



HAL
open science

Data-driven design in virtual reality : combining user perception with operational and physiological measures

Léa Saunier

► **To cite this version:**

Léa Saunier. Data-driven design in virtual reality : combining user perception with operational and physiological measures. Artificial Intelligence [cs.AI]. Institut Polytechnique de Paris, 2024. English. NNT : 2024IPPAS027 . tel-04849713

HAL Id: tel-04849713

<https://theses.hal.science/tel-04849713v1>

Submitted on 19 Dec 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



INSTITUT
POLYTECHNIQUE
DE PARIS

NNT : 2024IPPAS027

Thèse de doctorat



Data-driven Design in Virtual Reality: Combining User Perception with Operational and Physiological Measures

Thèse de doctorat de l'Institut Polytechnique de Paris
préparée à Télécom SudParis

École doctorale n°626 École doctorale de l'Institut Polytechnique de Paris (EDIPP)
Spécialité de doctorat : Signal, Images, Automatique et Robotique

Thèse présentée et soutenue à Évry-Courcouronnes, le 13 décembre 2024, par

LÉA SAUNIER

Composition du Jury :

Béatrice Pesquet-Popescu Directrice Innovation Alten, HDR	Présidente
Valérie Gouranton Professeur, INSA Rennes (Département Informatique)	Rapporteur
Alexis Paljic Professeur, Mines Paris PSL (CAOR)	Rapporteur
Philippe Guillotel Directeur de recherche émérite, InderDigital Inc.	Examineur
Catalin Fetita Professeur, Télécom SudParis (SAMOVAR)	Directeur de thèse
Marius Preda Docteur, Télécom SudParis (SAMOVAR)	Co-directeur de thèse
Michel Simatic Directeur d'étude, Télécom SudParis (Département Informatique)	Invité

Contents

List of Figures and Tables	5
Glossary	7
Acknowledgements	8
French Summary	9
1 Introduction	16
1.1 Designing Virtual Reality - Challenges and Gaps	18
1.2 VR-SOLUS, Sensory, Operational and User data Synthesis for Virtual Reality	19
1.2.1 Hypothesis	19
1.2.2 Experimental Setup	19
1.2.3 Materials	20
1.2.4 Summary of Findings	21
2 Related Work - Designing Virtual Reality Applications and Experiments	23
2.1 Designing Virtual Reality	23
2.2 Design Methodologies	24
2.2.1 Traditional Virtual Reality Design	24
2.2.2 Data-driven Virtual Reality Design	25
2.3 Proposed Methodology	27
3 Operational Measures	28
3.1 General Performance Indicators	28
3.1.1 Time	29
3.1.2 Errors	30
3.1.3 Multi-tasking	31
3.1.4 Implementation of Common Performance Indicators	32
3.2 Use-case Specific Measures	38
3.2.1 Implementation through Study - Earthwork Teleoperation	38
3.2.2 Implementation through Study - Dyslexia Therapy	43
3.3 Evaluation of Extrinsic Operational Measures	46
4 Physiological Measures	47
4.1 Physiological Sensors Integration	48
4.2 Eye-tracking	49
4.3 Electroencephalogram	51
4.4 Physical Load Sensors	53
4.5 Implementation of Physiological Measures	54
4.5.1 Hardware Selection	54
4.5.2 Software Optimisation	59
4.6 Challenges of Signal Processing	65
4.6.1 Eye-tracking and Intent-tracking	66
4.6.2 Electroencephalogram and Artefacts	67

4.6.3	Accelerometers and Heart Rate	69
4.7	Evaluation of Intrinsic Physiological Measures	71
5	Hybrid Approach and Correlations Analysis	72
5.1	VR-SOLUS Hybrid Pipeline	73
5.2	User Feedback vs Data-driven Method	74
5.2.1	Testing Methodologies - Implementation	74
5.2.2	Design Phase	75
5.2.3	Evaluation Phase	77
5.2.4	The Hybrid Method	85
5.3	Correlation Study	86
5.3.1	Operational Correlation	86
5.3.2	Operational - Physiological Correlation	87
5.3.3	Objective - Subjective Correlation	89
5.3.4	Evaluation of Hybrid Measures and Correlation	91
6	Discussion - Towards Further Virtual Reality Design	92
6.1	Contributions	92
6.1.1	VR-SOLUS	92
6.1.2	Earthwork Use-case	94
6.1.3	Therapeutic Use-case	98
6.2	Limitations	99
6.2.1	Experimental Setup	99
6.2.2	Software Improvement	100
6.2.3	Hardware Improvement	101
6.3	Further Developments	102
7	Conclusion	104
	References	106
A	Miscellaneous Figures	115
B	Experimental Surveys	118
C	Supplementary Data Analysis	123

List of Figures

1.1	Virtual Reality Timeline	16
1.2	Extended Reality (XR) Software Market 2025 Fields and Revenue (according to Statist)	17
1.3	The VR-SOLUS Framework	22
3.1	Operational Metrics Framework Highlighted into the Global Analysis Approach Proposed by VR-SOLUS	29
3.2	Example of Completion Time Metric*	30
3.3	Example of Input Errors Metric	31
3.4	Example of Multi-input Metric, Percentage of Time spent Using more than One Input	32
3.5	Controller Selection Spectrum	32
3.6	HTC Vive Pro Eye	34
3.7	Controller Selection Hardware	34
3.8	User Interface for the Controller Selection Experiment with Interchangeable Controller Mapping Scheme Display (Here the Joysticks Mappings)	34
3.9	Gaming, Hybrid and Professional Mappings	35
3.10	Completion Time Metric	36
3.11	Input Errors per Controller	37
3.12	Multi-inputs Percentage across Exercises	37
3.13	Comparing VR Interface and Real Machine, a HERACLES Robotics Partnered Study	40
3.14	Tutorial Exercise - Learning Basic Controls	40
3.15	Target Practice - Learning Precise Control	40
3.16	Digging/Unload Exercise - Learning Teleoperation Main Skill	41
3.17	Unity-ROS Excavator Control Architecture	41
3.18	Control Smoothness compared with Completion Time in the Real Machine interface (RM) and the Virtual Reality Interface (VR)	42
3.19	<i>Dys & Dragons</i> Exercises	44
3.20	Immersive Visual Rehabilitation for Neurological Troubles, a Proposed Framework	45
4.1	Physiological Metrics Framework Highlighted into the Global Analysis Approach Proposed by VR-SOLUS	49
4.2	Traditional vs VR Eye-tracking Setups	49
4.3	HTC Vive Pro Eye with Integrated Tobii 120Hz Eye-tracker	50
4.4	Brainwave Spectrum and their Associated Cognitive States	51
4.5	Common ERP Responses	52
4.6	Cerebral Cortex and its Associated Functions by Lobe	52
4.7	g.tec Unicorn Hybrid Black and its Electrodes Mapping	53
4.8	Physical Load Capture Setup	54
4.9	Physiological Data Pipeline v1	55
4.10	Eye-tracking Raycasting in Unity	56
4.11	Percentage of time Looking at Mappings required by Control Usage Reminder	56
4.12	Visual Representation of the Theta-Alpha Ratio	57
4.13	Cognitive Load with Regular Normalisation	58
4.14	Cognitive Load with Motion-Hard Normalisation	58
4.15	Cognitive Load with Rest-hard Normalisation	59
4.16	Virtual Environment Example for the VA Study	61

4.17 TAR Evaluation for each Set of Virtual Annotation	63
4.18 Head and Arm Motions (in mG)	64
4.19 Average Heart Rate per Set (in BPM)	64
4.20 VR-SOLUS - Physiological Data Pipeline v2	66
4.21 An Example of the Midas Touch Effect Issue in VR	67
4.22 EEG Pipeline - From Capture to Analysis Software Pipeline using Unity, Unicorn Suite, LSL, OpenVibe and Python	70
5.1 Objective Metrics Framework Highlighted into the Global Analysis Approach Proposed by VR-SOLUS	72
5.2 Operational + Physiological + User Feedback Pipeline	75
5.3 Design Phase Operational Results	76
5.4 Completion Time for each Methodology Set	78
5.5 Input Errors for each Methodology Set	78
5.6 Multi-input Percentage for each Methodology Set	79
5.7 Cognitive Load Evaluation for each VA Set	80
5.8 Physical Load Measures	81
5.9 World Instructions VA - Text vs Icon	83
5.10 Users Ranking of the VAs Sets, with the Number of First Places Attributed to each Set	83
5.11 Combining Performance, User Status and Perception through VR-SOLUS	86
5.12 Sweller's Cognitive Load Theory (CLT)	91
6.1 VR-SOLUS - Designing and Optimising Virtual Reality through a Hybrid Approach	93
6.2 HERACLES Robotics, French Earthwork Machinery Automation	94
6.3 Interface Comparison Setup Example - Real Machine (Left) versus VR Interface Remote Control (Right)	95
6.4 Hardware Selection Study Setup	96
6.5 Example of Interface used for the Software Optimisation Study	97
6.6 Dys&Dragons Interface Example	98
6.7 VR-SOLUS and Adaptive Design	102
6.8 Galea VR Headset, the Next Generation of Bio-sensors Integration VR Hardware	103
7.1 Sensory, Operational and User data Synthesis methodology for Virtual Reality - VR-SOLUS	105
A.1 Excavator Schematics	115
A.2 Real (HERACLES Robotics') and Virtual (VR-SOLUS') Excavator	116
A.3 Final Joysticks Mapping	117
A.4 Virtual Annotations Study Interface	117
B.1 NASA-Task Load Index Survey Template	119
B.2 System Usability Scale Inspired Survey Template (1st page)	120
B.3 System Usability Scale Inspired Survey Template (2nd page)	121
B.4 System Usability Scale Inspired Survey Template (3rd page)	122

List of Tables

2.1	Advantages and Limitations of Traditional Design Methods	25
2.2	Advantages and Limitations of Data-driven Design Methods	26
2.3	Advantages and Limitations of the Proposed Method	27
3.1	Validation of the Multi-input (MI%) Metric (Pearson correlation)	38
4.1	Existing Sensor-based Methodologies	48
4.2	Virtual Annotations and their Display Modalities Tested in our Study	60
4.3	Virtual Annotations Sets	61
4.4	Virtual Annotation Eye-tracking Selection	63
5.1	Design Phase NASA-TLX Results	76
5.2	Design Phase SUS Results	77
5.3	Virtual Annotations Sets in the Evaluation Phase	77
5.4	Eye-tracking Analysis: Modalities Matches between the Design and the Evaluation Phase Results (a cross marks when both the Design Phase and the Evaluation Phase modalities match)	80
5.5	Evaluation Phase NASA-TLX Results	82
5.6	SUS Analysis, Affinity, Usefulness, Modalities Matches in the Evaluation Phase Results (a cross marks when both the SUS Results and the Evaluation Phase modalities match)	82
5.7	Data Correlation for Operational Metrics (Pearson correlation)	87
5.8	Data Correlation for Physiological Metrics (Pearson correlation) - LookAt represents the User Attention and TAR the Cognitive Load	88
5.9	Correlation between User Feedback and Measured Data (Pearson correlation)	89
5.10	Correlation between User Profile and Physiologically Measured Cognitive Load (Pearson Correlation)	90
C.1	Statistical Analysis of the VR-SOLUS Measures Across the Different Studies (One-way ANOVA)	124

Glossary

- **VR: Virtual Reality** - «A “virtual reality” is defined as a real or simulated environment in which a perceiver experiences telepresence.» – Steuer 1992
- **ET: Eye-Tracking** - A process of measuring where and how long a person looks (gaze) at various points in a visual environment.
- **EEG: Electroencephalogram** - A method used to record electrical activity of the brain.
- **IMU: Inertial Measurement Unit** - A sensor that measures and reports on a body’s specific force, angular rate, and sometimes the magnetic field surrounding the body, used to track motion.
- **Input** - As in "user input", when used when talking about an action in an application, it means the interaction of the user with the application itself. Example: pressing a button, moving a joystick, clicking (and so on) are (user) inputs. User input, outside the scope of application interaction, is more synonymous to "user feedback".
- **TAR: Theta-Alpha Ratio** - A measure of cognitive load based on the study of the Theta and Alpha bandwave present in the brain.
- **Cognitive Load**: The amount of mental effort being used in the working memory.
- **BPM: Beats Per Minute** - A measure of heart rate, often used as an indicator of stress or physical exertion.
- **VA: Virtual Annotation** - Virtual marks or notes added within a VR environment, used to guide or inform users during tasks.
- **Teleoperation**: The remote control of machinery or robots from a distance, commonly used in fields such as earthwork robotics.
- **Digital Twin**: A virtual model designed to accurately reflect a physical object, often used in simulations and training.
- **Midas Touch**: A challenge in eye-tracking where every gaze is interpreted as an intentional selection, leading to false interactions.
- **Peripheral Vision**: The part of vision that occurs outside the very center of gaze, crucial in immersive environments like VR.
- **Cybersickness**: A form of motion sickness that occurs in virtual environments, often associated with VR use. Motion sickness induced by cyber-environments’ interactions.
- **Bio-sensors**: Biometric sensors recording data derived from measurements of the body, such as heart rate, brain activity, and movement patterns.
- **HMD: Head-Mounted Display** - A device worn on the head that has a small display optic in front of one (monocular HMD) or each eye (binocular HMD).

Acknowledgements

Completing this thesis has been a journey marked by intellectual challenges, personal growth, and the unwavering support of many individuals and organisations. It is with deep gratitude that I acknowledge those who have contributed to this work and supported me throughout this process.

First and foremost, I would like to express my sincere thanks to my director, Catalin Fetita and co-director Marius Preda, for their guidance, insight, and encouragement. Their expertise and mentorship have been invaluable in shaping this research and in helping me navigate the complexities of academic inquiry.

I am also deeply grateful to the members of the jury for their willingness to engage with this work, and especially to the reviewers who will now carefully examine the following hundred-plus pages. Your time and effort in evaluating this thesis are greatly appreciated.

A special thanks goes to Télécom SudParis and the Institut Polytechnique de Paris, my academic institution, for their investment in this project and for the support provided by their administration. Without this backing and the help of their personnel, the research presented here would not have been possible.

I would also like to extend my appreciation to HERACLES Robotics, the company that played a crucial role in our studies. The collaboration with their personnel has been essential in bringing this work to fruition. The same gratitude goes to the medical staff involved in this thesis and their invaluable expertise required for this thesis completion.

To my colleagues, thank you for the shared experiences, the discussions, the reviews, and the camaraderie that made the challenges of research more bearable. Your presence has been a constant source of motivation.

Finally, and most importantly, I want to thank my whole family, especially my parents, for their unconditional support and belief in me. To my friends, particularly spacey (TBI) and the Free Feathers (including their DM of course), and all those who supported me along this thesis, thank you for listening to my endless complaints over the past three (and some) years! Your patience, encouragement, and friendship have been my anchor, and for that, I am forever grateful.

Merci à tous !

*"Might go somewhere sunny. Sit on beach, look at ocean, collect sea shells.
Might run tests on the sea shells."*

Mordin Solus

French Summary

Introduction

La réalité virtuelle (RV) a connu une évolution significative depuis ses origines au XIXe siècle avec la création du stéréoscope. Cependant, la version moderne de la RV telle que nous la connaissons aujourd'hui a été conceptualisée dans les années 1960. Ce domaine en pleine expansion touche désormais divers secteurs, y compris l'industrie 4.0, les thérapies médicales, le jeu vidéo, et le métavers. La RV a révolutionné divers aspects de la société moderne, s'imposant comme une technologie clé dans de nombreux secteurs. Dans l'industrie 4.0, la RV est utilisée pour simuler des environnements de production, former les employés, et optimiser les processus de fabrication. Cette immersion virtuelle permet une visualisation précise des chaînes de production et une simulation des opérations avant leur mise en œuvre réelle, réduisant ainsi les risques et améliorant l'efficacité des opérations. En thérapie, la RV offre des solutions innovantes pour traiter divers troubles, allant des phobies aux troubles de l'anxiété et à la réhabilitation physique et motrice. Les environnements virtuels ainsi développés permettent des interventions adaptées dans un cadre contrôlé, facilitant ainsi la progression thérapeutique. Dans le domaine du jeu vidéo, la RV a transformé l'expérience de jeu, permettant aux joueurs de plonger dans des mondes interactifs avec un niveau d'immersion sans précédent. Les expériences immersives offrent une nouvelle dimension d'interaction et d'engagement, modifiant profondément les dynamiques du divertissement interactif. Par ailleurs, la RV trouve des applications dans des domaines émergents comme le métavers, où elle sert de plateforme pour la création de nouveaux espaces sociaux et économiques, intégrant des aspects de la vie quotidienne dans des environnements virtuels partagés. Ces diverses utilisations démontrent la polyvalence de la réalité virtuelle et son potentiel pour transformer non seulement le divertissement et la formation, mais aussi la manière dont nous abordons des défis complexes dans des contextes variés. L'essor de la RV dans ces domaines a entraîné une demande accrue pour des expériences immersives de haute qualité, ce qui a mis en lumière la nécessité d'optimiser la conception de ces expériences.

Historiquement, les méthodes de conception en RV reposent largement sur des questionnaires et des retours subjectifs venant des utilisateurs. Bien que ces méthodes fournissent des aperçus précieux sur l'expérience utilisateur, elles présentent des limites notables, notamment le biais subjectif et la difficulté à quantifier des aspects complexes de l'interaction en RV. Pour surmonter ces limitations, des approches basées sur des données objectives, telles que les mesures opérationnelles et physiologiques, ont émergé. Ces méthodes promettent de fournir une analyse plus complète et quantitative des interactions et de l'état des utilisateurs en RV, allant au-delà des évaluations subjectives. Dans ce contexte, cette thèse propose une méthodologie hybride appelée VR-SOLUS - Méthode de synthèse des données sensorielles, opérationnelles et de l'utilisateur pour la conception de la RV (en anglais Sensory, Operational and User data Synthesis methodology for Virtual Reality) - qui intègre des mesures opérationnelles, des mesures physiologiques et des retours utilisateurs pour optimiser la conception et l'optimisation des environnements RV. Cette approche vise à combiner les avantages des méthodes traditionnelles et des nouvelles approches basées sur les données pour fournir une vue plus complète des expériences en RV.

Travaux Similaires

Méthodes Traditionnelles

Les méthodes de conception traditionnelles reposent principalement sur des retours subjectifs des utilisateurs. Ces méthodes incluent des enquêtes, des questionnaires, et des interviews pour évaluer l'expérience utilisateur

en RV. Bien que ces méthodes puissent fournir des informations précieuses sur les préférences et les niveaux de satisfaction des utilisateurs, elles présentent plusieurs limitations :

- **Biais Subjectif** : les réponses des utilisateurs peuvent être influencées par des biais personnels et des perceptions individuelles, ce qui peut affecter l'objectivité des données collectées. Les questionnaires et interview ayant souvent lieu après les phases de test, il est aussi possible d'observer un biais de mémorisation : l'utilisateur peut ne pas se souvenir de certains éléments, ou d'en confondre certains entre eux;
- **Difficulté de Quantification** : certaines expériences et interactions en RV sont difficiles à quantifier de manière précise en utilisant uniquement des méthodes subjectives, dû à la difficulté inhérente à quantifier la perception humaine;
- **Variabilité des Réponses** : les retours des utilisateurs peuvent varier considérablement même au sein d'un même groupe (i.e. même profils) de participants, rendant difficile l'identification de tendances générales ou de problèmes spécifiques.

Méthodes Basées sur les Données

Les méthodes basées sur les données, quant à elles, se concentrent sur la collecte et l'analyse de données objectives pour évaluer les expériences en RV. Ces méthodes incluent:

- **Mesures Opérationnelles** : elles évaluent des indicateurs tels que le temps de complétion des tâches, les erreurs commises par l'utilisateur, et les interactions avec l'environnement RV de celui-ci. Ces mesures fournissent des données quantitatives sur la performance des utilisateurs;
- **Mesures Physiologiques** : les capteurs physiologiques, tels que le suivi du regard, l'électroencéphalogramme (EEG), les capteurs de charge physique (IMU), et les moniteurs de fréquence cardiaque, mesurent les réponses physiologiques des utilisateurs, telles que l'attention, la charge cognitive et le niveau de fatigue. Ces mesures offrent des informations sur l'état cognitif et physique des utilisateurs.
- **Analyse des Données** : l'analyse des données collectées ainsi que de leurs corrélations peut révéler des tendances et des modèles dans les interactions des utilisateurs, offrant une perspective plus objective sur l'expérience en RV.

Utilisation de Bio-capteurs

L'intégration de bio-capteurs dans la conception d'environnements de RV représente une avancée significative dans l'amélioration de l'expérience utilisateur et dans l'optimisation des interfaces homme-machine. Les bio-capteurs permettent de recueillir des données physiologiques en temps réel, offrant ainsi une compréhension plus approfondie des états cognitifs et émotionnels des utilisateurs lors de leur immersion dans des environnements virtuels. Les méthodologies de conception RV présentent des travaux récents d'exploration de divers bio-capteurs dans la conception RV.

Électroencéphalogramme (EEG)

L'EEG est l'un des bio-capteurs les plus couramment utilisés dans la recherche sur la VR. Il permet de mesurer l'activité électrique du cerveau et d'en déduire des indicateurs de charge cognitive, d'attention et d'engagement émotionnel. Plusieurs études ont démontré l'efficacité de l'EEG pour adapter dynamiquement les environnements virtuels en fonction de l'état mental de l'utilisateur. Ce capteur peut être utilisé pour extraire de nombreuses mesures : charge cognitive, concentration, stress, cinématose (ou motion sickness en anglais, mal au coeur lié à la désynchronisation entre la perception visuelle et physique des mouvements en RV), etc. Cependant, l'EEG présente des défis significatifs en termes de bruit et d'interférences des signaux, particulièrement dans des contextes immersifs où les utilisateurs se déplacent et interagissent avec leur environnement.

Suivi Oculaire

Le suivi oculaire est un autre bio-capteur crucial dans l'étude de l'interaction en RV. Il permet de capturer les mouvements des yeux, offrant des informations précieuses sur les points d'intérêt et les zones de focalisation de l'utilisateur. Cette technologie a été largement utilisée pour évaluer l'efficacité des interfaces RV et pour détecter des problèmes potentiels dans la conception des interfaces, tels que des éléments trop distrayants ou mal placés. Le suivi oculaire peut permettre d'évaluer l'attention de l'utilisateur, les zones d'intérêt visuel (et inversement les zones inutiles ou inutilisées) d'un environnement virtuel, le stress, la concentration, etc. Étudier les mouvements oculaires d'un utilisateur permet de créer des applications plus fluides, et qui répondent directement à l'attention de l'utilisateur.

Mesure de la Fréquence Cardiaque et de la Réponse Galvanique de la Peau (GSR)

Les capteurs de fréquence cardiaque et de réponse galvanique de la peau sont souvent utilisés pour évaluer le niveau de stress ou d'excitation de l'utilisateur. Ces mesures sont particulièrement utiles dans les applications RV où la gestion des émotions est cruciale, comme dans les simulations de formation ou les thérapies immersives. Dans cette thèse, nous nous concentrons surtout sur la fréquence cardiaque et les mouvement généraux du corps, qui peuvent indiquer le niveau de charge physique de l'utilisateur, la reconnaissance de geste ou l'évaluation de la douleur et du confort.

Défis et Limites

Malgré les avantages évidents de l'intégration des bio-capteurs dans la conception VR, plusieurs défis persistent. L'un des principaux défis est la synchronisation des données provenant de multiples capteurs, chacun ayant ses propres caractéristiques, fréquence d'échantillonnage et latence. De plus, l'intégration physique des capteurs dans le matériel VR existant, tel que les casques ou les manettes, peut poser des problèmes en termes de confort et de mobilité des utilisateurs. Les artefacts de mouvement et le bruit électronique dans les signaux enregistrés, particulièrement présents dans les mesures EEG, constituent également des obstacles importants à surmonter pour obtenir des données fiables et exploitables.

Les travaux récents se sont concentrés sur l'amélioration de la fiabilité et de l'intégration des bio-capteurs dans les environnements RV. Par exemple, l'émergence de casques VR équipés de capteurs intégrés, tels que des électrodes EEG directement intégrées dans le bandeau du casque, ouvre de nouvelles possibilités pour des expériences encore plus immersives et personnalisées. De plus, le développement de techniques avancées de traitement du signal, telles que les filtres adaptatifs et l'apprentissage automatique, permet une analyse plus fine et en temps réel des données physiologiques.

Méthode Proposée

La méthodologie VR-SOLUS propose une approche hybride en combinant des mesures opérationnelles, physiologiques, et des retours utilisateurs pour optimiser les environnements RV. Cette approche vise à surmonter les limitations des méthodes traditionnelles en fournissant une analyse plus complète et nuancée des expériences en RV. La méthodologie inclut des mesures opérationnelles qui forment des indicateurs de performances. Elles incluent le temps de complétion, les erreurs, la facilité de contrôle ainsi que des mesures spécifiques aux cas d'usage étudiés. VR-SOLUS inclut également des mesures physiologiques, comme l'attention (par suivi du regard), la charge cognitive (par étude EEG) ou la charge physique (grâce à des accéléromètre et la mesure du rythme cardiaque). Le suivi des réponses physiologiques sert à évaluer le statut d'un participant à un moment donné. Enfin, VR-SOLUS n'abandonne pas les retours Utilisateurs, mais en fait une de ses forces. En collectant des évaluations subjectives pour obtenir des informations sur la satisfaction et le confort des utilisateurs, VR-SOLUS peut ainsi évaluer la perception des utilisateurs, et apporter une dimension subjective à l'évaluation de l'expérience utilisateur.

Matériel et Conception

Matériel

Cette thèse s'inscrit dans le cadre de l'optimisation des interactions homme-machine en RV à travers le développement de la méthodologie VR-SOLUS. Pour ce faire, deux champs d'application principaux ont été choisis : les travaux de terrassement et la thérapie, chacun représentant des défis uniques en termes de conception d'interfaces et d'optimisation des performances utilisateur.

Dans le cadre de nos travaux, nous avons utilisé une combinaison d'équipements de RV et de dispositifs spécialisés. Les études ont principalement été réalisées avec le casque de réalité virtuelle HTC Vive Pro Eye, qui intègre un suivi oculaire pour la capture des mouvements des yeux des utilisateurs. Ce casque permet une immersion complète des participants dans les environnements virtuels. Pour les mesures physiologiques, un système EEG g.tec Unicorn Hybrid Black a été utilisé afin de capter l'activité cérébrale des participants. Cet appareil, équipé de 8 électrodes, permet de suivre en temps réel les signaux neuronaux, bien que sa résolution spatiale et sa sensibilité puissent être limitées par rapport à des systèmes EEG plus complexes (e.g. avec plus d'électrodes). Le suivi de la charge physique a été effectué à l'aide du capteur de fréquence cardiaque Polar OH1+, couplé à un accéléromètre pour enregistrer les mouvements du corps.

Les expériences ont été pilotées par des logiciels dédiés au contrôle des dispositifs RV et à l'analyse des données recueillies. Unity a été le moteur central pour la création et la gestion des environnements virtuels, facilitant l'intégration des différents capteurs et l'interaction avec l'utilisateur. Pour l'analyse des données EEG, nous avons utilisé la suite logicielle Unicorn Suite, complétée par OpenVibe, un outil open-source spécialisé dans le traitement et la visualisation des signaux EEG. Les mesures oculaires ont été traitées via le SDK SRanipal de HTC, qui s'intègre directement dans Unity pour fournir des données en temps réel sur les fixations et les saccades des utilisateurs. Pour le suivi des données de charge physique, l'application Polar Sensor Logger a été employée, permettant de synchroniser les données de fréquence cardiaque et les mouvements du corps avec le reste des mesures recueillies.

Conception des Études

Les études menées dans cette thèse sont structurées autour de deux grandes applications : l'entraînement et la téléopération dans le domaine des travaux de terrassement, et le développement d'une thérapie innovante et immersive pour les personnes atteintes de dyslexie.

Dans les études de terrassement, les participants étaient immergés dans un environnement de simulation RV où ils devaient accomplir diverses tâches, telles que l'excavation et la navigation d'engin de chantier. Ces tâches étaient conçues pour reproduire les conditions réelles de travail, tout en permettant la mesure continue de leurs performances opérationnelles (temps d'exécution, erreurs de manipulation), de leurs réponses physiologiques (activité cérébrale, rythme cardiaque), et de leurs retours subjectifs (questionnaires).

Pour la thérapie de la dyslexie, des exercices visuels spécifiques ont été développés, simulant les traitements classiques tout en explorant de nouvelles approches grâce à la RV. L'idée principale était d'exploiter les mesures physiologiques et opérationnelles pour ajuster dynamiquement les exercices en fonction des besoins individuels des patients.

Collecte et Analyse des Données

Les données collectées au cours des études étaient de nature variée, incluant des mesures physiologiques, opérationnelles et subjectives. Les mesures opérationnelles, comme le temps de tâche et les erreurs commises, ont été analysées pour évaluer l'efficacité et la performance des interfaces proposées. Les données physiologiques, notamment le suivi oculaire, et l'activité EEG, ont été synchronisées grâce au Lab Streaming Layer, traitées via le logiciel OpenVibe pour extraire des marqueurs cognitifs (charge mentale, attention) et analysées avec Python. Ces analyses ont impliqué l'utilisation de techniques avancées de filtrage et de synchronisation des signaux, en raison

des différentes fréquences d'échantillonnage et de la complexité des signaux enregistrés. Les questionnaires post-tâches ont permis de recueillir des retours qualitatifs sur l'expérience utilisateur, apportant un complément indispensable aux mesures objectives. La combinaison de ces trois types de données a été au cœur de la méthodologie VR-SOLUS, visant à proposer une approche hybride et complète pour l'optimisation des environnements VR.

Résultats

Les résultats des études réalisées dans le cadre de cette thèse ont confirmé l'efficacité de la méthodologie VR-SOLUS en fournissant des informations détaillées sur les expériences en RV. Cette section résume les principales découvertes et contributions apportées par nos diverses études. Toutes les expériences ont été menées dans le strict respect des normes éthiques et des politiques de confidentialité des données de notre institution.

Étude de Validation de l'Interface RV

Cette étude a comparé l'utilisation d'une interface RV pour le contrôle d'une pelleuse réelle (fournie par Heracles Robotics, une entreprise de terrassement autonome, partenaire de cette thèse) avec une pelleuse simulée en RV. Les résultats ont montré que l'interface RV offrait des performances comparables à celles de la machine réelle, tant pour les professionnels que pour les novices. Cette validation a confirmé que l'interface RV est une alternative viable pour les applications de téléopération, offrant une solution efficace pour des environnements industriels simulés.

Pour la méthodologie VR-SOLUS, cette étude a permis une première mise à l'épreuve des données opérationnelles que sont le temps de complétion, les erreurs de manipulation et l'aisance de contrôle du système. Grâce à cette étude ont également pu être extraites et validées des mesures spécifiques au cas d'usage, sous la forme de fluidité et maladresse des contrôles. Cela a pu démontrer l'importance de contextualiser la méthode de conception d'une application RV à ses utilisateurs mais également à ses cas d'usage.

Étude de Sélection de Matériel

L'étude a comparé l'efficacité de trois types de contrôleurs : les contrôleurs de jeu (manette), les contrôleurs RV hybrides (fournis avec le casque), et les joysticks professionnels. Les résultats ont révélé que les joysticks offraient les meilleures performances globales en termes de précision et de contrôle pour les tâches de formation en RV. Cette découverte a confirmé que les normes professionnelles surpassent la familiarité avec les contrôleurs pour les tâches d'excavation et de formation en RV.

Dans le cadre de la contribution à notre méthodologie, cette étude a validé les mesures opérationnelles évoquées précédemment, et a permis l'étude de leurs corrélations. Un premier essai de mise en place de capteurs physiologiques a également été effectuée pendant cette étude, présentant des résultats encourageant une intégration plus complète mais nécessitant une analyse de données plus approfondie.

Étude d'Optimisation de Logiciel

Cette étude a évalué différentes configurations d'annotations virtuelles (AVs) et leurs modalités d'affichage. Les résultats ont montré que l'approche hybride, combinant les données objectives et les retours utilisateurs, offrait les meilleurs résultats en termes de satisfaction et d'efficacité, comparativement à des méthodes purement objectives ou purement subjectives. Les méthodologies purement objectives ou subjectives n'ont pas capturé toutes les nuances de l'expérience utilisateur, tandis que la méthodologie hybride VR-SOLUS a fourni une vue plus complète des annotations virtuelles et aboutissant à de meilleures performances utilisateurs.

Étude Thérapeutique

Un prototype de thérapie dyslexique en RV a été développé et validé avec des orthoptistes et une neurologue. Bien que l'étude clinique n'ait pas été réalisée, les résultats préliminaires montrent un potentiel prometteur pour l'application de la méthodologie VR-SOLUS et les technologies de RV dans la thérapie de la dyslexie. L'intégration de mesures opérationnelles, physiologiques, et des évaluations médicales suggère que cette approche pourrait offrir une méthode efficace pour la thérapie de la dyslexie, bien que des études cliniques supplémentaires soient nécessaires pour valider pleinement cette approche.

En terme de méthodologie, cette étude offre une perspective sur l'utilisation à long-terme de VR-SOLUS. Non seulement cette méthodologie peut s'utiliser dans la phase de conception, au début du cycle de vie de l'application. Mais de par cette étude, une possibilité d'utilisation continue est envisageable, pour une mise à jour continue de l'application afin qu'elle réponde aux besoins de l'utilisateur. Dans ce contexte thérapeutique, cela signifie une potentielle personnalisation de la thérapie pour la rendre d'autant plus efficace et améliorer les soins prodigués en RV.

Limitations

Les études réalisées ont rencontré plusieurs limitations significatives qui doivent être prises en compte dans l'analyse de notre méthodologie, ainsi que dans ses futurs développements :

- **Configuration Expérimentale** : les études réalisées pendant cette thèse l'ont été dans un environnement qui ne correspondait pas aux conditions idéales d'un laboratoire EEG sans interférences. Cette limitation a pu affecter la qualité des signaux EEG, malgré les efforts pour réduire le bruit lors du filtrage et de l'analyse des signaux. De plus, l'absence de cadre clinique pour l'étude de thérapie de la dyslexie a limité la validation de cette étude en usage réel, malgré l'approbation des spécialistes médicaux;
- **Nombre Limité de Participants** : le nombre limité de participants a restreint la puissance statistique des analyses, réduisant la robustesse et la généralisabilité des conclusions. Les tests ANOVA effectués ont fourni des résultats acceptables, mais la force et la signification des résultats n'étaient pas optimales en raison du nombre insuffisant de participants;
- **Synchronisation des Données** : la synchronisation des données a été réalisée par logiciel plutôt que par matériel (qui est la technique traditionnellement employée), ce qui a pu affecter la précision des signaux enregistrés, notamment pour les données provenant de différentes modalités avec des timings précis et des fréquences différentes;
- **Limites des Capteurs** : les capteurs utilisés, en particulier pour l'EEG, ont nécessité un filtrage et une analyse poussés. Les capteurs actuels, avec seulement huit électrodes, sont limités en termes de précision et de confort. L'intégration de capteurs plus avancés ou de capteurs EEG intégrés directement dans le casque de RV pourrait améliorer la qualité des signaux et réduire les problèmes de synchronisation.

Futures Recherches

Pour améliorer la méthodologie VR-SOLUS et surmonter les limitations identifiées, plusieurs pistes de développement sont envisagées :

- **Augmentation du Nombre de Participants** : il est crucial d'augmenter le nombre de participants pour obtenir des résultats plus généralisables et plus robustes. Des études supplémentaires dans différents contextes d'application permettraient de tester la généralisabilité des résultats. Une étude clinique complète pour la thérapie de la dyslexie pourraient confirmer (ou infirmer) en usage réel l'utilité et l'efficacité de la RV pour le traitement médical;
- **Intégration de Nouveaux Bio-capteurs** : l'intégration de nouveaux capteurs, tels que des capteurs électromyogrammes (EM)G et des capteurs EEG intégrés au casque de RV et possédant plus d'électrodes, pourrait améliorer la synchronisation des données, réduire le bruit et augmenter le confort des utilisateurs. Cette

intégration permettrait de recueillir plus de données, de manière plus précise, afin de mieux comprendre les interactions en RV;

- **Analyse Avancée des Données** : l'exploration de techniques d'analyse et de filtrage plus avancées pour l'EEG et autres données physiologiques est nécessaire pour affiner les mesures obtenues et améliorer la précision des résultats. L'intégration et l'analyse conjointe de nouvelles mesures objectives et subjectives pourrait révéler de nouvelles informations sur l'expérience en RV;
- **Études à Long Terme** : la réalisation d'études à long terme utilisant la méthodologie VR-SOLUS dans divers contextes d'application permettrait de mieux comprendre l'efficacité et les implications de cette approche. Ces études pourraient offrir des perspectives supplémentaires sur l'optimisation des expériences en RV en continu.

Conclusion

Dans cette thèse, nous avons démontré que la méthodologie VR-SOLUS, en combinant des mesures opérationnelles, physiologiques et des retours utilisateurs, offre une approche plus complète pour la conception et l'optimisation des environnements RV. Les résultats des différentes études, allant de la validation d'interfaces de téléopération à l'optimisation des annotations virtuelles et à la thérapie dyslexique, ont confirmé que cette approche hybride est plus efficace que les méthodes purement objectives ou purement subjectives. En répondant aux questions fondamentales 1) quelles sont les mesures opérationnelles utiles, 2) quelle est l'efficacité des capteurs physiologiques et 3) quelles sont les méthodologies pour améliorer la conception de la RV, cette recherche ouvre la voie à de futures avancées dans le domaine.

La méthodologie VR-SOLUS propose une vision hybride de la conception RV qui dépasse les limites des méthodes traditionnelles et basées sur les données, offrant une analyse plus complète et précise des expériences en RV. Les contributions de cette recherche, en termes de validation des interfaces, de sélection du matériel, d'optimisation du logiciel et d'application thérapeutique, fournissent une base solide pour de futures explorations et développements dans le domaine de la réalité virtuelle. La poursuite du développement de nouveaux capteurs et de techniques d'analyse plus avancées permettra de perfectionner encore davantage les expériences en RV, offrant des solutions plus précises et adaptées aux besoins des utilisateurs.

Chapter 1

Introduction

We can trace the origins of virtual reality (VR) head-mounted displays back to the creation of the stereoscope in the 1830s, with the version we recognise today taking shape in the 1960s [1]. Two hundred years later, the global head-mounted displays market - including VR - is projected to reach more than \$370 billion by 2034 [2], due to its expanding use in industries such as healthcare, education and entertainment. The VR market has rapidly evolved from a niche technology to a mainstream tool with applications across various sectors (see Fig. 1.1). Initially associated with gaming and entertainment, VR has expanded its reach into fields as diverse as industry, healthcare, education, and social interaction, reshaping how we engage with both digital and physical environments. The expansion of VR technologies is not stopping, reaching further and further, into more complex fields and into a wider range of users.

Coates defined Virtual Reality in 1992 as *"electronic simulations of environments experienced via head-mounted eye goggles and wired clothing enabling the end user to interact in realistic three-dimensional situations"* [3]. That same year, Greenbaum described it as *"an alternate world filled with computer-generated images that respond to human movements. These simulated environments are usually visited with the aid of an expensive data suit which features stereophonic video goggles and fiber-optic data gloves"* [4]. While these technology-centric definitions remain relevant today, this thesis adopts a user-centric perspective, as expressed by Steuer: *"A 'virtual reality' is defined as a real or simulated environment in which a perceiver experiences telepresence"* [5]. Throughout our studies, we place the user at the center of the system, designing and optimising VR to enhance user experience.



Figure 1.1: Virtual Reality Timeline

In the context of Industry 4.0, VR plays a crucial role in simulating and optimising complex manufacturing processes. It allows for the creation of digital twins - virtual replicas of physical systems - enabling real-time monitoring, testing, training, remote operating and optimisation without the risks and costs associated with physical trials [6]. This integration of VR with advanced manufacturing techniques is driving significant efficiencies, from product design to maintenance and worker training. The healthcare sector has also embraced VR, particularly in therapy and rehabilitation. VR is being used as part of digital therapeutics, to treat conditions such as phobias, Post-Traumatic Stress Disorder, and chronic pain, offering immersive environments where patients can confront and manage their symptoms in controlled settings [7]. This technology is also revolutionising surgical training, providing surgeons with highly realistic simulations to practice complex procedures, improving outcomes, and reducing the risk of errors. In the gaming industry, VR continues to push the boundaries of immersive entertainment. With the development of more sophisticated hardware and software, players are experiencing games in more engaging and lifelike ways than ever before. The gaming industry's investment in VR is not just about enhancing gameplay but also about creating social spaces where players can interact within the virtual world, blending the lines between reality and fiction [8]. The recent rise of the metaverse represents another significant leap for VR. As companies invest heavily in building interconnected virtual worlds, VR is becoming central to social interaction, commerce, and even work. The metaverse aims to be a fully immersive digital universe where people can meet, collaborate, and create, fundamentally altering how we perceive and engage with the online world.

Through those fields, and the many others making use of VR nowadays, we see VR applications becoming more widespread and accessible to a broader range of consumers, see Fig. 1.2¹. Their use-cases and specific needs is also broadening, creating a demand for optimised, complex application with a user-centred design more prominently than ever. With such diverse and complex applications, from industrial simulations to healthcare treatments and immersive entertainment, the need to refine and perfect VR experiences is capital. One does not simply design a therapeutic tool the same way one designs a game or an industrial monitoring system. Effective VR design requires not only technological innovation, but also a deep understanding of user interaction, psychology, and ergonomics in the context of VR application development. As the user base for VR expands, so too does the need for sophisticated, well-designed systems that can answer to a wide array of needs, ensuring that VR experiences are not only immersive and effective, but also intuitive and accessible to all, despite the complexity of their use-cases. To produce that, design methodologies exist to help researcher and developers create VR applications that answer to their needs, but more importantly to their users' needs.

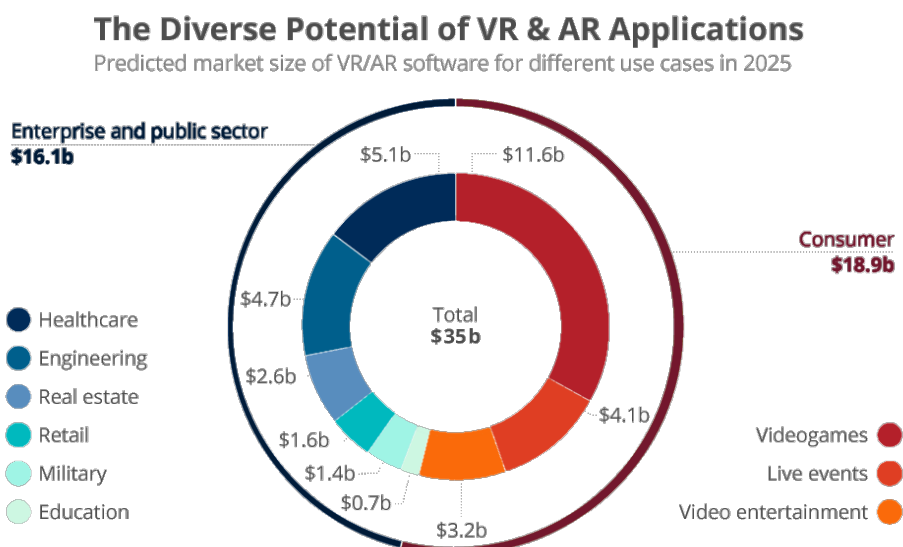


Figure 1.2: Extended Reality (XR) Software Market 2025 Fields and Revenue (according to Statist)

As VR continues to advance, so too do the methodologies used to design and optimise VR experiences [9]. Traditional design approaches have often relied on iterative development, where user feedback plays a crucial role. This

¹ Statista 2025 XR Market Prediction <https://www.statista.com/chart/4602/virtual-and-augmented-reality-software-revenue/>

feedback, typically collected through surveys, interviews, and observational studies, provides valuable insights into user preferences and pain points. However, these methods are inherently subjective and rely heavily on user self-reported experiences, which can introduce bias and limit the depth of analysis [10]. In response to these limitations, there has been a significant shift towards data-driven design methodologies. These approaches use objective data collected from user interactions within the VR environment, providing a more nuanced and accurate understanding of user behaviour [11]. By analysing metrics such as task completion times, error rates, and interaction patterns, designers can make informed decisions to enhance user experience and optimise system performance. This shift towards data-driven methods marks a new era in VR development, where decisions are grounded in objective measures rather than solely on subjective user perceptions. The emergence of biosensors has further revolutionised the design and optimisation process in VR. Biosensors, such as eye-tracking devices, electroencephalograms (EEG), and heart rate monitors, allow the collection of physiological data that reflects users' cognitive, physical and emotional states in real time [12]. These sensors provide insights into how users respond mentally and physically to the VR environment, revealing aspects of the experience that traditional methods might overlook. For example, eye-tracking can show where a user's attention is focused, while EEG data can indicate levels of cognitive load or emotional engagement.

