubauer, C., Saxby, D. J., Wohleber, R. W., & Lin, J. (2019). Dangerous intersections? A review of studies of fatigue and distraction in the automated vehicle. *Accident Analysis & Prevention, 126*, 85-94.

Maurino, D. E., Reason, J., Johnston, N., & Lee, R. B. (2017). *Beyond aviation human factors: Safety in high technology systems*: Routledge.

McNeese, N. J., Schelble, B. G., Canonico, L. B., & Demir, M. (2021). Who/What Is My Teammate? Team Composition Considerations in Human–AI Teaming. *IEEE Transactions on Human-Machine Systems, 51*(4), 288-299. doi:10.1109/THMS.2021.3086018

Meeks, R. K., Anderson, J., & Bell, P. M. (2023). *Physiology Of Spatial Orientation*: StatPearls Publishing, Treasure Island (FL).

Mehrali, M., Bagherifard, S., Akbari, M., Thakur, A., Mirani, B., Mehrali, M., . . . Emneus, J. (2018). Blending electronics with the human body: A pathway toward a cybernetic future. *Advanced science, 5*(10), 1700931.

Mehta, R. K., & Parasuraman, R. (2013). Neuroergonomics: a review of applications to physical and cognitive work. *Frontiers in Human Neuroscience, 7*. doi:10.3389/fnhum.2013.00889

Mélan, C., & Cascino, N. (2022). Effects of a modified shift work organization and traffic load on air traffic controllers' sleep and alertness during work and non-work activities. *Applied Ergonomics, 98*, 103596. doi:https://doi.org/10.1016/j.apergo.2021.103596

Meng, X., Zhao, Y., Lu, J. G., Ye, T., & Ma, X. Y. (2025). A Haptic Interface Concept for Highly Automated Vehicle Human-Machine Interfaces Based on a Mathematical Design Method. *International Journal of Human-Computer Interaction*. doi:10.1080/10447318.2025.2546043

Mengtao, L., Fan, L., Gangyan, X., & Su, H. (2023). Leveraging eye-tracking technologies to promote aviation safety- A review of key aspects, challenges, and future perspectives. *Safety Science, 168*, 106295. doi:https://doi.org/10.1016/j.ssci.2023.106295

Merat, N., Jamson, A. H., Lai, F. C. H., Daly, M., & Carsten, O. M. J. (2014). Transition to manual: Driver behaviour when resuming control from a highly automated vehicle. *Transportation Research Part F: Traffic Psychology and Behaviour, 27*, 274-282. doi:https://doi.org/10.1016/j.trf.2014.09.005

Miller, C. A., & Hannen, M. D. (1999). The Rotorcraft Pilot's Associate: design and evaluation of an intelligent user interface for cockpit information management. *Knowledge-Based Systems, 12*(8), 443-456. doi:10.1016/s0950-7051(99)00035-0

Miller, M., & Holley, S. (2023). *Deficits in Cognitive Resilience of Commercial Pilots: The Case for Adding a Computer Information and Automation Tenet in Digital Flight Operations*.

Millot, P., & Pacaux-Lemoine, M.-P. (2013). A common work space for a mutual enrichment of human-machine cooperation and team-situation awareness. *IFAC Proceedings Volumes, 46*(15), 387-394.

Morcos, M., Fishman, S., Saetti, U., Berger, T., Bachelder, E. N., & Godfroy-Cooper, M. (2025). Full-Body Haptic and Spatial Audio Cueing for Augmented Pilot Perception. *Journal of Guidance, Control, and Dynamics, 48*(6), 1334-1347. doi:10.2514/1.G008722

Mumaw, R. J., Billman, D., & Feary, M. S. (2020). Analysis of pilot monitoring skills and a review of training effectiveness.

Naderpour, M., Lu, J., & Zhang, G. (2016). A safety-critical decision support system evaluation using situation awareness and workload measures. *Reliability Engineering & System Safety, 150*, 147-

159. doi:https://doi.org/10.1016/j.ress.2016.01.024
Nakade, T., Fuchs, R., Bleuler, H., & Schiffmann, J. (2023). Haptics based multi-level collaborative steering control for automated driving. *Communications Engineering, 2*(1), 2. doi:10.1038/s44172-022-00051-2
Němcová, A., Svozilová, V., Bucsuházy, K., Smíšek, R., Mézl, M., Hesko, B., . . . Kolář, R. (2021). Multimodal Features for Detection of Driver Stress and Fatigue: Review. *IEEE Transactions on Intelligent Transportation Systems, 22*(6), 3214-3233. doi:10.1109/TITS.2020.2977762
Newman, D. G., & FAICD, A. (2007). *An overview of spatial disorientation as a factor in aviation accidents and incidents*: Australian Transport Safety Bureau Canberra City.
Nguyen, T., Lim, C. P., Nguyen, N. D., Gordon-Brown, L., & Nahavandi, S. (2019). A review of situation awareness assessment approaches in aviation environments. *IEEE Systems Journal, 13*(3), 3590-3603.
Nicholl, R. (2014). Airline head-up display systems: human factors considerations. *International Journal of Economics & Management Sciences*.
Ohneiser, O., De Crescenzio, F., Di Flumeri, G., Kraemer, J., Berberian, B., Bagassi, S., . . . Babiloni, F. (2018). Experimental simulation set-up for validating out-of-the-loop mitigation when monitoring high levels of automation in air traffic control. *International Journal of Aerospace and Mechanical Engineering, 12*(4), 379-390.
Olaverri-Monreal, C., & Jizba, T. (2016). Human Factors in the Design of Human–Machine Interaction: An Overview Emphasizing V2X Communication. *IEEE Transactions on Intelligent Vehicles, 1*(4), 302-313. doi:10.1109/TIV.2017.2695891
Pang, Y., Hu, J., Lieber, C. S., Cooke, N. J., & Liu, Y. (2023). Air traffic controller workload level prediction using conformalized dynamical graph learning. *Advanced Engineering Informatics, 57*, 102113. doi:https://doi.org/10.1016/j.aei.2023.102113
Papadimitriou, E., Schneider, C., Aguinaga Tello, J., Damen, W., Lomba Vrouenraets, M., & ten Broeke, A. (2020). Transport safety and human factors in the era of automation: What can transport modes learn from each other? *Accident Analysis & Prevention, 144*, 105656. doi:https://doi.org/10.1016/j.aap.2020.105656
Parasuraman, R., & Manzey, D. H. (2010). Complacency and Bias in Human Use of Automation: An Attentional Integration. *Human Factors, 52*(3), 381-410. doi:10.1177/0018720810376055
Perello-March, J. R., Burns, C. G., Woodman, R., Elliott, M. T., & Birrell, S. A. (2023). Using fNIRS to Verify Trust in Highly Automated Driving. *IEEE Transactions on Intelligent Transportation Systems, 24*(1), 739-751. doi:10.1109/tits.2022.3211089
Pilla, P., & Moreira, R. A. A. (2022). Cybernetic Systems: Technology Embedded into the Human Experience. In M. V. Albert, L. Lin, M. J. Spector, & L. S. Dunn (Eds.), *Bridging Human Intelligence and Artificial Intelligence* (pp. 171-183). Cham: Springer International Publishing.
Pires, L., Leitão, J., Guerrini, C., & Simões, M. R. (2018). Cognitive control during a spatial Stroop task: Comparing conflict monitoring and prediction of response-outcome theories. *Acta Psychologica, 189*, 63-75. doi:https://doi.org/10.1016/j.actpsy.2017.06.009
Post, J. (2021). The Next Generation Air Transportation System of the United States: Vision, Accomplishments, and Future Directions. *Engineering, 7*(4), 427-430. doi:https://doi.org/10.1016/j.eng.2020.05.026
Prinzel Iii, L. J., Comstock, J. J. R., Glaab, L. J., Kramer, L. J., Arthur, J. J., & Barry, J. S. (2004). The Efficacy of Head-Down and Head-Up Synthetic Vision Display Concepts for Retro- and Forward-Fit of Commercial Aircraft. *The International Journal of Aviation Psychology, 14*(1), 53-77. doi:10.1207/s15327108ijap1401_4
Qin, L., Shi, Y., He, Y., Zhang, J., Zhang, X., Li, Y., . . . Yan, H. (2022). ID-YOLO: Real-Time Salient Object Detection Based on the Driver's Fixation Region. *IEEE Transactions on Intelligent Transportation Systems, 23*(9), 15898-15908. doi:10.1109/TITS.2022.3146271
Ramanathan Parthasarthy, A., Mehrotra, S., Fitzpatrick, C., Roberts, S., Christofa, E., & Knodler, M. (2021). Driver behavior and performances on in-vehicle display based speed compliance. *Accident Analysis & Prevention, 162*, 106390. doi:https://doi.org/10.1016/j.aap.2021.106390
Richards, D., & Lamb, P. Functional Symbology - Evaluation of task-specific Head-Up Display information for use on a commercial flight deck. In *AIAA Modeling and Simulation Technologies Conference*.
Rodriguez-Paras, C., McKenzie, J. T., Choterungruengkorn, P., & Ferris, T. K. (2021). Severity-Mapped Vibrotactile Cues to Support Interruption Management with Weather Messaging in the General Aviation Cockpit. *Atmosphere, 12*(3), 341. Retrieved from https://www.mdpi.com/2073-4433/12/3/341