In this thesis, we study VR design and optimisation methodologies and evaluate their metrics to create our own hybrid methodology: VR-SOLUS, the Sensory, Operational and User data Synthesis method for Virtual Reality. By integrating biosensor data with operational metrics and user feedback, designers using VR-SOLUS can achieve an extensive understanding of user experience, combining objective measures of performance with physiological indicators of user state, as well as user perception. This synergy between data-driven methodologies, biosensors, and traditional feedback paves the way for creating more adaptive, customised, and effective VR systems, capable of responding to user needs in real time and improving overall usability and effectiveness across many fields.

1.1 Designing Virtual Reality - Challenges and Gaps

The evolution of Virtual Reality (VR) design and development has seen significant advancements, but several challenges remain. Addressing these challenges is essential for enhancing the effectiveness and user experience of VR applications.

Traditional VR design methodologies often rely heavily on subjective measures, including user surveys and feedback. These approaches, while invaluable, are not without limitations. Surveys and subjective feedback provide critical insight into user experiences and preferences, but they are inherently limited by their reliance on personal opinions and perceptions. This subjectivity can introduce biases and variability that may not always accurately reflect the true effectiveness or usability of the VR experience. Furthermore, traditional methods can be labour-intensive and time-consuming, making it challenging to gather comprehensive and actionable data. In response to these limitations, the field has increasingly adopted data-driven design methodologies. These approaches use objective data to inform VR design, using metrics such as user performance, physiological responses, and interaction logs. Although data-driven methods offer a more quantifiable and objective means of evaluating VR experiences, they also come with their own set of challenges. Objective data can sometimes overlook the nuanced aspects of user experience that are captured through subjective feedback. In addition, data-driven methods often require sophisticated technology and extensive data analysis, which can be complex and resource-intensive.

Our thesis aims to bridge the gaps between traditional subjective methods and data-driven approaches by introducing a hybrid methodology that integrates operational, physiological, and user feedback data. This comprehensive approach seeks to address the limitations of both traditional and data-driven design techniques:

- **Enhanced Accuracy and Reliability:** by combining subjective feedback with objective operational and physiological data, our methodology aims to provide a more extensive and accurate understanding of user experience. This integration helps mitigating the biases inherent in purely subjective assessments and the limitations of data-driven approaches alone;
- **Comprehensive User Insights:** our approach addresses the gap in capturing both the qualitative and quantitative aspects of user experience. While traditional methods may miss certain nuances, and data-driven methods may overlook user perception, our hybrid approach provides a richer and more complete picture of how users interact with VR applications;

- **Improved Design Iteration:** the hybrid methodology enables more informed and iterative design processes by using diverse data sources. This allows for more effective adjustments and refinements to VR applications, ultimately leading to better user experiences.

1.2 VR-SOLUS, Sensory, Operational and User data Synthesis for Virtual Reality

1.2.1 Hypothesis

To address these gaps, our investigation focusses on the combined use of operational and physiological data alongside traditional user feedback. This multimodal approach aims to provide a comprehensive understanding of user interactions and experiences within VR applications, allowing developers and researchers to make informed decisions in their VR design and optimisation. To achieve this goal, our methodology exploits the following:

- **Operational Data:** metrics such as task completion times, percentage of time using multiple inputs, and number of input errors are collected to objectively measure user performance and efficiency;
- **Physiological Data:** using advanced sensors, we gather physiological markers that provide insights into the user's cognitive and physical state:
 - Eye-Tracking: captures fixation points and gaze patterns to understand visual attention and identify areas of interest;
 - EEG (Electroencephalogram): measures cognitive load, allowing us to assess the mental effort required to complete tasks in the VR environment;
 - IMU (Inertial Measurement Unit) and BPM (Beats Per Minute) Trackers: monitor physical movements and heart rate to evaluate physical load and stress levels during interactions.
- **User Feedback:** traditional surveys and interviews are conducted to gather users' subjective experiences and satisfaction levels. This feedback is essential to understand the perceived usability and enjoyment.

In the framework of our thesis, we combine these data inside the VR-SOLUS (Sensory, Operational and User data Synthesis for Virtual Reality) methodology. This comprehensive approach not only highlights areas for improvement, but also helps in developing more intuitive and user-friendly VR applications. Our studies validate this methodology through various use cases, including VR teleoperation interfaces, controller selection, virtual annotation designs, and gaze-based therapy for dyslexia.

Through this thesis we answer three questions:

1. **Which operational measures are valuable for VR development?**
2. **How to employ physiological sensors efficiently?**
3. **What kind of methodology can further improve VR design and optimisation?**

1.2.2 Experimental Setup

Our research explores advancements in VR through studies in two distinct domains: earthwork teleoperation and therapy. This section details the experimental setup, including the design of our studies, the hardware and software used, and the measures and analysis techniques applied to develop the VR-SOLUS methodology. All experiments were carried out in strict accordance with our institution's ethical standards and data privacy policies.

Earthwork Studies

The earthwork teleoperation studies aim to refine VR interfaces and controls for excavation tasks. These studies were divided into three main investigations:

- **Controller Selection Study:** this investigation focused on identifying the optimal controller for VR-based earthwork training. We compared three types of controllers — gamepads (gaming standard), VR-specific controllers (hybrid), and joysticks (professional standard) — to determine their impact on training efficacy. The study included 13 participants and evaluated controller performance based on task execution and user feedback. This study helped us identifying and validating performance indicators;
- **Interface Validation Study:** this study compared a VR teleoperation interface with real-world machine control. Using a digital twin of an excavation machine, we linked the VR environment to a real Cat323 excavator through Unity, using an emulated digital-twin to control the real machine. The study involved 10 participants and assessed performance metrics such as task completion time and accuracy to validate the effectiveness of the VR interface. It is also through this study that we tested and validated use-case specific metrics as efficient operational measures, and where we first integrated bio-sensors into our framework;
- **Software Optimisation Study:** this study explored the impact of different virtual annotations (VAs) modalities on user performance in VR. Two phases were involved: the design phase, where the VA sets were empirically defined and tested, and the evaluation phase, which compared three VA sets — one based on user feedback, one on objective measures, and a hybrid approach. The study, involving 30 participants, aimed to refine VR training tools by assessing the effectiveness of the various VA modalities. Not only did this study assess the neurological part of our data analysis, it also provided a testing ground where we confronted purely subjective, purely objective, and hybrid methodologies. The hybrid methodology, VR-SOLUS, was shown to be superior and validated through this experiment.

Therapeutics Study

The therapy study aimed to develop and validate a VR prototype for dyslexia therapy. This study integrated operational measures, physiological metrics, user feedback, and clinical evaluation to create an adaptive VR tool. The prototype was validated with input from three orthoptists and one neurologist, focusing on visual exercises tailored to support dyslexia therapy.

1.2.3 Materials

Through this thesis, we employed diverse hardware and software, aimed for VR simulation, physiological data recording, and data analysis.

Hardware

- **VR Headset:** HTC Vive Pro Eye, a VR headset with a 1440 x 1600 pixels per eye resolution (2880 x 1600 pixels combined), with a field of view of 110° and a frequency of 90Hz. This headset provides quality VR simulation, hence its use for our studies;
- **Eye-Tracking:** the HTC Vive Pro Eye is equipped with an integrated Tobii eye-tracker with a 120Hz frequency, used for assessing gaze patterns and visual focus;
- **Electroencephalogram (EEG):** the g.tec Unicorn Hybrid Black is an EEG cap with eight electrodes, used to monitor cognitive load and brain activity with a 200Hz frequency. It is also equipped with IMU (gyroscopes and accelerometers);
- **Inertial Measurement Units (IMUs) and Heart Rate Monitor:** the Polar OH1+ armband was used for tracking physical load and physiological responses during VR tasks through its accelerometer and heart rate monitor.

Software

- **Unity**: the primary platform for developing and simulating VR environments and recording operational data. This game engine provided us all the tools needed for our VR applications, as well as calibration phases and data recording;
- **SRanipal**: package used in Unity for processing eye-tracking data from the HTC Vive Pro Eye;
- **Unicorn Suite**: software suite used for analysing EEG and IMUs data from the g.tec Unicorn Hybrid Black;
- **Polar Sensor Logger**: Android API used for logging physical data from the Polar OH1+ IMUs and heart rate monitors;
- **Lab Streaming Layer**: used for synchronising our multimodal streams of data;
- **OpenVibe**: physiological signal processing and analysis tool, used for EEG filtering and processing;
- **Python**: code used for post-processing filtering and analysis of our diverse data.

1.2.4 Summary of Findings

Our VR-SOLUS framework Fig. 1.3 yields the strengths of both subjective and objective data to provide an extensive methodology for designing and optimising VR applications. It explores **1) operational measures as performance indicators**. Multiple measures are presented and evaluated in Chapter 3. It also incorporates **2) diverse physiological sensors** and extracts their different measures, as presented in Chapter 4. Finally, through **3) a combination of objective and subjective measures**, VR-SOLUS proposes a hybrid method for VR design. This hybrid approach is extensively tested, and its correlation of measures is studied in Chapter 5. The VR-SOLUS approach addresses existing gaps in current design methodologies, ultimately aiming to enhance user experience and performance in VR environments.

In the following sections, we present the findings from our various studies that validate the VR-SOLUS approach. These studies, which span domains such as earthwork teleoperation and therapy, examine the integration of operational metrics, physiological signals, and user feedback to improve the design of VR applications. Four studies are presented in this thesis:

- **Interface Validation**: earthwork teleoperation study where users controlled a real excavator. The goal of this study was to validate VR as a potential teleoperation interface. Once validated, the follow-up earthwork studies aimed to improve said interface for training future teleoperators, using a simulated VR environment;
- **Controller Selection**: earthwork training study, aiming for the selection of a suitable controller for our VR interface. This study was done entirely in simulation and its goal was to optimise the hardware part of our VR interface;
- **Virtual Annotations**: earthwork training study, also done in simulation, serving both as a software optimisation of our VR interface and a validation of our optimisation methodology;
- **Dyslexia Therapy**: visual immersive therapy prototype for dyslexia rehabilitation, applying our methodology to another use-case than earthwork.

In the following chapters, our analysis focuses on task performance, user experience, and the role of physiological signals in measuring cognitive load and physical exertion in the above-mentioned studies.

One key aspect of our work is the exploration of correlations between operational data, physiological signals, and subjective user feedback, presented in Chapter 5. Our findings highlight how these diverse types of data interact, revealing valuable patterns. These correlations are critical to understanding the full user experience and reveal that mismatches between objective performance data and subjective feedback can offer unique insights into usability challenges [13].

This summary sets the stage for the detailed exploration of these findings throughout the manuscript, showing how the hybrid integration of multimodal data through VR-SOLUS holds the potential to significantly advance VR application design and optimisation across multiple fields.

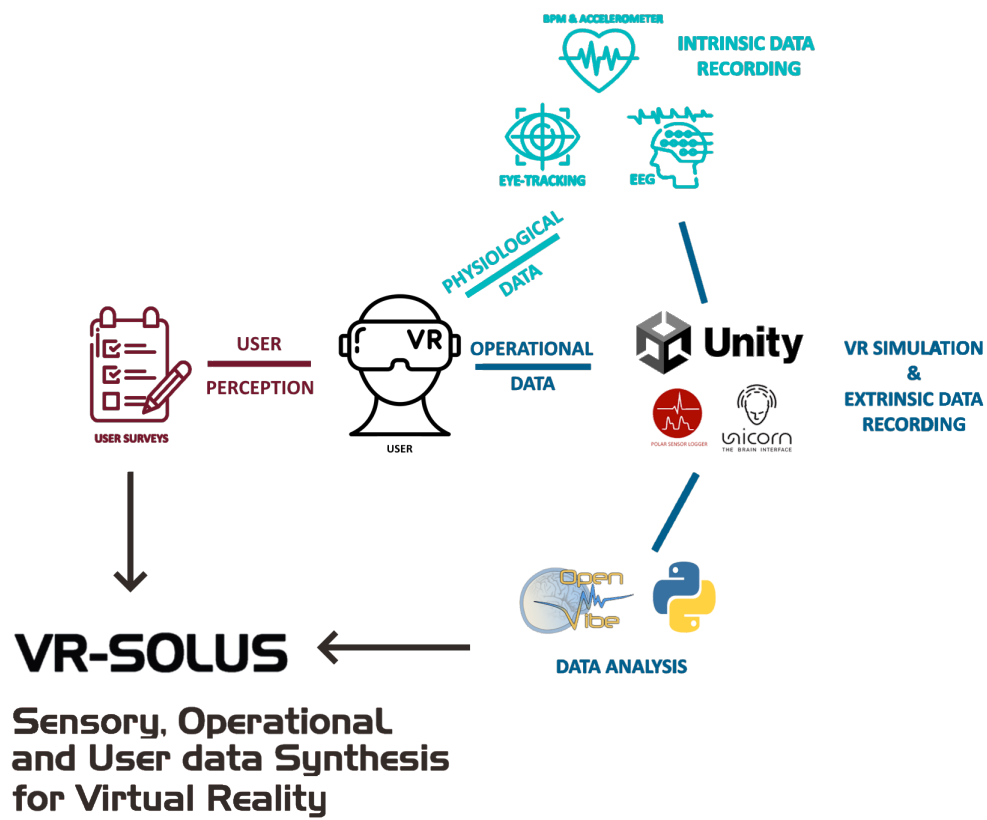


Figure 1.3: The VR-SOLUS Framework

Chapter 2

Related Work - Designing Virtual Reality Applications and Experiments

Virtual Reality (VR) applications have significantly transformed various sectors by offering immersive experiences that seamlessly integrate virtual environments with real-world applications. Industry 4.0, healthcare, education, entertainment, marketing, and so on make full use of VR technologies. To design these applications, developers rely on a variety of methodologies aimed at optimising performance and user experience.

Traditional approaches to VR design often involve gathering user feedback through methods such as surveys, interviews, and usability testing. This user-centric approach ensures that designers gain insight into user preferences, challenges, and interaction patterns. By incorporating this feedback iteratively throughout the design process, VR applications can evolve to better meet the needs and expectations of their users. For example, in therapeutic settings, understanding how patients interact with virtual environments can lead to improvements that enhance treatment efficiency and user engagement throughout therapy.

In addition to traditional methods, data-driven design methodologies are increasingly important in the development of VR applications. These methodologies combine real-time analytics derived from operational metrics (such as user interaction data and task completion rates) and physiological metrics (including cognitive load and emotional responses measured through biometric sensors or, in short, biosensors). Analysing these data enables designers to make informed decisions that aim to optimise user interfaces, refine virtual environments, and ultimately improve the overall user experience. If we keep the healthcare example, VR applications use physiological data to create simulations that allow medical professionals to practice and refine complex procedures in a controlled virtual environment.

By integrating these diverse design methodologies, VR developers ensure that applications are not only technically robust but also user-friendly and able to address specific use-case needs. As VR technology continues to advance, these methodologies will play a crucial role in shaping immersive experiences that push the boundaries of innovation across sectors.

In this thesis, we propose a methodology that uses both a traditional and a data-driven approach, combining standard surveys with operational and physiological data to improve VR design and optimisation.

2.1 Designing Virtual Reality

Whether designed for games, training tools, therapeutic devices, or any other application, virtual reality (VR) applications do have basic design guidelines. VR applications are meticulously crafted to optimise immersion, interaction, and mitigate issues such as cybersickness, to provide the best performance and user experience. These general guidelines presented in Jerald's *VR book* [14] are centred around two cores: the user first and foremost, and the use-case of the application.

Whichever use-case is chosen, VR design must encompass several critical aspects that collectively shape the user experience:

- **Creating immersion:** VR design focuses on creating a sense of presence within virtual environments. This is achieved through high-resolution graphics, realistic spatial audio, and accurate tracking of user movements [15]. Immersion leads to engagement, engagement leads to learning and improvement of the user [16]. For some applications, this requires highly-detailed digital-twins to have an environment as close to reality as possible [17]. But it is mainly having a coherent environment that keeps the user from breaking immersion and engagement [18], both in terms of visual and auditory identity;
- **Providing interaction:** interaction design in VR centres on intuitive controls and natural virtual movements that mirror real-world actions. Users traditionally interact with virtual environments using motion controllers, but also hand gestures [19], or gaze-based interfaces [20]. Designers aim to make interactions seamless and responsive, enabling users to manipulate objects, navigate spaces, and engage with virtual content effortlessly;
- **Mitigate cybersickness:** cybersickness is a form of motion sickness experienced in virtual reality environments typically caused by the offset between visual and the body inputs received by the brain [21]. Characterised by symptoms such as nausea, dizziness, and disorientation, it can be triggered for any user at any time and force them to completely stop their VR experience. As such, it is crucial to address cybersickness and mitigate it as much as possible. Techniques include offering multiple locomotion options (e.g., teleportation, smooth movement [22]), reducing acceleration speeds, and maintaining stable frame rates to prevent motion-induced nausea or disorientation [23];
- **Ensure user comfort:** in addition to reducing cybersickness, user comfort also involves optimising visual fidelity, reducing latency, and implementing ergonomic design principles. VR applications prioritise smooth rendering, minimal latency, and comfortable viewing angles to prevent eye strain and fatigue during extended use [24]. Designers also consider ergonomic factors in controller design and headset ergonomics to improve overall comfort [25];
- **Use adaptive design:** some VR applications incorporate adaptive design elements to respond to diverse user needs and preferences. This includes custom settings for movement sensitivity, comfort modes, and accessibility features [26]. Adaptive design allows users to adjust VR experiences according to their comfort levels, physical abilities, or specific preferences without compromising usability or immersion.

To effectively implement and analyse these key aspects of VR design, researchers and developers employ a range of methodologies, ranging from traditional approaches that use user feedback to data-driven methods using quantified data.

2.2 Design Methodologies

Design methodologies in VR are systematic approaches used to create, evaluate, and refine virtual experiences. These methods serve to ensure that VR applications are effective, engaging, and user-friendly. They typically involve a cycle of planning, prototyping, testing, and iterating.

Traditional methodologies often rely on qualitative feedback from users, gathered mostly through surveys. These provide valuable insights into user preferences and experiences. More recently, data-driven methodologies have gained popularity among virtual reality (VR) designers and developers. These approaches employ quantitative data collected from various sources, including user interactions within the VR environment, physiological responses, and performance metrics. By combining both traditional and data-driven methods, designers can gain a more comprehensive understanding of user needs and behaviours in virtual spaces, from both subjective and objective points of view.

2.2.1 Traditional Virtual Reality Design

VR design has long relied on conventional user-centred approaches adapted from human-computer interaction fields. These traditional methods rely on subjective means, primarily surveys and interviews. They form a cornerstone of VR design processes [27] and focus on gathering qualitative insights directly from users, emphasising

perceptions, opinions, and self-reported experiences to comprehend and optimise user-experience (UX), from the user's own inputs.

User surveys and questionnaires are fundamental tools in traditional VR design. These instruments collect structured feedback on various aspects of the VR experience, including usability, comfort, presence, and immersion [28]. Designers often employ standardised scales, such as the NASA Task-Load-Index (NASA-TLX) [29], the System Usability Scale (SUS) [30], the Kennedy Simulator Sickness Questionnaire (SSQ) [31] or the Presence Questionnaire [32], to allow comparisons across different VR applications. These subjective methods offer several advantages. They provide rich, contextual data about user experiences, preferences, and perception. This qualitative information is crucial to understanding the patterns behind user behaviours and identifying potential areas for improvement that might not be apparent from quantitative data alone [33].

However, self-reported data in VR design, while valuable, have inherent limitations. Users often struggle to accurately articulate their experiences, which can lead to incomplete or biased feedback [34]. Post-experience surveys and interviews can introduce recall bias, where recent events or emotions can influence user recollections [35]. The administration of surveys within the VR environment can partially mitigate recall bias [36]. However, this approach does not fully address the subjective nature of user feedback or the inherent biases in survey methodologies. Another challenge is the potential offset between users' reported experiences and their actual behaviour in VR environments. What users say they prefer or experience may not always align with their actions or physiological responses during VR use [37].

Despite these limitations, traditional survey-based and subjective methods remain valuable in VR design. They offer a human-centred perspective that helps ensuring that VR applications meet user needs and expectations. However, the field increasingly recognises the need to complement these approaches with more objective data-driven methodologies to create a more comprehensive understanding of VR user experiences [38].

Pros and Cons of Traditional Methodologies

Traditional methods have many advantages but also have limitations, summarised in Table 2.1:

Advantages	Limitations
Easy to use Insight into user perception Cheap	Recall bias Offset between perception and reality No extrinsic measures

Table 2.1: Advantages and Limitations of Traditional Design Methods

2.2.2 Data-driven Virtual Reality Design

Data-driven methodologies represent a paradigm shift in design and research practices, emphasising the use of quantitative data to enrich decision-making processes. These approaches rely on the collection, analysis, and interpretation of objective measurements, using statistical techniques to guide design and development [39]. In recent years, data-driven methods have spread across various fields, including software development, user experience design, and scientific research, offering a more empirical approach to problem solving and innovation [40].

In the realm of VR, data-driven methodologies have emerged as a powerful complement to traditional subjective design approaches. VR environments present unique challenges and opportunities for data collection, given their immersive and interactive nature. These methods employ a wide array of metrics to assess user experience, performance, and physiological responses within virtual environments [41]. By providing quantifiable and objective data about user interactions, data-driven approaches can reveal patterns, preferences, and issues that might not be apparent through subjective feedback alone. This objectivity is particularly valuable in VR design, where users may struggle to express their experiences fully or where subtle aspects of the virtual environment can significantly impact the overall experience [42].

Operational Measures

Operational measures in VR focus on quantifiable aspects of user performance and interaction within the virtual environment. They are extrinsic measures that provide concrete data on how users interact with the VR environment. Key operational measures include task completion times, which can indicate the efficiency of user interactions [43]; error rates [44], highlighting areas where users struggle; navigation patterns [45], revealing how users explore and move through virtual spaces; and input accuracy [46], assessing the precision of user actions within the VR environment [47]. By analysing these metrics, designers can identify areas of difficulty in the VR interface, optimise spatial layouts, and improve overall usability. For instance, tracking user movements within a virtual space might reveal that certain areas are underused or difficult to navigate, prompting redesigns to enhance user motion fluidity [48]. Similarly, analysing task completion times across different interface designs can provide empirical evidence for the most effective layout [49].

These operational measures offer a detailed view of user behaviour from an external and objective point of view, allowing designers to make data-informed decisions about everything from button placement to environmental design. However, interpreting this data requires careful consideration of the context and goals of the VR application [50]. Moreover these operational measures remain external to the user, giving insights but only from an external point of view. To obtain more objective data, the internal point of view must also be factored in the design of VR applications. To do so requires physiological measures of the users.

Physiological Measures

Physiological measures capture users' bodily responses during VR experiences, offering a window into the subconscious and involuntary reactions to virtual stimuli [51]. These intrinsic measures can provide insights into users' cognitive load, emotional states, and levels of engagement or stress, often revealing aspects of the user experience that may not be consciously perceived or easily articulated [52].

Common physiological metrics in VR studies include heart rate and heart rate variability, which can indicate changes in stress or emotional arousal [53]; eye movements and pupil dilation, which can reveal patterns of attention and cognitive processing [54]; and brain activity measured via electroencephalography (EEG), offering insights into cognitive states and emotional responses [55]. Other examples include, but are not limited to, changes in heart rate that might indicate heightened anxiety in certain virtual scenarios, allowing designers to identify and modify potentially stressful elements [56]. Eye-tracking data could reveal which elements of a virtual environment attract the most attention, informing decisions about information placement or environmental design [57]. EEG data might provide insights into cognitive load during complex tasks, helping to optimize the difficulty curve of VR applications [58].

The integration of these physiological measures into VR design processes offers unprecedented insights into the user experience, from an internal point of view. However, it also presents challenges in data collection, analysis, and interpretation, requiring specialised equipment and expertise [59].

Pros and Cons of Data-driven Methodologies

Like traditional methods, data-driven methodologies also possess multiple advantages and limitations, summarised in Table 2.2:

Advantages	Limitations
Intrinsic and extrinsic measures Numerous metrics to choose from Quantifiable data	Does not take into account user experience Advanced analysis required Expensive and complex setup

Table 2.2: Advantages and Limitations of Data-driven Design Methods

2.3 Proposed Methodology

The proposed methodology is founded on the hypothesis that integrating operational and physiological data with traditional user feedback can significantly enhance the design and optimisation of VR applications. Plano Clark [60] advocates combining different data sources to fully understand the research problem, which we aim for by combining subjective and objective data. On one hand, traditional methodologies rely primarily on subjective user feedback, which, while valuable, can be limited by individual perception and bias. This creates a gap where the ground truth, or the objective reality of user interactions and experiences, may differ from the perceived truth reported by users (Table 2.1). On the other hand, data-driven methodologies may rely on objective data, but said data present challenges when it comes to their collection, analysis, and interpretation. Moreover, the use of physiological data requires specialised hardware, which implies more complex and expensive experimental setups (Table 2.2).

The proposed methodology also considers a strong complementarity between subjective feedback and objective data, bringing the benefits of both approaches while mitigating their drawbacks. User feedback provides insights into user satisfaction and perceived usability, while operational data (e.g. task completion times, input errors) and physiological data (e.g., cognitive load from EEG, eye-tracking fixations, physical load from IMU and BPM trackers) offer precise, measurable indicators of user performance and experience. This combination allows for a more holistic understanding of user interactions in VR environments, as seen in Table 2.3. A limitation of our methodology is its potential complexity and cost, which depend largely on the specific sensors chosen for use. Integrating multiple high-end sensors can provide valuable and detailed datasets, but may also significantly increase both the expense and technical demands of the system. Professional grade biometric equipment, while offering precise data, can be costly to acquire and maintain. Additionally, the volume of data generated by these sensors requires advanced signal processing and analysis techniques, which can further complicate the implementation. However, it is important to note that the level of complexity and cost is directly linked to the number and type of sensors used, meaning that a more tailored selection of sensors can mitigate these challenges while still benefiting from the methodology.

Advantages	Limitations
Intrinsic and extrinsic measures Insight into user perception Numerous metrics to choose from Quantifiable data Possibility to compare subjective and objective data	Can be expensive and complex to use

Table 2.3: Advantages and Limitations of the Proposed Method

Chapter 3

Operational Measures

The field of virtual reality (VR) design has increasingly embraced data-driven approaches to enhance user experiences and optimise application performance. This shift towards quantitative methodologies reflects a growing recognition of the limitations inherent in relying solely on subjective user feedback [61]. Although qualitative insights remain valuable, the integration of objective and measurable data provides designers and researchers with a more comprehensive understanding of user interactions within virtual environments.

Data-driven design in VR capitalises on quantitative metrics to improve decision making processes, allowing for more precise and targeted improvements [62]. This approach is particularly crucial in VR, where the immersive nature of virtual environments can make it challenging for users to fully articulate their experiences. Operational measures, which quantify user actions and system performance, offer an objective lens through which to view and analyse VR interactions. The effectiveness of data-driven design in VR depends on the careful selection and implementation of appropriate metrics. This process requires a thorough understanding of both the specific use-case and the target audience for each VR application [14]. Different applications, whether for entertainment, education, healthcare, or professional training, may require distinct sets of operational measures to accurately capture relevant user behaviours and system performance indicators.

Our VR-SOLUS methodology (Sensory, Operational and User data Synthesis for Virtual Reality) recognises the critical role of operational data in VR design and optimisation. By systematically collecting and analysing operational measures, VR-SOLUS aims to provide designers with more quantitative insights. This approach enables a more nuanced understanding of user interactions, facilitating targeted improvements, and improving VR experiences through the use of extrinsic measures. The following sections will detail the specific operational measures employed in our methodology, their implementation, and their interpretation in the context of various VR applications. By examining these quantitative metrics, we aim to illustrate how data-driven approaches, especially the use of operational data, can significantly contribute to the advancement of VR design.

3.1 General Performance Indicators

Operational metrics provide a ground truth that complements subjective user feedback, allowing designers and researchers to track concrete improvements and identify specific areas for optimisation [63]. The reliability and objectivity of these metrics make them invaluable tools for evaluating performance, ease of use, and user progression in VR applications, thus their integration in the VR-SOLUS framework (Fig. 3.1). Unlike subjective measures, which can be influenced by user bias or inconsistent reporting, operational metrics offer consistent, reproducible data that can be analysed across multiple users and sessions.

In our VR-SOLUS methodology, we focus on three generic operational metrics:

- **Time:** the duration users take to complete specific tasks or actions within the VR environment;
- **Errors:** mistakes or unintended input actions (interactions of the user with the system, e.g. pressing a button) made by the user during their interaction with the VR environment;

- **Multi-input:** the complexity and coordination of user input actions.

Each of these metrics provides unique insights into user interaction and system performance. To illustrate their application, we used a virtual reality training system for earthwork teleoperation. In this scenario, different controller configurations are being evaluated for their effectiveness in training operators in virtual environments to use remote-controlled excavation equipment.

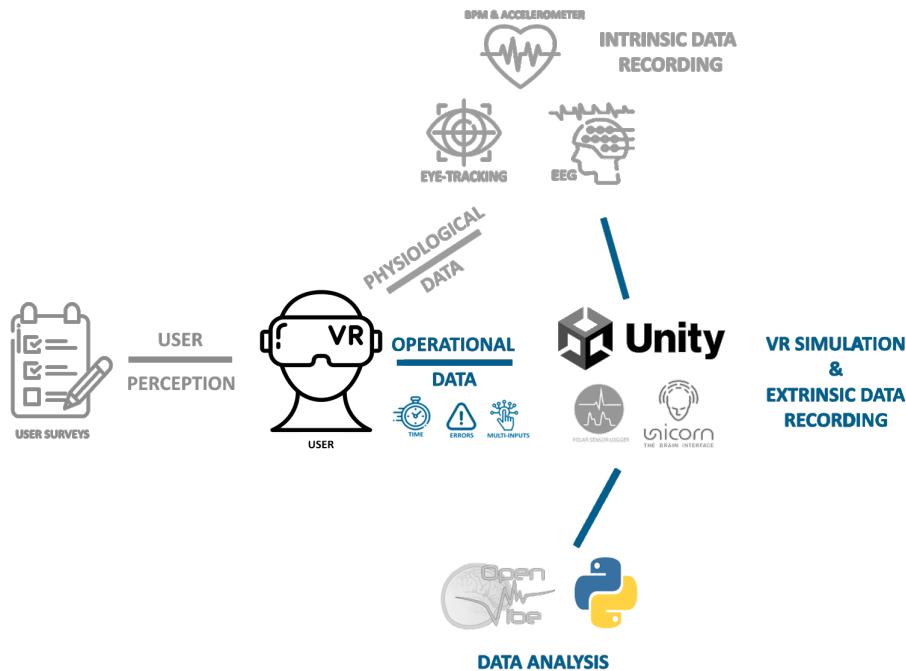


Figure 3.1: Operational Metrics Framework Highlighted into the Global Analysis Approach Proposed by VR-SOLUS

3.1.1 Time

Time is a common and versatile metric in VR performance evaluation. It presents a quantifiable measure of user efficiency and system responsiveness. As a key component of human-computer interaction evaluation [64], time measurements have been extensively studied and applied across various domains, VR not excluded. In the field of VR research, time metrics are typically used as total task completion time and/or sub-task duration measures. These measures have been widely used to evaluate interface designs, assess learning curves, and compare different interaction techniques. Recent studies have also explored more granular time metrics, such as decision-making time and reaction time, to gain deeper insights into user cognitive processes in virtual environments [65].

In our VR-SOLUS methodology, we employ two primary time measures: full-time completion and task completion time. Full-time completion captures the total duration from the start to the end of a defined experiment, providing a global view of user performance. This metric is particularly useful for comparing overall efficiency across different system configurations or user groups, as seen in the example in Fig. 3.2. Task completion time, on the other hand, focuses on the duration of specific exercises within the VR environment, allowing for detailed analysis of user performance at different stages of interaction and at executing different actions. It also helps to identify particular areas that may require optimisation. From these raw time data, we can derive several meaningful measures. The average completion time indicates general task difficulty or system efficiency, while the time variability reflects consistency of performance across users or sessions. By plotting these measures over multiple sessions, we can visualise learning curves and assess improvement rates.

Time metrics provide a reliable ground truth for performance assessment, as long as it is correctly implemented. The validity of these measures depends on the careful design of the experiment and the interpretation of the data.

To ensure the integrity of time data, it is important to clearly define the start and end points for each measured task or exercise, control for external factors that might influence timing, use consistent measurement methods across all test conditions and participants, and consider the context of the task when interpreting results. If the experiment design allows for it, inter-participant time can be compared as is. But more often than not, intra-participant normalisation is needed before comparing results, in order to ensure the validity of time data.

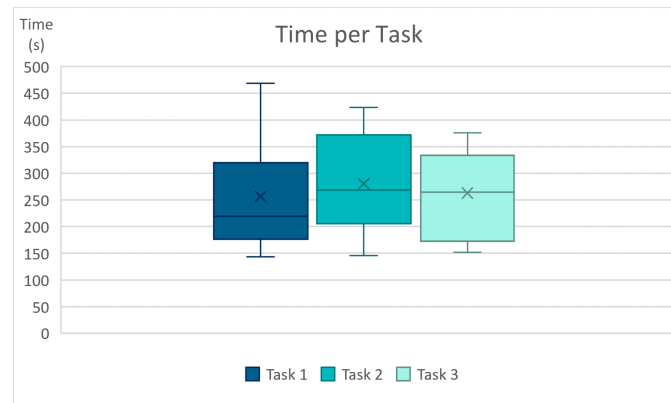


Figure 3.2: Example of Completion Time Metric*

*Box plots (also called a box-and-whisker plots) visualise the distribution of a dataset. For all the box plots presented in this thesis, here is a breakdown of their elements:

- **Box:** represents the middle 50% of the data, known as the interquartile range (IQR). The lower edge is the first quartile (Q1), and the upper edge is the third quartile (Q3);
- **Whiskers:** extend from the box to the minimum and maximum data points within 1.5 times the IQR from Q1 and Q3, representing the range of most of the data;
- **Median (line inside the box):** a line inside the box that shows the median (50th percentile) of the data;
- **Cross/Marker:** marks the mean (average) value;
- **Outliers:** data points outside the whiskers, shown as dots or small markers, representing unusually high or low values.

3.1.2 Errors

In VR and human-computer interaction (HCI) in a more general way, error measures are central for the evaluation of the effectiveness and usability of the system. These measures enable researchers to quantify user performance and compare different interfaces or interaction techniques. Recent studies have expanded our understanding of various error types in VR environments.

In our VR-SOLUS methodology, we focus on input errors, which can be particularly relevant depending on the application. These errors relate to the accuracy and precision of user inputs - interactions - in the virtual environment, as seen in the example Fig. 3.3. Scientific studies highlight the importance of considering input errors - incorrect interaction compared to the expected one - in the context of the overall user experience, rather than as isolated metrics [66]. Errors can also be used as performance measures, ergonomic measures, and evolution measures studied over time. Performance measures evaluate task efficiency, while ergonomic measures reflect physical and cognitive load during VR interactions, and are essential to provide satisfying user experience [67]. Input errors validity as a metric also depends on careful implementation, recording, and interpretation. They must be correctly and precisely defined to avoid variability between participants. As they are often recorded manually, what is an input error and what is not must not fluctuate during experiments (inter and intra participants) to ensure metric integrity.

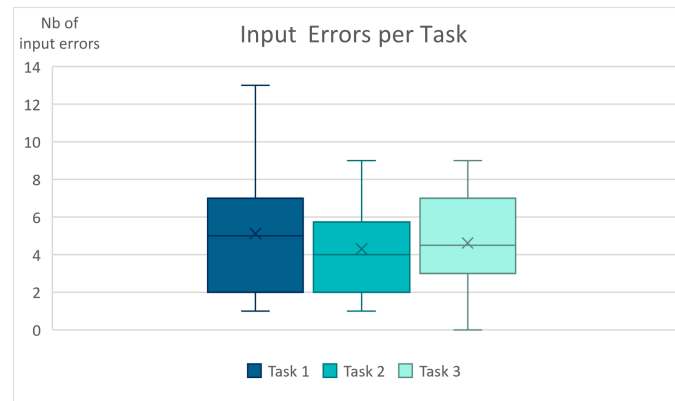


Figure 3.3: Example of Input Errors Metric

3.1.3 Multi-tasking

The multi-tasking measure evaluate the capacity and performance of a user performing multiple interactions/tasks at a time. These actions can be multimodal (e.g. motion and cognitive actions) or with the same modality (e.g. two different motion tasks). This measure is essential for understanding how users interact with complex VR systems that allow simultaneous manipulation of various inputs, require the use of multiple body parts, or evaluate brain-body coordination. By capturing the frequency and duration of multi-tasking data, it can provide insights into user engagement, performance, ergonomics, cognitive load, and interaction efficiency within the VR application.

Research in multi-tasking performance typically focuses on the ability to manage multiple tasks simultaneously, often measured through time-on-task, error rates, and subjective workload assessments. Traditional metrics include the Dual-Task Performance Paradigm, where participants perform two tasks simultaneously, and their performance is compared to when tasks are done individually [68]. Another common measure is the Multi-Attribute Task Battery (MATB), which assesses the ability to handle several simultaneous tasks by recording response times, accuracy, and overall task management [69].

In our VR-SOLUS approach, we focus on the multi-inputs measure. This metric is centered on the simultaneous use of multiple input devices within a VR setting. This measure records the simultaneous activation of different controls, providing insight on how users manage complex interactions. By analysing these patterns, we can infer the user's cognitive strategies and their ability to effectively integrate multiple streams of input data. The multi-inputs measure captures the duration of simultaneous inputs, the frequency of multi-input episodes, and the context in which multi-inputs occur, correlating with specific tasks in the VR environment.

To ensure the validity of the multi-inputs measure, we mostly depend on metrics correlation and experimental setting. As the recording is done directly by the software registering the user inputs, we need to define precisely the start and end of the range of interest. The duration of multi-inputs usage is an exploitable measure as is, but it can also be exploited as a multiple input rate during a task, i.e. extract the percentage of time a user spends using multiple inputs during a task (normalise the multi-input value) before inter-participant comparison, as illustrated in the example Fig. 3.4. To ensure that the multi-input metrics is valid, it must be correlated with other metrics such as time and inputs errors: a strong (inverted) correlation with time is expected (the more inputs combination, the less time it takes to finish a task); while a poor or non-significant correlation with inputs errors is necessary to ensure the multiple inputs use is a result of conscious decision and not an unconscious error.

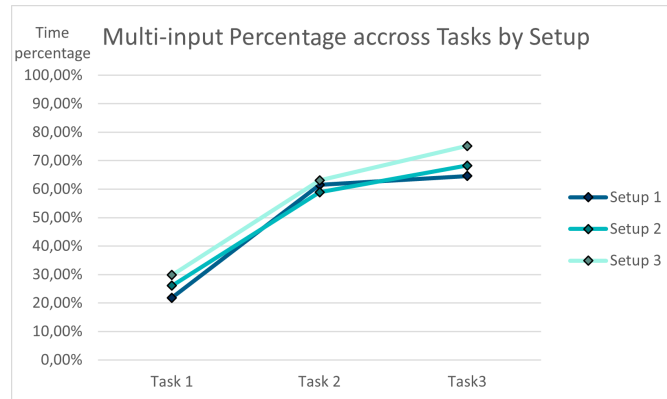


Figure 3.4: Example of Multi-input Metric, Percentage of Time spent Using more than One Input

3.1.4 Implementation of Common Performance Indicators

In this section, we present the implementation of operational data within our proposed methodology, highlighting their practical applications and effectiveness. Specifically, we focus on the use-case of controller selection process for an earthwork teleoperation and training tool in virtual reality. This use-case aims to identify the most suitable controller for the VR application from a diverse range of user-grade controllers, encompassing gamer, hybrid, and professional options, see Fig. 3.5. By examining these controllers through the scope of operational data, we seek to enhance the operational efficiency and user experience of the VR tool, ultimately contributing to the optimisation of VR application design and functionality.

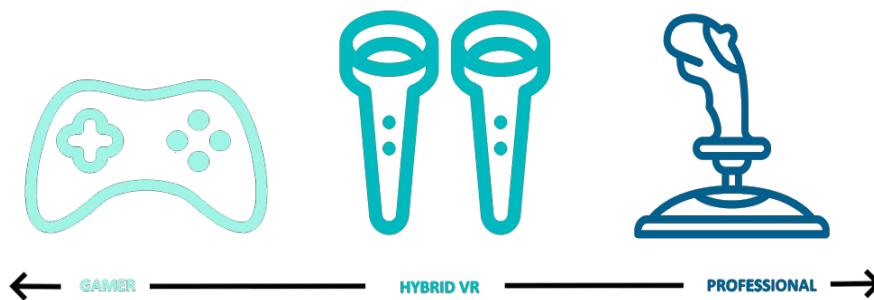


Figure 3.5: Controller Selection Spectrum

Earthwork Training

Teleoperation and training within the field of robotics has seen significant advancements in the latest years, particularly in earthwork robotics. Training in this domain focuses on providing operators with the skills necessary to handle complex machinery efficiently and safely. The integration of digital twin technology and simulation has changed the training landscape, offering numerous advantages over traditional methods.

Simulations play a crucial role in the training of earthwork robotics by creating immersive and realistic scenarios that operators may encounter on actual job sites. These simulations enable trainees to experience a wide range of situations, from routine tasks to emergency responses, thus preparing them for the challenges of real-world operations. The effectiveness of simulation-based training has been demonstrated in various studies, highlighting improvements in both operator performance and confidence [70]. To reinforce these simulations, digital twins, which are virtual replicas of physical systems, allow for real-time monitoring and analysis of robotic operations [71]. This technology provides a dynamic, realistic and interactive training environment where operators can practice and hone their skills without the risks associated with real-world scenarios. The use of digital twins in training not only enhances learning outcomes but also reduces operational costs and downtime.

Another enhancement of training simulation comes from the medium itself: the use of VR environments. VR has emerged as a powerful tool in the realm of robotic training, reinforcing immersiveness and providing engaging learning experiences. VR applications provide a safe and controlled environment where trainees can naturally interact with virtual models of industrial equipment, practice operational procedures, and receive instant feedback on their performance. They safely and cost-effectively learn the correct gestures and patterns needed for their future operations. The advantages of VR-based training include enhanced spatial awareness, improved hand-eye coordination, and the ability to conduct repetitive practice without wear and tear on actual machinery [72]. Existing applications of VR in robotic training span various sectors, including construction, mining, earthwork or medical robotics. These applications use the immersive nature of VR to their advantage to replicate real-world conditions and provide hands-on experience. For instance, VR training modules for excavator operators have been developed to simulate digging, loading, and maneuvering tasks, enabling operators to gain practical skills in a virtual setting [73].

Controller Selection

The selection of an appropriate controller is crucial in the context of VR earthwork training due to its direct impact on user performance, learning efficiency, and overall experience [74]. Controllers are the first direct interface between the user and the virtual environment, making their design and functionality essential. An intuitive and responsive controller enhances task performance by allowing users to operate machinery with greater precision and speed, thereby reducing the likelihood of errors and improving operational efficiency [75]. This is especially important in training applications where the goal is to provide realistic and practical experience that closely mimics real-world teleoperations, to be later used in real-life use-cases. A well-designed controller can bridge the gap between virtual training and actual fieldwork, ensuring that skills learned in the virtual environment transfer effectively to real-world teleoperation work done by the trainee.