Rothfuß, S., Wörner, M., Inga, J., Kiesel, A., & Hohmann, S. (2023). Human–Machine Cooperative Decision Making Outperforms Individualism and Autonomy. *IEEE Transactions on Human-Machine Systems, 53*(4), 761-770. doi:10.1109/THMS.2023.3274916

Salvucci, D. D., & Goldberg, J. H. (2000). *Identifying fixations and saccades in eye-tracking protocols*. Paper presented at the Proceedings of the 2000 symposium on Eye tracking research & applications, Palm Beach Gardens, Florida, USA. https://doi.org/10.1145/355017.355028

Sánchez-Tena, M. A., Alvarez-Peregrina, C., Valbuena-Iglesias, M. C., & Palomera, P. R. (2018). Optical Illusions and Spatial Disorientation in Aviation Pilots. *Journal of Medical Systems, 42*(5). doi:10.1007/s10916-018-0935-4

Sarter, N. B., & Woods, D. D. (1995). How in the world did we ever get into that mode? Mode error and awareness in supervisory control. *Human Factors, 37*(1), 5-19.

Sauer, J., Kao, C.-S., & Wastell, D. (2012). A comparison of adaptive and adaptable automation under different levels of environmental stress. *Ergonomics, 55*(8), 840-853. doi:10.1080/00140139.2012.676673

Sehoon, Y., Gyuwang, K., Sejun, K., Cheol, P., & Youngjin, C. (2024). Machine Learning-Based Spatial Disorientation Detection In Rotary-Wing Aircraft. *Nanotechnology Perceptions*. doi:10.62441/nano-ntp.v20is11.93

Sikander, G., & Anwar, S. (2019). Driver Fatigue Detection Systems: A Review. *IEEE Transactions on Intelligent Transportation Systems, 20*(6), 2339-2352. doi:10.1109/Tits.2018.2868499

Silva, P., Garganta, J., Araújo, D., Davids, K., & Aguiar, P. (2013). Shared Knowledge or Shared Affordances? Insights from an Ecological Dynamics Approach to Team Coordination in Sports. *Sports Medicine, 43*(9), 765-772. doi:10.1007/s40279-013-0070-9

Spaccarotella, C., Polimeni, A., Mancuso, C., Pelaia, G., Esposito, G., & Indolfi, C. (2022). Assessment of Non-Invasive Measurements of Oxygen Saturation and Heart Rate with an Apple Smartwatch: Comparison with a Standard Pulse Oximeter. *Journal of Clinical Medicine, 11*(6). doi:10.3390/jcm11061467

Spicer, R. P., Russell, S. M., & Rosenberg, E. S. (2017, Apr 10-11). *The Mixed Reality of Things: Emerging Challenges for Human-Information Interaction.* Paper presented at the 5th Conference on Next-Generation Analyst, Anaheim, CA.

Srivastava, D., Lilly, J. M., & Feigh, K. M. (2024). Exploring the role of judgement and shared situation awareness when working with AI recommender systems. *Cognition, Technology & Work*. doi:10.1007/s10111-024-00771-9

Stanton, N. A. (2016). Distributed situation awareness. *Theoretical Issues in Ergonomics Science, 17*(1), 1-7. doi:10.1080/1463922X.2015.1106615

Stanton, N. A., Plant, K. L., Roberts, A. P., Allison, C. K., & Howell, M. (2019). Seeing through the mist: an evaluation of an iteratively designed head-up display, using a simulated degraded visual environment, to facilitate rotary-wing pilot situation awareness and workload. *Cognition, Technology & Work, 22*(3), 549-563. doi:10.1007/s10111-019-00591-2

Stanton, N. A., Plant, K. L., Roberts, A. P., Allison, C. K., & Howell, M. (2020). Seeing through the mist: an evaluation of an iteratively designed head-up display, using a simulated degraded visual environment, to facilitate rotary-wing pilot situation awareness and workload. *Cognition, Technology & Work, 22*(3), 549-563. doi:10.1007/s10111-019-00591-2

Stanton, N. A., Plant, K. L., Roberts, A. P., Harvey, C., & Thomas, T. G. (2016). Extending helicopter operations to meet future integrated transportation needs. *Applied Ergonomics, 53*, 364-373. doi:10.1016/j.apergo.2015.07.001

Stefana, E., Marciano, F., Rossi, D., Cocca, P., & Tomasoni, G. (2021). Wearable Devices for Ergonomics: A Systematic Literature Review. *Sensors, 21*(3), 777. Retrieved from https://www.mdpi.com/1424-8220/21/3/777

Strand, N., Nilsson, J., Karlsson, I. C. M., & Nilsson, L. (2014). Semi-automated versus highly automated driving in critical situations caused by automation failures. *Transportation Research Part F: Traffic Psychology and Behaviour, 27*, 218-228. doi:10.1016/j.trf.2014.04.005

Stróżak, P., Francuz, P., Lewkowicz, R., Augustynowicz, P., Fudali-Czyż, A., Bałaj, B., & Truszczyński, O. (2018). Selective Attention and Working Memory Under Spatial Disorientation in a Flight Simulator. *The International Journal of Aerospace Psychology, 28*(1-2), 31-45. doi:10.1080/24721840.2018.1486195

Studio, P. (2016). HUD Aircraft GUI. Retrieved from https://assetstore.unity.com/packages/tools/gui/hud-aircraft-gui-55389

Sun, B. H., Shan, Y. J., Wu, G. P., Zhao, S., & Xie, F. (2024). Personalized Shared Control for Automated Vehicles Considering Driving Capability and Styles. *Sensors, 24*(24). doi:10.3390/s24247904

Tan, Z., Dai, N., Su, Y., Zhang, R., Li, Y., Wu, D., & Li, S. (2022). Human–Machine Interaction in Intelligent and Connected Vehicles: A Review of Status Quo, Issues, and Opportunities. *IEEE Transactions on Intelligent Transportation Systems, 23*(9), 13954-13975. doi:10.1109/TITS.2021.3127217

Teshima, T., Niitsuma, M., & Nishimura, H. (2024). Determining the onset of driver's preparatory action for take-over in automated driving using multimodal data. *Expert Systems with Applications, 246*, 123153. doi:https://doi.org/10.1016/j.eswa.2024.123153

Tran, T. H., Behrend, F., Funning, N., & Arango, A. (2018). Single Pilot Operations with AR-Glasses using Microsoft HoloLens. *2018 Ieee/Aiaa 37th Digital Avionics Systems Conference (Dasc)*, 407-413. Retrieved from <Go to ISI>://WOS:000455217900057

van den Hoed, A., Landman, A., Van Baelen, D., Stroosma, O., van Paassen, M. M., Groen, E. L., & Mulder, M. (2022). Leans Illusion in Hexapod Simulator Facilitates Erroneous Responses to Artificial Horizon in Airline Pilots. *Human Factors, 64*(6), 962-972. doi:10.1177/0018720820975248

Van der Burg, E., Ledegang, W. D., Kooi, F. L., Houben, M. M. J., & Groen, E. L. (2024). Attentional Tunneling in Pilots During a Visual Tracking Task With a Head Mounted Display. *Human Factors, 0*(0), 00187208241236395. doi:10.1177/00187208241236395

Van der Burg, E., Ledegang, W. D., Kooi, F. L., Houben, M. M. J., & Groen, E. L. (2025). Attentional Tunneling in Pilots During a Visual Tracking Task With a Head Mounted Display. *Human Factors, 67*(1), 63-78. doi:10.1177/00187208241236395

Verstappen, V. J. M. P., Pikaar, E. N., & Zon, R. G. D. (2022). Assessing the impact of driver advisory systems on train driver workload, attention allocation and safety performance. *Applied Ergonomics, 100*, 103645. doi:https://doi.org/10.1016/j.apergo.2021.103645

Ververs, P. M., & Wickens, C. D. (1998). Head-up displays: Effects of clutter, display intensity, and display location on pilot performance. *International Journal of Aviation Psychology, 8*(4), 377-403. doi:DOI 10.1207/s15327108ijap0804_4

Wang, A., Huang, C., Wang, J., & He, D. (2024). The association between physiological and eye-tracking metrics and cognitive load in drivers: A meta-analysis. *Transportation Research Part F: Traffic Psychology and Behaviour, 104*, 474-487. doi:https://doi.org/10.1016/j.trf.2024.06.014

Wang, P., Wei, X., Hu, F., & Han, W. (2024, 19-21 July 2024). *TransGPT: Multi-modal Generative Pre-trained Transformer for Transportation.* Paper presented at the 2024 International Conference on Computational Linguistics and Natural Language Processing (CLNLP).