To test the efficiency of operational measures in VR design, we propose a study that evaluates the performance and user experience of three different controllers: gamepad, VR controllers, and professional simulator joysticks, within a VR excavator teleoperation training setup. While other studies already examined controllers for earthwork teleoperation and training [76], they often use either the default VR controllers provided with the VR headset [77] or extensive custom hardware [78]. While both approaches have their advantages - respectively cost-efficiency and precision - they also have their drawbacks - namely the need for custom interactions development and expensive hardware. Our study bridges two gaps: can user-grade, plug-and-play controllers provide the necessary precision and interaction for complex machinery control? And are operational metrics enough in their selection method?

Study

The study involved 13 participants aged between 22 and 41 years, with an average age of 28.1 years, 12 males, 1 non disclosed. These participants had diverse backgrounds, with varying levels of experience in both VR technology and earthwork operations, as well as different degrees of familiarity with the controllers being tested: gamepad, VR controllers, and professional simulator joysticks [79]. They were tested on a VR application designed to train the future excavation teleoperators, in complete immersive simulation.

The VR headset used in the study is the HTC Vive Pro Eye (Fig. 3.6). The three types of controllers evaluated in the study were a standard Xbox One gamepad, the HTC Vive Wands (the default VR controllers associated with the HTC Vive Pro Eye), and simulator Thrustmaster T.16000M Duo joysticks mapped to match the actual heavy machinery real-life joysticks (Fig. 3.7). The experimental setup included an advanced VR environment, developed with Unity and the Open Construction Simulator package for realistic terrain simulation¹. It includes a 1:1 digital twin of a Cat323 excavator (see Appendix A.1 for the excavator schematics), with emulated control, as seen from the user point of view Fig. 3.8.

¹Open Construction Simulator <https://github.com/Field-Robotics-Japan/OpenConstructionSimulator>



Figure 3.6: HTC Vive Pro Eye



Xbox One gamepad



HTC Vive Wand



Thrustmaster T.16000M

Figure 3.7: Controller Selection Hardware

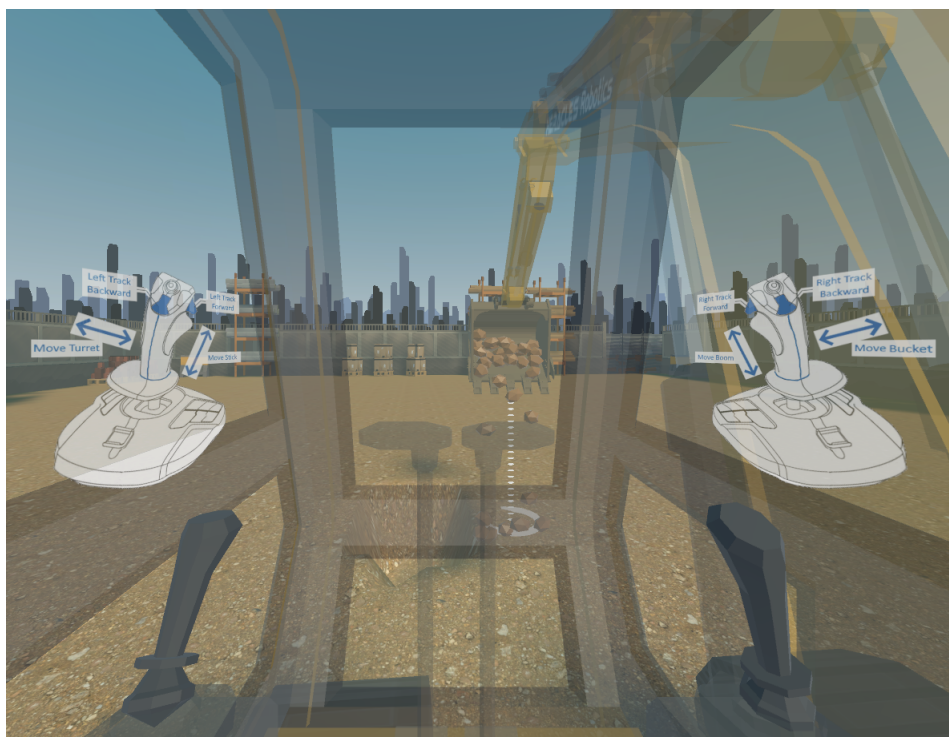


Figure 3.8: User Interface for the Controller Selection Experiment with Interchangeable Controller Mapping Scheme Display (Here the Joysticks Mappings)

The operational performance metrics were recorded through Unity, namely task completion time, input errors, and multiple inputs usage. All participants began the experiment with a general explanation of the process, signing a consent form for data collection and completing a user profile questionnaire. They then went through three cycles of testing, one for each controller. Each cycle consisted of four exercises, done entirely in a virtual simulation,

designed to test our three controllers and their mappings (as seen in Fig. 3.9):

- **Exercise 1 - Tutorial:** the participant learns how to use each articulation of the machine, one at a time. Mappings are always displayed on the side;
- **Exercise 2 - Dig:** a simple digging exercise, where the user had to dig in front of them, unload on the side, dig again on the side, and unload at the front. No navigation (vehicle travel) was needed for this exercise;
- **Exercise 3 - Navigation:** a pure navigation exercise, where the user had to go to multiple positions, learning how to travel using independent tracks control;
- **Exercise 4 - Complete Training:** a compound exercise encompassing everything learned before. The user had to navigate to a digging spot, dig, navigate to an unload spot, unload, repeat this process twice, and finally reposition the excavator in its original location.

At the end of all three cycles, the user had to rank the controllers in their preferred order. To avoid a learning bias, the order of the controller cycles was varied for each participant.

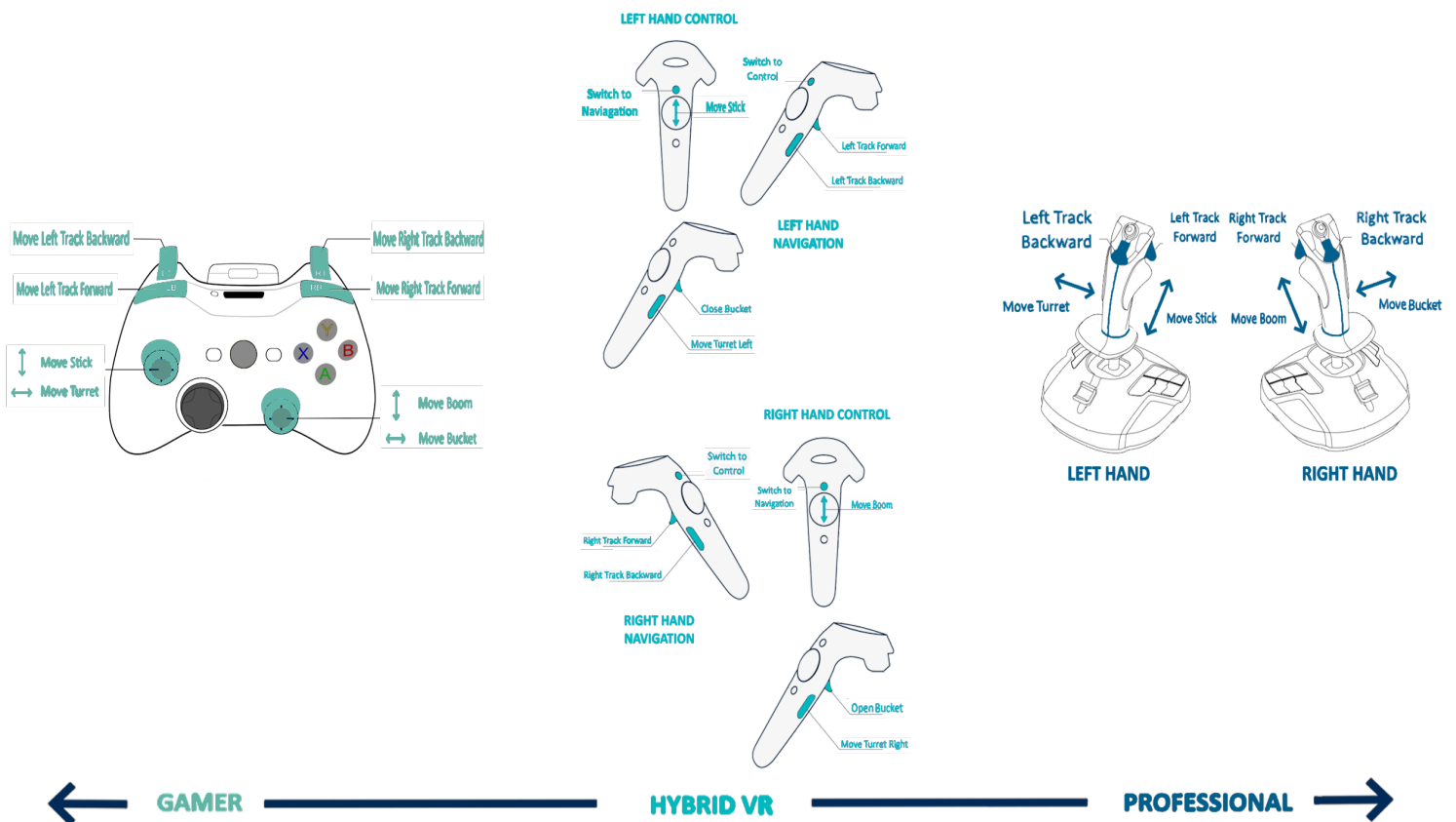


Figure 3.9: Gaming, Hybrid and Professional Mappings

Results

Over the cycles of exercises for each controller we evaluated three common operational metrics: completion time, inputs errors and multi-inputs usage. All three serve as performance metrics, and each one brings supplementary information on the user experience.

Completion Time

Each exercise in the sequence of tasks needed to be completed and validated by the software and/or the supervisor. Because of this step-by-step validation, we consider exercise completion time as the "ground truth" of user performance. As seen in Fig. 3.10, evaluating time not only shows actual performance (Fig. 3.10a) but also progression between exercises (Fig. 3.10b): EX1 - tutorial, requiring the longest duration; EX2 - digging exercise, short; EX3 - navigation exercise, longer because it involves unusual independent track control; EX4 - the comprehensive exercise involving all previous tasks. Each exercise is defined by the controller used, with the gamepad in light blue, the VR controllers in medium blue, and the joysticks in dark blue. Not only do the joysticks take on average less time, but there is also a larger gap between the tutorial and the other exercises compared to the VR controllers or the gamepad.

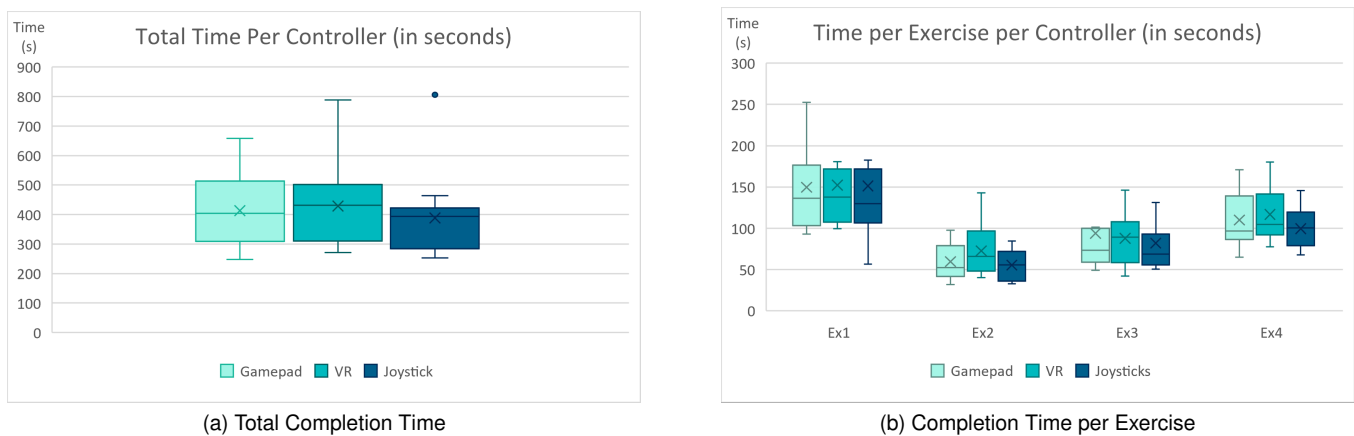


Figure 3.10: Completion Time Metric

Time is obtained directly through software recording, namely Unity. The completion time metric - both total time and per exercise time - provides not only a performance assessment but also progression information. By dividing completion time in task/exercises, comparing the user evolution becomes possible, inferring their progression across different type of actions, or in this use-case, different type of controllers.

Input Errors

Input errors, in this study, describe either moving the wrong articulation or moving the right articulation the wrong way. Results, shown in Fig. 3.11, once again put the joysticks on top with the fewest errors across all controllers.

Input errors are evaluated through video recording and supervisor judgement in a two-times evaluation (during the experiment and after the experiment using video recording). Studying input errors provides a measure of performance, but also a learning and an ergonomics measure. In this particular study, the joysticks were the less familiar controller of the three for the participants, and yet still produced fewer errors. Therefore, this metric shows that not only the user learned the controls faster with this specific controller, but also that the controller itself was better adapted to the application, the more ergonomic option out of the ones tested.

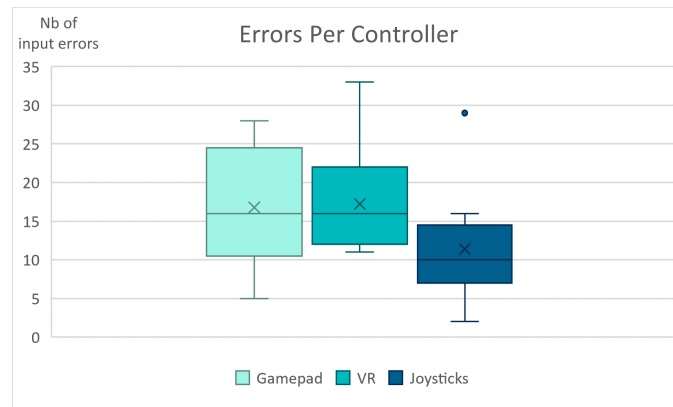


Figure 3.11: Input Errors per Controller

Multi-inputs

The multi-inputs metric corresponds to the percentage of time spent using more than one input at a time. In our use-case, this means moving multiple articulations of the virtual machine. As an excavator requires using both tracks at the same time to navigate, it would automatically increase the percentage of time spent using multiple inputs and potentially bias the measures. However, as the exercises are identical and involve the same amount of navigation for all controllers, this potential bias is taken into account. The tutorial exercise requires moving only one articulation at a time and thus is excluded from the data analysis. According to the results, it should be noted that the joysticks not only involve the most usage of multiple inputs at a time, showing the ease of use of the controller, but also show faster acclimatization and better progression across the exercises, as shown in Fig. 3.12.

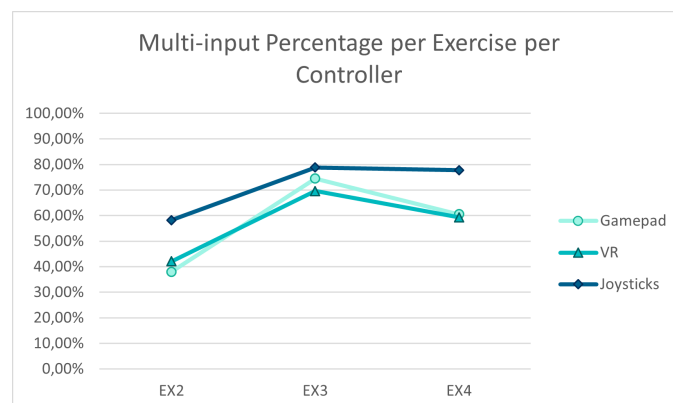


Figure 3.12: Multi-inputs Percentage across Exercises

The multi-input metric is recorded through software, directly in Unity. This metric once again reports performance, but also ease-of-use and user progression, provided it is valid, which needs to be checked by comparing multi-inputs data with time and errors data. Multi-inputs can be a great source of information as long as it does reflect a voluntary action from the user, instead of an involuntary action attributed to the user or to controller encoding. The validation of the multi-input metric is investigated by means of Pearson correlation [80], computed with respect to time and errors, as reported in Table 3.1. We found that the percentage of multi-inputs (MI%) is significantly negatively correlated with time (our ground truth), meaning the more inputs combination leads to less time to finish the task, which fits our expectation. Interestingly, the joysticks which are the best performers among all controllers show high but non-significant correlation between MI% and time. One possible explanation could be that being the less familiar controller, the use of joysticks is more sequential (less MI%) but still with a faster and more accurate outcome. Now for the question of voluntary or involuntary usage of multiple inputs, we see that errors and multi-inputs do not correlate significantly, and tend to be rather negatively correlated, showing that multi-inputs do not result from errors, rather the opposite. In conclusion, this analysis validates the multi-input metric as a strong operational measure.

Data correlation analysis	Correlation	p-values
MI% vs Time (Gamepad)	-0.77	0.002
MI% vs Time (VR)	-0.75	0.003
MI% vs Time (Joysticks)	-0.42	0.16
MI% vs Time (Average)	-0.65	0.05
MI% vs Errors (Gamepad)	-0.26	0.4
MI% vs Errors (VR)	-0.52	0.06
MI% vs Errors (Joysticks)	-0.34	0.2
MI% vs Errors (Average)	-0.37	0.22

Table 3.1: Validation of the Multi-input (MI%) Metric (Pearson correlation)

3.2 Use-case Specific Measures

While common operational metrics such as completion time, input errors, and multi-input usage provide a fundamental understanding of user performance and interaction efficiency, it is essential to consider the unique demands and characteristics of each specific use case. In the presented study, we see that time, inputs errors and multi-inputs percentage already give an answer as to "which controller should be used for an excavator training application?". These operational metrics established a baseline for performance evaluation across our scenario, but other metrics, very specific to this use-case could have provided further insights on different aspects of the application. To gain a comprehensive and nuanced insight into the effectiveness of VR applications, we must also incorporate use-case specific metrics tailored to the particular requirements and objectives of each context.

In this section, we will delve into these specialised metrics, which are tailored to address the unique aspects and challenges of different use cases. Earthwork applications, while seemingly reliant on operational measures, and therapeutic applications, which often prioritise physiological measures, both benefit from being studied with a comprehensive hybrid methodology and an open-minded approach. This perspective allows us to uncover insights that may not emerge from a single type of measure alone. Additionally, the specific demands and constraints of these fields necessitate professionally defined metrics to ensure accurate and meaningful evaluation. By integrating these specialised metrics with common operational measures, our approach becomes both targeted and context-sensitive, providing a robust framework for assessing and optimising VR applications across diverse domains.

Using use-case specific data is already a practice in place in various fields, where those specific metrics have been identified and applied to different VR scenarios, such as medical research, industrial applications and more. For example, using metrics such as precision of movements, adherence to procedural steps, and time spent on critical tasks in medical training can be used to gauge effectiveness and skill acquisition [81]. In VR-based rehabilitation, specific metrics might include range of motion, exercise repetition counts, and patient engagement levels [82]. For VR applications in architectural design and urban planning, user interaction with design elements, navigation patterns, and spatial understanding are key metrics [83].

By integrating these specialised metrics, we can better address the distinct requirements and challenges of each VR application, leading to a more accurate and relevant assessment. In the following sections, we will explore two specific use-cases and their associated metrics: the validation of a VR interface for earthwork teleoperation and a prototype of visual rehabilitation for dyslexia. Both studies demonstrate how tailored metrics can contribute to a deeper understanding and more effective optimisation of a VR system.

3.2.1 Implementation through Study - Earthwork Teleoperation

Teleoperation in robotics, and within the domain of earthwork, has significantly advanced with the development of various interface technologies, as well as its combination with robotic automation. These interfaces facilitate remote control of machinery, enhancing operators' safety, control precision, and cost-efficiency. The integration of VR into teleoperation offers immersive and intuitive controlled environments that traditional interfaces may not provide.

Traditional teleoperation interfaces typically include physical controls such as joysticks, levers, and touchscreens. These methods, while effective, often come with limitations in terms of user engagement and spatial awareness. Operators must rely heavily on visual feedback from 2D monitors, which can hinder depth perception and situational awareness [84]. In contrast, VR interfaces for teleoperation provide a three-dimensional, immersive environment that can enhance the operator's spatial awareness and control precision. By combining head-mounted displays (HMDs) and motion-tracking systems, VR interfaces allow operators to experience a more natural and intuitive interaction with the remote environment. For example, operators can perform complex maneuvers with greater confidence and accuracy, as the VR system provides a more realistic representation of the remote site [85]. The use of VR in earthwork teleoperation has been explored in several studies. One approach involves digital-twin, i.e. the use of VR to simulate the operator's cabin, replicating the controls and providing a virtual view of the worksite. This method allows operators to practice and execute tasks in a controlled virtual environment before performing them in the real world, thereby reducing the risk of errors and accidents [86]. Moreover, VR interfaces can incorporate additional data layers, such as real-time sensor feeds and augmented reality (AR) overlays, to provide comprehensive situational awareness. This integration enables operators to access vital information, such as machine status, environmental conditions, and potential hazards, directly within the VR environment [87].

Despite these advancements, challenges remain in the widespread adoption of VR for teleoperation. Issues such as cybersickness, high equipment costs, and the need for robust and low-latency communication links are significant barriers [88]. Furthermore, the effectiveness of VR interfaces is highly dependent on the fidelity of the virtual environment and the accuracy of the motion-tracking systems. Thus the need for extensive design and optimisation. To this end, we propose using operational data such as completion time, inputs errors and multiple inputs usage, but recognise that for such a specific and complex use-case, those general metrics might not reflect all the necessary aspect for VR design. Consequently, we propose additional specific measures for evaluation.

Study

Through the lens of operational measures, we studied the usability of a VR interface for excavator teleoperation. This study was done in partnership with HERACLES Robotics, a French company focused on earthwork machinery automation. It was actually performed before the controller selection study presented previously, and it is through this study that user feedback advocated to perform a controller selection study. The use of autonomous and semi-autonomous earthwork machines has significantly increased in recent years due to their benefits, including reduced labor costs, increased productivity, and improved safety [89]. However, while these machines provide significant improvements, they are not immune to unforeseen events and technical failures, which can lead to potential safety risks, delays, and additional costs. Their deployment can also be delayed in certain country due to legislative reasons.

To provide a bridge between traditional machine control and autonomous vehicle, we focused on a middle-ground solution: teleoperation. Our primary objective was to provide a fail-safe solution in situations where unforeseen events occur during automated operations. We aimed to create a universal and user-friendly interface for remote earthwork that minimises the need for extensive training and complex skill requirements. VR training for teleoperation has already proven its efficiency, as well as for direct control. By enabling operators to assume direct control of the machines from a remote location, we eliminate the need for constant on-site presence, while paving the way for semi-autonomous and autonomous machinery. In this study we let 10 participants - professional excavator operators and non-professional - evaluate a VR interface compared to the real machine itself, Fig. 3.13.

Our primary investigation focused on comparing the performance of operators using our VR interface and direct control of the real machine, with operators having varying levels of experience. The preliminary results indicated that the VR interface delivered equivalent and satisfactory performance for experts while enhancing the efficiency of non-experts. Therefore, not only does everyone performed well in the virtual environment, but the training time could also be significantly reduced, as non-experts can achieve similar performance under the same conditions. The complete study was published in Electronics [90].

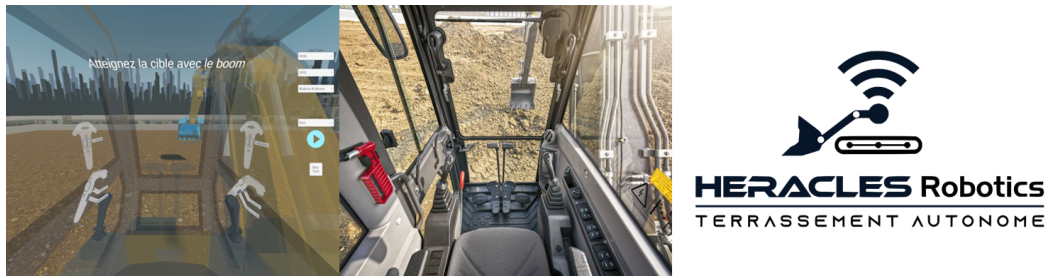


Figure 3.13: Comparing VR Interface and Real Machine, a HERACLES Robotics Partnered Study

Results

In our experiment, both professional and non-professional excavator operators were tasked with evaluating two interfaces: a VR teleoperation interface and a real CAT323 machine. The participants performed the same set of exercises on both interfaces to ensure a fair comparison:

- **Tutorial:** participants first completed a tutorial designed to teach them the basic controls of each interface (Fig. 3.14);
- **Target Practice:** in this exercise, participants had to reach designated ground targets with the machine's bucket (Fig. 3.15).
- **Digging/Unloading Exercise:** participants performed a typical operator task involving digging and unloading (Fig. 3.16).

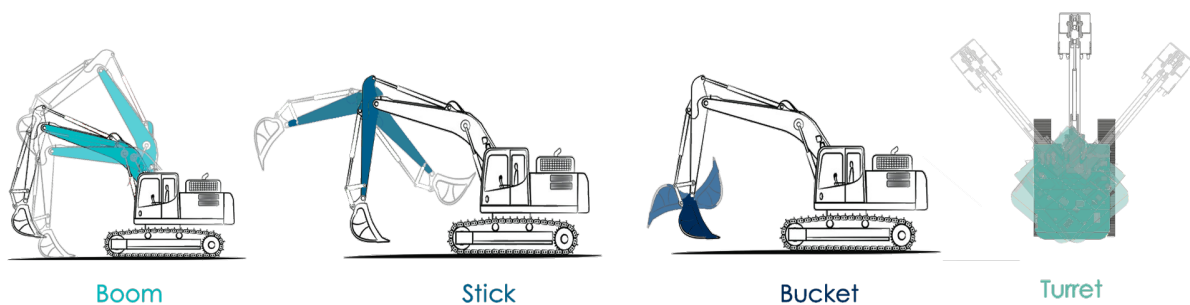


Figure 3.14: Tutorial Exercise - Learning Basic Controls

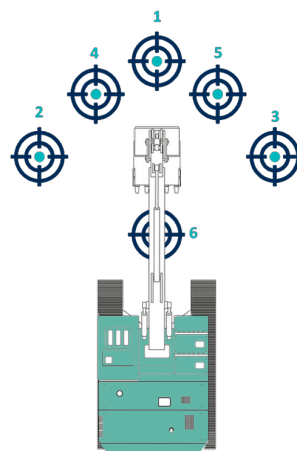


Figure 3.15: Target Practice - Learning Precise Control

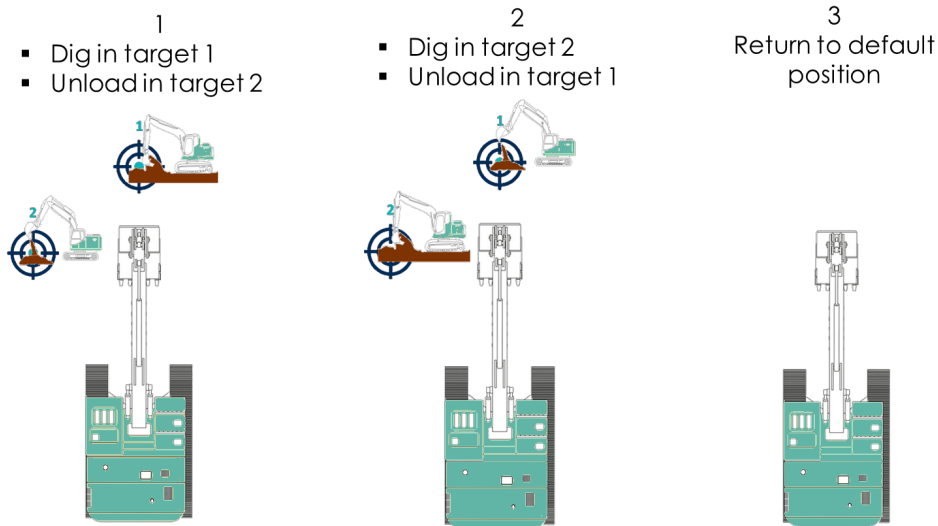


Figure 3.16: Digging/Unload Exercise - Learning Teleoperation Main Skill

The VR interface was used to directly control the real CAT323 machine through Unity (see Appendix A.2): the user could visualise and give commands to the virtual machine and the Robot Operating System (ROS) actualised the real machine articulations with the Unity commands and sent back the real machine's proprioception and terrain perception - as illustrated in Fig. 3.17.

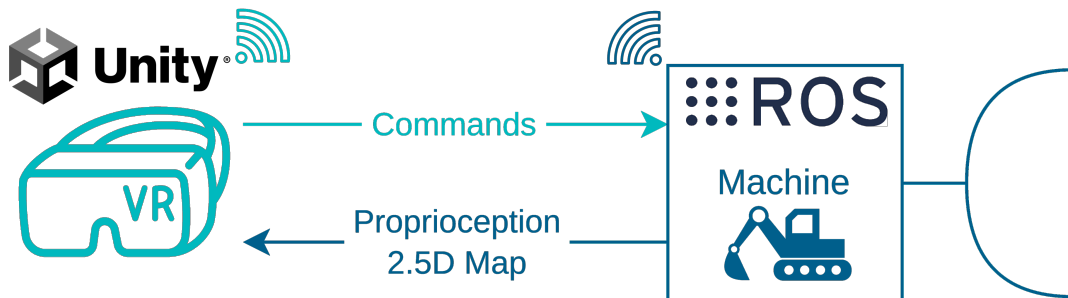


Figure 3.17: Unity-ROS Excavator Control Architecture

In this specific use-case, by using a real machine, its associated software (ROS) and sensors (e.g. LiDAR, cameras, etc.) we can extract brand new specialised data: machine proprioception, machine consumption, terrain perception, and so on. In this study, we were able to extract a specialised operational measure: the control smoothness.

Novice operators tend to be clumsy, slow, or brutal with the machine, wearing it down quickly. This concern is major in the conception of the teleoperation controls, as both improving future operators skills and preserving the machinery is paramount. To assess control smoothness, we measure both the clumsiness and the dexterity of the participants:

- **Clumsiness:** the clumsiness of a participant p with an articulation a , noted $Clumsiness(p, a)$ is monitored through the acceleration of the articulation, data provided by the real machine (through ROS). Clumsiness correspond to rare and extreme events (e.g. brutally stopping an articulation when hitting the ground or a mechanical stop) and is defined as the standard deviation of the acceleration of an articulation over all exercises, noted $StdAcceleration(p, a)$. For one articulation, it is calculated as:

$$ClumsinessRaw(p, a) = StdAcceleration(p, a) \quad (3.1)$$

In order to aggregate the total clumsiness of a participant for each articulation, we normalise their amplitudes with respect to the average value over all participants:

$$Clumsiness(p, a) = \frac{ClumsinessRaw(p, a)}{Average_p(ClumsinessRaw(p, a))} \quad (3.2)$$

We can measure a meaningful feature of the participant by averaging his clumsiness for all articulations (in this experiment, the boom, the stick and the bucket, see Appendix A.1), as normalised clumsiness values of a given participant for each articulation are strongly correlated:

$$Clumsiness(p) = (Clumsiness(p, Boom) + Clumsiness(p, Stick) + Clumsiness(p, Bucket))/3 \quad (3.3)$$

We finally normalise the clumsiness metric:

$$NormalisedClumsiness(p) = \frac{Clumsiness(p) - \min_p(Clumsiness(p))}{\max_p(Clumsiness(p)) - \min_p(Clumsiness(p))} \quad (3.4)$$

- **Dexterity:** we define the dexterity of a participant, $Dexterity(p)$, as the opposite of his clumsiness:

$$Dexterity(p) = 1 - NormalisedClumsiness(p) \quad (3.5)$$

Through clumsiness and dexterity, we can evaluate the control smoothness metric. The best excavator operators are not only the fastest (in terms of completion time) - completing a lot of work quickly - but also the smoothest (in dexterity) - preserving the machine in the long run. The idea of smoothness is deeply embedded in the design and implementation of the remote control system, and by analysing this smoothness metric Fig. 3.18, we showed in this study that our interface allows even the most novice operators to operate the excavator in complete safety with regards to the machine.

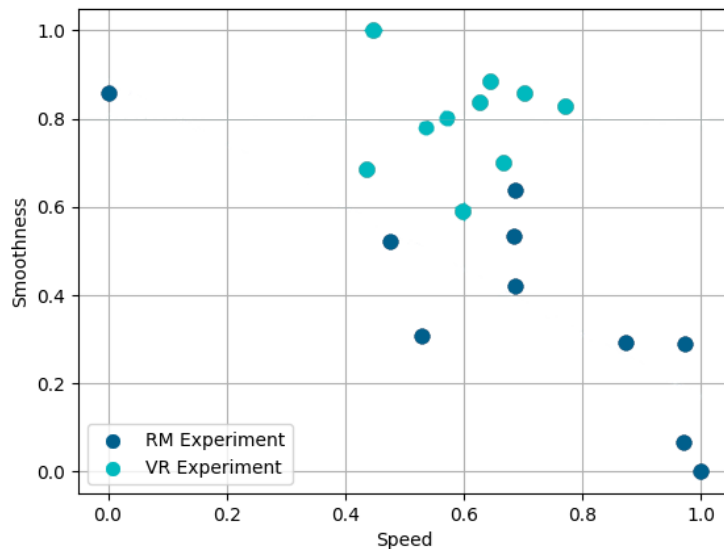


Figure 3.18: Control Smoothness compared with Completion Time in the Real Machine interface (RM) and the Virtual Reality Interface (VR)

The ideal control behaviour, which optimises both speed and smoothness, is situated at the upper-right corner of the graph. As depicted in the graph, users operating the real machine at higher speeds tend to exhibit lower smoothness in their movements. While this allows for excellent performance in terms of completing tasks quickly, it poses a risk of damaging the machine over time. In contrast, the VR interface, on average, operates at a slightly slower pace but maintains a consistently high level of smoothness.

Only by studying specialised metric can we answer this specific challenge of our use-case, namely combining user performance and machine safety. Through this analysis, we saw that tasks performed in VR show longer-lasting benefits for the machine's condition, as well as enhanced precision, albeit with a minor trade-off in execution speed compared to the real-machine. This validates the design of the VR interface and gives insight into its future axes of optimisation for earthwork teleoperation.

3.2.2 Implementation through Study - Dyslexia Therapy

Dyslexia, derived from the Greek words "Dys" (meaning difficult) and "Lexia" (meaning words), is a reading disorder affecting approximately 8 to 12% of the global population. It is characterised by difficulties in reading that can impact spelling, reading speed, word comprehension, and pronunciation [91]. Dyslexia poses significant learning challenges from a young age, that can persist throughout an individual's academic career and beyond, as the symptoms tend to affect individuals for life. Early diagnosis of dyslexia often occurs in school, typically around the age of six, and involves a series of assessments that include vision, memory, spelling, and reading exercises. Although dyslexia therapies can alleviate some symptoms and reduce their impact on daily life, they do not eliminate the disorder entirely. The origins of dyslexia are still debated: is it a neurological disorder present at birth showing symptoms when learning to read and write; or is it developed in the brain when children learn to read and write? While this question will sadly not be answered in this thesis, dyslexia rehabilitation will nonetheless be elaborated.

Traditional rehabilitation methods often involve repetitive pencil-and-paper exercises [92]. These methods can be tedious and demotivating, which can affect therapy adherence and overall efficacy. Typically conducted in medical offices under the supervision of specialists, these exercises may also be performed at home with parental oversight when possible. The monotonous nature of these tasks frequently leads to a lack of engagement and, consequently, reduced therapeutic benefits as patients tend to give up before completion. New rehabilitation methods such as digital therapeutics (DT) use digital applications to enhance or provide treatment for various conditions [93]. These technologies offer innovative solutions without chemical side effects, and include features such as data collection, monitoring, and engaging, game-like elements that can be particularly motivating for children [94]. In recent years, the integration of DT with VR has emerged as a promising approach. Unlike traditional methods, VR provides a mobile, immersive experience that can be used at home, reducing the need for frequent medical visits. The immersive nature of VR has demonstrated effectiveness in various therapeutic contexts [95].

Study

Although the debate on the origins of dyslexia is still open (nature or nurture), it is undeniable that both the brain and the eyes are at the heart of this pathology. In partnership with medical specialists, we developed *Dys & Dragons*, a prototype digital therapy focused on visual rehabilitation through gamified virtual reality [96]. This prototype's goals were multiple:

- **Provide engaging therapy:** using new technologies and gamification to provide an engaging rehabilitation interface for patients;
- **Assess the contribution of 3D:** while current visual dyslexia therapies are based mostly on 2D exercises, using VR provides an environment where 3D exercises can be developed and analysed;
- **Extend traditional therapy:** transposing and extending current therapeutics exercises to VR, while also creating new ones.

Due to the medical nature of this use-case, other operational data are needed to evaluate patients performance during the exercises, and more importantly their progression during the therapy.

Following the advice of medical professionals - neurologists and orthoptists, three types of exercises were first developed in this prototype (Fig. 3.19):

- **Visual exercises:** only based on eye movements, these exercises include smooth pursuit (following a continuously moving object), saccadic pursuit (following a teleporting object) and focused gaze (focusing on a stationary object while the environment is filled with distractors) (Fig. 3.19a). Their aim is to train directly the eye movements which are impacted during reading by dyslexia;

- **Ecological exercises:** in the orthoptist terminology, those exercises concern environmental tasks, such as finding an object in a scene (Fig. 3.19b). These exercises work on memory, visual perception, and by using 3D virtual environments, spatial awareness and perception;
- **Reading exercises:** the most traditional exercises in dyslexia therapy (Fig. 3.19c). It is used as a session evaluation rather than a repetitive exercise (like the previous one) to assess the current patient's progress.



Figure 3.19: *Dys & Dragons* Exercises

Results

Currently, no clinical trial results are available for this study. However, in collaboration with medical specialists, we have extracted operational metrics that offer valuable insights into the effectiveness and requirements of such interventions.

In the context of rehabilitation, specifically dyslexia, integrating specialised operational data is essential. This includes physiological sensors and supervision by medical professionals. Our application serves a dual purpose: it acts as both a "therapeutic game" for patients and a "therapeutic supervision tool" for practitioners. In addition, the focus of this rehabilitation tool is visual training, making eye-tracking a key component. Eye-tracking provides the following benefits:

- **Smooth Pursuit Quality Assessment:** measures how well the patient can follow a smoothly moving object during visual exercises;
- **Saccadic Motion Analysis:** evaluates saccadic eye movements as the patient tracks a teleporting object in the virtual environment;
- **Focus:** assesses the patient's ability to maintain attention on an object in the presence of environmental distractions;
- **Spatial Search:** reveals 3D spatial visual search patterns during ecological exercises;
- **Visual Attention:** provides an overview of the patient's visual attention linked to spatial search during ecological exercises;
- **Reading Patterns:** analyses reading patterns to assess dyslexia progression through specific reading exercises.

Another key set of metrics in this use-case are neurological metrics. By using sensors such as electroencephalograms (EEG), it is possible to monitor the patient's state throughout different stages of the therapy. This includes tracking cognitive load, stress levels, engagement, and other factors that allow supervising doctors to tailor the therapy to the specific needs of the patient. For example, if precise eye-tracking analysis (e.g., micro-saccadic motions during pursuit) reveals that the patient performs better with smooth pursuit exercises but struggles with saccadic movements, the practitioner can increase the focus on saccadic pursuit exercises. Similarly, if the cognitive load is higher during dynamic ecological exercises compared to stationary ones, the practitioners can adjust the difficulty

accordingly. To enable this optimisation, a setup such as the one proposed in Fig. 3.20 could be implemented.

This study could not be led to its end during this thesis due to the disengagement of our partner who contracted the collaboration with the medical staff. However, despite the lack of clinical results, this study shows a transition to incorporating physiological data, such as eye-tracking and cognitive data inside operational data. It underscores the importance of capturing not only external behavioural metrics, but also intrinsic data. Intrinsic data refers to the underlying physiological responses that can offer deeper insights into a user's interaction with the VR environment.

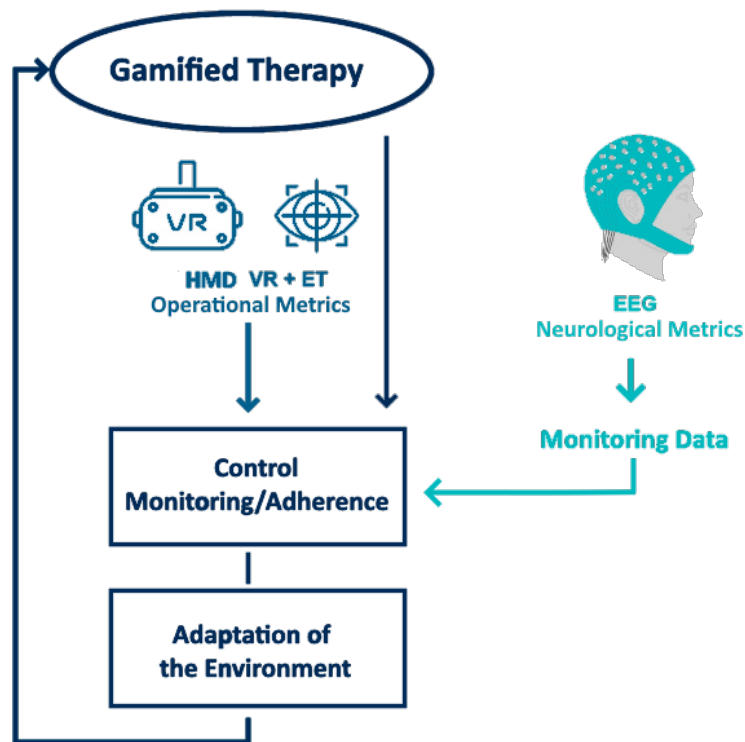


Figure 3.20: Immersive Visual Rehabilitation for Neurological Troubles, a Proposed Framework

3.3 Evaluation of Extrinsic Operational Measures

In this chapter, we have detailed our contributions to the analysis of operational data for VR design and optimisation. Our research focused on several key metrics: completion time, error rates, and the use of multiple input methods. These metrics provided insights on user experience in the form of performance, progression, ergonomics, and ease-of-use. We also showed that specific use-cases call for specific metrics, such as machinery data when working on teleoperation, or medical data when designing therapeutics applications.

Our findings indicate that, while the operational data collected is valuable, it might not be sufficient on its own. Each experiment presented in this chapter also had a user feedback phase. When confronted to user feedback, we observed that the operational measures provided similar answers to the fundamental design questions of each study:

- **Which controller for earthwork excavation training?** All our operational metrics showed that the joysticks performed better and provided better ergonomics, showing for this study the superiority of professional design over controller familiarity (users were more familiar with the other controllers but performed better with the joysticks). It matches with the user's perception who preferred the joysticks to the other controller, found them more efficient while requiring less effort and generating less frustration;
- **Can a VR interface be used for excavation teleoperation?** According to our generic metrics, while the performance of the real machine are slightly better in terms of time, VR performs admirably, and even shows better control smoothness, optimising both the user performance and the machine safety. When asked, the users would also guarantee that the VR interface was viable for teleoperation;
- **Is VR an efficient support for dyslexia therapy?** While there was no user study, in-laboratory testing showed that it was possible to apply metrics such as completion time or errors. It also showed that physiological data were the most suited for medical applications. According to the medical specialists working on this project, the *Dys & Dragons* application correctly replicated the existing current exercises and could provide insight into the use of 3D in rehabilitation.

These insights slightly enhanced our understanding of user performance and experience, but overall the results from both methods largely coincided. This suggests that while operational data add some value, they might not be enough when used alone.

Although they are easy to use, obtain, and analyse, our operational measures have one major flaw: they are completely external to the user. Even when setting aside user perception and other subjective metrics and keeping to objective metrics, operational measures are extrinsic and judge the user experience from an external point of view only. This does give the advantage of judging not only the user but also the external parameters (e.g. excavator consumption when teleoperating compared to direct control). However, in return they do not provide internal insight into the user, which underscores a limitation in relying solely on external evaluations of user performance. We saw that in the case of medical application, physiological - intrinsic - data are needed. We also believe that medical applications would not be the only ones benefitting from such data.