Webb, C. M., Estrada Iii, A., & Kelley, A. M. (2012). The Effects of Spatial Disorientation on Cognitive Processing. *The International Journal of Aviation Psychology, 22*(3), 224-241. doi:10.1080/10508414.2012.689211

Weiland, M. Z., Roberts, D. M., Fine, M. S., & Caywood, M. S. (2013). Real Time Research Methods. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 57*(1), 1615-1619. doi:10.1177/1541931213571359

Wen, S., Middleton, M., Ping, S., Chawla, N. N., Wu, G., Feest, B. S., . . . Zahabi, M. (2025). AdaptiveCoPilot: Design and Testing of a NeuroAdaptive LLM Cockpit Guidance System in both Novice and Expert Pilots. *arXiv preprint arXiv:2501.04156*.

Wen, S., Middleton, M., Ping, S., Chawla, N. N., Wu, G., Feest, B. S., . . . Silva, C. T. (2025). *AdaptiveCoPilot: Design and Testing of a NeuroAdaptive LLM Cockpit Guidance System in both Novice and Expert Pilots*.

White, A., & O'Hare, D. (2022). In plane sight: Inattentional blindness affects visual detection of external targets in simulated flight. *Appl Ergon, 98*, 103578. doi:10.1016/j.apergo.2021.103578

Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science, 3*(2), 159-177. doi:10.1080/14639220210123806

Wickens, C. D. (2008). Multiple resources and mental workload. *Human Factors, 50*(3), 449-455. doi:10.1518/001872008X288394

Wickens, C. D., & Alexander, A. L. (2009). Attentional Tunneling and Task Management in Synthetic Vision Displays. *The International Journal of Aviation Psychology, 19*(2), 182-199. doi:10.1080/10508410902766549

Wickens, C. D., Goh, J., Helleberg, J., Horrey, W. J., & Talleur, D. A. (2017). Attentional models of multitask pilot performance using advanced display technology. In *Human Error in Aviation* (pp. 155-175): Routledge.

Wohleber, R. W., Stowers, K., Barnes, M., & Chen, J. Y. C. (2023). Agent transparency in mixed-initiative multi-UxV control: How should intelligent agent collaborators speak their minds? *Computers in Human Behavior, 148*, 107866. doi:https://doi.org/10.1016/j.chb.2023.107866

Wojciechowski, P., & Blaszczyk, J. (2019). SIMULATOR SICKNESS IN THE AIRCRAFT TRAINING OF MILITARY AND CIVIL PILOTS OF VARIOUS TYPES OF AIRCRAFT. *Medycyna Pracy, 70*(3), 317-325. doi:10.13075/mp.5893.00766

Wu, E. Q., Cao, Z., Sun, P. Z. H., Li, D., Law, R., Xu, X., . . . Yu, M. (2022). Inferring Cognitive State of Pilot's Brain Under Different Maneuvers During Flight. *IEEE Transactions on Intelligent Transportation Systems, 23*(11), 21729-21739. doi:10.1109/TITS.2022.3189981

Wu, E. Q., Deng, P. Y., Qiu, X. Y., Tang, Z., Zhang, W. M., Zhu, L. M., . . . Sheng, R. S. F. (2021). Detecting Fatigue Status of Pilots Based on Deep Learning Network Using EEG Signals. *IEEE Transactions on Cognitive and Developmental Systems, 13*(3), 575-585. doi:10.1109/TCDS.2019.2963476

Wu, E. Q., Zhu, L. M., Li, G. J., Li, H. J., Tang, Z. R., Hu, R. H., & Zhou, G. R. (2022). Nonparametric Hierarchical Hidden Semi-Markov Model for Brain Fatigue Behavior Detection of Pilots During Flight. *IEEE Transactions on Intelligent Transportation Systems, 23*(6), 5245-5256. doi:10.1109/tits.2021.3052801

Wu, Q., Dey, N., Shi, F., Crespo, R. G., & Sherratt, R. S. (2021). Emotion classification on eye-tracking and electroencephalograph fused signals employing deep gradient neural networks. *Applied Soft Computing, 110*. doi:10.1016/j.asoc.2021.107752

Xia, Z., Chen, C.-H., Hsieh, M.-H., Ma, G., Liu, B., Yu, X., . . . Dong, L. (2025). A systematic review on human-AI hybrid systems and human factors in air traffic management. *Journal of Engineering Design*, 1-49. doi:10.1080/09544828.2025.2509056

Xiao, F., Ding, W., & Pedrycz, W. (2025). A Generalized f-Divergence With Applications in Pattern Classification. *IEEE Transactions on Knowledge and Data Engineering, 37*(4), 1556-1570. doi:10.1109/TKDE.2025.3530524

Xing, Y., Lv, C., Cao, D. P., & Hang, P. (2021). Toward human-vehicle collaboration: Review and perspectives on human-centered collaborative automated driving. *Transportation Research Part C-Emerging Technologies, 128*. doi:10.1016/j.trc.2021.103199

Xu, G. J. W., Pan, S., Sun, P. Z. H., Guo, K., Park, S. H., Yan, F., . . . Wu, E. Q. (2025). Human-Factors-in-Aviation-Loop: Multimodal Deep Learning for Pilot Situation Awareness Analysis Using Gaze Position and Flight Control Data. *IEEE Transactions on Intelligent Transportation Systems, 26*(6), 8065-8077. doi:10.1109/TITS.2025.3558085

Xu, T., Wang, H., Lu, G., Wan, F., Deng, M., Qi, P., . . . Sun, Y. (2023). E-Key: An EEG-Based Biometric Authentication and Driving Fatigue Detection System. *IEEE Transactions on Affective Computing, 14*(2), 864-877. doi:10.1109/TAFFC.2021.3133443

Xu, W., Dainoff, M. J., Ge, L., & Gao, Z. (2023). Transitioning to Human Interaction with AI Systems: New Challenges and Opportunities for HCI Professionals to Enable Human-Centered AI. *International Journal of Human–Computer Interaction, 39*(3), 494-518. doi:10.1080/10447318.2022.2041900

Yamin, P. A. R., Park, J., & Kim, H. K. (2024). In-vehicle human–machine interface guidelines for augmented reality head-up displays: A review, guideline formulation, and future research directions. *Transportation Research Part F: Traffic Psychology and Behaviour, 104*, 266-285. doi:https://doi.org/10.1016/j.trf.2024.06.001

Yan, S., Ahmed, W., & Saeed, H. A. (2024). Alert and on task: decoding how mental alertness and workload influence maritime operators task performance using task network modeling. *Cognition, Technology & Work, 26*(3), 535-553. doi:10.1007/s10111-024-00769-3

Yang, J., Liang, N., Pitts, B. J., Prakah-Asante, K. O., Curry, R., Blommer, M., . . . Yu, D. (2023). Multimodal Sensing and Computational Intelligence for Situation Awareness Classification in Autonomous Driving. *IEEE Transactions on Human-Machine Systems, 53*(2), 270-281. doi:10.1109/THMS.2023.3234429

Yang, J. H., Tang, M. J., Cong, L., Sun, J. C., Guo, D. L., Zhang, T. H., . . . Hu, W. D. (2024). Development and validation of an assessment index for quantifying cognitive task load in pilots under simulated flight conditions using heart rate variability and principal component analysis. *Ergonomics, 67*(4), 515-525. doi:10.1080/00140139.2023.2229075

Yang, Y., Liu, X., Easa, S. M., Huang, L., & Zheng, X. (2024). Impact of roadside advertisements near traffic signs on driving safety. *Transportation Research Part F: Traffic Psychology and Behaviour, 100*, 254-272. doi:https://doi.org/10.1016/j.trf.2023.10.028

Yin, L., Chahine, M., Wang, T. H., Seyde, T., Liu, C., Lechner, M., . . . Rus, D. (2023, 1-5 Oct. 2023). *Towards Cooperative Flight Control Using Visual-Attention.* Paper presented at the 2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS).