To address this gap, we propose the integration into our VR-SOLUS framework of intrinsic data obtained through physiological sensors. By incorporating physiological metrics, such as those captured by eye-tracking technology, electroencephalogram, motion sensors, and so on, we can gain deeper insights into the user's internal states and responses. This approach promises to improve the accuracy of our evaluations, providing a more extensive understanding of user interactions within VR environments.

Chapter 4

Physiological Measures

Incorporating physiological measures into the evaluation and design of virtual reality (VR) applications is a significant advancement over traditional operational metrics such as task completion time or error rates. Although operational metrics provide valuable information on the effectiveness and efficiency of VR systems, they do so only from an external point of view. Physiological measures, on the other hand, offer a more comprehensive understanding of user interactions by providing objective and intrinsic data on users' physical and cognitive states.

Traditional operational metrics, though useful, can sometimes overlook critical aspects of user experience and engagement due to their extrinsic nature. For instance, a user may complete a task efficiently but experience high levels of cognitive load or discomfort, which can negatively impact their overall experience and long-term usability of the system. Physiological measures can capture these internal dimensions by recording real-time data on various physiological responses, such as heart rate, brain activity, and eye movements. This data helps to understand the user's emotional and cognitive states, providing a deeper insight into their interactions with the VR system [97]. Physiological measures contribute to VR design and optimisation in several key ways:

- **Enhanced User Experience:** by monitoring physiological responses, designers can identify and mitigate sources of discomfort, stress, or fatigue, thus enhancing overall user experience. This is particularly important in VR environments, where prolonged exposure can lead to issues such as motion sickness or eye strain [98];
- **Objective Cognitive Load Assessment:** cognitive load, which reflects the mental effort required to perform a task, can be objectively assessed through measures like electroencephalography (EEG) [99]. Understanding cognitive load from an internal and objective point of view helps in designing VR applications that are not only engaging but also cognitively appropriate for users, preventing overload and improving task performance;
- **Real-Time Feedback:** physiological measures provide real-time feedback, providing data not only for the initial design phase, but also for dynamic adjustments to the VR environment based on the user's current state [100];
- **Personalised Experiences:** by using the real-time feedback provided by physiological data, VR systems can be tailored to individual users, offering personalised experiences that cater to their specific needs (e.g. adaptive medical therapy or difficulty level in professional training) and preferences [101]. This enhances engagement and effectiveness, making VR applications more user-centric;
- **Validation or Substitution of Subjective Reports:** user feedback collected through surveys or interviews can be biased or incomplete. Physiological data serves as an objective complement to these subjective reports, helping to validate and refine user feedback [102]. Physiological measures are sometimes used to substitute surveys, but a dual approach could ensure a more accurate and comprehensive understanding of user experiences.

Physiological measures provide a complementary layer of data to our VR-SOLUS framework that enhances the design and optimisation of VR applications. By capturing the subtle and often overlooked intrinsic aspects of user experience, these measures ensure that VR systems are not only operationally efficient, but also user-friendly and cognitively appropriate. The integration of physiological data with traditional operational metrics represents the second step in our VR-SOLUS method, to develop sophisticated, responsive, and immersive VR environments.

4.1 Physiological Sensors Integration

The integration of physiological sensors into VR applications represents a significant advancement in the field of human-computer interaction. They capture a wide range of physiological markers, providing a deeper understanding of user experiences and interactions. In this thesis, we focus mainly on eye trackers, electroencephalograms (EEGs), inertial measurement units (IMUs), and heart rate monitors, each offering unique insights into different aspects of user behavior and state:

- **Eye-tracking:** eye-tracking technology involves the measurement of eye positions and movements, which can reveal where a user is focusing their attention, their processing of visual information, and what elements in the VR environment are most engaging. Eye-tracking data are important for understanding user interaction with the virtual world and can help optimise the placement of virtual objects and interfaces [103]. For instance, eye-tracking can identify areas that are frequently overlooked or cause visual discomfort, guiding adjustments to improve user experience;
- **Electroencephalography (EEG):** EEG sensors measure electrical activity in the brain, providing real-time data on cognitive processes such as attention, workload, and emotional states. In VR, EEG can be used to assess how users mentally interact with the virtual environment, identifying periods of high cognitive load or stress [104]. This information is valuable for designing VR applications that are not only engaging, but also cognitively efficient for users. EEG data can also be used to personalise VR experiences by adapting the complexity of tasks based on the user's current cognitive state;
- **Inertial Measurement Units (IMU):** IMUs are sensors that measure acceleration, rotation, and other aspects of body movement. In VR, IMUs, including accelerometers and gyroscopes, provide detailed data on user physical interactions with the virtual environment, such as gestures, posture changes, and locomotion [105]. These sensors are even essential for the VR headset to work, as they allow user movements inside of the VR environment. As external sensors, they are useful for creating more natural and intuitive VR experiences, as it allows the system to respond accurately to user movements;
- **Heart Rate Monitors:** heart rate monitors measure the user's cardiovascular activity, providing indicators of physiological arousal, stress, and overall physical state. In the context of VR, heart rate data can be used to measure the user's emotional and physical responses to different scenarios [106]. Heart rate data can also be used to adapt VR environments in real-time, such as by reducing the intensity of stimuli during periods of high stress.

As shown above, physiological sensors already have roots in VR research. However, most of the existing studies typically employ only a single type of sensor and/or measure one specific metric, and they seldom incorporate these sensors during the application's design phase. A non-exhaustive list of recent studies that use the same sensors as our VR-SOLUS method is available in Table 4.1.

Sensor	Description	Methodology
Eye-tracking	Monitors eye movements to determine where the user is looking and for how long	Attention [107], [108], [109] Cognitive Load and Fatigue [110], [111], [112] Performance and Learning [113],[114], [115] Stress [116], [117], [118]
Electroencephalogram (EEG)	Measures electrical activity of the brain through sensors placed on the scalp	Cognitive load [119], [120], [121] Cyber-sickness [122], [123], [124] Attention [125], [126], [109] Stress [127], [128], [129]
Inertial Measurement Unit (IMU) & Heart Rate	Measures acceleration, rotation, orientation & cardiac rhythm	Physical Load [130], [131], [132] Behaviour and Gesture Analysis [133], [134], [135] Pain [136], [137], [138]

Table 4.1: Existing Sensor-based Methodologies

In our VR-SOLUS method, we focus on eye-tracking, EEG and IMUs (see the framework Fig. 4.1) to extract new intrinsic measures, contributing further to our VR design and optimisation.

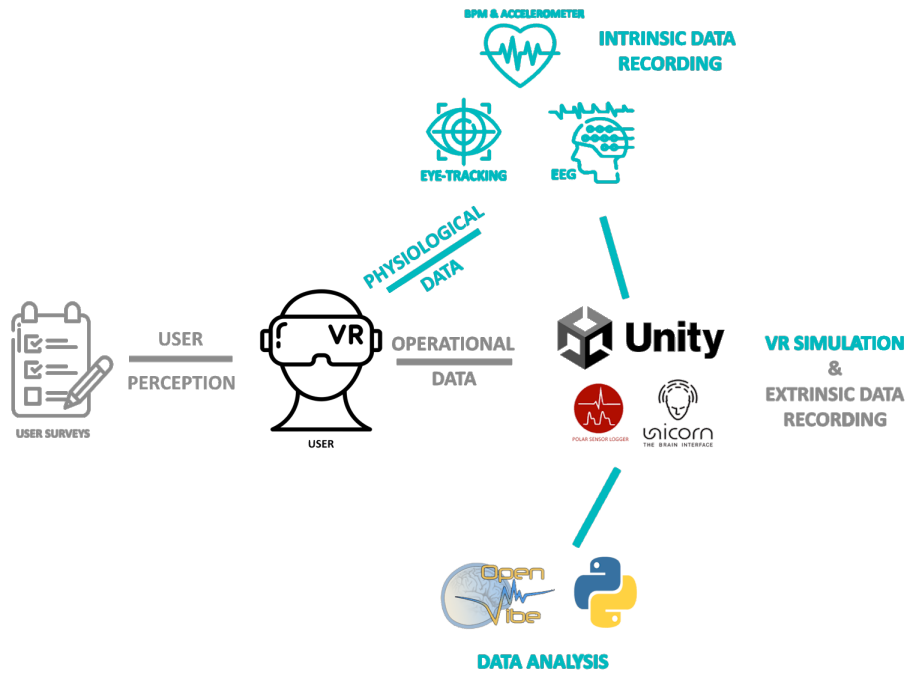


Figure 4.1: Physiological Metrics Framework Highlighted into the Global Analysis Approach Proposed by VR-SOLUS

4.2 Eye-tracking

Eye-tracking technologies measure and record eye movements to determine where and for how long a person is looking at a specific point. This technology uses sensors and cameras to track the position and movement of the eyes, providing information on visual attention and cognitive processes. Eye tracking is widely used in various fields such as psychology, neuroscience, marketing, and human-computer interaction, helping researchers understand how individuals interact with visual stimuli. When comparing regular eye-tracking with VR eye-tracking, the primary difference lies in the context and setup. Regular eye-tracking typically involves a stationary setup in which participants view stimuli on a monitor, and their eye movements are tracked using cameras placed around the screen. This setup restricts head movement and limits the interaction to a two-dimensional space. In contrast, VR eye-tracking is integrated into virtual reality headsets, allowing eye movements to be tracked within immersive, three-dimensional environments. This integration enables the capture of more dynamic and interactive visual behaviours, as users can move their heads and interact with the virtual world naturally, providing a more comprehensive analysis of visual attention in real-time scenarios, see Fig. 4.2.

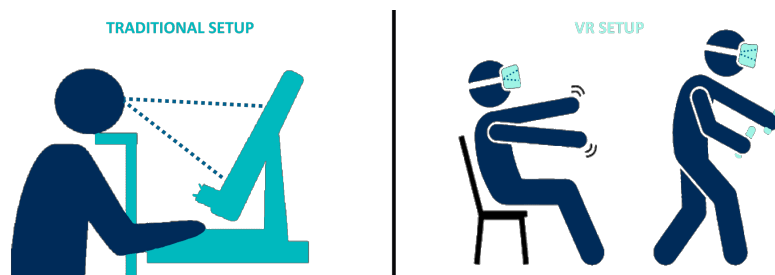


Figure 4.2: Traditional vs VR Eye-tracking Setups

In VR, eye-tracking has several common applications. It is used in usability testing to assess how users interact with VR interfaces and identify areas for improvement [139]. In gaming, it enhances the gameplay experience by adapting the game environment based on the location where players are looking, making interactions more intuitive and immersive [140]. In training and education, eye-tracking analyses the attention and the focus of the trainees, helping to optimise training programs and educational content [141]. Psychological research benefits from eye-tracking in VR by studying visual attention, cognitive load, and emotional responses in controlled virtual environments. Additionally, medical applications use eye-tracking for diagnosing and treating visual and neurological disorders by analysing eye movement patterns within simulated scenarios [142].

To answer those various use-cases, eye-tracking provides a range of metrics that can be used to analyse visual attention and cognitive processes in VR. Among them, we can mention:

- **Fixation:** refers to the periods when the gaze is held steady on a particular object or area, indicating the level of interest or importance of that object;
- **Saccades:** rapid movements of the eyes between fixations, reflecting the scanning patterns and visual search behavior of the user;
- **Pursuit:** a tracking measure evaluating the smooth following of moving objects, providing insights into the user's ability to track motions;
- **Zones of interest (ZOIs):** specific areas within the virtual environment that are predefined for detailed analysis, such as areas where critical interactions occur. The duration and frequency of fixations within these zones can reveal user engagement and the effectiveness of the VR design.

In our study, we primarily focus on fixation metrics to evaluate on one hand the user attention and engagement, and on the other, the zones of interest (and improvement) of our VR applications. We use the HTC Vive Pro Eye headset, which integrates a 120Hz Tobii eye-tracker (see Fig. 4.3), allowing for precise measurement of where and for how long users focus within the virtual environment. The software used to capture eye-tracking data is Unity, with the integration of the SRanipal SDK¹, which provides comprehensive eye-tracking.

Integrating eye-tracking in our VR-SOLUS framework enables us to capture detailed and accurate eye movement data, providing valuable intrinsic measures into user visual behavior and interactions in VR.

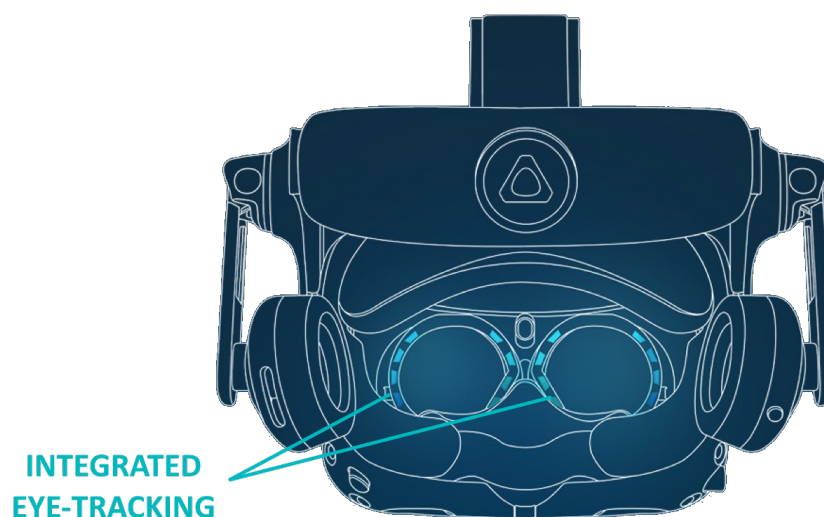


Figure 4.3: HTC Vive Pro Eye with Integrated Tobii 120Hz Eye-tracker

¹SRanipal SDK for HTC Vive Eye-tracking <https://developer.vive.com/resources/vive-sense/eye-and-facial-tracking-sdk/tutorials/sranipal-getting-started-steps/?site=kr>

4.3 Electroencephalogram

Electroencephalography (EEG) is a non-invasive method used to record electrical activity of the brain. This technique involves placing electrodes on the scalp to detect and measure the electrical signals produced by neural activity. EEG is widely used in clinical and research settings to study brain function, diagnose neurological disorders, and monitor cognitive states. The recorded data provide insights into brain activities such as attention, perception, motor control and cognitive load. Integrating EEG with Virtual Reality (VR) allows for exploration of brain-behaviour relationship in immersive multimodal environments. Traditional EEG setups, just like traditional eye-tracking setups, often require subjects to remain relatively stationary. Combining VR and EEG allows for the study of brain activity during dynamic and interactive experiences, reinforced with hardware integration of EEG directly into VR headsets. This combination enables researchers to investigate how the brain responds to complex, realistic stimuli and how cognitive processes unfold in real-time within virtual environments. A major drawback comes from the added motion artefacts that need to be removed with signal processing. This setback does not stop EEG to be used in various VR applications such as neurofeedback training [143], cognitive and motor rehabilitation [144], and user experience enhancement in gaming and simulations [145].

EEG provides numerous data that can be analysed to extract diverse markers indicative of different cognitive and physiological states. Some key markers are often distributed in three study categories:

- **Brainwave Study:** electric signals in the brain range in certain frequency bands [146]. Some of these bands have been associated with different cognitive states, see Fig. 4.4. By studying the power of such bands, users' cognitive status can be precisely evaluated. Among them, the most commonly studied are:
 - Gamma Waves (30-100 Hz): the fastest of the brainwave frequencies and associated with higher mental activity, including perception, problem-solving, and consciousness;
 - Beta Waves (12-30 Hz): linked to active thinking, problem-solving, and concentration. Increased beta activity can signal heightened alertness and cognitive engagement;
 - Alpha Waves (8-12 Hz): often associated with relaxed, wakeful states and decreased cognitive load. Changes in alpha wave activity can indicate levels of relaxation and attention;
 - Theta Waves (4-8 Hz): associated with drowsiness, meditation, and light sleep. Theta activity can provide insights into cognitive workload and memory processes;
 - Delta Waves (0.5-4 Hz): predominantly observed during deep sleep but also can indicate high cognitive load when awake.

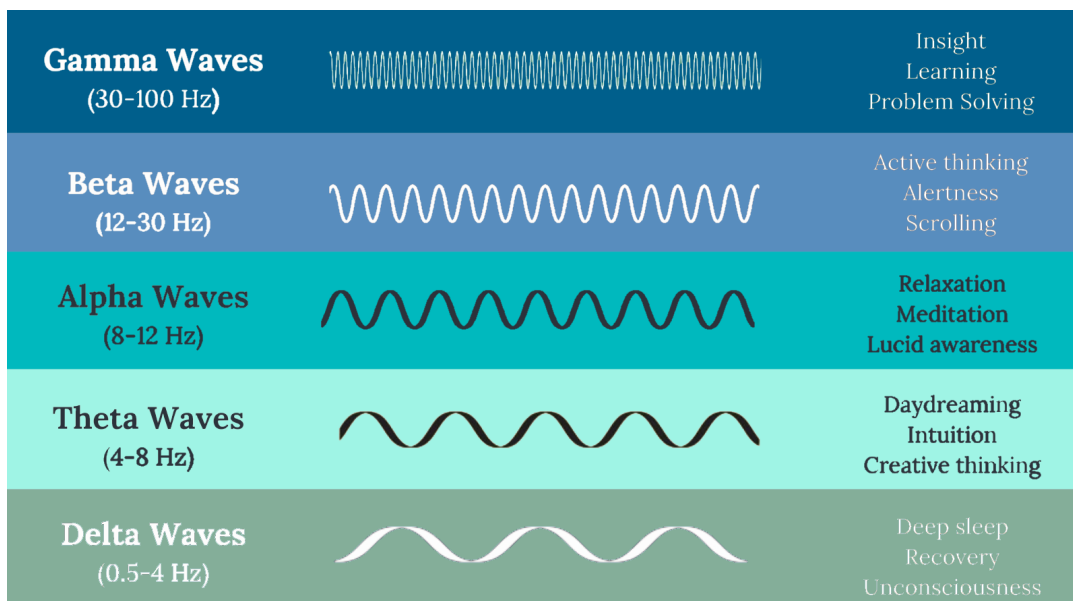


Figure 4.4: Brainwave Spectrum and their Associated Cognitive States

- **Event-Related Potentials (ERPs) Study:** examine the brain's electrical response to specific sensory, cognitive, or motor events [147]. ERPs are time-locked to these events (e.g., the P-300 ERP, appearing 300ms after an event), providing a detailed look at how the brain processes information in real-time (Fig. 4.5). By analysing the latency and amplitude of ERP components, researchers can investigate various cognitive processes such as attention, perception, and memory. This technique helps in understanding the temporal dynamics of brain activity and is widely used in cognitive neuroscience and clinical research;

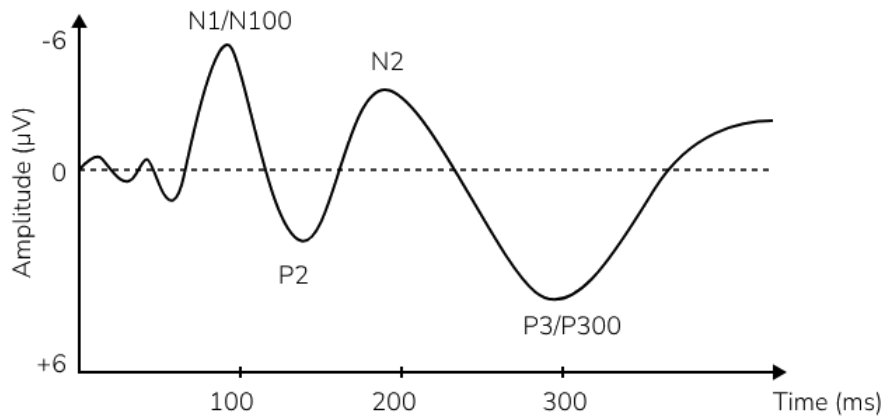


Figure 4.5: Common ERP Responses

- **Spatial Study:** focuses on identifying which regions of the brain are active during spatial tasks and the intensity of this activity [148]. Key areas involved include the parietal lobes, particularly the posterior parietal cortex, which is critical for spatial awareness and navigation. The hippocampus also plays a vital role in spatial memory and the encoding of spatial environments (Fig. 4.6). Functional neuroimaging techniques like fMRI can be used in addition to EEG to measure brain activity and power during spatial tasks, revealing patterns of neural activation and connectivity.

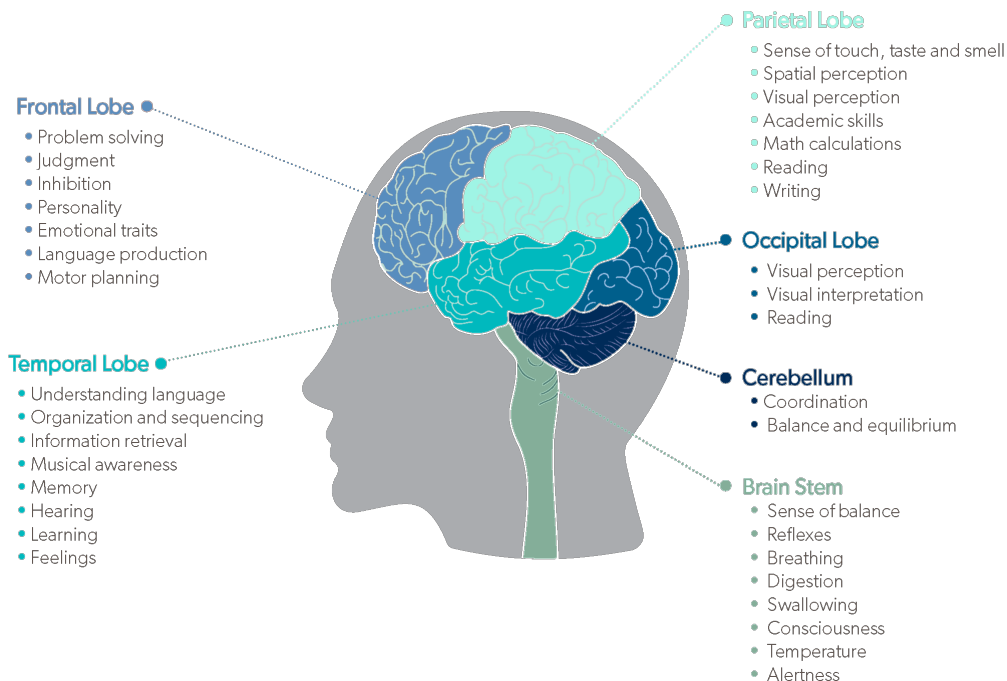


Figure 4.6: Cerebral Cortex and its Associated Functions by Lobe

In this thesis, we focus mostly on brainwave study to extract a cognitive load metrics through the integration of a EEG device inside of the VR-SOLUS framework. We use the g.tec Unicorn Hybrid Black EEG headset, an EEG cap designed for research and practical applications. This headset features 8 electrodes, recording at 250Hz, in a 10-20 system (Fz, Cz, Pz, Oz, C3, C4, PO7, PO8) positions, compatible with our VR headset, see Fig. 4.7. In terms of software, we employ a combination of Unity, the Unicorn Suite provided with the EEG cap, and OpenVibe² to record and process our brain data.

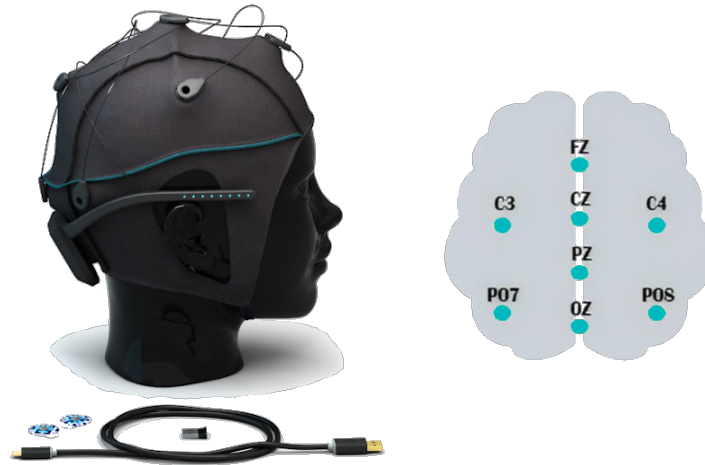


Figure 4.7: g.tec Unicorn Hybrid Black and its Electrodes Mapping

By adding EEG device to our VR-SOLUS framework, we can capture and analyse neurophysiological data effectively within the VR context, providing greater internal insight into user cognitive states, and enhancing the overall user experience.

4.4 Physical Load Sensors

Physical load sensors are essential tools for measuring the physiological load and stress experienced by individuals during various activities. These sensors provide critical data for understanding exertion levels, optimising performance, and preventing injuries. Common physical load sensors include inertial measurement units (IMUs) [149], heart rate monitors [150] and electromyography (EMG) sensors. These devices provide valuable data on movement patterns, muscle activity, and cardiovascular response.

In this thesis, we used IMUs as well as heart rate monitors. These monitors are particularly valuable in the context of physical load measurement due to their precision and applicability across different fields. IMUs consist of accelerometers and gyroscopes (and sometimes magnetometers) capturing data on linear acceleration, angular velocity, and orientation. These information helps analysing movement patterns, which in our case is used for evaluating physical load, but can also be used to identify abnormal motion - an important knowledge when working on physical rehabilitation [151]. Heart rate monitors, on the other hand, measure the number of heartbeats per minute, providing insights into cardiovascular exertion and fitness levels. These monitors are widely used in athletic training, health monitoring, and research to assess physical stress and endurance [152].

The data obtained from physical load sensors can be analysed to extract various markers indicative of physical performance and stress:

- **From IMUs:** key markers include acceleration and deceleration, which indicate movement intensity and sudden changes in motion; angular velocity, which measures the rate of rotation and is useful for analyzing joint movements and balance; and orientation angles, which provide information about body posture and alignment;
- **From heart rate monitors:** important markers include heart rate (HR), indicating cardiovascular load, and heart rate variability (HRV), reflecting autonomic nervous system activity and recovery status.

²OpenVibe software <https://openvibe.inria.fr/>

For our thesis, we focus on physical load extraction through acceleration values obtained from head and arm movements, as well as heart rate. To this end, we use the Polar OH1+ optical fitness tracker - integrating a heart rate monitor and an IMU setup - placed on the user's arm, as well as the integrated IMU inside of our Unicorn Hybrid Black EEG cap. The data collected by the Polar OH1+ is logged using the Polar Sensor Logger APK³. For the head IMUs, the accelerometers and gyroscopes values are captured alongside the EEG values inside the Unicorn Suite, as seen in the setup Fig. 4.8.

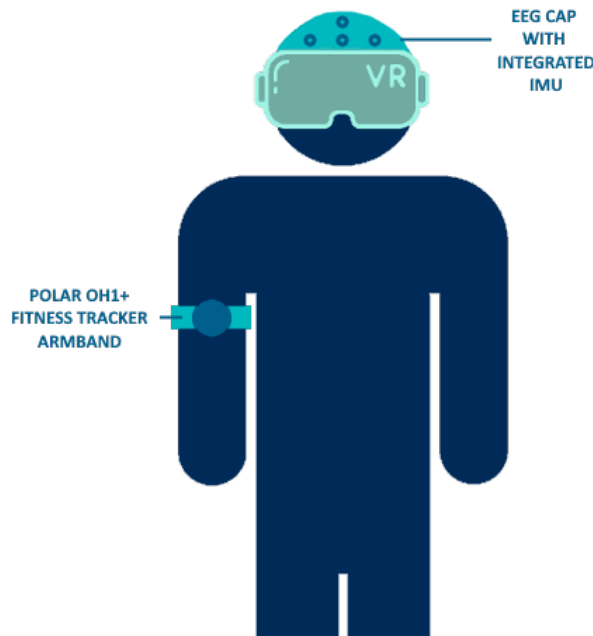


Figure 4.8: Physical Load Capture Setup

By integrating IMUs and heart rate monitors, our VR-SOLUS methodology achieves an assessment of physical load in VR environments. Not only does this provide observations into user performance and well-being, it also gives insight into medium and long-term physical usability of the VR applications and setups we design.

4.5 Implementation of Physiological Measures

In this section, we present our implementation of physiological data for VR design. We showcase their use through two earthwork training studies: a hardware selection study (introduced in Section 3.1.4 for operational data implementation) and its follow-up software optimisation study. Both studies make use of our intrinsic physiological metrics: visual interest (through eye-tracking), cognitive load (through EEG) and physical load (through IMUs and heart rate monitor).

4.5.1 Hardware Selection

Study

Presented through the scope of operational metrics in section 3.1.4, we now analyse it through the scope of physiological measures [153]. The controller selection study focused on evaluating three different types of controllers—gamepad (gaming), VR (hybrid), and joysticks (professional)—within an excavation training application. The objective was to determine the most effective hardware for this specific VR application. Thirteen participants (on average 28.1 years old, 12 males, 1 non disclosed) were involved in the study, each undergoing three rounds of experiments, corresponding to the three different controllers. During each round, participants completed four

³Polar Sensor Logger APK for the Polar OH1+ Fitness Tracker https://play.google.com/store/apps/details?id=com.j_ware.polarsensorlogger&hl=fr

exercises: a tutorial to familiarise themselves with the controller, a simple digging exercise, a navigation exercise, and a compound exercise (combining tasks from all the previous exercises). These exercises were designed to comprehensively assess the performance and usability of each controller type.

The results indicated that joysticks provided the best performance based on operational metrics, including completion time, input errors, and the ability to manage multiple inputs simultaneously. This suggests that professional-grade joysticks, despite being more complex and less commonly used outside of professional settings, offer superior control and precision for tasks requiring intricate manipulation and coordination.

This study also integrated physiological sensors, namely eye-tracking and EEG. Physical load evaluation was not integrated at that time, although head accelerometer values were used for EEG signal filtering. Users were equipped with the HTC Vive Pro Eye, recording their gaze data, and the Unicorn Hybrid Black EEG cap, recording their physical load. The VR simulation and eye-tracking capture was done through Unity. The brain activity was recorded through the Unicorn Suite. Both sets of data was synchronised through the Lab Streaming Layer⁴, then filtered and analysed through the OpenVibe software, as presented in Fig. 4.9.

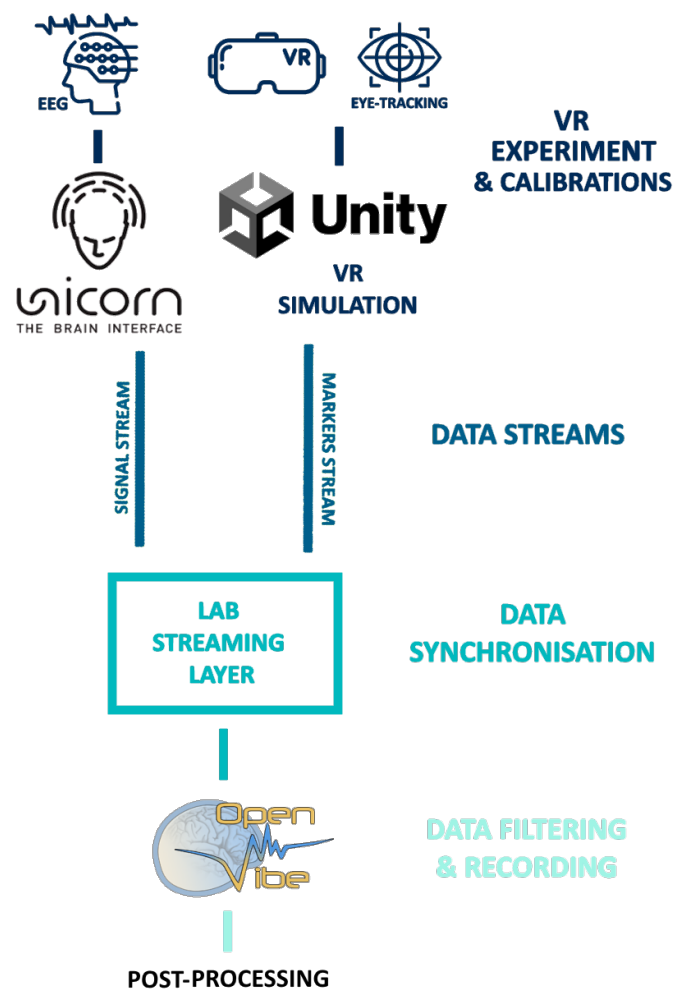


Figure 4.9: Physiological Data Pipeline v1

⁴Lab Streaming Layer for Data Synchronisation <https://labstreaminglayer.org/#/>

Physiological Results

In this study, we extracted intrinsic metrics from the eye-tracker and the EEG sensors.

Eye-tracking

From the eye-tracking, we extracted the time spent looking at interest zones. In particular, the time spent looking at the mappings of the different controllers, which is considered as both a learning metric and an ergonomics one. This is done in Unity by sending rays (aka "raycasting") from the original position of the user's eyes in the virtual environment in the direction of their gaze (given by the eye-tracker). When those rays enter in collision with one of the mappings display (when they hit the "collider" component of the mapping sprites), a collision alert is registered ("collision hit") which gives the object of the collision and allows to register its duration, as represented Fig. 4.10. Then we divide the time spent "colliding" with the mappings to obtain a percentage of time spent looking at controls, normalised inter-participants.

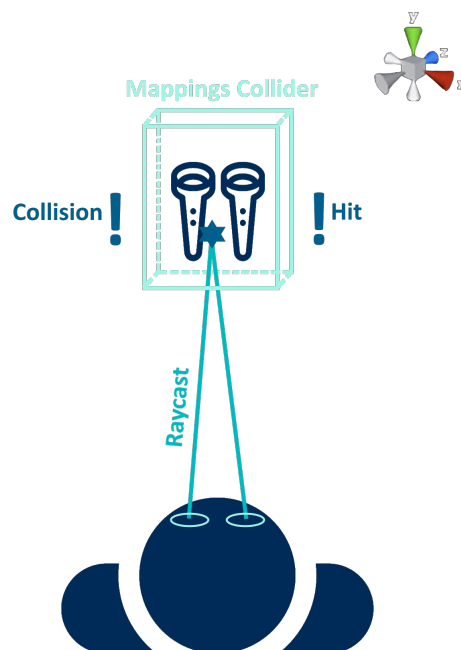


Figure 4.10: Eye-tracking Raycasting in Unity

We observe that despite being the most familiar controller among the three, the gamepad is the one that required the most visual reminders according to Fig. 4.11.

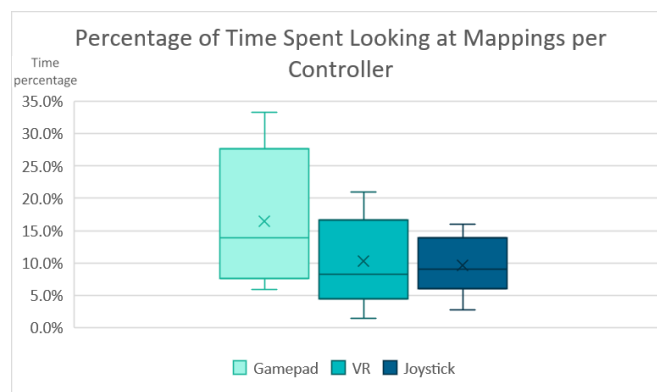


Figure 4.11: Percentage of time Looking at Mappings required by Control Usage Reminder

Electroencephalogram

We used the EEG sensor to evaluate the cognitive load based on the Theta Alpha Ratio (TAR) method. The human brain produces signals with different frequency bands, called brainwaves, each attributed to their own function. In particular, the Alpha band (8-12Hz) and the Theta band (4-8Hz) are both linked to relaxation and focus. In 1999, Klimech observed that cognitive load could be evaluated by the variation of those bands' spectral power in the Frontal and Parietal region of the brain [154]. Respectively, mental load augments when the Theta power increases in the Frontal Fz position while Alpha power diminishes in the Parietal Pz position (Fig. 4.12). By studying the ratio of those spectral bands, one can evaluate the cognitive load of a person. This TAR technique is still to this day one of the prominent cognitive load evaluation method, and we based our evaluation on it, inspired by the work of Cabañero et al. [155].



Figure 4.12: Visual Representation of the Theta-Alpha Ratio

To obtain the TAR, EEG signals requires filtering, as EEG devices record all the electrical potential at the surface of the scalp [156]. They not only detect brain activity, but also head muscles activity. Head movements, neck movements, facial movements such as talking or blinking, and so on, create motion artefacts in the EEG signals recorded that need to be removed. Movements are usually restricted in traditional EEG experiments, but VR studies rarely limit the users motions, and neither are we, hence the need for motion calibration and in-depth filtering. Both motion and cognitive load are calibrated. Motions are assessed in a head and face movement exercise, while recording cognitive load. Accelerometer values are then used for motion filtering. Pure cognitive load is assessed through two calibration exercises, a hard 2-back test (based on the standard n-back test, a stationary sequential memory test [157]) and a resting exercise (1 minute of relaxing in silence with eyes closed).

For our signal processing, we first used pass-bands filters to isolate our Alpha (8-12Hz) and Theta (4-8Hz) bands. Doing so allowed us to reject the 50Hz line noise as well as the bulk of the motion artefacts, usually situated in the Delta band (0-4Hz) [158]. We then segmented the signal in 1 second windows with an overlap of 0.5 second. In post-processing, accelerometer data was used to identify and reject windows contaminated by too much motion artefacts. Then, to account for variability between participants, TAR values were normalised using three min-max techniques:

- **Regular Normalisation:** taking the minimum $\min(\text{TAR}_{\text{ex}})$ and maximum $\max(\text{TAR}_{\text{ex}})$ TAR values of the exercises to normalise each average TAR value from the time windows, noted X, as seen in equation (4.1).
- **Motion-Hard Normalisation:** using the regular formula, but changing the inferior boundary to the maximum obtained value in the head motion calibration $\max(\text{TAR}_{\text{motion}})$, and the superior boundary to the maximum value obtained in the hard cognitive calibration $\max(\text{TAR}_{\text{hard}})$, as seen in equation (4.2). This allows us to take into account both motions values and pure cognitive values (theoretically with no motion pollution) into our normalisation, removing the motion pollution even more from the data;
- **Rest-Hard Normalisation:** using the regular formula with the inferior boundary being the lowest recorded cognitive load value during the resting calibration $\min(\text{TAR}_{\text{rest}})$, and the superior boundary being the maximum recorded cognitive load value during the high level cognitive load calibration $\max(\text{TAR}_{\text{hard}})$, as seen in equation (4.3). This method only takes into account the lowest and highest mental load values of each individual, to scale their values based on pure cognitive data.

$$RegularNormalisation(X) = \frac{X - \min(TAR_{ex})}{\max(TAR_{ex}) - \min(TAR_{ex})} \quad (4.1)$$

$$MotionHard(X) = \frac{X - \max(TAR_{motion})}{\max(TAR_{hard}) - \max(TAR_{motion})} \quad (4.2)$$

$$RestHard(X) = \frac{X - \min(TAR_{rest})}{\max(TAR_{hard}) - \min(TAR_{rest})} \quad (4.3)$$

While regular normalisation shows more variance (Fig. 4.13), it does not take into account motion or cognition range of each individual. The other normalisation techniques (Fig. 4.14 and Fig. 4.15) do consider them, but tend to squash data, while also providing better outliers identification. Interestingly, even if shown as the most efficient and easiest-to-use by the operational metrics, the joysticks appear to physiologically trigger slightly more cognitive load on average (the cross mark on each plot).

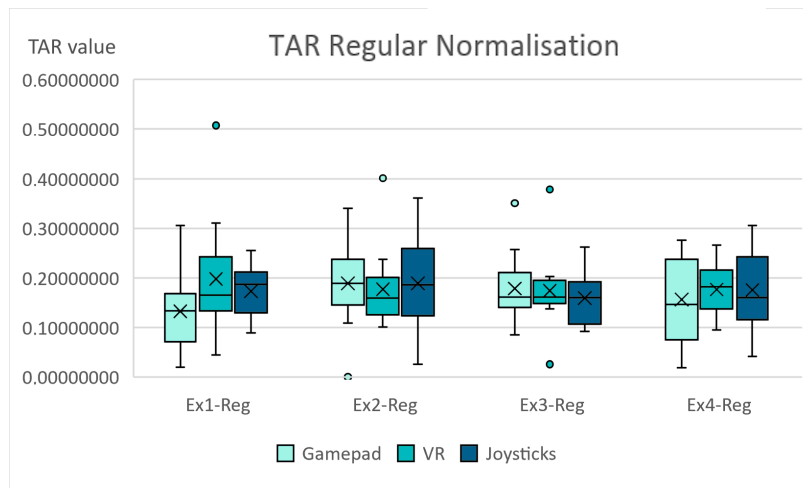


Figure 4.13: Cognitive Load with Regular Normalisation

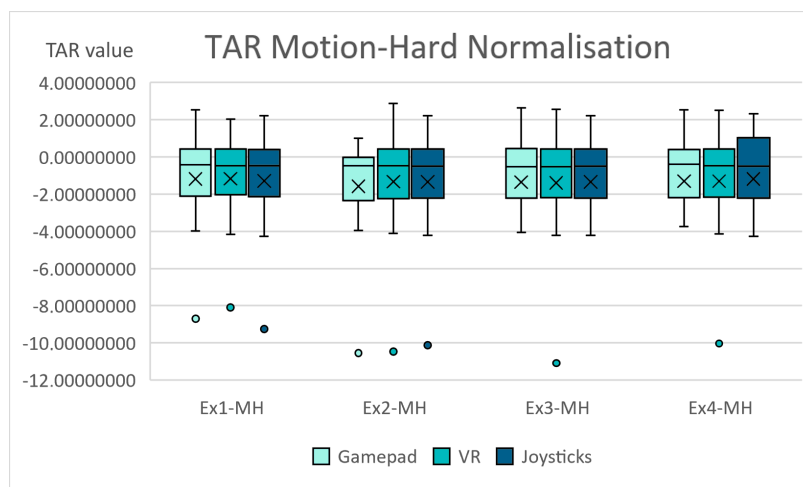


Figure 4.14: Cognitive Load with Motion-Hard Normalisation

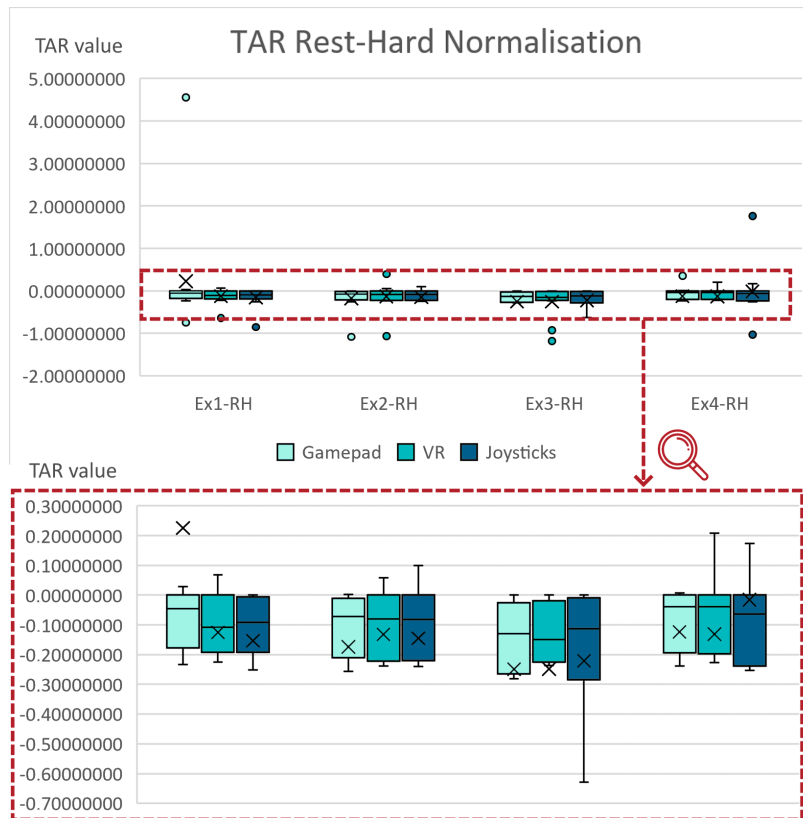


Figure 4.15: Cognitive Load with Rest-hard Normalisation

4.5.2 Software Optimisation

Study

Design and optimisation in VR can be done on two aspects: hardware and software. We presented previously our work on hardware selection, and how professional mapped joystick would match our use-case the best. In our continuing exploration of earthwork teleoperation and training applications, we now focus on software optimization. In this new study, we aimed to design and evaluate different virtual annotations (VAs) to determine their effectiveness in enhancing user experience and performance within a virtual reality environment. Virtual annotations, which include visual and audio cues, are integral to guiding and informing users during complex tasks in VR settings. The concept of virtual annotations has been extensively studied in various fields, particularly in augmented reality and VR, where they are used to provide contextual information, enhance situational awareness, and improve task performance [159]. In the context of earthwork teleoperation, effective VAs can significantly reduce cognitive load, minimize errors, and enhance the overall efficiency of operations [160].