Yiu, C. Y., Ng, K. K. H., Li, Q., & Yuan, X. (2024, 15-17 May 2024). *How do the levels of automation*

in flight operations affect pilots' cognitive workload, reaction time, and EEG brain waves in cruising flights? Paper presented at the 2024 IEEE 4th International Conference on Human-Machine Systems (ICHMS).
Yiu, C. Y., Ng, K. K. H., Li, Q., & Yuan, X. (2025a). Keeping Pilots in the Loop: An Explainable Spatiotemporal EEG-Driven Deep Learning Framework for Adaptive Automation in Cruising Flight Phase. *IEEE Transactions on Intelligent Transportation Systems, 26*(7), 9838-9851. doi:10.1109/TITS.2025.3567987
Yiu, C. Y., Ng, K. K. H., Li, Q., & Yuan, X. (2025b). Keeping Pilots in the Loop: An Explainable Spatiotemporal EEG-Driven Deep Learning Framework for Adaptive Automation in Cruising Flight Phase. *IEEE Transactions on Intelligent Transportation Systems*, 1-14. doi:10.1109/TITS.2025.3567987
Yiu, C. Y., Ng, K. K. H., Li, Q. B., & Yuan, X. (2025c). Gaze behaviours, situation awareness and cognitive workload of air traffic controllers in radar screen monitoring tasks with varying task complexity. *International Journal of Occupational Safety and Ergonomics*. doi:10.1080/10803548.2025.2453312
Yiu, C. Y., Ng, K. K. H., Li, X., Zhang, X., Li, Q., Lam, H. S., & Chong, M. H. (2022). Towards safe and collaborative aerodrome operations: Assessing shared situational awareness for adverse weather detection with EEG-enabled Bayesian neural networks. *Advanced Engineering Informatics, 53*, 101698. doi:https://doi.org/10.1016/j.aei.2022.101698
Yiu, C. Y., Ng, K. K. H., Li, X. Y., Zhang, X. G., Li, Q. B., Lam, H. S., & Chong, M. H. (2022). Towards safe and collaborative aerodrome operations: Assessing shared situational awareness for adverse weather detection with EEG-enabled Bayesian neural networks. *Advanced Engineering Informatics, 53*. doi:10.1016/j.aei.2022.101698
Yu, C. S., Wang, E. M. Y., Li, W. C., Braithwaite, G., & Greaves, M. (2016). Pilots' Visual Scan Patterns and Attention Distribution During the Pursuit of a Dynamic Target. *Aerospace Medicine and Human Performance, 87*(1), 40-47. doi:10.3357/Amhp.4209.2016
Yu, R., Sun, Y., Zhang, X., & Guo, C. (2025). Effect of number of interaction modalities and complexity on pilot performance for civil aircraft flying tasks. *Journal on Multimodal User Interfaces, 19*(4), 391-410. doi:10.1007/s12193-025-00466-z
Yu, W., Jin, D., Cai, W., Zhao, F., & Zhang, X. (2022). Towards tacit knowledge mining within context: Visual cognitive graph model and eye movement image interpretation. *Computer Methods and Programs in Biomedicine, 226*, 107107. doi:https://doi.org/10.1016/j.cmpb.2022.107107
Yu, W., Jin, D., Zhao, F., & Zhang, X. (2023). Towards pilot's situation awareness enhancement: A framework of adaptive interaction system and its realization. *ISA Transactions, 132*, 109-119. doi:https://doi.org/10.1016/j.isatra.2022.12.005
Yuan, X., Li, Q., Gao, Y., Liu, H., Fan, Z., & Bu, L. (2022). Age-related changes in brain functional networks under multisensory-guided hand movements assessed by the functional near – Infrared spectroscopy. *Neuroscience Letters, 781*, 136679. doi:https://doi.org/10.1016/j.neulet.2022.136679
Yuan, X., Ng, K. K. H., Lau, C. K., & Fung, K. H. (2024). Enhancing Single-Pilot Operations With Augmented Reality HUD: Improving Situational Awareness and Performance in a Human-Centric Cockpit Design. In *AIAA AVIATION FORUM AND ASCEND 2024*.
Yuan, X., Ng, K. K. H., Li, Q., Yiu, C. Y., Lau, C. K., Fung, K. H., & Ng, L. H. (2024). Exploring the Human-Centric Interaction Paradigm: Augmented Reality-Assisted Head-Up Display Design for Collaborative Human-Machine Interface in Cockpit. *Advanced Engineering Informatics, 62*, 102656. doi:https://doi.org/10.1016/j.aei.2024.102656
Yuan, X., Ng, K. K. H., Li, Q., Yiu, C. Y., Zhou, Q., & Bu, Y. (2025). From unrecognized to incapacitating: Neuro-ergonomic differentiation of spatial disorientation typologies induced by visual illusions for pilot cognitive training and cockpit design. *Chinese Journal of Aeronautics*, 103992. doi:https://doi.org/10.1016/j.cja.2025.103992
Yuan, X., Ng, K. K. H., Yiu, C. Y., & Li, Q. (2025, 2025//). *Optimising Pilot-Aircraft Interaction: A Low-Cost Projection-Enhanced Head-Up Display (PE-HUD) with Neuro-Ergonomic Validation.* Paper presented at the HCI International 2025 Posters, Cham.
Zhang, C., Wang, Z., Zhou, G., Chang, F., Ma, D., Jing, Y., . . . Zhao, D. (2023). Towards new-generation human-centric smart manufacturing in Industry 5.0: A systematic review. *Advanced Engineering Informatics, 57*, 102121. doi:https://doi.org/10.1016/j.aei.2023.102121
Zhang, J., & Edwards, T. E. (2017). Guest editorial for special issue on modeling and analysis of human–machine systems in transportation. *Cognition, Technology & Work, 19*(4), 543-544. doi:10.1007/s10111-017-0449-8

Zhang, M., Liu, D., Wang, Q., Zhao, B., Bai, O., & Sun, J. (2022). Detection of alertness-related EEG signals based on decision fused BP neural network. *Biomedical Signal Processing and Control, 74*, 103479. doi:https://doi.org/10.1016/j.bspc.2022.103479

Zhang, S., Fu, D., Liang, W., Zhang, Z., Yu, B., Cai, P., & Yao, B. (2024). TrafficGPT: Viewing, processing and interacting with traffic foundation models. *Transport Policy, 150*, 95-105. doi:https://doi.org/10.1016/j.tranpol.2024.03.006

Zhang, T., Yang, J., Liang, N., Pitts, B. J., Prakah-Asante, K. O., Curry, R., . . . Yu, D. (2020). Physiological Measurements of Situation Awareness: A Systematic Review. *Human Factors*, 0018720820969071. doi:10.1177/0018720820969071

Zhang, Y., Li, F., Man, S. S., & Han, S. (2025). Enhancing Readback Verification in ATCOs-Pilot Communication Using Audio-Visual Interaction: Insights from fNIRS and Eye-Tracking. *International Journal of Human–Computer Interaction*, 1-19. doi:10.1080/10447318.2025.2564270

Zhang, Y., Lou, S., Hu, Z., Zhu, J., & Lv, C. (2025). Socially-Compliant Hierarchical Human–Vehicle Collaboration With Multimodal Haptic Steering. *IEEE Transactions on Intelligent Transportation Systems, 26*(8), 11809-11822. doi:10.1109/TITS.2025.3575593

Zhang, Z., Feng, S., Yang, T., Huang, R., Wang, H., Wang, F., & Li, F. (2026). AviationCopilot: Building a reliable LLM-based Aviation Copilot inspired by human pilot training. *Advanced Engineering Informatics, 69*, 103806. doi:https://doi.org/10.1016/j.aei.2025.103806

Zhou, F., Yang, X. J., & Winter, J. C. F. d. (2022). Using Eye-Tracking Data to Predict Situation Awareness in Real Time During Takeover Transitions in Conditionally Automated Driving. *IEEE Transactions on Intelligent Transportation Systems, 23*(3), 2284-2295. doi:10.1109/TITS.2021.3069776

Zhou, Y. Y., Huang, S., Xu, Z. M., Wang, P. P., Wu, X., & Zhang, D. Q. (2022). Cognitive Workload Recognition Using EEG Signals and Machine Learning: A Review. *IEEE Transactions on Cognitive and Developmental Systems, 14*(3), 799-818. doi:10.1109/tcds.2021.3090217

Ziv, G. (2017). Gaze Behavior and Visual Attention: A Review of Eye Tracking Studies in Aviation. *The International Journal of Aviation Psychology, 26*(3-4), 75-104. doi:10.1080/10508414.2017.1313096

Zuo, Y., Qi, J., Maksimovic, C., Chen, B., Hu, J., & Zhao, Q. (2025). Visual Attention Distribution of Pilot Flying vs. Pilot Monitoring During Different Flight Phases. *International Journal of Human–Computer Interaction*, 6495–6514. doi:10.1080/10447318.2024.2380949