Our study tested multiple VAs, as well as multiple display modalities, including always visible/audible annotations, those accessible via a menu, and event-based annotations triggered by specific actions or conditions. This comprehensive approach allowed us to explore how different presentation methods affect user interaction and task performance. All our VAs and their modalities are presented in Table 4.2.

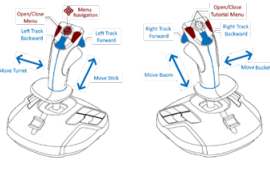
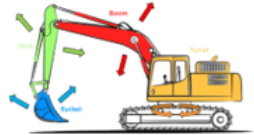
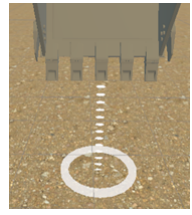
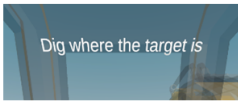


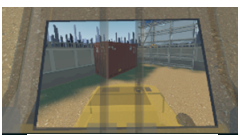



Virtual annotation	Description	Display Modalities Investigated
	Mappings	Eye-tracked / in a menu / always visible
	Tutorial	Always in a menu
	Reticle, ground projection of the bucket position	Eye-tracked / in a menu / always visible
	Cabin Instructions	Eye-tracked / in a menu / always visible
	World Instructions	Eye-tracked / in a menu / special (icon instead of text) / always visible
	Arrow to Objective	Eye-tracked / in a menu / always visible
	Rear View Camera	In a menu / special (only when going backward) / always visible
	Tracks Directions, ground projection of the tracks position and direction	In a menu / special (only when going moving) / always visible
	Engine Sound	In a menu / always audible
	Backing Sound	In a menu / special (when going backward)

Table 4.2: Virtual Annotations and their Display Modalities Tested in our Study

A total of 17 participants (on average 34.1 years old, 12 males, 5 females) were involved in the study. After completing a tutorial exercise to familiarise themselves with the VR environment (as seen in Fig. 4.16) and the joysticks control, they proceeded through three sets of exercises.



Figure 4.16: Virtual Environment Example for the VA Study

Each set involved the same tasks but featured a different configuration of VAs, summarised in Table 4.3:

- **Set A:** the first set primarily relies on visual annotations displayed at head level, providing constant feedback and guidance directly in the user's line of sight;
- **Set B:** the second set incorporated more world-integrated VAs, blending visual cues into the environment to enhance immersion and contextual relevance;
- **Set C:** the third set combined both visual and audio VAs, aiming to evaluate the impact of multi-modal feedback on user performance and perception.

Those three sets were designed empirically to fulfill two objectives: first, allow the users to test all the modalities for each VA across the sets. Second, obtain some insight on global display method - head-level visual display (Set A), world integrated display (Set B), and multimodal (visual and audio) display (Set C). The order of the sets tested was changed between each participant to avoid a learning bias.

Virtual Annotations	Set A	Set B	Set C
Mappings	Always visible	Eye-tracked	Deleted
Tutorial	In a menu	In a menu	In a menu
Reticle	In a menu	In a menu	In a menu
Cabin Instructions	Always visible	Deleted	Eye-tracked
World Instructions	In a menu	In a menu	In a menu
Arrow to Objective	Eye-tracked	In a menu	Always visible
Rear View	Always visible	In a menu	Special (only when backing up)
Tracks Directions	Special (only when moving)	Always visible	In a menu
Engine Sound	In a menu	Deleted	Always audible
Backing Sound	In a menu	Special	Special (only when backing up)

Table 4.3: Virtual Annotations Sets

Participants were evaluated using a range of physiological measures to gain insights into their cognitive and physical responses to the different VA configurations. Eye-tracking data (from the HTC Vive Pro Eye) was used to assess fixation on visual annotations, providing a direct measure of attention and engagement. EEG data (from the Unicorn Hybrid Black) was collected to monitor cognitive load, offering insights into how different VA modalities affected mental effort. Physical load measurement was added to this study and assessed using a combination of IMU data (from the head with the Unicorn Hybrid Black and the arm with the Polar OH1+) and heart rate monitoring (also from the Polar OH1+ armband), providing a global view of the participants' physical exertion during the tasks.

A key objective of this study was to investigate whether a methodology based solely on bio-signals (i.e. eye-tracking, EEG, and physical load sensors) could be effectively employed for VR design and optimisation.

Results

In this study, we extracted intrinsic metrics from eye-tracking, EEG and physical sensors (IMUs and heart rate monitors).

Eye-tracking

The eye-tracking data were used to evaluate fixations on the diverse (visual) VAs. As gaze does not equate intent (e.g. looking at something while thinking about something else), eye-tracking filtering is needed to remove involuntary fixations. According to Trevathen's theory [161], there is a duration threshold (50-150ms) under which fixations can be classified as "ambient" (i.e. no intent), and one (300-500ms) over which they can be classified as "focal" (i.e. possible intent and/or understanding). To remove all certified "unintended" fixation, we got rid of all fixations under the 300ms. While this does not guarantee that the remaining fixation are all intentional (and more importantly translate the understanding of the looked-at VA), it does remove parasitic data due to eye movements alone, giving us an exploitable metric through gaze assessment.

We obtain our attention metric by dividing the duration of the (filtered) fixations on a VA by the completion time of the exercise. Then we look at the distribution of the fixation to estimate which visual VAs are the most useful (looked at) and thus should be kept visible at all time, while the less useful (less looked at) are hidden in a menu or even deleted. We considered the modality selection as follows:

- **Q0 (0%) -> DELETE:** if the percentage of fixation is minimal, the VA can be deleted as it is not useful;
- **Q0 (0%) - Q1 (2%) -> MENU:** if the percentage of fixation is in the first quartile (in our study, between 0% and 2% of fixation time), they are not very useful, but should still be kept in the application. Therefore we put them in a menu where the user can choose to enable or disable them at will;
- **Q1 (2%) - Q2 (3%) -> EYE-TRACKED OR EVENT-BASED:** if the percentage of fixation is in-between the first quartile (2% of fixation time) and the median (3% of fixation time), the VA is deemed useful enough to be displayed "when needed" - either when the user is looking directly at them (they stay semi-transparent when not looked at and become opaque when looked at) or when an event justify their use (e.g. displaying the rearview camera when moving backward only);
- **> Q2 -> ALWAYS VISIBLE:** if the fixation time of the VA is above the median time (3% of fixation time), then it is deemed useful enough to be displayed at all time.

The tutorial VA was the only exception, placed in its own menu and displayed when pressing a button. On the aforementioned VAs, the results are as follow in Table 4.4.

Virtual Annotations	Fixation Time Percentage	Distribution	Selected Modality
Mappings	9.7%	> Q2	ALWAYS VISIBLE
Reticle	2.4%	Q1-Q2	EYE-TRACKED
Cabin Instructions	1.5%	Q0-Q1	MENU
World Instructions	0.0%	Q0	DELETED
Arrow to Objective	1.6%	Q0-Q1	MENU
Rear View	5.2%	> Q2	ALWAYS VISIBLE
Tracks Directions	3.9%	> Q2	ALWAYS VISIBLE

Table 4.4: Virtual Annotation Eye-tracking Selection

Electroencephalogram

We employ the same methodology for EEG treatment as seen in Section 4.5.1. We once again extract a cognitive load measure from the EEG data. First we filter the raw EEG data to isolate our Alpha (8-12Hz) and Theta (4-8Hz) bands, then we remove motion polluted data thanks to head accelerometers and gyroscopes data. Then, we calculate the theta-alpha ratio (TAR) for each participant. As cognitive load is not generalisable inter-participants [162], we normalise the TAR values, using a regular min-max formula (see Equation 4.1).

To assess our different regularisation techniques, we used one-Way ANOVA analysis [163] to determine whether there are statistically significant differences in the means of our data. As a result of this statistical analysis performed on the hardware selections and software optimisation studies, we decided to keep only the regular normalisation technique, and not the motion-hard or rest-hard regularisation, which performed poorly. Finally we average the normalised TAR values and compare them for each VA sets. The results are shown in Fig. 4.17.

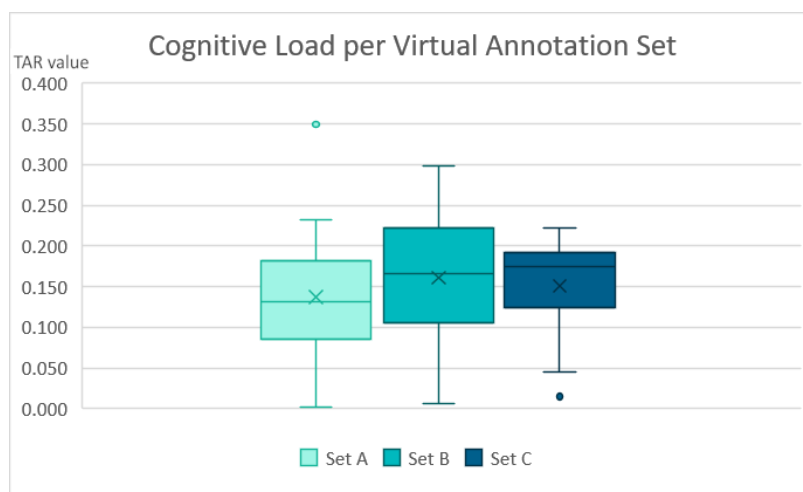


Figure 4.17: TAR Evaluation for each Set of Virtual Annotation

These results seem to indicate that head-level visual annotations (Set A) require less mental demand, probably because the users do not need to search for the information as everything is directly displayed in their immediate field of view. Another observation is that multimodal annotations (namely adding sound as in the sets B and C) generates more cognitive load. This is to be expected, but another less expected finding is that it is better to have all sounds (constant engine sound and backward beeping sound) as in Set C than some punctual sounds (just the backing sound) as in Set B, as the average cognitive load is on average lower in Set C than Set B. This might reflect the effect of better audio immersion on lessening the cognitive load of the user [164]. Through the cognitive load metric, we confirm the eye-tracking measure on always displaying some annotations or not, especially those at head level. Another design choice extracted from this measure is to keep both audio elements (constant engine sound and punctual backing sound) for better immersion and lesser mental load.

Physical Load Sensors

We use two types of sensors for evaluating physical load: IMUs (placed on the head and arm) and heart rate monitors.

From IMUs, we extract mainly the acceleration values from head and arm movements, from all three 3D axes (X, Y, Z), in absolute values, to obtain a compound "motion value", in milli-gravity (mG), in 0.5 seconds windows (t).

$$MotionValue(t) = |Xacceleration(t)| + |Yacceleration(t)| + |Zacceleration(t)| \quad (4.4)$$

We then get the average values for head and arms for the whole set and compare them inter-participants, see Fig. 4.18.

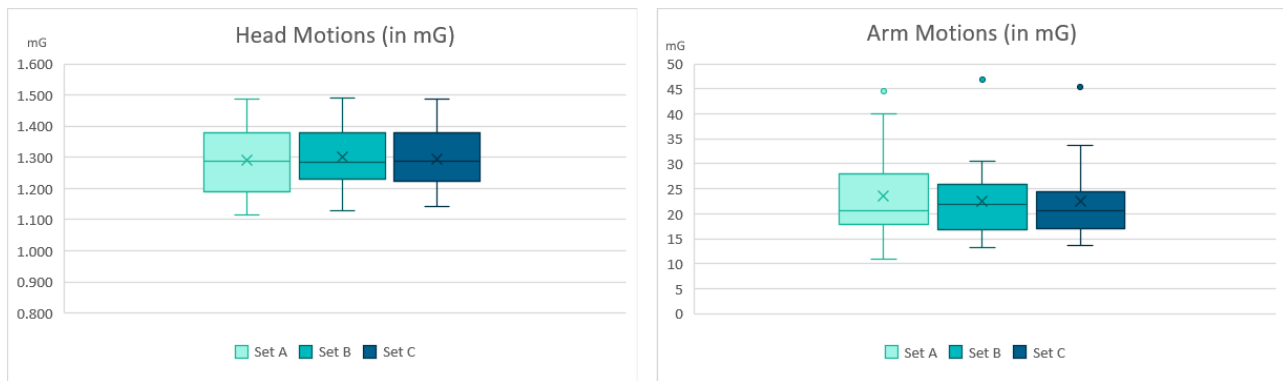


Figure 4.18: Head and Arm Motions (in mG)

As the experiment is spent seated, arms resting on a table, the differences between each set are quite small. Logically, head motions are slightly lesser in Set A, were annotations are displayed at head level (requiring less head movement to find the information). In return, arms motions are slightly higher in Set A, most annotations are displayed at eye-level, eliminating the need to search for them (less head movements), but the in-world objectives are not indicated, requiring users to move the vehicle more frequently (more arm movements) to see and reach them.

From the heart rate monitor (the Polar OH1+ armband), we extract the average heart rate per set, in beats per minute (BPM), as seen in Fig. 4.19.

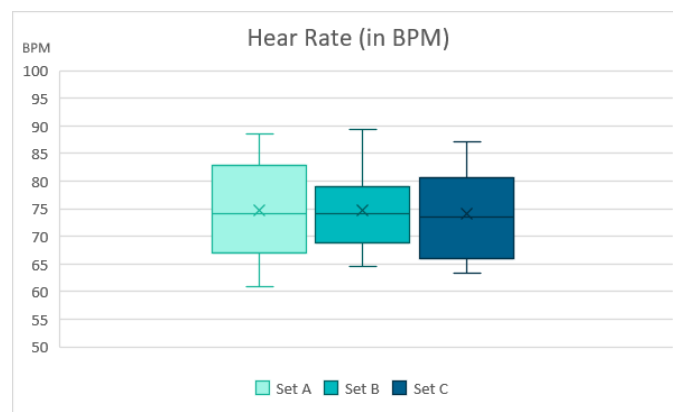


Figure 4.19: Average Heart Rate per Set (in BPM)

Once again, due to the low physical intensity of the experiment, there is little divergence in heart rate between sets, especially when looking at the average heart rate. Heart rate can also be a stress indicator [165], but the low divergence does not yield significant results on that front either.

While physical load is an important factor in VR design, especially in medium and long-term usage, our experimental setting did not permit an in-depth evaluation of this metric. However, we do think that it is worthy to keep in our VR-SOLUS framework, as we want a generic methodology that can be applied to various use-cases, including the more physically demanding ones. As we can see, the use-case/experimental setup can be a limitation to physiological measure. Another challenge that appeared for every biosensors is the importance of advanced signal processing in the analysis of each set of data.

4.6 Challenges of Signal Processing

The processing of physiological signals in virtual reality (VR) environments presents several challenges due to the multimodal nature of the sensors involved. Each type of sensor — whether it is eye-tracking, electroencephalogram (EEG), or inertial measurement units (IMU) — captures data at different frequencies and in distinct modalities, requiring a specific approach to signal treatment for each sensor.

Before delving into the signal treatment necessity for each sensor, multimodal physiological data approaches face a first major challenge: the need for precise data synchronisation. Given that each sensor operates at its own sampling rate, integrating the data streams into a cohesive analysis framework requires meticulous alignment [166]. This is traditionally done through hardware, but can also be done through software management, which is the solution we chose. In our VR-SOLUS pipeline, we integrate eye-tracking data with a 120Hz frequency, EEG data with a 250Hz rate and IMUs recording at 250Hz (for the head) and 100Hz (for the arm). Without proper synchronisation, any analysis that combines these data sources would suffer from temporal misalignment, leading to inaccurate or misleading interpretations (e.g. a cognitive reaction to a wrong visual cue if the data timing is not aligned).

To make sure our VR-SOLUS pipeline is synchronised, we make use of different software for recording, filtering and processing for our different sensors:

- **Eye-tracking:** recorded in Unity through the SRanipal SDK⁵. While Unity exploits the eye-tracking 120Hz rate, it registers only the punctual fixations events happening when the users look at predefined zones of interest (ZOI). Only those punctual events are sent to the rest of the pipeline and synchronised with the other signals;
- **EEG:** recorded through the Unicorn Suite. They provide a continuous flow of data which is considered the main data stream, while the others are discrete event streams;
- **IMU - head:** the head motions are recorded through the same hardware (EEG cap) and software (Unicorn Suite) as the EEG, and therefore are integrated into the main data stream;
- **IMU - arm and Heart Rate Monitor:** recorded separately on Android through the Polar Sensor Logger. Due to the separate recording hardware, the data are synchronised after all the other signals through timestamp alignment in Python. In our studies, due to the rarity and low intensity of the physical events, we rather use an averaged, independent physical load measure than a synchronised one.

To summarise, our data is divided into two streams: the main data stream (continuous flow of EEG and hand IMU data) and the event data stream (discrete timestamps marking important events of eye-tracking and simulation events such as beginning and end of certain tasks). These flows are synchronised through the Lab Streaming Layer⁶, then treated through OpenVibe⁷ and analysed through Python, as shown in Fig. 4.20.

To achieve data synchronisation and coherent analysis, independent signal treatment is also needed for each sensor.

⁵SRanipal SDK for HTC Vive Eye-tracking <https://developer.vive.com/resources/vive-sense/eye-and-facial-tracking-sdk/tutorials/sranipal-getting-started-steps/?site=kr>

⁶Lab Streaming Layer for Data Synchronisation <https://labstreaminglayer.org/#/>

⁷OpenVibe software <https://openvibe.inria.fr/>

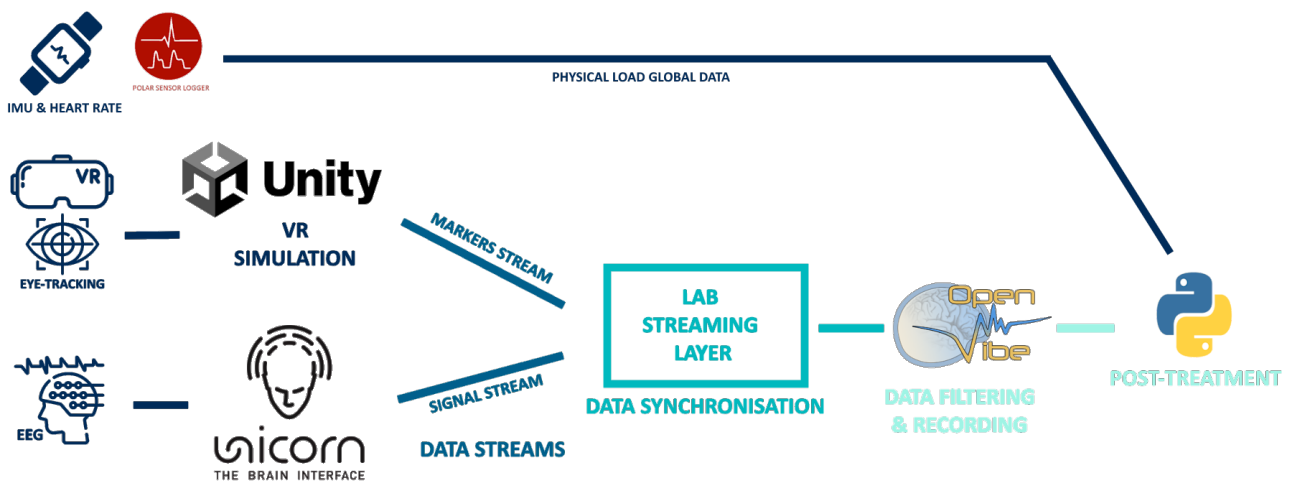


Figure 4.20: VR-SOLUS - Physiological Data Pipeline v2

4.6.1 Eye-tracking and Intent-tracking

Challenges

Eye-tracking technology in VR offers profound insights into user behavior, interaction patterns, and cognitive processes. However, it also presents a set of unique challenges that must be carefully managed to ensure accurate and meaningful data collection. The first and more prominent one is the "Midas Touch" problem [167]. When it comes to gaze, what you see is what you get, and eye-trackers get everything. Every movement is recorded, either voluntary, involuntary, with intent or not, real or fake fixation, understanding of what is seen or not and so on. In VR environments, this problem is even more notable, as the user's gaze is often used as an input method to interact with virtual objects or interfaces [168]. As the eyes naturally move constantly, even without the user's intention to interact with the environment, this can lead to unintended selections or actions, as the system might interpret a casual glance as an intentional command. In our study, as we try to detect the fixations of the users' gaze on specific object, the Midas Touch effect can lead to wrong detection, as illustrated in Fig. 4.21.

Proposed Solution

In this thesis, the main eye-tracking metric is fixation. Due to the Midas Touch effect, another challenge is differentiating between voluntary and involuntary fixation. According to Trevathen's theory [161], fixations can be divided into "ambient" and "focal" fixations, as evoked earlier. Ambient fixation occurs when the user is broadly scanning their environment without focusing on any specific object - lasting between 50ms and 150ms, while focal fixation involves concentrated attention on a particular item - lasting at least between 300ms and 500ms. In VR, both types of fixation are common, but they carry different implications for user intent and engagement. For example, focal fixations can indicate interest or cognitive load, while ambient fixations may suggest exploration or searching behavior. To accurately distinguish between these types of fixations is crucial for our VR-SOLUS method, as we want to extract relevant fixations on our ZOI. Which is why we filter and remove any fixation below 300ms: we make sure to remove all ambient (involuntary) fixations, and keep only those that can translate intent and understanding.

Another element that needs to be taken into account when using eye-tracking is peripheral vision. It also plays a significant role in how users experience VR, but it is often challenging to account for in eye-tracking studies. Eye-trackers typically focus on where the central point of vision is directed, potentially overlooking important stimuli detected in the periphery. If one were to enlarge this central point zone of detection, more peripheral information could be detected, but with a price of elevated number of false-positives. Understanding and incorporating the role of peripheral vision into the design and analysis of VR experiences is essential but technically challenging. We do not explore it in this thesis but encourage its inclusion in further research.

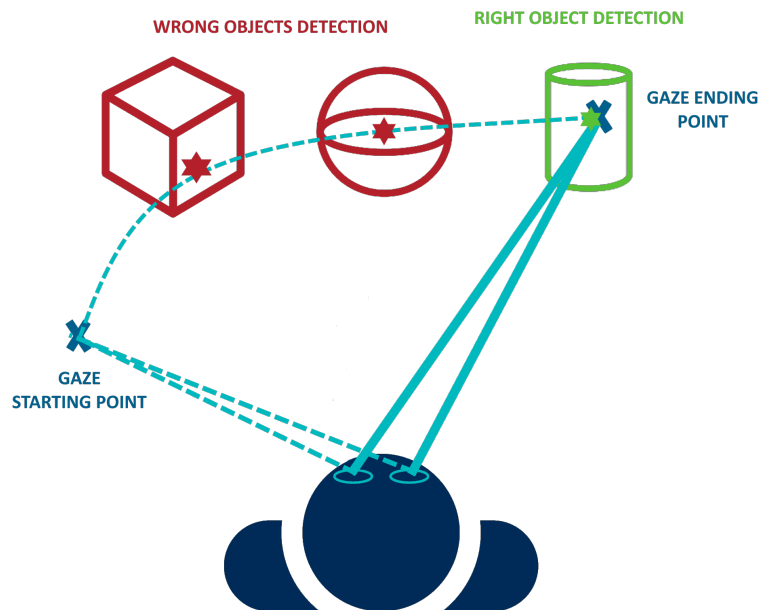


Figure 4.21: An Example of the Midas Touch Effect Issue in VR

4.6.2 Electroencephalogram and Artefacts

Challenges

EEG is a powerful tool for measuring brain activity and understanding cognitive and emotional states. However, even in traditional experimental settings, EEG data can be challenging to work with due to its sensitivity to various forms of noise and interference, forming polluting artefacts in the recorded signal. When applied in a VR environment, these challenges are exacerbated, and new obstacles emerge that complicate further data collection and analysis.

In conventional EEG experiments, participants are typically seated in a controlled environment, where movements are minimised to reduce the introduction of noise into the EEG signals. As EEG records all electrical signals at the surface of the scalp, head and face motions are also recorded. Therefore participants are often instructed to avoid talking, moving, or even blinking excessively, as these actions can generate artefacts that obscure the true brain activity being measured. Additionally, EEG experiments are usually conducted in specialised laboratories designed to minimize electromagnetic interference from surrounding electronic devices. Despite these precautions, traditional EEG recordings still face challenges such as:

- **Muscle Artefacts:** movements of the head, neck, and face muscles can produce electrical signals that contaminate the EEG data [169]. Even slight movements, such as swallowing or clenching the jaw, can generate significant noise;
- **Eye Movements:** blinks and saccades (rapid eye movements) create electrical potentials that can overshadow the brain signals of interest, requiring careful filtering and artifact correction techniques [170];
- **Environmental Interference:** electromagnetic noise from electronic devices, lights, or other equipment in the lab can introduce unwanted signals into the EEG data [171], necessitating the use of shielding or specialized equipment to mitigate these effects.

When EEG is combined with VR, the challenges inherent in traditional EEG experiments are amplified, and new ones arise due to the dynamic and immersive nature of VR, such as:

- **Motion Artefacts:** in VR, participants are encouraged to move their heads and bodies as they interact with the virtual environment. These movements introduce substantial motion artefacts into the EEG data, complicating

the identification of genuine brain activity [172]. Headset motion, especially rapid or jerky movements, can produce significant noise in the EEG recordings;

- **Headset Friction and Pressure:** the physical interaction between the EEG cap and the VR headset can create friction and pressure on the scalp, leading to signal disturbances [173]. The weight and fit of the VR headset can also shift the electrodes, causing inconsistent contact with the scalp and resulting in intermittent signal quality, as well as discomfort for the user;
- **Electronic Interference:** VR systems involve multiple electronic components, including displays, sensors, and wireless communication devices, which can generate electromagnetic interference [174]. This interference can manifest as high-frequency noise or low-frequency drifts in the EEG data, obscuring the underlying brain signals.

To manage these challenges, researchers have developed and implemented various strategies. A lot of them rely on extensive advanced signal processing. To remove artefacts, algorithm using independent component analysis (ICA) [175] or wavelet transformation [176] can help identifying and removing noise generated by muscle movements, eye blinks, and other sources. Specific motion artefacts algorithms are also often employed [177] to differentiate between brain signals and artifacts caused by head and body movements. These algorithms often incorporate data from additional sensors, such as accelerometers and gyroscopes, to help filter out movement-related noise. It is rather common actually to combine multiple physiological measures with EEG for signal treatment: IMUs for motion artefacts [178], eye-tracking for eye movements (e.g. blinking or saccadic movements) artefacts [179], as well as other sensors and brain imagery like functional magnetic resonance imaging (fMRI) [180]. Another approach to improve EEG data quality is through hardware itself. To mitigate electromagnetic interference, shielding material can be used around the EEG electrodes [181]. To avoid headset friction, some manufacturers now directly integrate EEG inside of their headset, which improve not only the EEG signal itself, but also the VR user experience by integrating neurofeedback.

Proposed Solution

In our studies, we extract a cognitive load metric from our EEG cap. We do this using non integrated electrodes (on the hardware) and follow the theta/alpha ratio (TAR) technique for analysis (on the software side). We set up our EEG pipeline based mainly on the work of Cabañero et al. [155], because of their similar work on ambulatory EEG analysis. We focused our signal treatment on motion artefacts which were the more prominent ones in our signal. We faced many of the challenges above in our research, and chose to mitigate them through the following steps:

1. **Signal Capture:** first of all, the EEG are recorded through our 8 channel EEG cap (Unicorn Hybrid Black), with a 10-20 mapping. This cap is also equipped with IMU (namely gyroscopes and accelerometers) which values we also record at the same 250Hz frequency. To mitigate friction with the headset and improve user comfort, we used padding between the EEG cap and the VR headset;
2. **Synchronisation:** the EEG (and IMU) data stream is synchronised through the Lab Streaming Layer with the VR simulation markers stream, containing timestamps of the different events (e.g. tasks, exercise, eye-tracking events) of the experiment;
3. **Brainwave Isolation:** in OpenVibe, we record the signals coming from our two streams. From the data stream, we isolate our frontal (Fz) and parietal (Pz) channels, where we respectively study the theta and alpha bands. We apply pass-band filters to keep only the theta (4-7Hz) and alpha (8-12Hz) bands. By doing so, we also remove the bulk of motion noise, usually present in the 0-4Hz delta band [158] and the 50Hz line noise;
4. **Signal Segmentation:** the filtered values are segmented in 1 second windows, with an overlap of 0.5 second, creating a series of exploitable dataframes;
5. **Theta/Alpha Ration (TAR) Calculation:** for each dataframe, we divide the obtained theta power by the alpha power to obtain our TAR values;
6. **Artefacts removal:** done with internal IMU exploitation. We segmented our accelerometer values the same way we did the EEG. Thanks to motion calibration exercises done in the VR experiment, we extracted for each participant a motion threshold above which we considered the EEG data "too polluted" by motions to be

exploitable - a technique based on the amplitude approach [182] - and removed those data from our analysis. This method makes use of individual motion data for each participant;

7. **Normalisation:** TAR values cannot directly be compared inter-participants as they are not generalisable [162]. Therefore we normalise those data beforehand by using a min-max scale per exercise. We also evaluated other regularisation techniques using calibrated values of low and high cognitive load and motions (to take them into account in the analysis), but their statistical significance was lesser than regular normalisation;
8. **Evaluation:** for each task/exercise/phase we want to compare between participants, we extract an average TAR value thanks to our events timestamps, obtained from the event stream from the VR application. We then confront those values inter-participants for our different setups (controllers, virtual annotations, etc.) to study the cognitive load impact on each participant.

This EEG pipeline, integrated into our VR-SOLUS framework, is represented in Fig. 4.22.

4.6.3 Accelerometers and Heart Rate

Physical sensors, such as IMUs and heart rate monitors, are commonly used in VR environments to capture data related to a user's physical state and movements. These sensors provide valuable insights that can be used to filter and enhance other physiological data, such as EEG and eye-tracking, by accounting for physical artifacts. However, when it comes to assessing physical exertion or load in VR, the application of these sensors for physical load evaluation can be limited by the experimental use-cases.

In many VR studies, physical sensors are employed mainly to support the interpretation of other physiological data. For example, IMUs can detect head or body movements that might introduce noise into EEG signals, allowing researchers to filter out these artifacts more effectively [178]. Similarly, heart rate monitors can help distinguish between changes in heart rate due to emotional stress versus physical activity [165].

Given the relatively sedentary nature of most VR experiences, including ours, the data from physical sensors serves a supplementary role rather than being the primary focus. The controlled, virtual nature of the tasks in many VR studies means that physical exertion is minimal, and the data from these sensors is not as critical for evaluating user performance or experience as it might be in more physically demanding contexts. In our proposed VR-SOLUS framework, we want to keep physical load metric, and for this use both head and arm motion sensors, as well as heart rate monitor. While the physical load extracted in our studies had little variations due to the nature of our research, we encourage the use of such sensors, for both physical load and stress evaluation. We also suggest the use of EMG for more in-depth and direct muscle activity measures.

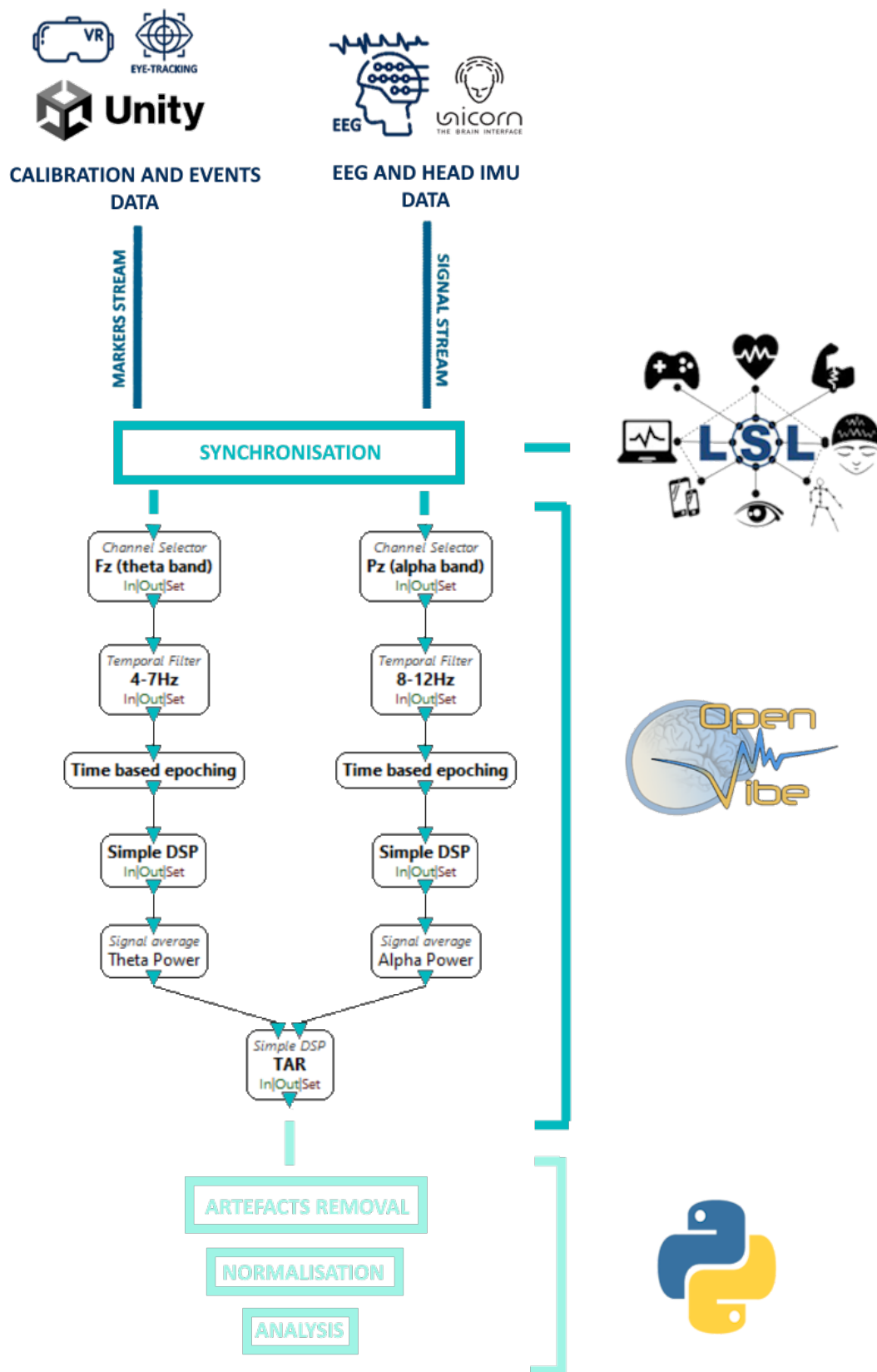


Figure 4.22: EEG Pipeline - From Capture to Analysis Software Pipeline using Unity, Unicorn Suite, LSL, OpenVibe and Python

4.7 Evaluation of Intrinsic Physiological Measures

In this chapter, we showed the contribution of physiological measures integration into VR design and optimisation. Physiological measures provide intrinsic insights that are highly valuable and serve as a crucial complement to operational metrics in the study of VR systems and user experience. While operational data can offer a clear picture of user performance, physiological measures reveal the underlying state of the user, such as cognitive load, attention, and physical engagement. This focus on the user's internal experience introduces a layer of analysis that is otherwise inaccessible through purely operational metrics, highlighting the complementary nature of these two types of data.

Physiological measures have distinct advantages. They provide objective, quantitative data that can reflect the user's state in a detailed and nuanced manner. This data can be particularly valuable for optimising VR experiences to better match user needs and capabilities, on the short, medium and long term. However, these benefits come with certain drawbacks. The analysis of physiological data often requires advanced techniques and careful interpretation. Additionally, the effectiveness of different sensors can vary depending on the specific use-case (e.g. IMU or heart rate monitors for physical load for seated, low movement experiments), with some sensors performing better or worse depending on the environment, task, or user. By integrating physiological data to our VR-SOLUS framework, we obtained relevant intrinsic data, that complement well with our other objective metrics, the operational measures. Interestingly, when confronted to user feedback (also present in each experiment of this chapter), besides the fact that both showcase intrinsic measure, objective and subjective results are not always the same. When answering the fundamental questions of each study, the results, while pointing to the same direction, showed more nuances, and even sometimes contradictions, than when using operational data only:

- **Which controller for earthwork excavation training ?** Operational data and user feedback placed joysticks on top of the competition with a large margin. While eye-tracking showed similar results - the joysticks needed less visual reminder of their mappings, making them more intuitive, cognitive load evaluation showed that despite being the preferred controllers, the joysticks were the more mentally demanding. Not only the bio-sensors brought more depth to the controllers evaluation, they even showed a contradiction between the perceived mental load (rated the lowest by the users) and the measured mental load (being on average the highest), linking physiological measures with initial users' skill level rather than perception;
- **What is the best VR interface for earthwork teleoperation ?** When it comes to rating virtual annotations (VAs) and their modalities, operational data could only provide performance metrics for a whole set of VAs, and not the individual ones. The results in terms of time, errors and multiple inputs were quite divided between sets and did not reveal much on individual VAs. On the other hand, physiological data were much more targeted: eye-tracking allowed for precise evaluation of the visual VAs, cognitive load allowed the assessment of audio VAs and physical load sensors showed that naturally, head-level display requires less head movement and thus less physical exertion. Users also assessed the VAs on their usefulness ("is this VA useful?") and their Affinity ("do you like this VA?") and interestingly, while some results matched the physiological measures, some VAs were deemed "not very useful" while also being "used a lot" (e.g. looked at for visual VAs), according to the bio-sensors. Another singular observation was how sounds were deemed "not useful" by the users, but most of them "liked them", saying it helps with immersion, a perception that is translated with a matching lower cognitive load.

These findings raise the question of correlation of data. Physiological and operational data are both objective but stem from different perspectives — one being rooted in the user's internal state (intrinsic) and the other in observable performance (extrinsic). On the other hand, while both physiological data and user feedback are intrinsic data, they differ in their respective objectivity and subjectivity. Understanding the interconnections between these three sources of data — physiological, operational, and user feedback — and how to exploit them, forms the basis of our VR-SOLUS method, as well as the layout of our next chapter, where we delve into the correlations and synergies between them.

To conclude on physiological measures, while they bring a new depth to the understanding of VR experiences, they also demand careful analysis and thoughtful integration with other types of data. Their intrinsic value lies in their ability to offer insights into the users' state, making them an essential tool in the broader toolkit of the VR-SOLUS method for VR design and optimisation.

Chapter 5

Hybrid Approach and Correlations Analysis

So far we have presented our use of objective measures - namely the operational and physiological metrics, their advantage and associated challenges (Fig. 5.1). But our pipeline would not be complete without its third metric, the subjective but no less informative user feedback.

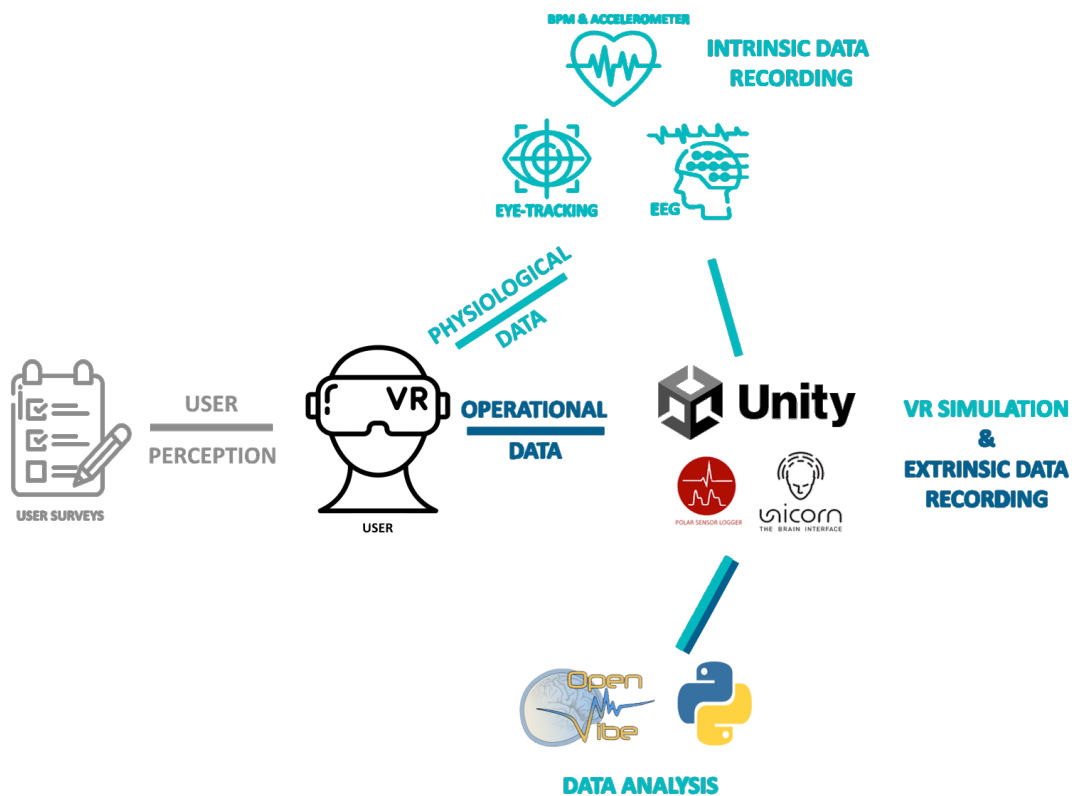


Figure 5.1: Objective Metrics Framework Highlighted into the Global Analysis Approach Proposed by VR-SOLUS

In the following sections, we explore the integration of user subjective data through survey and feedback, their synergy and complementarity with objective data, as well as their correlation, which makes the strength of the proposed VR-SOLUS method.

5.1 VR-SOLUS Hybrid Pipeline

Our attempt to understand user interactions within VR environments consisted until now to integrate both operational and physiological metrics, providing a comprehensive framework for evaluating VR interfaces.

Operational measures are extrinsic metrics, derived directly from the user's interactions within the VR system. These metrics include in our work:

- **Completion Time:** this measures how long it takes a user to finish a specific task within the VR environment. As a direct performance indicator, completion time provides insights into the efficiency and ease of use of the interface;
- **Input Errors:** this metric tracks inaccuracies in user inputs, such as incorrect button use or misaligned movements. Input errors serve as a measure of the interface's intuitiveness and the user's mastery over the control mechanisms;
- **Multi-Input Usage Percentage:** this examines the extent to which users employ multiple inputs simultaneously, which can reflect their comfort level and familiarity with the VR system. A higher percentage may indicate more advanced control and effective interaction;
- **Use-case Specific Metrics:** specific use-cases require specific metrics that measure the adequacy between the user and the task at hand.

On the other hand, physiological measures, our intrinsic metrics offer a window into the user's internal states, capturing data that reflects cognitive and physical responses during VR tasks. We included:

- **Eye-Tracking Fixation:** eye-tracking technology records where users focus their gaze and for how long. This information can highlight areas of interest or difficulty within the VR environment, providing valuable feedback on the design's effectiveness and usability;
- **Electroencephalogram (EEG) Cognitive Load:** using EEG data, the cognitive load is evaluated through the assessment of the Theta Alpha Ratio (TAR). This metric helps gauge the mental effort required to complete tasks, offering insights into the user's cognitive state and how demanding they find the VR experience;
- **Inertial Measurement Units (IMUs) and Heart Rate Monitors for Physical Load:** IMUs and heart rate monitors track physical activity and exertion. These metrics assess the physical demands placed on the user, helping to understand how the VR experience affects the body and whether it aligns with the intended physical load.

The integration of both extrinsic (operational) and intrinsic (physiological) metrics offers a dual perspective on user interaction. Operational measures provide clear, objective measures of task performance, helping to assess the efficiency, accuracy, and intuitiveness of the VR interface, making them reliable performance indicators. Physiological metrics offer deeper insights into the user's cognitive and physical states, revealing how the VR environment affects their mental load, attention, and physical effort. This data can identify areas where the user may be struggling or excelling, which might not be evident from performance metrics alone. By synthesising both, the VR-SOLUS method allows for broader understanding of user experience and thus better guidelines for VR design and optimisation. This is not without challenges, such as the synchronisation of multimodal data and the advanced signal treatment and analysis.

However, a key question arises: can the VR-SOLUS method be used with only its objective measures? Can objective metrics fully replace user feedback, or is there still a crucial role for traditional subjective input in the design and optimisation of VR experiences? While objective data provides quantifiable insights into performance and user state, user feedback offers personal perspectives that can capture nuances and subjective experiences that objective metrics alone may miss. In the following section, we interrogate this aspect via a test implemented in a software design study. The results of this study, detailed in the following sections, suggest that both objective and subjective data are necessary for optimal results, and that the VR-SOLUS framework would not be complete without user feedback. While operational and physiological metrics can objectively guide design choices and reveal underlying issues, user feedback remains invaluable for understanding personal experiences, preferences, and satisfaction, which are essential for creating truly immersive and user-friendly VR environments.

5.2 User Feedback vs Data-driven Method

In the field of VR design and optimisation, two main approaches are often employed: traditional user feedback methods and novel data-driven techniques. Each has its own set of advantages and drawbacks. Traditional user feedback, typically gathered through surveys and interviews, provides direct insights into user preferences and experiences, but can be subjective and limited in scope. On the other hand, data-driven methods, which use software metrics and physiological measurements, offer objective and quantifiable data but may lack the context provided by user perspectives.

This raises several key questions: Is one approach inherently superior to the other? Can these methods be used independently, or is a hybrid approach more effective? To address these questions, we conducted a study using our proposed methodology, integrating both traditional and data-driven techniques. We explore how these methods compare and whether combining them leads to better outcomes.

5.2.1 Testing Methodologies - Implementation

To investigate the cross benefits of objective and subjective measures, we designed a software optimisation study where three methodologies were evaluated: a pure user feedback one, a pure data-driven (operational and physiological measure) one, and a hybrid one. This study is the follow-up of the one presented in section 4.5.2.

Experimental Protocol

In this study, 30 participants from 20 to 61 years old (average 33.4 years old) participated in a VR excavator training experiment. In this experiment, participants first completed the calibration of physiological sensors; followed by a tutorial on operating the virtual excavator with joysticks (see mappings Appendix A.3); and then performed three similar excavation exercises, each featuring a different set of virtual annotations (VA) with identical tasks to perform (see Table 4.2 for the list of all VAs). The participants were separated in two groups:

- **The Design group:** this group tested three different sets of VA empirically designed to encompass all possible display modalities (see Section 4.5.2). Participants in this group evaluated all VAs using NASA-Task Load Index (NASA-TLX) surveys [29] (template available Appendix B.1), System Usability Scale (SUS) inspired surveys [30] (template available Appendix B.2, Appendix B.3, Appendix B.4), open feedback, and physiological measures from eye-tracking, EEG, fitness tracker, and accelerometer data. This part of the study is presented in section 4.5.2, through the scope of physiological analysis;
- **The Evaluation group:** this group tested three new sets of VAs based on the results from the Design group. One set was designed solely based on user feedback (the User Set), another set was based exclusively on objective data (the Data-driven Set), and the third set was developed using a hybrid approach that combined both objective and subjective data (the Hybrid Set), corresponding to the VR-SOLUS methodology).

The Design group consisted of 17 participants, while the Evaluation group included 16 participants, comprising 13 new participants and 3 from the Design group. Including some individuals from the first group aimed to assess user evolution with the different methodologies.

During this study, the participants used the HTC Vive Pro Eye VR headset, as their main head mounted device and as our eye-tracking provider. They also used the Unicorn Hybrid Black, a 8 channels EEG cap providing both brain signals and head acceleration values. Finally, they were equipped with the Polar OH1+ armband, providing arm acceleration values and heart rate. The virtual environment was designed in Unity, and controlled by the users with Thrustmaster 16000M Duo joysticks. Operational and eye-tracking data were recorded in Unity, EEG and head movement data were recorded through the Unicorn suite, arm movement and heart rate were recorded with the Polar Sensor Logger application. Data were synchronised through Lab Streaming Layer (LSL), filtered in OpenVibe and analysed in Python. User feedback were done on paper and anonymised when registered on computer. This setup is shown in Fig. 5.2.

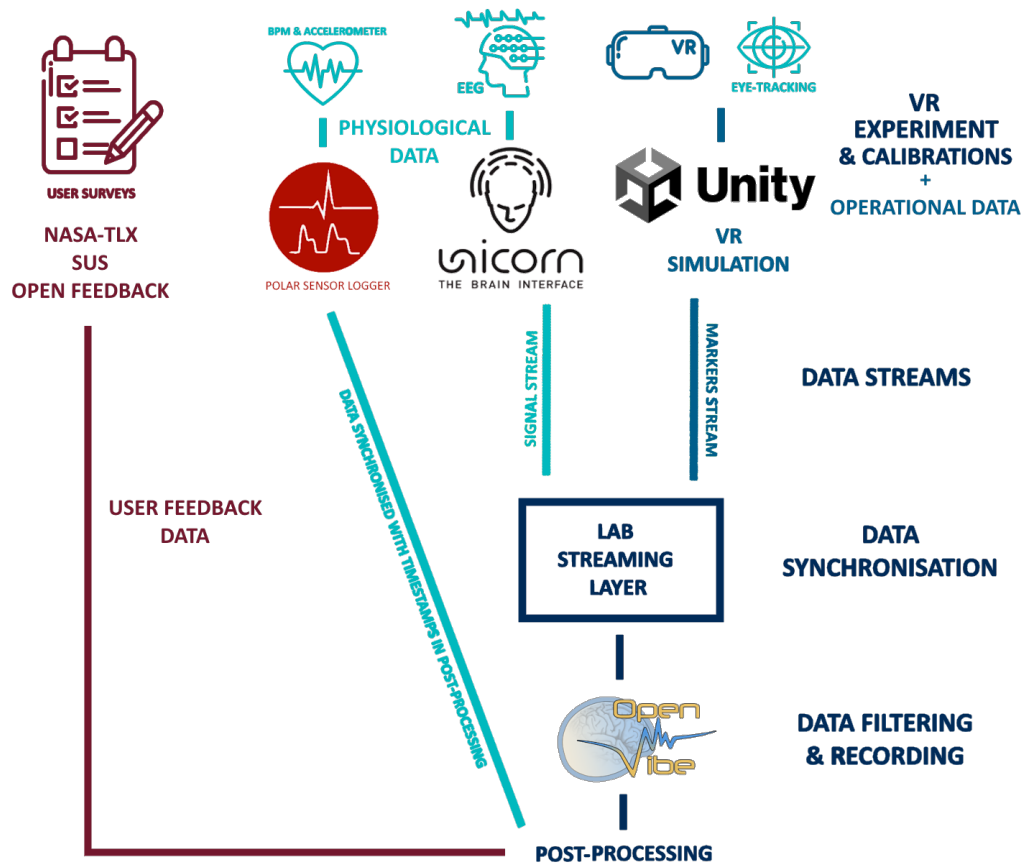


Figure 5.2: Operational + Physiological + User Feedback Pipeline

5.2.2 Design Phase

In the first stage of this study, the Design phase, we defined three sets of VAs: visual VAs at head level, VAs integrated in the world (and not just in the vehicle cabin), and more immersive VAs with sound and event-based activation (the "special" modalities for the VAs). The design of those sets allowed the participants to experience all modalities of the VAs.

During the experiments, users data were recorded and the VA sets evaluated using multiple modalities:

- **Operational Measures:** completion time, inputs errors and multi-inputs percentage, those measures provided performance indicators for each set of VAs;
- **Eye-tracking:** fixations on visual VAs were recorded and used as one of the main metrics for usefulness of visual VAs;
- **EEG:** cognitive load was evaluated for each set and revealed that head level information produced less mental demand (less visual search required), and also that it was better to have immersive sound (continuous and intermittent) than just intermittent sound;
- **IMUs and Heart Rate:** the experiment had very little physical demand, and thus those measures did not give significant information, other than confirming that the head physical demand is lessened when information is displayed at head-level (but produces more arm movement, to align the virtual vehicle with its environment);
- **User Feedback:** a NASA-TLX survey was conducted after each set; a SUS-inspired survey was conducted at the end of the experiment regarding the perceived usefulness and appreciation of each VA, as well as which modality should be kept for display according to users (always visible/audible, eye-tracked, event-based, in a

menu, deleted); and finally users had to rate their preferred set in order. Open feedback on the experiment was also recorded.

Some evaluation modalities (operational data, physical data, EEG, NASA-TLX) evaluated whole sets, giving global (but still exploitable) information on the user experience. Other modalities (eye-tracking, SUS) allowed more targeted analysis on specific VAs. The physiological results and their analysis is available in section 4.5.2.

Operational Results

Operational results evaluate the whole sets. They are performance indicators that tend to show that head-displayed information produce better completion time and more multi-input percentage, but also produce more errors. On the other hand, integrating information in the world requires more completion time (more information search needed) and less multi-input percentage (more focus on the information than the control) but less errors (slow and steady control). Multimodal feedback finally combines both previous sets advantage and drawbacks, showing middle time, middle errors and middle multi-input percentage, as shown in Fig. 5.3.

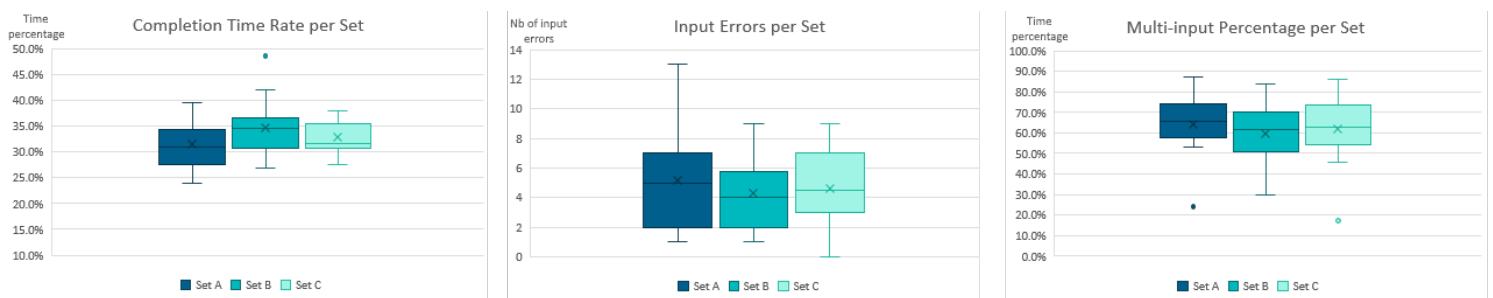


Figure 5.3: Design Phase Operational Results

User Feedback Results

The user-centric results are derived from two surveys and open feedback. The first survey is the classic NASA-TLX questionnaire, which evaluates (on a 20 points scale) mental demand, physical demand, temporal demand, performance, effort, and frustration based on user insights. The second survey is inspired by the SUS. While the NASA-TLX provides insights into various aspects of the VA sets, the SUS focuses specifically on the VAs, assigning each a usability score (out of 10), an affinity score (out of 10), and a modality.

The NASA-TLX results are fairly consistent across all three sets (see Table 5.1). However, Set C (integrated/immersive set) shows slightly better performance results. Additionally Set A (purely visual at head-level) exhibits slightly higher frustration levels.

Parameter	Set A	Set B	Set C
Mental demand	8.8	8.3	8.3
Physical demand	2.6	2.6	2.5
Temporal demand	3.6	3.7	3.5
Performance	13.1	13.6	14.2
Effort	8.9	8.3	8.4
Frustration	5.6	4.9	4.6

Table 5.1: Design Phase NASA-TLX Results

The SUS provided a precise score (on a 10 points scale) for each VA in terms of affinity and usefulness, along with a modality score. This modality score, weighted by the affinity and usefulness ratings, served as the primary parameter for modality selection in the Evaluation round, see Table 5.2.

VA	Usefulness	Affinity	Modality
Mappings	8.6	6.9	In a menu
Tutorial	5.8	7.3	In a menu
Reticle	8.4	8.5	Always visible
Cabin Instructions	7.9	5.6	Always visible
World Instructions	7.4	6.3	Eye-tracked
Arrow to Objective	8.4	8.3	Always visible
Rear View	8.7	8.6	Special
Tracks Directions	6.9	6.4	Special
Engine Sound	4.1	5.3	In a menu
Backing Sound	6.9	7.6	Special

Table 5.2: Design Phase SUS Results

5.2.3 Evaluation Phase

After testing our different VAs in the Design phase, we extracted three new VAs sets for our evaluation phase, based on three methodologies:

- **The User Set:** based only on user feedback, NASA-LTX survey, SUS-inspired questionnaire and open feedback;
- **The Data-driven Set:** based only on operational and physiological metrics, from the eye-tracker, the EEG, accelerometers and inputs study;
- **The Hybrid Set:** created by combining both methodology. Each modality is chosen following the rule: if both the User Set and the Data-driven Set VA modality matches, we keep it as is; if both modalities do not match, we balance our choice by performance and load (metrics obtained in both methodology) and try to establish a middle ground modality.

The final sets to be evaluated in the Evaluation phase are described in Table 5.3 and a visual example of the interfaces is available Appendix A.4. In the case of the World Instructions VA, the "special" display modality refers to an in-world icon above the objective, while the "eye-tracked" modality refers to a in-world text with instruction above the objective with its opacity depending on the gaze.

Annotations	User Set	Data-driven Set	Hybrid Set
Mappings	In a menu	Always visible	Eye-tracked
Tutorial	In a menu	In a menu	In a menu
Reticle	Always visible	Eye-tracked	Always visible
Cabin Instructions	Always visible	Always visible	Always visible
World Instructions	Eye-tracked	Deleted	Special (icon)
Arrow to Objective	Always visible	In a menu	Eye-tracked
Rear View	Special	Always visible	Always visible
Tracks Directions	Special (only when moving)	Always visible	Always visible
Engine Sound	In a menu	Always audible	Always audible
Backing Sound	In a menu	Special (only when backing up)	Special (only when backing up)

Table 5.3: Virtual Annotations Sets in the Evaluation Phase

A new set of participants (and a few from the Design phase) formed the Evaluation group and assessed these new sets with the same experiment as before, and with the same sensors and modalities.

Results on the Evaluation Phase

We used three types of measures to evaluate our methodologies: operational measures, physiological measures and user feedback.

Operational Results

We evaluated the three sets of VAs - User Set, Data-driven Set and Hybrid Set - with performance indicators, in the form of completion time, input errors and multi-input percentage.

Completion Time

Participants spent on average less time on the Hybrid set exercises (30.9%) compared to the User set (32.5%) and the Data-driven Set (36.6%) as shown in Fig. 5.4.

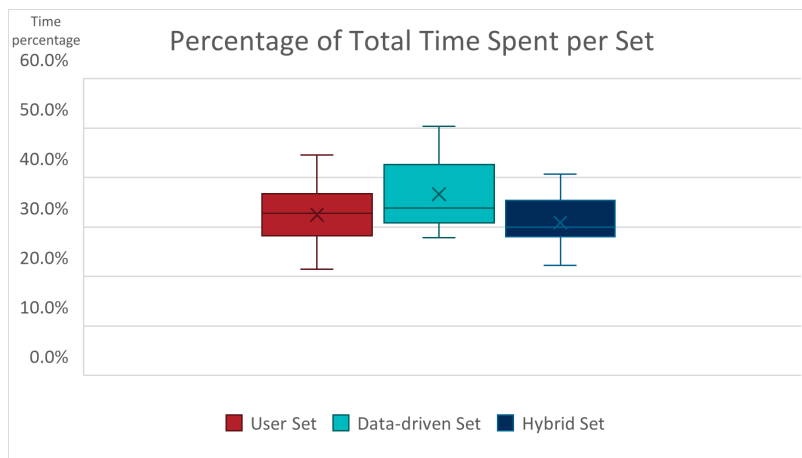


Figure 5.4: Completion Time for each Methodology Set

Input Errors

Participants made on average fewer errors in the User Set (average of 5.25) than in the Hybrid Set (5.69) and the Data-driven Set (7.31), as seen in Fig. 5.5.

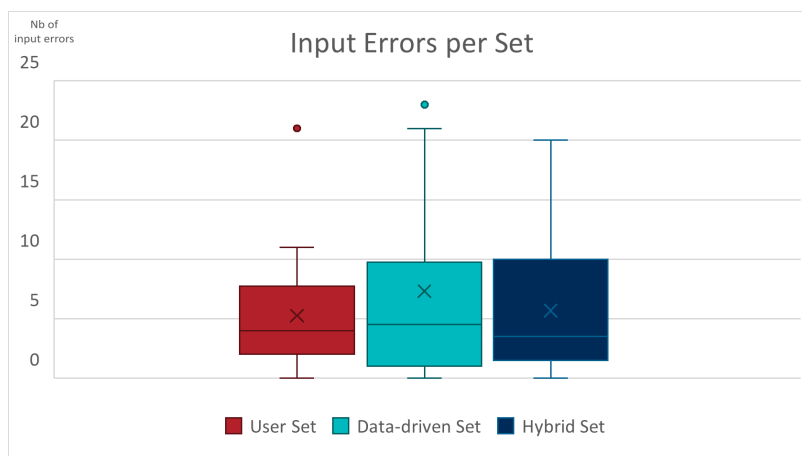


Figure 5.5: Input Errors for each Methodology Set

Multi-input Percentage

Participants demonstrated on average better multi-input usage in the Hybrid Set (74.3%) than in the User Set (74.2%) and the Data-driven Set (68.3%), see Fig. 5.6.

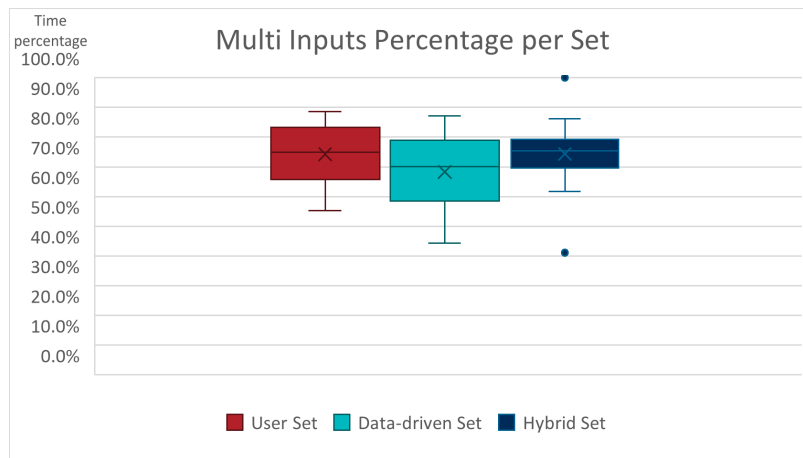


Figure 5.6: Multi-input Percentage for each Methodology Set

Operational Results Conclusion

In terms of pure operational measures, the Hybrid Set and the User Set show very close performance. The Hybrid Set is slightly better in terms of completion time and multi-input usage, showing a slightly better performance and ease-of-use overall, while the User Set produces slightly less errors, showing a faster integration of the controls. While there is no clear better methodology based on operational measures alone, the Data-driven Set, based only on objective measures, has already shown lesser performance than the other two.

Physiological Results

Our three methodological sets of VAs were also evaluated through bio-sensors. Namely, we used eye-tracking fixations (for visual VAs), EEG cognitive load, as well as physical load sensors (head and arm IMUs and heart rate monitor).

Eye-tracking Results

For the eye-tracking, we analyse the fixations on the different visual VAs. We then study the distribution and choose modalities based on it: if the percentage of fixation is equal to the minimum (Q0), the annotation is deleted; between Q0 and Q1 (25% quartile), the annotations is put in a menu; between Q1 and Q2 (25% quartile and the median) the annotation is event-based (either eye-tracked or special condition); and above Q2 (median), the annotation is always visible. The tutorial is an exception, we keep it in its specific menu. Based on the gaze fixations observation we obtain the results shown in Table 5.4.

The choices of display modalities of the visual annotations for the Data-driven Set relied heavily on the fixations analysis done in the Design phase. As thus, it was expected to observe in the Evaluation Phase a close match between the gaze analysis results from the Design Phase (5 out of 7 modalities corresponding). The observation that really stands out is the divergence between user-selected modalities and their actual visual use. Users were asked which modality they thought was the best for each VA in the Design phase, along with a "usefulness" and "affinity" score. Even though they "use" (look at) the VAs, the users do not deem them as useful or likeable as they appear through physiological analysis. Finally, the Hybrid Set showed medium matches (3 out of 7). This was expected, as the Hybrid Set was created as a middle ground between the Data-driven Set modalities (based on Design phase eye-tracking) and the users' choices.

Annotations	Eye-tracking Results	User Set	Data-driven Set	Hybrid Set
Mappings	Always visible		X	
Reticle	Always visible	X		X
Cabin Instructions	Special or Eye-tracked			
World Instructions	Deleted		X	
Arrow to Objective	Menu		X	
Rear View	Always visible		X	X
Tracks Directions	Always visible		X	X
Matches		1	5	3

Table 5.4: Eye-tracking Analysis: Modalities Matches between the Design and the Evaluation Phase Results (a cross marks when both the Design Phase and the Evaluation Phase modalities match)

Electroencephalogram Results

We used the Theta-Alpha Ratio (TAR) metric as our cognitive load measure. The signal was filtered with pass-band to isolate the Theta (4-8Hz) and Alpha (8-12Hz) bands, as well as eliminate the bulk of noise created (e.g. motion noise (0-4Hz), 50Hz line noise, etc.). The data was then segmented and filtered with the head motion data to remove all the motion polluted segments. Finally, each participant TAR values were normalised using regular normalisation (as seen in Equation 4.1) before comparison. The results are shown in Fig. 5.7.

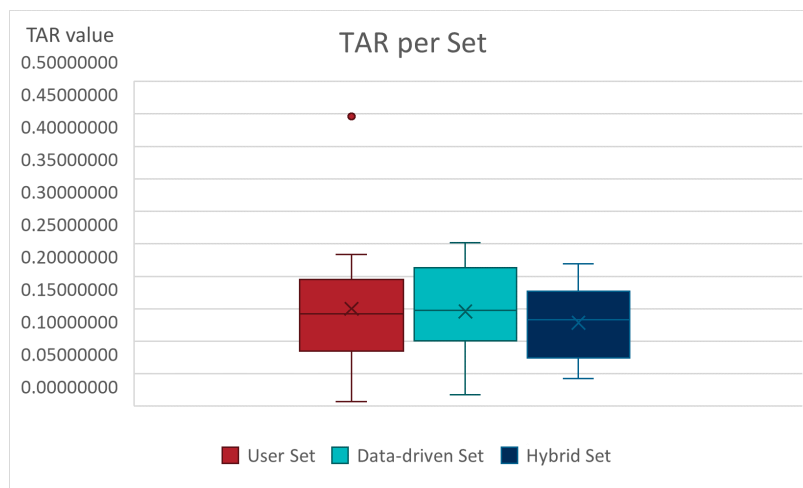


Figure 5.7: Cognitive Load Evaluation for each VA Set

Participants experienced on average lower cognitive load with the Hybrid Set (average TAR value of 0.1250) compared to the Data-driven Set (0.1423) and the User set (0.1475). Those results are quite paradoxical: the Data-driven Set, while based (among others) on the cognitive load measured in the Design Phase, is not the most optimal in the Evaluation Phase. This can be explained by the fact that cognitive load are applied to whole sets and not specific VAs, therefore not accounting for every specificities of each initial sets. Likewise, the User set once again shows the most discrepancy between users perception and the actual physiological results. The VAs' modalities chosen during the Design phase were those deemed the less mentally demanding according to feedback, but turned out to be the most mentally demanding according to the physiological sensors measures. A notable point is that while being a bridge between user perception and physiological measures, the Hybrid set does not show a middle ground between the User set and the Data-driven set results, but rather the best results out of the three sets instead.

Physical Load Results

Due to the stationary nature of the tasks, motion measures were very similar across sets. Head motions (Fig. 5.8a) were slightly higher for the Data-driven Set (1.255mG) than the User Set (1.233mG) and the Hybrid Set

(1.230mG). Arm motions (Fig. 5.8b) followed a similar pattern, with the User set requiring slightly more arm movements (25.165mG) than the Hybrid Set (24.410mG) and the Data-driven Set (24.345mG). Heart rates (Fig. 5.8c) were also very close for each set (71.6BPM for the User Set, 70.9BPM for the Hybrid Set, and 70.8BPM for the Data-driven Set). We can outline a slightly better performance tendency coming from the Hybrid Set across all physical load measures, although not statistically significant.

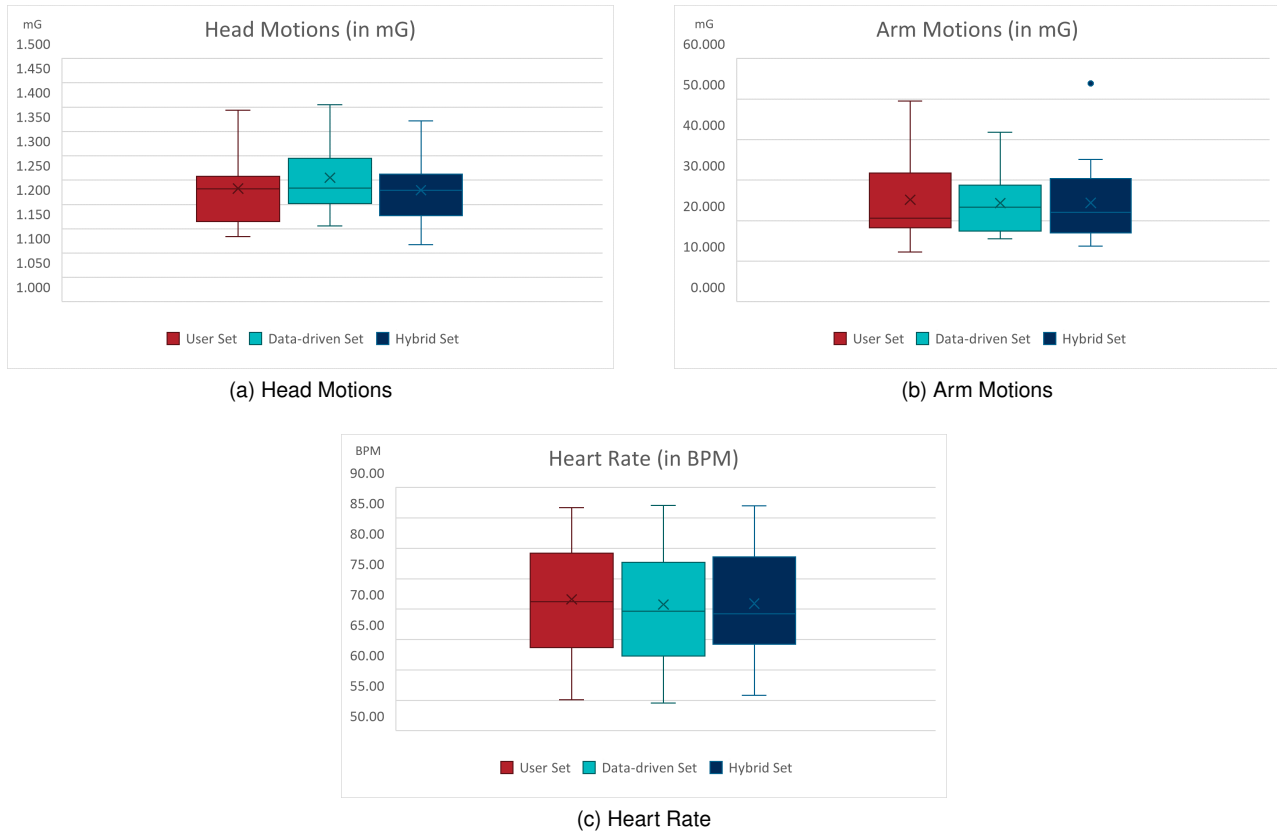


Figure 5.8: Physical Load Measures

User Feedback Results

User feedback was collected through NASA-TLX survey after testing each set (evaluating the whole set), through a SUS-inspired survey at the end of the whole experiment (evaluating each VA separately) and finally through open feedback, notably the preferred ranking of the different sets.

NASA-TLX Results

The Evaluation group used the same surveys as the Design group to assess the new VA sets. As shown in Table 5.5, mental load and performance significantly improved across all methodology-selected sets compared to the empirically chosen ones from the Design round (Table 5.1), indicating an efficient optimisation effect on those main points. Additionally, the results are similar across all three techniques, demonstrating comparable capability.

Perceived cognitive load and EEG-measured cognitive load (referred as "mental demand" in the NASA-TLX test) between the Design phase and the Evaluation phase did not seem to match, according to the EEG results of the Evaluation phase (the chosen VA modalities were not the most optimal), when comparing Table 5.1 and Fig. 4.17. However, inside of the Evaluation phase, we see that perceived (Table 5.5) and measured cognitive load (Fig. 5.7) completely match. The Data-driven Set, albeit partially based on cognitive load measures, is notably once again shown as the more mentally demanding.

Parameter	User Set	Data-driven Set	Hybrid Set
Mental demand	5.9	6.4	5.3
Physical demand	4.3	4.3	4.4
Temporal demand	5.7	5.6	5.8
Performance	14.4	14.3	14.8
Effort	7.8	7.9	7.2
Frustration	5.0	4.9	5.3

Table 5.5: Evaluation Phase NASA-TLX Results

SUS Results

The SUS results, which assess user affinity and the perceived usefulness of each VA, reveal an unexpected evolution from the Design Phase. Despite the fact that the User Set VA modalities were designed almost entirely based on the SUS results from the Design Phase, the Evaluation Phase (Table 5.6) shows that the User Set is not the one with the most alignment with the previous results. Instead, the Hybrid Set features the display modalities most preferred by users. This highlights that different groups of users exhibit varying preferences, and basing a design methodology solely on user feedback may only suit the group tested, rather than others. Interestingly, three participants from the Design Phase also took part in the Evaluation Phase. Their scores for usefulness and affinity not only shifted but, on average, they suggested changing 1.3 out of the 10 display modalities compared to their previous choices. This suggests that user feedback evolves over time and with repeated exposure, emphasising the need for continuous testing and potentially ongoing adjustments in VR design to accommodate the changing preferences of users.

Annotations	Usefulness	Affinity	SUS Results	User Set	Data-driven Set	Hybrid Set
Mappings	7.2	7	Menu	X		
Reticle	8.4	8.4	Always visible	X		X
Cabin Instructions	6.7	6.9	Eye-tracked			
World Instructions	6.7	6.9	Special (Icon)			X
Arrow to Objective	6.9	6.9	Always visible	X		
Rear View	8.2	8.4	Always visible		X	X
Tracks Directions	7.6	7.2	Always visible		X	X
Engine Sound	5.9	6.7	Menu	X		
Backing Sound	7	8	Special (when backing up)		X	X
Matches				4	3	5

Table 5.6: SUS Analysis, Affinity, Usefulness, Modalities Matches in the Evaluation Phase Results (a cross marks when both the SUS Results and the Evaluation Phase modalities match)

Open Feedback Results

Open feedback, as the name implies, let the users give any feedback they could think of, whether on VAs, sets, or even hardware setup and software simulation. On the subject of VAs and VAs set, two remarks were recurrent:

- **Sounds create immersion:** having both engine sound and backing sound was preferred because it created immersion. However, when we look at the SUS-inspired answers (Table 5.6), we see that the users chose not to keep them present at all time. This suggests that users favored usefulness over affinity (without being prompted to) when choosing the VAs' modalities;
- **A picture is worth a thousand words:** users preferred when the World Instructions texts were replaced with just an icon with an arrow to show where they should do their task (see Fig. 5.9). Having both Cabin Instructions and World Instructions in the form of text was redundant, while having only Cabin Instructions

and no World Instructions required more visual search for the objective position (using the Arrow to Objective or not). Having a big, visible, in-world icon allowed to find the objective fast, without having to think about it, producing less mental load.

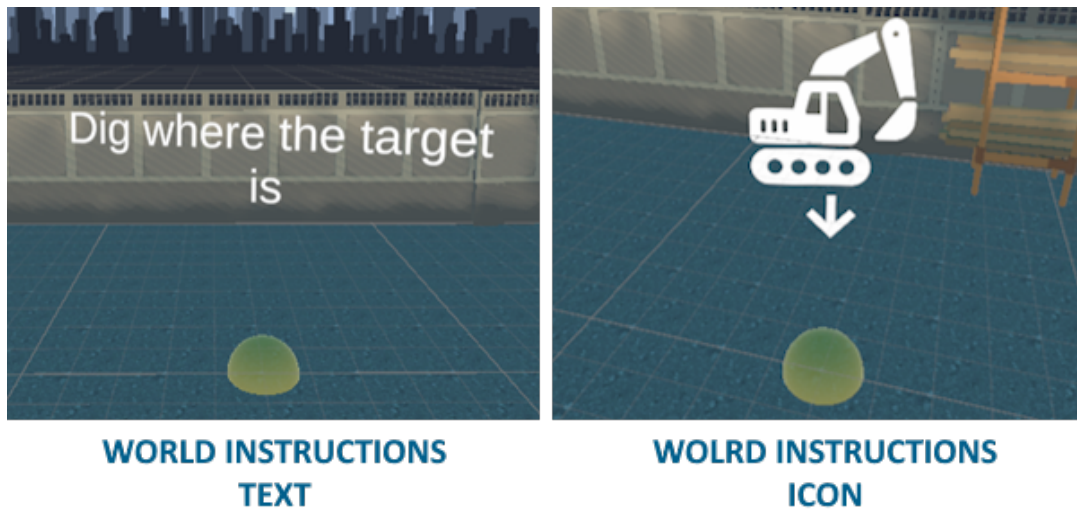


Figure 5.9: World Instructions VA - Text vs Icon

A final question was asked to all participants: "How would you rank the VAs Sets?". Users had to rank first, second and third the three sets. First place gives two points, second place one point and last place zero. Through this system we obtained the following results Fig. 5.10: the User Set obtained 18 points with 8 first places, the Data-driven Set obtained 12 points with 2 first places and the Hybrid Set obtained 18 points with 6 first places. The User Set got the most first places, which is to be expected: ranking the sets is a pure subjective question on user appreciation. It is thus likely that a set designed by users for users would provide the most appreciated result. But the Hybrid Set has overall the same amount of points, making it a more "averagely liked" set (6 second places, 4 third places), while the User Set is more a "either loved or hated" Set (2 second places, 6 third places). This illustrates the subjectivity of user feedback, showing that a group can love a feature while another can hate it at the same time. These results also show how a hybrid methodology can bridge this gap in user preferences.

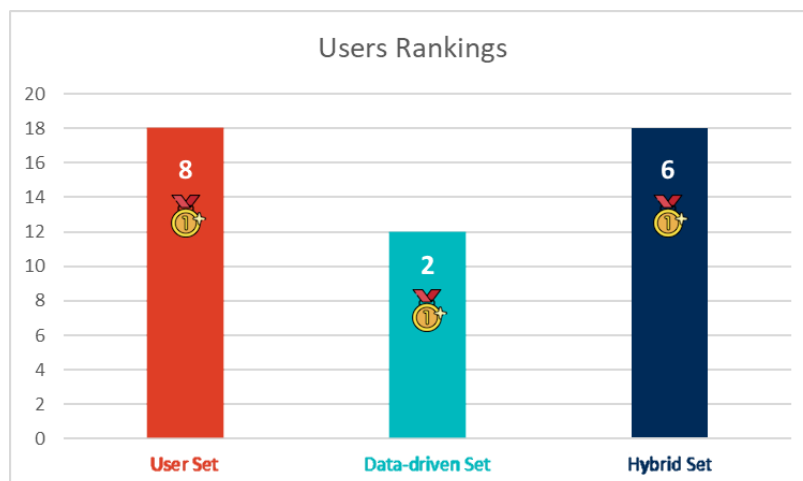


Figure 5.10: Users Ranking of the VAs Sets, with the Number of First Places Attributed to each Set

Results Conclusion

While there is no clear winner between the User Set and the Hybrid Set based on users' preferences, there is clearly a loser: the Data-driven Set. This is a clear answer to our previous question *"Can the VR-SOLUS method be used with only its objective measures?"* and this answer is **NO**. In terms of pure users' choice, a pure objective data-driven method does not provide the best user experience. And thus the answer to our second question *"Is there still a necessary role for traditional subjective input in the design and optimisation of VR experiences?"* is **YES**, user feedback is still crucial in VR design, as our study showed.

Overall, when looking at all our metrics, we see that a fully data-driven methodology is not optimal, while traditional user feedback methodology performs largely better. That leaves the question of the hybrid methodology: is there really a need for all these additional data, or is the traditional methodology sufficient? If we look at all our measures:

- **Operational Measures:** the difference is small, but the Hybrid Set is on top when it comes to operational metrics. It is slightly better in terms of completion time (overall performance) and multi-inputs (ease-of-use and learning), and slightly below the User Set in terms of errors (understanding);
- **Physiological Measures:** eye-tracking shows that the Hybrid Set aligns with the users gaze behaviour and preferences. It has the most matches between preferred modalities and actual physiological results. In terms of cognitive load, the EEG analysis showed that the Hybrid Set required less mental demand to use than the user set. In terms of physical demand, the difference is too small to be significant, but leans towards the Hybrid set being less physically demanding, an observation to confirm with more participant and more long-term experiments;
- **User Feedback:** user feedback is also divided. The NASA-TLX survey shows similar results as the bio-sensors and operational data (small differences between the Hybrid and User sets). The SUS survey however, while being the key component of the User Set design, shows that the Hybrid Set is more fit to the users tastes. Finally, users' ranking and open feedback show both the User Set and the Hybrid Set on top, one important observation is how the User Set is either first or last in rankings but rarely second (keep it or leave it), while the Hybrid Set is more globally appreciated (first and second choices but more rarely last).

In this study, the hybrid method proved to be superior to both the data-driven and user feedback-based approaches. While operational data alone did not indicate a clear preference for any specific method, the combination of physiological and user data clearly favored the hybrid approach. Physiological results demonstrated that the hybrid method outperformed even a purely physiological-based method, indicating better cognitive load management and user engagement. Similarly, when evaluating user feedback, the hybrid method also surpassed the user feedback-based approach, showing higher user satisfaction and perceived usability. These findings suggest that the hybrid approach effectively balances objective and subjective measures.

We included participants in this Evaluation phase that were also in the Design phase. While globally consistent in their feedback, we observed changes, while their performance and status measures were coherent (similar between phases, with a learning effect naturally taking place). Those observations do not make user feedback unreliable, but just what it is: subjective. Users can change opinion from one session to another, one person's tastes might not match another person's. On the other hand, objective metrics are reliable and do evolve with time in a consistent pattern, making them easier for researcher and developer to use in the long term. But the results provided by those objective measures do not always appeal to the users.

We advocated this whole thesis that a hybrid methodology brings the best of both objective and subjective methodology, and proved it through this study. Of course, the results are not set in stone: we lack a pool of participants large enough to have higher margin between methodologies and the necessary hindsight of long term studies. However, we still have results worth saying: a hybrid method is profitable for VR design and optimisation.

5.2.4 The Hybrid Method

By using a hybrid methodology combining operational measures, physiological measures as well as user feedback we combine the best of both objective and subjective data. This hybrid approach allows the proposed VR-SOLUS methodology access to various aspects of user experience.

Operational Performance

Operational metrics provide performance indicators, extrinsic measures such as:

- **Completion Time:** basic of performance metrics, can be divided by task, exercise, experiment. Also useful to employ during repetitive tasks to observe user progress;
- **Input Errors:** give insight into ease-of-use, user understanding and learning;
- **Multi-input Percentage:** time spent using more than one input, it gives essential data on ease-of-use, ergonomics and user progression;
- **Use-case Specific Measures:** gaming, industry, training, therapy, and so on, each use-case has its own design problematics and associated metrics. Integrating those measures is paramount for an efficient specialised design methodology.

Physiological User State

Physiological measures, done through biosensors, form intrinsic measures providing information on the user state. Among them:

- **Eye-tracking:** attention, focus, zones of interest, all are important when using a VR headset to know what the user actually sees and looks at during their experience;
- **Electroencephalogram:** neurological data becomes essential today when VR applications are more complicated and specialised than ever. Knowing the user cognitive state, in particular their mental load, ensure not only better user experience but also more healthy usage of VR;
- **Physical Load Sensors:** such as IMUs or heart monitor, they allow the assessment of the physical state of the user, their physical engagement, an assessment necessary for long-time use of VR application.

Surveyed Perception

User feedback, the traditional approach to VR design and optimisation, gives insight into user perception. This subjective intrinsic measure provide information of perceived user experience. While it is not ground truth, as it is dependent on the user, in the end, a VR application is addressed to users. Besides surveys like the NASA-TLX or the SUS that give quantitative perception (through scoring systems), there is also open feedback. Users tend to give useful information out of the scope of surveys (e.g. improvement of the software simulation, changes to the hardware setup, etc.) when they can answer freely. And with each user profile comes a new point of view, with new insights that sensors, user score or software metrics might not capture.

VR-SOLUS

Our VR-SOLUS methodology synthesises performance indicators provided by operational measures, user status through physiological analysis, as well as user perception thanks to user feedback (Fig. 5.11).

To be a complete hybrid methodology, correlation analysis is a critical factor in the validation and reliability of our VR-SOLUS method, as it ensures that the various metrics we collect — operational, physiological, and user feedback — are not only individually meaningful but also consistently related to each other. Establishing strong correlations among these diverse data demonstrates that the physiological responses captured by sensors, such as EEG or eye-tracking, align with both the operational performance and user perceptions. This interconnection is needed to reinforce the methodology's ability to provide a comprehensive and accurate assessment of VR applications, and to ultimately enhance the confidence in our findings and the validity of our approach.

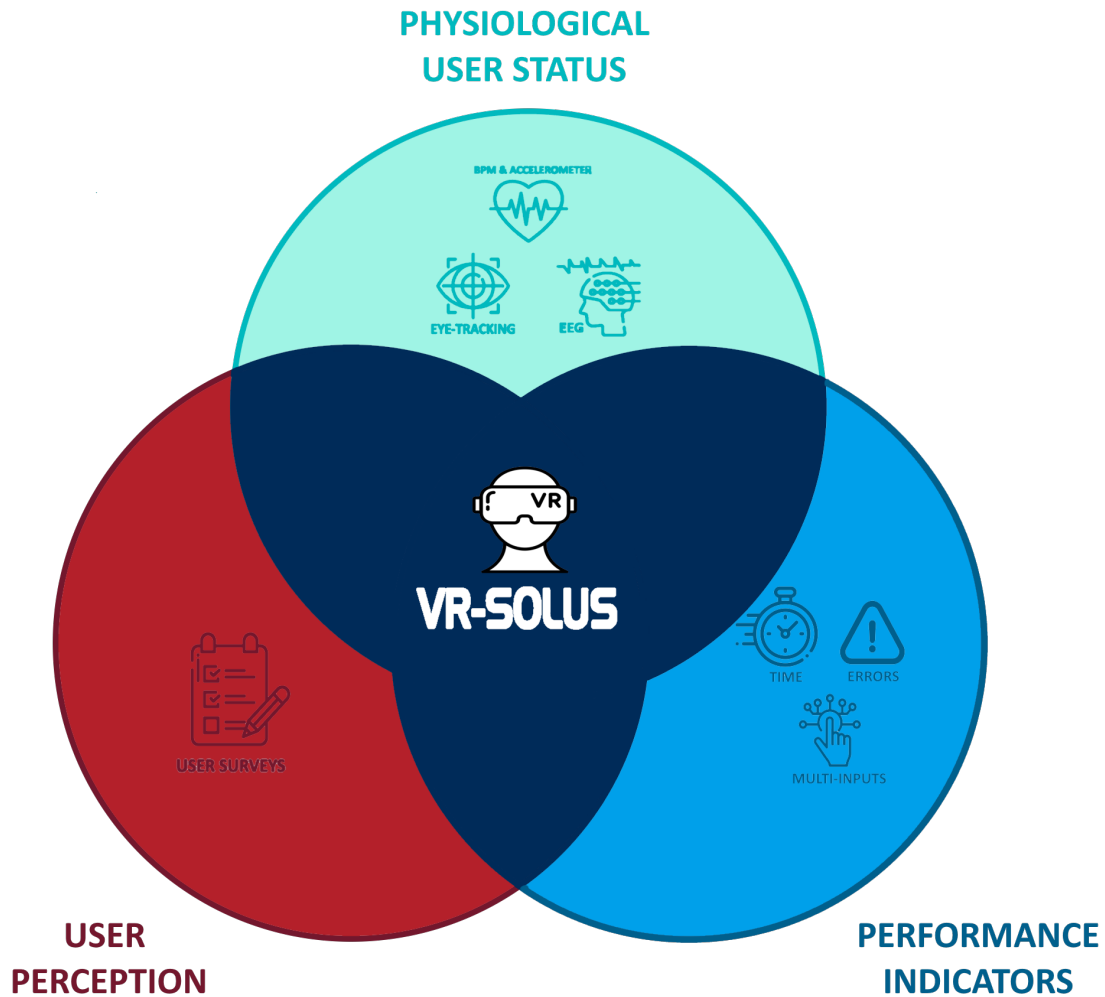


Figure 5.11: Combining Performance, User Status and Perception through VR-SOLUS

5.3 Correlation Study

In the analysis of VR design, understanding the various types of data and their relationship is essential for drawing meaningful conclusions. One key area of focus is the correlation of operational measures (which assess user performance) between themselves and with physiological measures (which reveal the user's internal state). Exploring how these two objective data streams interact can provide valuable insights into how performance and user experience are linked. Furthermore, the correlation between these objective measures (operational and physiological) and subjective data (user feedback) is also needed. This correlation helps bridging the gap between measurable performance metrics and the user's perceived experience, offering an extensive view of the VR environment's effectiveness and its impact on the user. These interconnections between operational, physiological, and subjective data form not only the validation of our methodology's metrics, but also the foundation for a more nuanced and accurate evaluation of VR applications.

5.3.1 Operational Correlation

To validate our operational data and thus consolidate their utility in the VR-SOLUS methodology, they need to be correlated together. To assert that our performance indicators do indeed indicate performance and not some other unrelated element, a correlation must be established between completion time (the ground truth of performance, should the experimental setup allow it), input errors and multi-input percentage. This issue was quickly evoked in Section 3.1.4, to confirm that multi-input percentage was indeed a performance indicator, and not a result of involuntary actions.

If we take a look at our methodologies study Section 5.2.3, we can investigate the correlation between our different metrics using the Pearson correlation (Table 5.7). We observe:

- **Error and Time:** a strong significant positive correlation between input errors and completion time. The more errors made, the more time it takes to complete the tasks, which is an expected result that validates input errors as a solid performance indicator;
- **Multi-input and Time:** a strong significant negative correlation between multi-input percentage (MI%) and completion time. The more compound movements used (higher MI%), the less time it takes to complete the tasks. This is also a logical results, that confirms MI% as a suitable performance and ease-of-use indicator, as well as a progression one when studied after periods of time;
- **Multi-input and Error:** a strong significant negative correlation between MI% and input errors. The more compound movements used, the less errors made. Not only this means that compound movements are not a result of involuntary actions (errors), but also that users employing compound movements are less likely to make errors.

These results show that our operational metrics are solid and reliable performance indicators. This stays true across all three methodologies - user-driven, data-driven and hybrid methodologies - showing the robustness of such metrics to evaluate user performance, whichever design choices were made prior.

Data Correlation Analysis	Correlation	p-value
Error vs Time (User Set)	0.68	0.004
Error vs Time (Data-driven Set)	0.91	8.3E-7
Error vs Time (Hybrid Set)	0.67	0.005
Error vs Time (Average)	0.75	0.003
MI% vs Time (User Set)	-0.65	0.007
MI% vs Time (Data-driven Set)	-0.84	4E-5
MI% vs Time (Hybrid Set)	-0.57	0.021
MI% vs Time (Average)	-0.69	0.007
MI% vs Error (User Set)	-0.62	0.01
MI% vs Error (Data-driven Set)	-0.85	2.7E-5
MI% vs Error (Hybrid Set)	-0.55	0.028
MI% vs Error (Average)	-0.67	0.013

Table 5.7: Data Correlation for Operational Metrics (Pearson correlation)

5.3.2 Operational - Physiological Correlation

Despite evaluating another aspect of the user experience - the user status, some correlations can be observed between physiological and operational measures, Table 5.8 (we excluded physical load metrics due to their low variance between users):

- **Gaze Attention and Time:** a significant rather strong positive correlation. The more time spent looking at virtual annotation (LookAt), the more time spent doing the tasks. This to be expected, as people rarely do two things at the same time, especially when they need to read/understand visual cue while controlling a virtual machine;
- **Gaze Attention and Error:** a significant rather strong positive correlation. The more users make errors, the more they need to look at the mappings to remind themselves of the correct control schemes. This explains why neophytes users tend to prefer having visual reminders displayed at all time, while experienced users tend to prefer having them less visually prominent, as they do not need them as much/anymore;

- **Cognitive Load and Time:** a non-significant and almost neutral correlation. There is little to no correlation between cognitive load (TAR) and time completion. There are too many elements to take into account during an experiment to directly correlate time and cognitive load. During an exercise, a user controls a machine, look at and listen to VAs, learn and progress. It is thus difficult to correlate one time value with one cognitive load average value;
- **Cognitive Load and Error:** a non-significant and almost neutral correlation. Inputs (and by extension vehicle control) is only one of the many elements to take into account in the cognitive load, therefore finding a direct correlation between cognitive load and error is rather hard;
- **Cognitive Load and Gaze Attention:** a non-significant small positive correlation. The more annotations, the more focus needed to understand them and include them into ones behaviour, thus the suggested augmented cognitive load.

We observe that performance indicators and user states have poor correlation. Gaze attention shows that adding a lot of annotations augments the time spent doing the task, which was already observed in literature [159]. However, one would expect that with more annotations - including inputs mappings - one would perform better. However the gaze-error correlation seems to suggest the opposite. This can be explained by the lack of classification of the gaze fixations: we do not know if there is understanding behind the attention, or even true intent. There is a small (but not significant) positive correlation with cognitive load, suggesting that indeed more virtual annotations take more mental resources to process, explaining the lesser performances. Cognitive load on its own does not show correlation with the performance metrics. This measure is not significantly linked to neither time of completion, nor to errors. This suggests that our cognitive load metrics cannot be a measure of performance from the user, or must be evaluated from another angle. One such angle is comparing neurological data to user-data.

Data Correlation Analysis	Correlation	p-value
LookAt vs Time (User Set)	0.49	0.052
LookAt vs Time (Data-driven Set)	0.68	0.003
LookAt vs Time (Hybrid Set)	0.77	4.5E-4
LookAt vs Time (Average)	0.65	0.018
LookAt vs Error (User Set)	0.47	0.062
LookAt vs Error (Data-driven Set)	0.61	0.012
LookAt vs Error (Hybrid Set)	0.63	0.009
LookAt vs Error (Average)	0.57	0.028
TAR vs Time (User Set)	-0.29	0.33
TAR vs Time (Data-driven Set)	-0.26	0.39
TAR vs Time (Hybrid Set)	0.17	0.57
TAR vs Time (Average)	-0.03	0.43
TAR vs Error (User Set)	-0.08	0.79
TAR vs Error (Data-driven Set)	-0.3	0.32
TAR vs Error (Hybrid Set)	0.41	0.163
TAR vs Error (Average)	0.01	0.42
TAR vs LookAt (User Set)	0.24	0.436
TAR vs LookAt (Data-driven Set)	0.006	0.985
TAR vs LookAt (Hybrid Set)	0.12	0.686
TAR vs LookAt (Average)	0.12	0.70

Table 5.8: Data Correlation for Physiological Metrics (Pearson correlation) - LookAt represents the User Attention and TAR the Cognitive Load

5.3.3 Objective - Subjective Correlation

In our study, we showed that associating user perception (user feedback), performance indicators (operational measures) and observed user states (physiological measures) brought better results. However, on paper, there is some redundancy between metrics (e.g. mental load or performance are evaluated by both user feedback and objective measures). Thus the question of true correlation (repetition of metrics) between perceived measures and observed measures is raised.

Perception and Observation

Through the NASA-TLX and the SUS-inspired surveys, we obtained quantified user feedback through parameter scores (respectively out of 20 and out of 10). The NASA-TLX offers insights on mental demand, physical demand, temporal demand, performance, effort and frustration. The SUS-inspired survey displays scores for usefulness and affinity. Among those, we can observe some redundancy with our operational and physiological measures, with interesting correlation results (Table 5.9):

- **Perceived Performance vs Performance Indicators:** when it comes to performance, users seem to evaluate quite accurately and objectively, as all the performance indicators have (almost) significant moderate correlation with the user performance scores. It seems that this user performance evaluation leans more heavily on errors and multi-input usage. This can be explained by the fact that those are measures "easy to register" as a user (e.g. participants can count how many errors they did, or at least approximate their number, they can feel how much they use different inputs, etc.), as opposed to completion time (which has the worst p-value) which is more difficult to grasp, as there is no displayed timer;
- **Perceived Mental Load vs TAR Values:** there is a tremendous gap between perceived mental load and evaluated TAR cognitive values. Not only the correlation is weak (and not significant due to the lack of participants), it even tends to be negative, suggesting an opposition between perceived mental demand and actual brain activity. This might come from the fact that it is quite complicated to rate one's own mental demand during a task (e.g. For a user, what points of comparison can they use? What can they consider the minimum and maximum cognitive load they ever experienced?);
- **Perceived Usefulness and Affinity vs Gaze-tracked Attention:** when it comes to users' rating of their affinity and their perceived usefulness of VAs, there is little to no correlation with participants' gaze behaviour. This suggests that affinity and usefulness assessment might be anchored in the domain of taste (subjective opinions) rather than actual behaviour, although usefulness is slightly more correlated with attention (albeit not significantly).

Data correlation analysis	Correlation	p-value
NASA-TLX Performance Score vs Time	-0.49	0.21
NASA-TLX Performance Score vs Errors	-0.46	0.075
NASA-TLX Performance Score vs MI%	0.56	0.038
NASA-TLX Mental Load Score vs TAR	-0.15	0.68
NASA-TLX Effort Score vs TAR	-0.16	0.47
SUS Usefulness Score vs Eye-tracked Attention	0.1	0.62
SUS Affinity Score vs Eye-tracked Attention	0.01	0.59

Table 5.9: Correlation between User Feedback and Measured Data (Pearson correlation)

While there seem to be redundancies in the data, there are actually few of them. Even performance, where the correlation between perception and observation is the strongest, does have nuances between what the user perceives and what was observed during the experiment. Which is why, even when similar questions are answered by different metrics, we advocate in favour of keeping all of them, as they add refinement to the design analysis and optimisation. Some metrics are also non-correlated (despite answering the same question) or inversely correlated: this is another argument in favour of a hybrid methodology, which provides both objective and subjective points of

view of the user experience, allowing developers and researchers to choose according to their need (e.g. a marketing application might want to follow user perception to appeal to its audience, while a therapeutic application meant to be used in the long term might want to lean on physiological measures).

Cognitive Load and User Profile

We have seen that not all objective metrics have strong correlation together. In particular, cognitive load does not seem to strongly correlate with our performance indicators (operational metrics). This does not make our TAR cognitive load metrics irrelevant, as it might stem from another element: subjective measures, and more precisely, user profiles. Users were asked to rate their level (on a scale of 1 to 10) of practical knowledge with VR technologies, excavation operations and joystick control. If we compare those with the cognitive load evaluations, we see Table 5.10 that on average, there is an appreciable negative correlation, although the low number of participants impacts the significance of this observation.

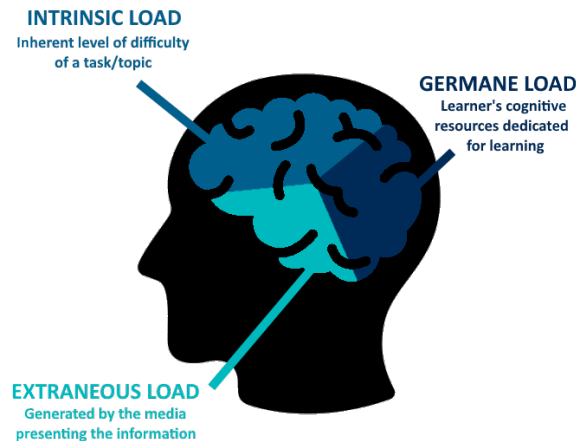
Data correlation analysis	Correlation	p-value
TAR vs VR Level	-0.20	0.326
TAR vs Excavation Level	-0.41	0.27
TAR vs Joystick Level	-0.54	0.06

Table 5.10: Correlation between User Profile and Physiologically Measured Cognitive Load (Pearson Correlation)

According to Sweller's Cognitive Load Theory (CLT) [183], when someone is learning something new, they create mental schemes, "problem solving templates" that they reuse later in similar situations. To learn something, mental load is actually divided in three categories (Fig. 5.12):

- **Intrinsic Cognitive Load:** induced by the complexity of the task/information someone is trying to learn (in our case, earthwork tasks);
- **Extraneous Cognitive Load:** induced by the medium used to learn (in our case VR and the different annotations) and the external parameters surrounding the task;
- **Germane Cognitive Load:** induced by the complexity for the brain to put in place a new schema or reinforce an existing one.

In our study, while time performance did not correlate with cognitive load, user profile seems to have a more tangible impact on cognitive behaviour. We can thus hypothesise that the cognitive load evaluated during the experiment is actually mostly the Germane Load. With training (i.e. user experience in the technical expertise), the cognitive load decreases and matches the user feeling because the cognitive schemes are already in place and need only be reinforced, not created from scratch.



SWELLER'S COGNITIVE LOAD THEORY

Figure 5.12: Sweller's Cognitive Load Theory (CLT)

5.3.4 Evaluation of Hybrid Measures and Correlation

The hybridisation of methodologies in our VR research offers a comprehensive and multifaceted approach to understanding and optimising user experience. By combining operational measures, physiological data, and user feedback, the hybrid method addresses limitations inherent in relying on a single type of data. Operational metrics capture task performance, physiological signals reveal user states, and feedback provides subjective insight, together offering a complete evaluation framework. As demonstrated in our study, traditional user feedback methods remain valuable, often reflecting strong preferences. However, they also exhibit polarisation, with users either strongly favouring or disliking specific approaches based on personal taste. In contrast, the hybrid methodology tends to yield moderate but consistent approval, balancing user satisfaction and performance optimisation.

Regarding correlations, our findings reveal key insights. Operational data correlated well across different measures, supporting their reliability in evaluating task performance. Eye-tracking metrics, particularly gaze data, exhibited relationships with operational data, indicating the potential of visual behaviour to infer task performance. However, no significant correlations were observed between operational data and brain activity (through EEG), suggesting distinct and independent contributions of cognitive and task-performance measures. Clear differences were also observed between perceived and measured performance, highlighting the divergence between subjective perception and objective measures. These findings reveal the necessity of hybridising data sources to gain a balanced understanding of user experience.

Cognitive load, as assessed through EEG, did not correlate strongly with other metrics but displayed a non-significant tendency to align with user profiles, hinting at its potential use as a learning or adaptability metric. This observation warrants further investigation to clarify its implications and possible applications in assessing user expertise and system learning curves. A reflection on the reliability and precision of physiological data in VR contexts is essential to assess whether or not those observations are correct. While the insights they offer are invaluable, the challenges of signal quality and interpretation demand careful attention, and their integration into hybrid methodologies must be both rigorous and adaptive.

Chapter 6

Discussion - Towards Further Virtual Reality Design

In this thesis, we presented the VR-SOLUS - Sensory, Operational and User data Synthesis for Virtual Reality - methodology, a comprehensive approach designed to enhance the process of developing and optimising virtual reality (VR) applications. Our work combined operational, physiological, and user feedback data to create a robust framework for VR design. By integrating diverse measures such as eye-tracking, EEG, and physical sensors with user perception, we developed a data-driven methodology that addresses both the objective and subjective aspects of VR experiences. Through a series of studies, we demonstrated the practical application of this methodology and our different metrics in various use-cases, focusing on both hardware and software optimisation. Our approach contributes to a deeper understanding of how to effectively design and optimise VR systems, ensuring that they meet the needs of users while maintaining high levels of performance and immersion.

6.1 Contributions

6.1.1 VR-SOLUS

The VR-SOLUS method represents a novel approach to designing and optimising VR applications by integrating multiple dimensions of data—operational metrics, physiological responses, and user feedback. It offers a combination of traditional, user-based, methodologies with the novelty of data-driven approaches. The core of this methodology lies in its hybrid nature, blending objective data with subjective insights to create a more comprehensive understanding of the VR experience.

The VR-SOLUS method is based on the collection and analysis of multimodal data:

- **Operational Metrics:** these include performance-based measures such as task completion time, input errors, and multiple inputs usage rates. They also integrate use-case specific measures to answer the needs of distinct applications. These metrics provide quantitative extrinsic insights into how effectively a user interacts with the VR environment, highlighting areas where the system may need usability or efficiency enhancement;
- **Physiological Data:** using biometric sensors, such as eye-tracking, EEG, and IMUs, the method captures the user's physiological responses during interaction. This data reveals the cognitive and emotional states of the user, offering intrinsic insights into attention, cognitive load and physical engagement;
- **User Feedback:** subjective data is collected through surveys and other forms of user feedback to capture the user's perceptions, preferences, and overall satisfaction with the VR experience. These data adds a qualitative layer, which is essential for understanding the user's subjective experience.

These three streams of data are synchronised and analysed together, allowing for a multimodal evaluation of the VR application. The integration of these diverse data sources ensures that the VR-SOLUS method can identify correlations and disparities between how a user feels and how they perform, providing a comprehensive view of the VR experience (Fig. 6.1).

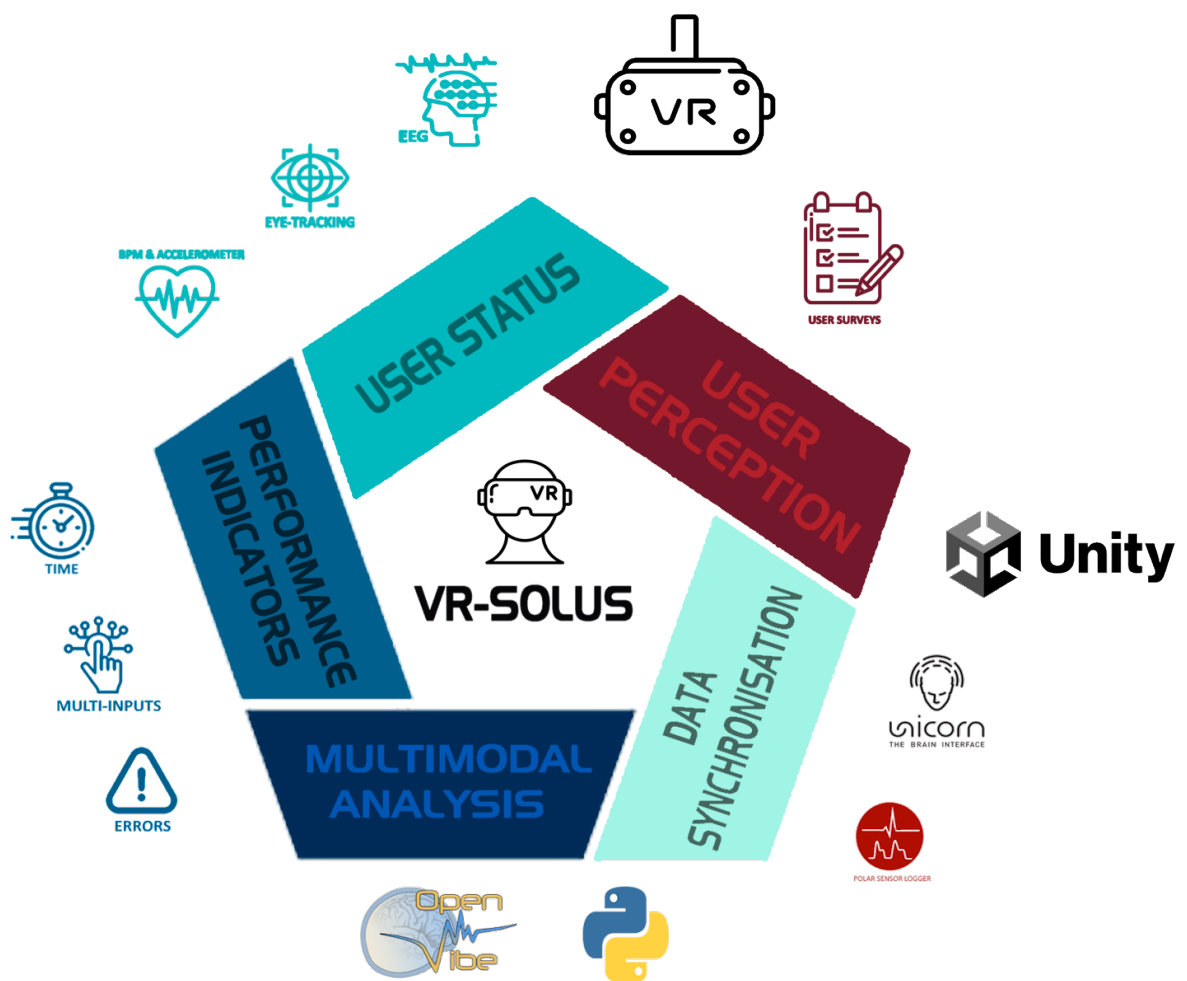


Figure 6.1: VR-SOLUS - Designing and Optimising Virtual Reality through a Hybrid Approach

The VR-SOLUS method offers several key advantages:

Comprehensive Analysis

By combining operational metrics with physiological data and user feedback, VR-SOLUS provides a well-rounded analysis that captures performance, user status and user perception. This enables developers to identify not only functional issues but also areas where the user experience may be enhanced;

Objective and Subjective Integration

The method bridges the gap between objective measures (like task performance and physiological responses) and subjective perceptions (such as user satisfaction). This integration is crucial for creating VR environments that are not only efficient but also engaging and comfortable for the user;

Data-Driven Decision Making

VR-SOLUS facilitates informed decision-making during the design and optimisation phases of VR development. By relying on concrete data from multiple sources, developers can prioritise changes and improvements that are backed by evidence rather than intuition alone;

Enhanced User Experience

Through its focus on physiological data, VR-SOLUS can uncover hidden factors that impact user experience, such as cognitive overload or discomfort, physical exertion or attention troubles, which might not be immediately apparent from performance metrics or user feedback alone;

Flexibility and Applicability

The methodology is versatile and can be adapted to a wide range of VR applications, from training simulations to gaming environments. It allows for the tailored analysis of different user groups and use-cases, ensuring that the VR application meets specific needs and contexts. Thanks to the multimodality of the used data, time projection is made possible, leading to an adaptive design along the life cycle of the VR application.

The VR-SOLUS method, therefore, represents a significant advancement in VR design and optimisation, providing a powerful toolkit for creating VR experiences that are both effective and user-friendly.

Through this thesis, we presented the VR-SOLUS method, its metrics and their use through multiple studies. Those studies belonged to two main use-cases: earthwork teleoperation and training, and dyslexia therapy. In the following section, we present the individual contributions of each study and their use in the final methodology.

6.1.2 Earthwork Use-case

The earthwork use-case section delves into a series of studies conducted in collaboration with HERACLES Robotics (Fig. 6.2, an innovative earthwork machinery automation company. This partnership has focused on exploring the potential of VR environments to enhance teleoperation and training within the earthwork sector. Through this collaboration, we have undertaken three distinct studies: an interface validation study, a controller selection study, and a software optimisation study centred on virtual annotations (VAs). These studies aimed to evaluate and optimise the tools and interfaces used in VR to ensure they meet the demands of both teleoperation and training in this specialised field. We made use of those studies for integrating, testing and improving the VR-SOLUS framework measures.



Figure 6.2: HERACLES Robotics, French Earthwork Machinery Automation

Interface validation

The interface validation study aimed to assess the effectiveness of a VR teleoperation interface by comparing it to the control of an actual excavator, courtesy of Heracles Robotics and their Cat323 excavator. The VR interface was developed using a digital twin of the excavator, which was designed and operated within the Unity game engine. This digital twin was seamlessly integrated with a SEMIL-1700 (Simulation Environment for Machine Interface Learning) system, which in turn communicated with the real machine via ROS (Robot Operating System).

The study included 10 participants, with an average age of 29.2 years, comprising both professional operators and non-professional users to ensure a comprehensive evaluation. Throughout the study, participants were tasked with executing a series of excavation actions in both the real and virtual environments (Fig. 6.3). The findings revealed that the VR interface produced performance that closely mirrored that of the actual machine, and closed the gap between professional (current operators) and non-professional users (future teleoperators). This outcome was consistent across both professional and non-professional users, thus confirming the viability of the VR interface for future applications in teleoperation development. The success of this validation highlights the potential of VR as a tool for not only training but also for practical teleoperation tasks in earthwork robotics.



Figure 6.3: Interface Comparison Setup Example - Real Machine (Left) versus VR Interface Remote Control (Right)

Study Contributions

First Integration of Operational Metrics

Testing completion time, inputs errors and multi-inputs percentage in an experimental setting for the first time. Validation of the metrics, leading to their definitive integration into the VR-SOLUS framework;

Testing of Use-case Specific Measures

Metrics specific to excavation operation were tested in both the real and virtual environments, with positive results, leading to the integration, when possible, of use-case specific measures into the VR-SOLUS method;

User Feedback on Hardware

A major flaw in our interface reported by the users was the VR controller and their mappings used during the experiment. This sparked the creation of a new study: the hardware selection study;

Validation of the VR Interface

Confirmation that our VR interface was operational on the use case of excavation teleoperation, providing a stable environment for future studies.

Associated Publications

SAUNIER, Léa, HOFFMANN, Nicolas, PREDA, Marius, et al. Virtual Reality Interface Evaluation for Earthwork Teleoperation. *Electronics*, 2023, vol. 12, no 19, p. 4151.

HOFFMANN, Nicolas, SAUNIER, Léa, PROUTEN, Samuel, et al. Industrial use-case: Digital twin for autonomous earthwork in virtual-reality. In : *Proceedings of the 27th International Conference on 3D Web Technology*. 2022. p. 1-4.

Hardware Selection

The hardware selection study aimed to identify the most effective control interface for earthwork training within a simulated virtual environment. The study evaluated three consumer-grade controllers positioned along a spectrum from gamer to professional. The study involved a standard gamepad representing common gaming controllers, an improved version of our VR controller offering enhanced interaction capabilities, and simulation joysticks with a mapping nearly identical to the real machine (Fig. 6.4).

A total of 13 participants with an average age of 28.1 years were involved in the study. Each participant performed a series of standardised earthwork tasks using all three controllers, allowing for a comprehensive comparison of performance and user experience across the different interfaces. The tasks were designed to mimic real-world excavation and earth-moving activities, providing relevant and practical insights into each controller's effectiveness.

The results demonstrated that the professional-grade joysticks outperformed both the gamepad and the improved VR controller across various performance metrics and physiological measures. Participants achieved greater precision, efficiency, and control when using the joysticks, despite lower levels of prior familiarity with such equipment. This finding underscores the importance of using industry-standard abiding hardware in training simulations to better prepare users for actual operational scenarios. It also highlights that the benefits of professional design can outweigh the comfort and familiarity associated with more common controllers.

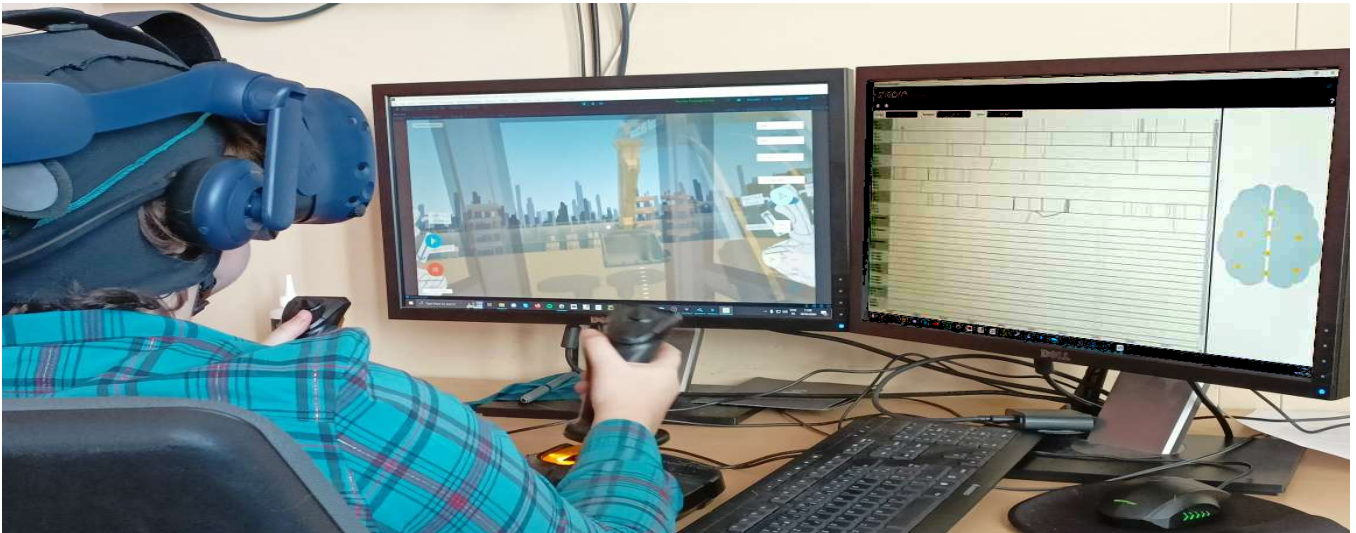


Figure 6.4: Hardware Selection Study Setup

Study Contributions

First Integration of Physiological Metrics

This study not only featured the previous operational measures, but also the integration of bio-sensors: eye-tracker and electroencephalogram (EEG). This enhancement of the VR-SOLUS framework showed promising results, as well as new axes for development (e.g. improved eye-tracking and EEG filtering, no current measure of physical load to be remedied);

Metrics Correlation

More metrics lead to more analysis and the question of correlation for synergy study and validation. This study has already shown the strong correlation between operational measures, and the lack of correlation between cognitive load and performance indicators. However the correlation between user feedback and mental load measures was already manifesting;

Hardware Selection

Joysticks were the winner of this hardware competition across multiple measures, and thus kept for the following study on software optimisation.

Associated Publications

SAUNIER, Léa, PREDA, Marius, et FETITA, Catalin. Controller Evaluation for Earthwork Teleoperation and Training in Virtual Reality. In : *2024 16th International Conference on Human System Interaction (HSI)*. IEEE, 2024. p. 1-6.

SAUNIER, Léa, PREDA, Marius, et FETITA, Catalin. Hybrid Methodology Using Electroencephalogram and Eye-tracking for Virtual Reality Design and Optimization. In : *2024 23rd IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*

Software Design

The software optimisation study focused on the selection and refinement of virtual annotations (VAs) and their display modalities within a VR earthwork training simulation (Fig. 6.5). The research was divided into two distinct phases:

- **Design Phase:** the VAs were empirically defined, with users participating in tests to provide comprehensive data. During this phase, operational metrics, physiological signals, and user feedback were collected to evaluate the effectiveness of the various VAs;
- **Evaluation Phase:** the second part of the study took this process further by assessing three different VA sets, each developed using a unique methodological approach. One set was based entirely on user feedback, another set was derived solely from objective measures, and the final set employed a hybrid approach, combining both objective and subjective data to create a more balanced and potentially effective solution.

The study involved 30 participants, with an average age of 33.4 years, who engaged in earthwork training using joysticks within the VR simulation. The results revealed significant insights: methodologies relying exclusively on objective measures were insufficient in capturing the full extent of user experience, often missing subtle but important aspects of user interaction and satisfaction. On the other hand, relying solely on user feedback proved to be too subjective, leading to difficulties in generalisation. The hybrid methodology demonstrated moderate but consistent improvements across all measures. This balanced approach ultimately validated the VR-SOLUS methodology, confirming its potential for enhancing VR training applications by providing a more comprehensive understanding of user experience.

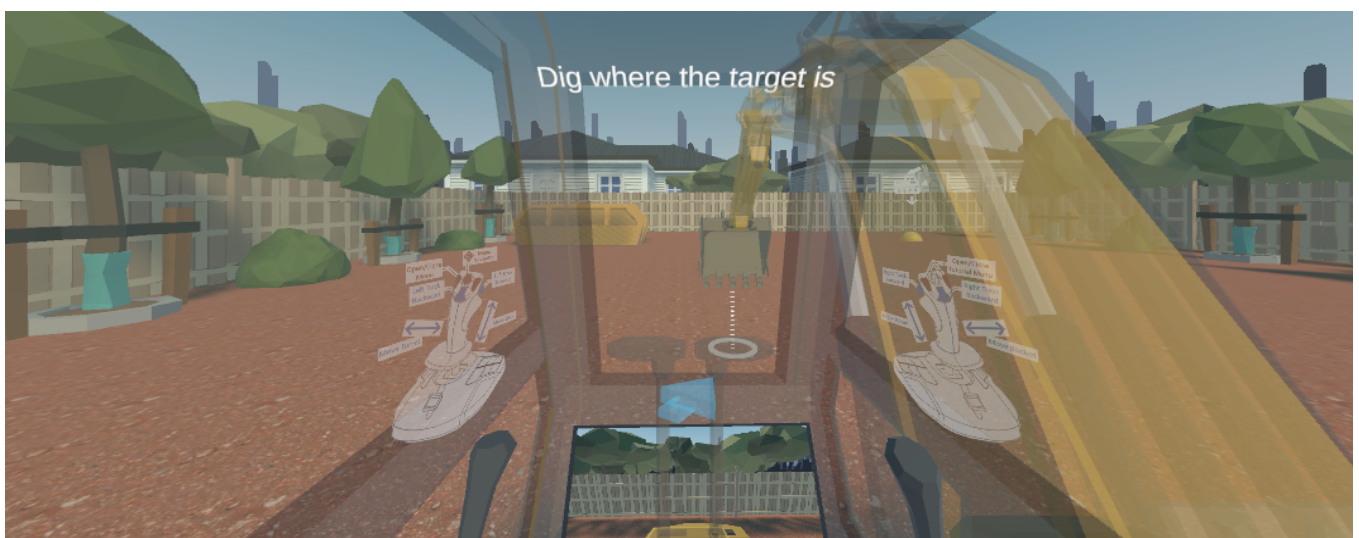


Figure 6.5: Example of Interface used for the Software Optimisation Study

Study Contributions

First Integration of Physical Load Metrics

Through the Polar OH1+ armband, physical activity was recorded for arm movement, as well as heart rate monitoring. Associated to head motions, this allowed for the first integration of physical load assessment into the VR-SOLUS framework;

Advanced Data Analysis

Better filtering of EEG and eye-tracking data, addition of physical data, more user feedback surveys, more data to correlate, this study allowed for the advanced analysis present in the VR-SOLUS methodology;

Software Interface

Selection of VAs and their modalities improved greatly the earthwork training and teleoperation interface developed during this thesis;

Hybrid Methodology Validation

Confronting traditional user feedback methodology with objective data only method and hybrid methodology confirmed the superiority of a hybrid methodology on many aspects, ultimately validating our VR-SOLUS methodology.

6.1.3 Therapeutic Use-case

The dyslexia therapy study focused on developing a prototype aimed at enhancing therapeutic interventions through an integrated approach. Although no formal clinical trials were conducted, the prototype - named Dys&Dragons - was developed and validated by three orthoptists and one neurologist. The therapy itself was centred around visual exercises (e.g. eye movements, visual search, reading) (Fig. 6.6) designed to aid individuals with dyslexia, targeting the specific challenges they face in processing visual information, making use of VR environment to assess 3D perception.

The core idea of this prototype was the study of the VR-SOLUS methodology combined with expert evaluations from medical professionals in adaptive design (i.e continuously evolving design). By incorporating these diverse data sources, the system was intended to dynamically adjust the therapy to meet the individual needs of each patient in real-time, ensuring a more personalised and effective treatment process. This approach was aimed at creating a therapeutic environment that not only responded to the user's immediate performance but also adapted over time, providing an increasingly tailored and effective therapy experience.



Figure 6.6: Dys&Dragons Interface Example

Study Contributions

Study of a Different Use-case

This study being placed in the therapeutic field, a completely distinct use-case from the earthwork field, allows us to gain insight on the generalisation of our methodology to very different fields, with very different needs. This study, although only in the prototype phase, was partnered with medical specialists, who provided very useful feedback. This showed that user feedback can come from multiple actors of VR applications, whether they are using the application themselves (patients) or monitoring it (doctors);

Impact of 3D on Dyslexia Therapy

Dys&Dragons first step was the "translation" of current traditional visual exercises used by doctors in a 2D setting into a new virtual 3D environment, done in the presented prototype. Once the value of 3D technologies has been proven, the idea was to develop brand new exercises with medical professionals to open a new page in dyslexia therapy;

Long-term Adaptive Design

Therapy is a long process. This study was meant to evaluate our VR-SOLUS methodology on the long term, seeing if it was as efficient in the design phase of a VR application as during its life cycle. Through this experiment, VR-SOLUS method could be used as a continuously adaptive design method.

Associated Publication

SAUNIER, Lea, PANOUILLERES, Muriel, FETITA, Catalin, et al. Visual Rehabilitation for Learning Disorders in Virtual Reality: Visual Rehabilitation for Learning Disorder in VR. In : *Proceedings of the 27th International Conference on 3D Web Technology. 2022.* p. 1-4.

6.2 Limitations

In this section, we will discuss the limitations encountered throughout the study, focusing on areas that may have impacted the outcomes or the general applicability of the findings. These limitations are manifold, and can be grouped around three axes: the experimental setup, the software used, and the hardware employed in the research. Each of these axes presents unique challenges that need to be considered when interpreting the results and understanding the scope of the conclusions drawn from this work. By examining these aspects, we aim to provide a balanced view of the study's constraints, offering insights into the potential areas for future improvement and refinement.

6.2.1 Experimental Setup

One of the primary limitations of this study lies within the constraints of the experimental setup, particularly in two significant areas: the environment in which the experiments were conducted and the characteristics of the participant pool.

Experimental Environment

Firstly, the laboratory environment in which we conducted our experiments was not a dedicated, interference-free EEG lab. This setting inevitably introduced noise and external interference that could have impacted the quality and reliability of our EEG signals. Although we implemented various denoising techniques to mitigate these effects, the lack of a controlled EEG environment may have influenced the results. The integrated eye-tracker inside of the VR headset not only presents filtering challenges, but also lacks peripheral vision analysis. Also, all our studies required very low physical engagement, making physical load hard to measure. Changing the setup to allow more movement, including whole body movements, would greatly help with physical demand assessment. Additionally, for the dyslexia therapy study, we did not have access to a clinical setting, which would have provided a robust environment for evaluating the therapy's effectiveness in a real-world context.

Participants

The statistical one-way ANOVA tests conducted in our studies (see Table C.1 for the full ANOVA values) yielded non-significant results across most measures, which can primarily be attributed to the relatively small sample size of participants. With only 10 to 16 individuals involved, the statistical power of the tests was inherently limited, making it challenging to detect subtle differences between the experimental conditions. Despite this limitation, the studies still provides valuable insights, particularly in the context of eye-tracking data, where significant results were observed in the virtual annotation study. This suggests that while the sample size may have constrained the overall effectiveness of the ANOVA, the methodology and experimental design are robust, and the findings remain relevant. In addition, operational measures demonstrated strong correlations with time, as evidenced by Pearson correlation coefficients shown in Section 5.3. The cognitive load evaluation with the Theta Alpha Ratio (TAR), which lacked statistical significance, exhibited a low correlation with time and a moderate one with user level. These findings underscore the need for caution in interpreting the statistical significance of our physiological measures and the need for larger scale evaluation before generalising our methodology. Therefore, our results should be viewed more as initial explorations or proof-of-concept rather than definitive conclusions. Increasing the participant pool in future studies will be crucial to enhancing the validity and generalisability of our findings.

6.2.2 Software Improvement

Another limitation concerns the software part of this thesis on two aspects: the VR simulation and the data analysis.

Simulation

Our VR simulation exhibits certain limitations that warrant further development and refinement. While the physics engine behind the simulation effectively replicates many aspects of the earthwork tasks, there is room for significant improvement. The physics could be more precise, especially in handling complex interactions between the machinery and the terrain. Furthermore, a more detailed and accurate digital twin would enhance the realism and reliability of the training and teleoperation tasks. One of the critical gaps in our current earthwork simulation is the lack of real-time environmental data during teleoperation tasks. At present, the simulation relies heavily on the proprioception of the machine and a basic model of terrain deformation (courtesy of Open Construction Simulator). However, in real-world applications, teleoperation would benefit from integrating data from LiDAR [184], cameras [185], and other sensors mounted on the machinery. These sensors could provide a real-time (or near real-time) rendering of the terrain, offering the operator a more accurate and dynamic view of the working environment. This enhancement is crucial for transitioning from simulated tasks to actual teleoperation scenarios safely.

In the therapeutic study focused on dyslexia, our VR simulation was limited by its reliance on translating existing exercises into a 3D virtual format. While this approach allowed us to test the viability of VR for therapy, it constrained our ability to innovate and develop exercises specifically tailored to the VR environment. To overcome this limitation, we would need more extensive involvement from medical experts, particularly in the design phase, to create new types of exercises that leverage the unique capabilities of VR. Additionally, conducting a clinical study would be necessary to validate these exercises and ensure their effectiveness in a therapeutic context.

Data Analysis

Our data analysis, while robust, faces several limitations that impact the precision and reliability of our findings. Each of our sensors — ranging from eye-tracking and IMU devices to heart rate monitors and EEG — could benefit from more advanced filtering and analysis techniques. Improved filtering would help isolate relevant signals from noise, while more sophisticated analysis could yield deeper insights into the data. Advanced correlation study could also benefit our method. One of the primary challenges in our data analysis process is the synchronisation of data from different sensors. In traditional setups, synchronisation is often managed with dedicated hardware, ensuring precise alignment of data streams from different modalities. However, in our study, synchronisation was handled through software, which, while functional, introduces potential variations, especially when dealing with data that requires very precise timing. These disparities could lead to slight but significant misalignment in the data, affecting the accuracy of our analysis.

The sensor that poses the greatest challenge in terms of filtering and analysis is the EEG. Our EEG system, which uses only 8 electrodes, is particularly susceptible to noise (e.g. motion, friction with the VR headset, experimental setup, etc.). This noise complicates the extraction of clear signals and requires extensive processing to achieve usable data. Our filtering methodology revolved around frequency band isolation (and thus unwanted band removal) [155] as well as amplitude threshold technique [182] combined with IMU data to remove motion noise. Plenty other analysis technique exist and could be applied to our study to improve our EEG signal, such as (not limited to):

- **Independent Component Analysis (ICA):** ICA is a statistical method that separates EEG signals into independent sources, which can help isolate brain activity from artifacts such as eye movements or muscle activity [175];
- **Wavelet Transform:** this technique decomposes the EEG signal into different frequency components, which can then be analysed or filtered separately [176]. It is particularly useful for detecting short-term events and removing noise;
- **Artifact Subspace Reconstruction (ASR):** ASR is a method for removing brief artifacts from EEG data by reconstructing the clean signal from a lower-dimensional subspace, effectively filtering out noise [186];
- **Common Average Referencing (CAR):** CAR is a spatial filtering technique that re-references the EEG signals to the average of all electrodes, reducing the influence of common noise sources across the scalp [187].

6.2.3 Hardware Improvement

Our study's hardware limitations have a significant impact on both the quality of data collected and the effectiveness of the software solutions we employ. The current setup, while functional, highlights several areas where improvements in hardware could substantially enhance our research outcomes. One of the primary limitations is the EEG system, which is equipped with only 8 electrodes. This limited number of electrodes constrains our ability to capture detailed and comprehensive brain activity data. More electrodes would provide finer spatial resolution and a more accurate representation of the neural signals, leading to more precise insights into cognitive states, but at a higher financial cost. Additionally, the current EEG setup does not integrate seamlessly with the VR headset, resulting in discomfort and potential signal noise due to friction and motion. An EEG system that is better designed to fit with the VR headset, or ideally, one that is fully integrated into the headset, would mitigate these issues, enhancing both the quality of the data and the participant experience.

Beyond EEG, the overall sensor setup could benefit from enhancements to improve data capture and synchronisation. While our software manages the synchronisation of various data streams, hardware-based synchronisation would offer greater precision, especially when dealing with the fine temporal resolution required for analysis. This would reduce potential offsetting in the data, improving the reliability of the results. This would also allow to record the different data on different device, lessening the load on the unique computer used in this thesis, that had to run the VR simulation and record all the multimodal data at the same time, impacting the performance. Moreover, more sophisticated sensors, with better motion compatibility and integrated synchronisation, would allow for more complex and accurate analyses. The need for better-integrated hardware is also evident in the potential for using advanced sensors that can be directly integrated into the VR hardware, enabling more real-time and accurate data collection.

The reliability and precision of biosensors in VR are widely debated in the research community, with several issues limiting their effectiveness for research [188]. Key challenges include signal quality variability, noise susceptibility, and difficulties in ensuring synchronisation across multiple data streams. These factors affect the usability of biosensors like EEG and eye-tracking systems in real-world VR applications. Research has shown that EEG signals, for instance, are prone to motion artifacts, making their use in highly interactive VR settings challenging. Efforts to improve artifact rejection and enhance signal clarity have been explored but remain insufficient for all applications [189]. Eye-tracking systems, while generally robust for measuring attention and gaze patterns, can face calibration inconsistencies and inter-individual variability that undermine their precision [190]. Multimodal setups often require advanced synchronisation protocols to ensure that signals from various biosensors align temporally and contextually, adding complexity to experimental designs. These challenges point to the necessity for more standardised analysis and enhanced data processing algorithms to improve the reliability of biosensor data in VR research.

6.3 Further Developments

As we look to advance our research and resolve our method's limitations, several key areas for further development have been identified that will enhance both the robustness and generalisability of our findings. First, the primary focus for future work is the expansion of our participant pool. Increasing the number of participants across different use-cases will allow us to obtain more reliable data, leading to results that are not only statistically significant but also generalisable to a broader range of applications. This expansion could also include testing additional measures to provide a more comprehensive understanding of the various factors at play in our studies.

Another promising path for future research lies in studies involving long-term monitoring, particularly in adaptive design applications such as was the idea of our therapeutic proposition. Long-term monitoring would enable us to track changes over time, providing insights into the sustained impact of different interventions and the potential for real-time adjustments based on user feedback and physiological data (Fig. 6.7). This is particularly important for application with a transformative goal, notably therapy and training.

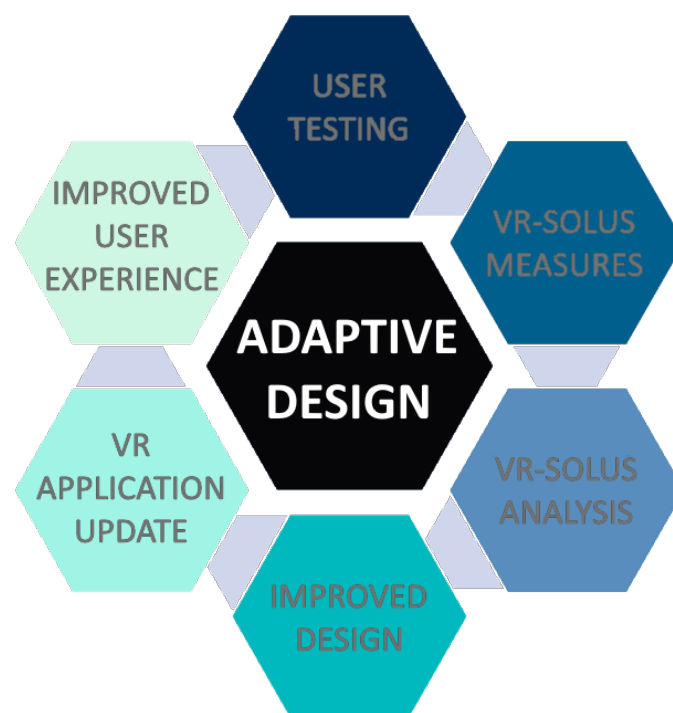


Figure 6.7: VR-SOLUS and Adaptive Design

In terms of data analysis, our biggest identified limitation has been the current handling of EEG data. Future work should explore alternative EEG cognitive load assessment techniques such as (not limited to):

- **Event-Related Potentials (ERPs):** ERPs involve analyzing the brain's response to specific sensory, cognitive, or motor events. The technique identifies changes in EEG signals that are time-locked to these events, providing insights into cognitive processing;
- **Power Spectral Density (PSD) Analysis:** this technique examines the distribution of power across various frequency bands (e.g., alpha, beta, gamma) to assess cognitive load. Different tasks and cognitive states are associated with shifts in power across these bands. While we focused only on the Alpha and Theta band, the exploration of other bands could provide a more in-depth cognitive load evaluation;

- **Task-Evoked EEG:** this method focuses on changes in EEG signals evoked by specific tasks, such as those requiring attention or working memory. It is often used to assess cognitive load during complex tasks. It could synchronise well with our other sensors data (e.g. eye-tracking, IMUs, heart rate monitor).

Our EEG setup could also benefit from more advanced EEG filtering methods, as stated in the Section 6.2. Testing and comparing these new techniques could provide the next step for the improvement of the accuracy of our cognitive load measurements and the reduction of the noise inherent to our current setup. This could include experimenting with more sophisticated algorithms or making use of machine learning techniques to better isolate meaningful signals from noise.

Another significant area for development is the hardware we use, particularly in relation to EEG and other physiological sensors. Integrating more sensors into the VR-SOLUS, could provide more exploitable data: for instance, adding electromyogram (EMG) into the framework could provide a more refined physical load evaluation. Integrating those sensors directly into the VR headset, especially the EEG sensors, would also greatly enhance data quality. Such integration would reduce issues related to synchronisation and signal noise, as well as improve user comfort by minimizing the bulk and complexity of the equipment. This approach could lead to more accurate data collection, less interference from external factors, and ultimately, a better overall experience for the user. This perspective on the future of VR headset is already a reality, with more and more manufacturers integrating diverse bio-sensors in their headset, such as the Galea headset ¹ with integrated eye-tracking, EEG and EMG (Fig. 6.8).

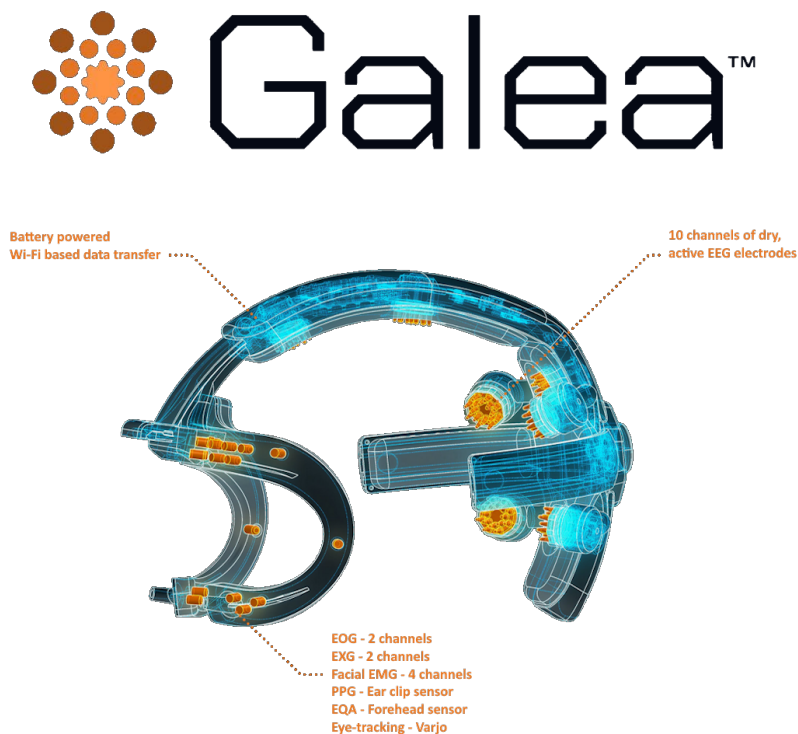


Figure 6.8: Galea VR Headset, the Next Generation of Bio-sensors Integration VR Hardware

To conclude, further development in these areas will not only address the limitations of our current research but also pave the way for more sophisticated and generalisable studies in the future. By expanding our participant base, refining our analytical techniques, and improving our hardware, VR-SOLUS will be well-positioned to push the boundaries of what is possible in VR-based physiological studies.

¹Galea Next-gen VR Headset <https://galea.co/>

Chapter 7

Conclusion

The design and optimisation of user experience and human-machine interaction in virtual reality (VR) represents a rapidly evolving field, driven by the need to create immersive, intuitive, and effective virtual environments, adapted to multiple use-cases. Traditional design methodologies in this domain rely on user feedback, obtaining subjective measures primarily through surveys, to gauge the efficiency of VR applications. While user feedback is invaluable for capturing personal experiences and preferences, it can lack the objectivity needed to provide comprehensive insights into the usability and performance of VR applications. In contrast, newer data-driven methodologies emphasise on objective measures such as operational metrics and physiological data, offering a more quantifiable approach to understanding user interactions within VR. However, these methodologies can overlook the nuanced and personal aspects of user experience that are crucial for refining VR systems.

In response to these challenges, this thesis presented the Sensory, Operational and User data Synthesis methodology for Virtual Reality, VR-SOLUS (Fig. 7.1). This novel hybrid approach integrates operational data - completion time, inputs errors, multiple inputs usage rate, as well as use-case-specific data, physiological measures - eye-tracking for attention, electroencephalogram (EEG) for cognitive load, Inertial Measurement Units (IMUs) and heart rate monitor for physical load, and user feedback. These measures provide performance indicators, user status insights and user perception assessments to create a more extensive framework for VR design and optimisation. This methodology aims to combine the strengths of both traditional and data-driven approaches, offering a broader spectrum of analysis that can capture both the quantitative and qualitative aspects of user interaction in VR.

The studies conducted within this research, particularly in the context of earthwork teleoperation and training and dyslexia therapy, demonstrate the effectiveness of the VR-SOLUS hybrid methodology. Our results indicate that a hybrid approach outperforms purely objective or purely subjective methodologies. Data correlation shows that even if objective and subjective data answer the same question (e.g. how mentally demanding is a task?), they show nuanced results, sometimes similar (e.g. low according to user feedback and EEG), sometimes opposite (e.g. low according to user feedback and high according to EEG), but always complementary. By integrating operational metrics, physiological data, and user feedback, the VR-SOLUS methodology offers researchers and developers a more refined analysis of user experience. This broader perspective leads to VR systems that are not only more efficient and effective but also more attuned to the needs and preferences of users.

To further showcase the contributions of this thesis, we revisit our three fundamental questions:

Which operational measures are valuable for VR development?

Our studies have shown that operational measures, such as task completion time, input errors, input usage and specific use-case measures are essential for evaluating the performance and efficiency of VR systems. These metrics provide performance indicators and concrete data on how users interact with VR environments, offering extrinsic insights into usability and identifying areas for improvement.

How to employ physiological sensors efficiently?

Physiological sensors, such as eye-tracking, EEG, IMUs and heart rate monitors, have been shown to offer valuable insights into the user's attention, cognitive state and physical state during VR interactions.

However, their effective use requires careful consideration of sensor placement, signal processing, and data interpretation. Our research emphasises the importance of integrating intrinsic physiological data with operational metrics and user feedback to gain a more complete understanding of user experience.

What kind of methodology can further improve VR design and optimisation?

The VR-SOLUS hybrid methodology represents a significant advancement in VR design and optimisation. By combining objective and subjective measures, this approach provides a more nuanced and comprehensive analysis of user interaction in VR. Far from invalidating traditional user feedback methods or data-driven method, this thesis identifies their strong points, and synergise them in a new hybrid methodology. The VR-SOLUS method has been validated across various studies, providing use-cases from different fields, demonstrating its potential to enhance the development and optimisation of VR systems.

Looking forward, there are several paths for future research that can build upon the findings of this thesis. One promising direction is the integration of new bio-sensors, such as EMG and headset-integrated EEG systems, which could offer more detailed and reliable physiological data while reducing issues related to sensor synchronisation and user discomfort. Additionally, further advancements in data analysis techniques, particularly in EEG signal processing, could lead to new insights into cognitive load and user experience in VR. Another potential area of exploration is the synergy between different data types — operational, physiological, and subjective — which could reveal new patterns and correlations that are currently not fully understood. Once those new measures and their analysis are in place in the VR-SOLUS framework, it could be used in a long-term manner, to serve as an adaptive design method for the whole application life-cycle, endlessly improving VR design and optimisation.

In conclusion, this thesis contributes to the field of VR design and optimisation by proposing and validating a hybrid methodology that applies the strengths of both traditional and data-driven approaches. The VR-SOLUS methodology has been shown to provide a more comprehensive and effective framework for understanding and enhancing user experience in VR, paving the way for future advancements in this dynamic and rapidly growing field.

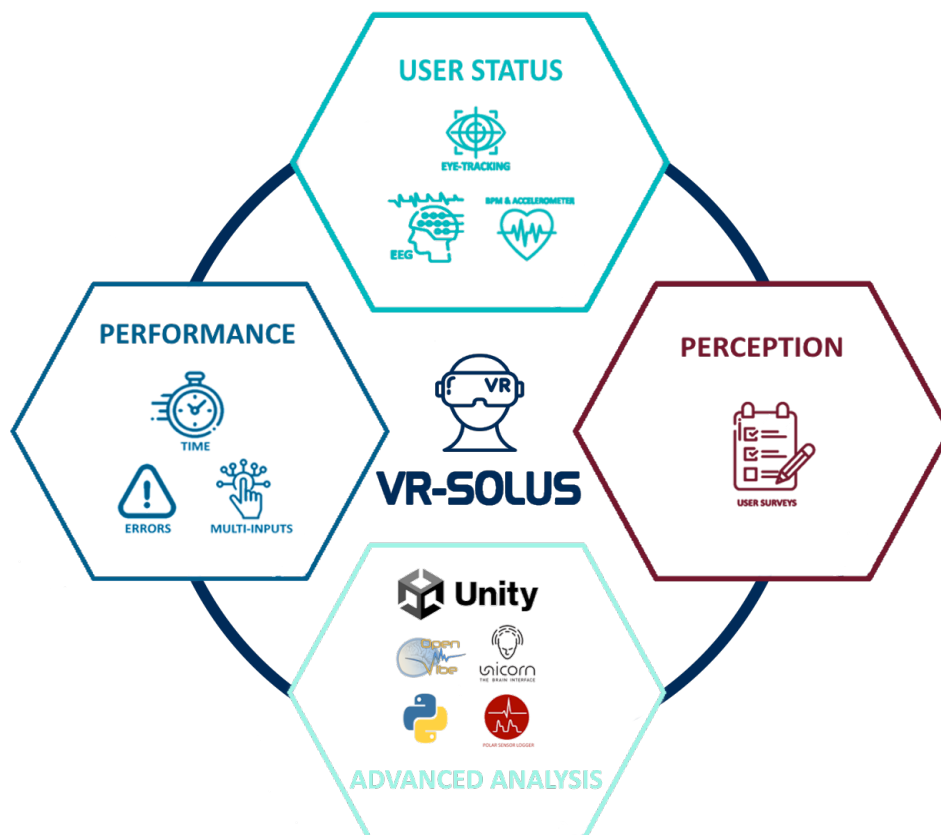


Figure 7.1: Sensory, Operational and User data Synthesis methodology for Virtual Reality - VR-SOLUS

References

- [1] Grigore C Burdea and Philippe Coiffet. *Virtual reality technology*. John Wiley & Sons, 2024.
- [2] Gartner Research. *Emerging Tech: Revenue Opportunity Projection for AR and VR Head-Mounted Displays*. 2024. URL: <https://www.gartner.com/en/documents/5179263>.
- [3] George Coates Performance Works and George Coates. *Program from Invisible Site—a Virtual Sho, a multimedia performance work*. Presented in March 1992. San Francisco, CA: George Coates Performance Works, 1992.
- [4] Peter Greenbaum. “The Lawnmower Man”. In: *Film and Video* 9.3 (Mar. 1992), pp. 58–62.
- [5] J Steuer. “Defining virtual reality: Dimensions determining telepresence”. In: *Communication in the age of virtual reality/Lawrence Erlbaum and Associates* (1995).
- [6] M Eswaran and MVA Raju Bahubalendruni. “Challenges and opportunities on AR/VR technologies for manufacturing systems in the context of industry 4.0: A state of the art review”. In: *Journal of Manufacturing Systems* 65 (2022), pp. 260–278.
- [7] Amit Dang, Dimple Arora, and Pawan Rane. “Role of digital therapeutics and the changing future of healthcare”. In: *Journal of Family Medicine and Primary Care* 9.5 (2020), pp. 2207–2213.
- [8] Kar-Long Chan et al. “Cloud-based VR gaming: our vision on improving the accessibility of VR gaming”. In: *2017 International Symposium on Ubiquitous Virtual Reality (ISUVR)*. IEEE. 2017, pp. 24–25.
- [9] Nicolás Dozio et al. “A design methodology for affective Virtual Reality”. In: *International Journal of Human-Computer Studies* 162 (2022), p. 102791.
- [10] Dmitry Alexandrovsky et al. “Examining design choices of questionnaires in VR user studies”. In: *Proceedings of the 2020 CHI conference on human factors in computing systems*. 2020, pp. 1–21.
- [11] Jing Qu et al. “Developing a virtual reality healthcare product based on data-driven concepts: A case study”. In: *Advanced Engineering Informatics* 57 (2023), p. 102118.
- [12] Henar Guillen-Sanz et al. “A systematic review of wearable biosensor usage in immersive virtual reality experiences”. In: *Virtual Reality* 28.2 (2024), p. 74.
- [13] Rosemarie Anderson. “Intuitive inquiry: Interpreting objective and subjective data”. In: *ReVision* 22.4 (2000), pp. 31–31.
- [14] Jason Jerald. *The VR book: Human-centered design for virtual reality*. Morgan & Claypool, 2015.
- [15] Thomas Potter, Zoran Cvetković, and Enzo De Sena. “On the relative importance of visual and spatial audio rendering on vr immersion”. In: *Frontiers in Signal Processing* 2 (2022), p. 904866.
- [16] Brendan Mouatt et al. “The use of virtual reality to influence motivation, affect, enjoyment, and engagement during exercise: A scoping review”. In: *Frontiers in Virtual Reality* 1 (2020), p. 564664.
- [17] Shiqian Ke et al. “A enhanced interaction framework based on VR, AR and MR in digital twin”. In: *Procedia Cirp* 83 (2019), pp. 753–758.
- [18] Wen Huang et al. “Motivation, engagement, and performance across multiple virtual reality sessions and levels of immersion”. In: *Journal of Computer Assisted Learning* 37.3 (2021), pp. 745–758.
- [19] SN David Chua et al. “Hand gesture control for human–computer interaction with Deep Learning”. In: *Journal of Electrical Engineering & Technology* 17.3 (2022), pp. 1961–1970.
- [20] Nikolaos Sidorakis, George Alex Koulouris, and Katerina Mania. “Binocular eye-tracking for the control of a 3D immersive multimedia user interface”. In: *2015 IEEE 1St workshop on everyday virtual reality (WEVR)*. IEEE. 2015, pp. 15–18.

- [21] Simon Davis, Keith Nesbitt, and Eugene Nalivaiko. "A systematic review of cybersickness". In: *Proceedings of the 2014 conference on interactive entertainment*. 2014, pp. 1–9.
- [22] Bernhard E Riecke and Daniel Zielasko. "Continuous vs. discontinuous (teleport) locomotion in vr: How implications can provide both benefits and disadvantages". In: *2021 IEEE conference on virtual reality and 3D user interfaces abstracts and workshops (VRW)*. IEEE. 2021, pp. 373–374.
- [23] Iqra Arshad et al. "Reducing cybersickness in 360-degree virtual reality". In: *Multisensory Research* 35.2 (2021), pp. 203–219.
- [24] Steven Vi, Tiago Silva da Silva, and Frank Maurer. "User experience guidelines for designing HMD extended reality applications". In: *Human-Computer Interaction–INTERACT 2019: 17th IFIP TC 13 International Conference, Paphos, Cyprus, September 2–6, 2019, Proceedings, Part IV* 17. Springer. 2019, pp. 319–341.
- [25] Alexis D Souchet et al. "A narrative review of immersive virtual reality's ergonomics and risks at the workplace: cybersickness, visual fatigue, muscular fatigue, acute stress, and mental overload". In: *Virtual Reality* 27.1 (2023), pp. 19–50.
- [26] Mohammadhossein Moghim et al. "Adaptive virtual environments: A physiological feedback HCI system concept". In: *2015 7th Computer Science and Electronic Engineering Conference (CEECE)*. IEEE. 2015, pp. 123–128.
- [27] Renee Wever, Jasper Van Kuijk, and Casper Boks. "User-centred design for sustainable behaviour". In: *International journal of sustainable engineering* 1.1 (2008), pp. 9–20.
- [28] Dmitry Alexandrovsky et al. "Examining design choices of questionnaires in VR user studies". In: *Proceedings of the 2020 CHI conference on human factors in computing systems*. 2020, pp. 1–21.
- [29] Sandra G Hart and Lowell E Staveland. "Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research". In: *Advances in psychology*. Vol. 52. Elsevier, 1988, pp. 139–183.
- [30] John Brooke. "SUS: a retrospective". In: *Journal of usability studies* 8.2 (2013), pp. 29–40.
- [31] Robert S Kennedy et al. "Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness". In: *The international journal of aviation psychology* 3.3 (1993), pp. 203–220.
- [32] Bob G Witmer and Michael J Singer. "Measuring presence in virtual environments: A presence questionnaire". In: *Presence* 7.3 (1998), pp. 225–240.
- [33] Guillaume Vailland et al. "User-centered design of a multisensory power wheelchair simulator: towards training and rehabilitation applications". In: *2019 IEEE 16th International Conference on Rehabilitation Robotics (ICORR)*. IEEE. 2019, pp. 77–82.
- [34] Andrew Bell et al. "Assessing recall bias and measurement error in high-frequency social data collection for human-environment research". In: *Population and Environment* 40 (2019), pp. 325–345.
- [35] Rodrigo Moreno-Serra et al. "Addressing recall bias in (post-) conflict data collection and analysis: lessons from a large-scale health survey in Colombia". In: *Conflict and health* 16.1 (2022), p. 14.
- [36] Martin Feick et al. "The virtual reality questionnaire toolkit". In: *Adjunct Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*. 2020, pp. 68–69.
- [37] Roy F Baumeister, Kathleen D Vohs, and David C Funder. "Psychology as the science of self-reports and finger movements: Whatever happened to actual behavior?" In: *Perspectives on psychological science* 2.4 (2007), pp. 396–403.
- [38] Sara Vlahovic, Mirko Suznjevic, and Lea Skorin-Kapov. "A survey of challenges and methods for Quality of Experience assessment of interactive VR applications". In: *Journal on Multimodal User Interfaces* 16.3 (2022), pp. 257–291.
- [39] Kasper Hornbæk and Morten Hertzum. "Technology acceptance and user experience: A review of the experiential component in HCI". In: *ACM Transactions on Computer-Human Interaction (TOCHI)* 24.5 (2017), pp. 1–30.
- [40] Foster Provost and Tom Fawcett. *Data Science for Business: What you need to know about data mining and data-analytic thinking*. " O'Reilly Media, Inc.", 2013.
- [41] Doug A Bowman and Larry F Hodges. "Formalizing the design, evaluation, and application of interaction techniques for immersive virtual environments". In: *Journal of Visual Languages & Computing* 10.1 (1999), pp. 37–53.

- [42] Mel Slater. "How colorful was your day? Why questionnaires cannot assess presence in virtual environments". In: *Presence* 13.4 (2004), pp. 484–493.
- [43] Doug A Bowman, Donald B Johnson, and Larry F Hodges. "Testbed evaluation of virtual environment interaction techniques". In: *Proceedings of the ACM symposium on Virtual reality software and technology*. 1999, pp. 26–33.
- [44] Paul Lubos, Gerd Bruder, and Frank Steinicke. "Analysis of direct selection in head-mounted display environments". In: *2014 IEEE Symposium on 3D User Interfaces (3DUI)*. IEEE. 2014, pp. 11–18.
- [45] Eike Langbehn, Paul Lubos, and Frank Steinicke. "Evaluation of locomotion techniques for room-scale vr: Joystick, teleportation, and redirected walking". In: *Proceedings of the Virtual Reality International Conference-Laval Virtual*. 2018, pp. 1–9.
- [46] Anil Ufuk Batmaz, Michel de Mathelin, and Birgitta Dresch-Langley. "Getting nowhere fast: Trade-off between speed and precision in training to execute image-guided hand-tool movements". In: *BMC psychology* 4 (2016), pp. 1–19.
- [47] Christoph Anthes et al. "State of the art of virtual reality technology". In: *2016 IEEE aerospace conference*. IEEE. 2016, pp. 1–19.
- [48] Ferran Argelaguet Sanz et al. "Virtual proxemics: Locomotion in the presence of obstacles in large immersive projection environments". In: *2015 IEEE virtual reality (vr)*. IEEE. 2015, pp. 75–80.
- [49] Frederik Winther et al. "Design and evaluation of a VR training simulation for pump maintenance". In: *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*. 2020, pp. 1–8.
- [50] Manargul Mukasheva et al. "Contextual structure as an approach to the study of virtual reality learning environment". In: *Cogent Education* 10.1 (2023), p. 2165788.
- [51] Andreas Halbig and Marc Erich Latoschik. "A systematic review of physiological measurements, factors, methods, and applications in virtual reality". In: *Frontiers in Virtual Reality* 2 (2021), p. 694567.
- [52] Ting Zhang et al. "Physiological measurements of situation awareness: a systematic review". In: *Human factors* 65.5 (2023), pp. 737–758.
- [53] Aaron Frederick Bulagang, James Mountstephens, and Jason Teo. "Multiclass emotion prediction using heart rate and virtual reality stimuli". In: *Journal of Big Data* 8 (2021), pp. 1–12.
- [54] Rui Liu et al. "Impacts of cues on learning and attention in immersive 360-degree video: an eye-tracking study". In: *Frontiers in Psychology* 12 (2022), p. 792069.
- [55] Christoph Tremmel et al. "Estimating cognitive workload in an interactive virtual reality environment using EEG". In: *Frontiers in human neuroscience* 13 (2019), p. 401.
- [56] Chai-Fen Tsai et al. "The effect of augmented reality and virtual reality on inducing anxiety for exposure therapy: a comparison using heart rate variability". In: *Journal of Healthcare Engineering* 2018.1 (2018), p. 6357351.
- [57] Juliana Melendrez-Ruiz et al. "An exploratory study combining eye-tracking and virtual reality: Are pulses good "eye-catchers" in virtual supermarket shelves?" In: *Frontiers in Virtual Reality* 2 (2021), p. 655273.
- [58] Ping-Keng Jao, Ricardo Chavarriaga, and José del R Millán. "EEG-based online regulation of difficulty in simulated flying". In: *IEEE Transactions on Affective Computing* 14.1 (2021), pp. 394–405.
- [59] Raphael P Weibel et al. "Virtual reality experiments with physiological measures". In: *JoVE (Journal of Visualized Experiments)* 138 (2018), e58318.
- [60] Vicki L Plano Clark. "Mixed methods research". In: *The Journal of Positive Psychology* 12.3 (2017), pp. 305–306.
- [61] Doug A Bowman, Joseph L Gabbard, and Deborah Hix. "A survey of usability evaluation in virtual environments: classification and comparison of methods". In: *Presence: Teleoperators & Virtual Environments* 11.4 (2002), pp. 404–424.
- [62] Kay M Stanney et al. "Usability engineering of virtual environments (VEs): identifying multiple criteria that drive effective VE system design". In: *International Journal of Human-Computer Studies* 58.4 (2003), pp. 447–481.
- [63] Corinna E Lathan et al. "Using virtual environments as training simulators: Measuring transfer". In: *Handbook of Virtual Environments*. CRC Press, 2002, pp. 443–454.

- [64] I Scott MacKenzie. "Fitts' law as a research and design tool in human-computer interaction". In: *Human-computer interaction 7.1* (1992), pp. 91–139.
- [65] Carolin Wienrich et al. "Developing and proving a framework for reaction time experiments in VR to objectively measure social interaction with virtual agents". In: *2018 IEEE conference on virtual reality and 3D user interfaces (VR)*. IEEE. 2018, pp. 191–198.
- [66] Ayoung Suh and Jane Prophet. "The state of immersive technology research: A literature analysis". In: *Computers in Human behavior 86* (2018), pp. 77–90.
- [67] Adailton Gonçalves da Silva et al. "Ergonomic analysis supported by virtual reality: a systematic literature review". In: *2020 22nd Symposium on Virtual and Augmented Reality (SVR)*. IEEE. 2020, pp. 463–468.
- [68] Steven P Broglio, Phillip D Tomporowski, and Michael S Ferrara. "Balance performance with a cognitive task: a dual-task testing paradigm". In: *Medicine & Science in Sports & Exercise 37.4* (2005), pp. 689–695.
- [69] J Raymond Comstock Jr and Ruth J Arnegard. *The multi-attribute task battery for human operator workload and strategic behavior research*. Tech. rep. 1992.
- [70] Andreas Gegenfurtner, Carla Quesada-Pallarès, and Maximilian Knogler. "Digital simulation-based training: A meta-analysis". In: *British Journal of Educational Technology 45.6* (2014), pp. 1097–1114.
- [71] Tero Kaarlela, Sakari Pieskä, and Tomi Pitkääho. "Digital twin and virtual reality for safety training". In: *2020 11th IEEE international conference on cognitive infocommunications (CogInfoCom)*. IEEE. 2020, pp. 000115–000120.
- [72] Pooya Adami et al. "Effectiveness of VR-based training on improving construction workers' knowledge, skills, and safety behavior in robotic teleoperation". In: *Advanced Engineering Informatics 50* (2021), p. 101431.
- [73] Lacey Duckworth, Tulio Sulbaran, and Andrew Peter Strelzoff. "Usability of a collaborative virtual reality environment earthwork exercises". In: *2012 ASEE Annual Conference & Exposition*. 2012, pp. 25–1405.
- [74] Jin Sol Lee et al. "Challenges, tasks, and opportunities in teleoperation of excavator toward human-in-the-loop construction automation". In: *Automation in Construction 135* (2022), p. 104119.
- [75] Daniel J Rea and Stela H Seo. "Still not solved: A call for renewed focus on user-centered teleoperation interfaces". In: *Frontiers in Robotics and AI 9* (2022), p. 704225.
- [76] Jin Sol Lee and Youngjib Ham. "Exploring human-machine interfaces for teleoperation of excavator". In: *Construction Research Congress 2022*. 2022, pp. 757–765.
- [77] Abdeldjallil Naceri et al. "Towards a virtual reality interface for remote robotic teleoperation". In: *2019 19th International Conference on Advanced Robotics (ICAR)*. IEEE. 2019, pp. 284–289.
- [78] Dongsu Shin et al. "Development of remote controlled manipulation device for a conventional excavator without renovation". In: *2012 IEEE/SICE International Symposium on System Integration (SII)*. IEEE. 2012, pp. 546–551.
- [79] Léa SAUNIER, Marius PREDA, and Catalin FETITA. "Controller Evaluation for Earthwork Teleoperation and Training in Virtual Reality". In: *2024 16th International Conference on Human System Interaction (HSI)*. IEEE. 2024, pp. 1–6.
- [80] Philip Sedgwick. "Pearson's correlation coefficient". In: *Bmj 345* (2012).
- [81] Yeshwanth Pulijala et al. "Effectiveness of immersive virtual reality in surgical training—a randomized control trial". In: *Journal of Oral and Maxillofacial Surgery 76.5* (2018), pp. 1065–1072.
- [82] Kate E Laver et al. "Virtual reality for stroke rehabilitation". In: *Cochrane database of systematic reviews 11* (2017).
- [83] Michelle E Portman, Asya Natapov, and Dafna Fisher-Gewirtzman. "To go where no man has gone before: Virtual reality in architecture, landscape architecture and environmental planning". In: *Computers, Environment and Urban Systems 54* (2015), pp. 376–384.
- [84] Shunsuke Okishiba et al. "Tablet interface for direct vision teleoperation of an excavator for urban construction work". In: *Automation in Construction 102* (2019), pp. 17–26.
- [85] Abdeldjallil Naceri et al. "Towards a virtual reality interface for remote robotic teleoperation". In: *2019 19th International Conference on Advanced Robotics (ICAR)*. IEEE. 2019, pp. 284–289.

- [86] Jin Sol Lee and Youngjib Ham. "Exploring human-machine interfaces for teleoperation of excavator". In: *Construction Research Congress 2022*. 2022, pp. 757–765.
- [87] Jing Du et al. "Zero latency: Real-time synchronization of BIM data in virtual reality for collaborative decision-making". In: *Automation in construction* 85 (2018), pp. 51–64.
- [88] Jin Sol Lee et al. "Challenges, tasks, and opportunities in teleoperation of excavator toward human-in-the-loop construction automation". In: *Automation in Construction* 135 (2022), p. 104119.
- [89] Hamdan Sulaiman, Mohd Noor Asril Saadun, and Ahmad Anas Yusof. "Modern manned, unmanned and teleoperated excavator system". In: *Journal of Mechanical Engineering and Technology (JMET)* 7.1 (2015).
- [90] Léa Saunier et al. "Virtual Reality Interface Evaluation for Earthwork Teleoperation". In: *Electronics* 12.19 (2023), p. 4151.
- [91] Sally E Shaywitz. "Dyslexia". In: *New England Journal of Medicine* 338.5 (1998), pp. 307–312.
- [92] Gerd Schulte-Körne. "The prevention, diagnosis, and treatment of dyslexia". In: *Deutsches Ärzteblatt International* 107.41 (2010), p. 718.
- [93] Simon Makin. "A smarter way to treat". In: *Nature* 573.7775 (2019), S106–S109.
- [94] Sally Radovick et al. "Gamification concepts to promote and maintain therapy adherence in children with growth hormone deficiency". In: *J* 1.1 (2018), pp. 71–81.
- [95] Corey J Bohil, Bradly Alicea, and Frank A Biocca. "Virtual reality in neuroscience research and therapy". In: *Nature reviews neuroscience* 12.12 (2011), pp. 752–762.
- [96] Lea Saunier et al. "Visual Rehabilitation for Learning Disorders in Virtual Reality: Visual Rehabilitation for Learning Disorder in VR". In: *Proceedings of the 27th International Conference on 3D Web Technology*. 2022, pp. 1–4.
- [97] Federica Pallavicini, Alessandro Pepe, and Maria Eleonora Minissi. "Gaming in virtual reality: What changes in terms of usability, emotional response and sense of presence compared to non-immersive video games?" In: *Simulation & Gaming* 50.2 (2019), pp. 136–159.
- [98] Andreas Halbig and Marc Erich Latoschik. "A systematic review of physiological measurements, factors, methods, and applications in virtual reality". In: *Frontiers in Virtual Reality* 2 (2021), p. 694567.
- [99] Mohammad Ahmadi et al. "Comparison of Physiological Cues for Cognitive Load Measures in VR". In: *2023 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. IEEE. 2023, pp. 837–838.
- [100] Oscar I Caldas et al. "Behavioral and Psychophysiological Measures of Engagement During Dynamic Difficulty Adjustment in Immersive Virtual Reality." In: *JUCS: Journal of Universal Computer Science* 29.1 (2023).
- [101] Guillermo Bernal et al. "PhysioHMD: a conformable, modular toolkit for collecting physiological data from head-mounted displays". In: *Proceedings of the 2018 ACM international symposium on wearable computers*. 2018, pp. 160–167.
- [102] Eric Krokos and Amitabh Varshney. "Quantifying VR cybersickness using EEG". In: *Virtual Reality* 26.1 (2022), pp. 77–89.
- [103] Viviane Clay, Peter König, and Sabine Koenig. "Eye tracking in virtual reality". In: *Journal of eye movement research* 12.1 (2019).
- [104] Sebastian Ocklenburg and Jutta Peterburs. "Monitoring brain activity in VR: EEG and neuroimaging". In: *Virtual Reality in Behavioral Neuroscience: New Insights and Methods*. Springer, 2023, pp. 47–71.
- [105] Mochamad Hilman, Dwi Kurnia Basuki, and Sritrusta Sukaridhoto. "Virtual hand: VR hand controller using IMU and flex sensor". In: *2018 International Electronics Symposium on Knowledge Creation and Intelligent Computing (IES-KCIC)*. IEEE. 2018, pp. 310–314.
- [106] Mochamad Hilman, Dwi Kurnia Basuki, and Sritrusta Sukaridhoto. "Virtual hand: VR hand controller using IMU and flex sensor". In: *2018 International Electronics Symposium on Knowledge Creation and Intelligent Computing (IES-KCIC)*. IEEE. 2018, pp. 310–314.
- [107] Murtada Dohan and Mu Mu. "Understanding user attention in VR using gaze controlled games". In: *Proceedings of the 2019 ACM International Conference on Interactive Experiences for TV and Online Video*. 2019, pp. 167–173.

- [108] Nayeon Kim and Hyunsoo Lee. “Visual attention in retail environments: design analysis using HMD based VR system integrated eye-tracking”. In: *25th International Conference on Computer-Aided Architectural Design Research in Asia, CAADRIA 2020*. The Association for Computer-Aided Architectural Design Research in Asia . . . 2020, pp. 631–640.
- [109] Rui Liu et al. “Impacts of cues on learning and attention in immersive 360-degree video: an eye-tracking study”. In: *Frontiers in Psychology* 12 (2022), p. 792069.
- [110] Alexis D Souchet et al. “Measuring visual fatigue and cognitive load via eye tracking while learning with virtual reality head-mounted displays: A review”. In: *International Journal of Human–Computer Interaction* 38.9 (2022), pp. 801–824.
- [111] Mohsen Parisay, Charalambos Poullis, and Marta Kersten-Oertel. “Felix: Fixation-based eye fatigue load index a multi-factor measure for gaze-based interactions”. In: *2020 13th International Conference on Human System Interaction (HSI)*. IEEE. 2020, pp. 74–81.
- [112] Somnath Arjun et al. “VR cognitive load dashboard for flight simulator”. In: *2022 Symposium on Eye Tracking Research and Applications*. 2022, pp. 1–4.
- [113] Natasha Anne Rappa et al. “The use of eye tracking technology to explore learning and performance within virtual reality and mixed reality settings: a scoping review”. In: *Interactive Learning Environments* 30.7 (2022), pp. 1338–1350.
- [114] Rustam Shadiev and Dandan Li. “A review study on eye-tracking technology usage in immersive virtual reality learning environments”. In: *Computers & education* 196 (2023), p. 104681.
- [115] Liat Liberman and Ilana Dubovi. “The effect of the modality principle to support learning with virtual reality: An eye-tracking and electrodermal activity study”. In: *Journal of Computer Assisted Learning* 39.2 (2023), pp. 547–557.
- [116] Christian Hirt, Marcel Eckard, and Andreas Kunz. “Stress generation and non-intrusive measurement in virtual environments using eye tracking”. In: *Journal of Ambient Intelligence and Humanized Computing* 11 (2020), pp. 5977–5989.
- [117] Chandran Jyotsna and Joseph Amudha. “Eye gaze as an indicator for stress level analysis in students”. In: *2018 International conference on advances in computing, communications and informatics (ICACCI)*. IEEE. 2018, pp. 1588–1593.
- [118] Maike Stoeve et al. “Eye tracking-based stress classification of athletes in virtual reality”. In: *Proceedings of the ACM on Computer Graphics and Interactive Techniques* 5.2 (2022), pp. 1–17.
- [119] Archana Hebbar et al. “Cognitive load estimation in VR flight simulator”. In: *Journal of Eye Movement Research* 15.3 (2023).
- [120] Eric Redlinger and Claudia Shao. “Comparing brain activity in virtual and non-virtual environments: a VR & EEG study”. In: *Measurement: Sensors* 18 (2021), p. 100062.
- [121] Marta Mondellini et al. “A Multimodal Approach Exploiting EEG to Investigate the Effects of VR Environment on Mental Workload”. In: *International Journal of Human–Computer Interaction* (2023), pp. 1–13.
- [122] Eric Krokos and Amitabh Varshney. “Quantifying VR cybersickness using EEG”. In: *Virtual Reality* 26.1 (2022), pp. 77–89.
- [123] Caglar Yildirim. “A review of deep learning approaches to EEG-based classification of cybersickness in virtual reality”. In: *2020 IEEE international conference on artificial intelligence and virtual reality (AIVR)*. IEEE. 2020, pp. 351–357.
- [124] Kyoung-Mi Jang et al. “Estimating objective (EEG) and subjective (SSQ) cybersickness in people with susceptibility to motion sickness”. In: *Applied ergonomics* 102 (2022), p. 103731.
- [125] Rhaíra Helena Caetano e Souza and Eduardo Lázaro Martins Naves. “Attention detection in virtual environments using EEG signals: a scoping review”. In: *frontiers in physiology* 12 (2021), p. 727840.
- [126] Baek Hwan Cho et al. “Attention enhancement system using virtual reality and EEG biofeedback”. In: *Proceedings IEEE virtual reality 2002*. IEEE. 2002, pp. 156–163.
- [127] Rateb Katmah et al. “A review on mental stress assessment methods using EEG signals”. In: *Sensors* 21.15 (2021), p. 5043.

- [128] Dorota Kamińska, Krzysztof Smółka, and Grzegorz Zwoliński. "Detection of mental stress through EEG signal in virtual reality environment". In: *Electronics* 10.22 (2021), p. 2840.
- [129] Tee Yi Wen and SA Mohd Aris. "Electroencephalogram (EEG) stress analysis on alpha/beta ratio and theta/beta ratio". In: *Indones. J. Electr. Eng. Comput. Sci* 17.1 (2020), pp. 175–182.
- [130] Frano Giakoni-Ramírez et al. "Physical Activity in Immersive Virtual Reality: A Scoping Review". In: *Healthcare*. Vol. 11. 11. MDPI. 2023, p. 1553.
- [131] Richard P Troiano et al. "Evolution of accelerometer methods for physical activity research". In: *British journal of sports medicine* 48.13 (2014), pp. 1019–1023.
- [132] Reymon Rotikan et al. "Virtual Runner: A Virtual Reality-based Exergaming Application using Accelerometer". In: *Virtual Runner: Sebuah Exergaming berbasis Virtual Reality Menggunakan Accelerometer*. In: ().
- [133] Sergio Caro-Alvaro et al. "Gesture-Based Interactions: Integrating Accelerometer and Gyroscope Sensors in the Use of Mobile Apps". In: *Sensors* 24.3 (2024), p. 1004.
- [134] Tiago Martins Andrade, Max Smith-Creasey, and Jonathan Francis Roscoe. "Discerning user activity in extended reality through side-channel accelerometer observations". In: *2020 IEEE international conference on intelligence and security informatics (ISI)*. IEEE. 2020, pp. 1–3.
- [135] Jie Li et al. "Real-time hand gesture tracking for human–computer interface based on multi-sensor data fusion". In: *IEEE Sensors Journal* 21.23 (2021), pp. 26642–26654.
- [136] Sammeli Liikkanen et al. "Body movement as a biomarker for use in chronic pain rehabilitation: An embedded analysis of an RCT of a virtual reality solution for adults with chronic pain". In: *Frontiers in Pain Research* 3 (2022), p. 1085791.
- [137] Tommi Gröhn et al. "Quantifying movement behavior of chronic low back pain patients in virtual reality". In: *ACM Transactions on Computing for Healthcare* 4.2 (2023), pp. 1–24.
- [138] Catherine Pollard. "Virtual reality for improving activity levels in people with chronic pain: A feasibility study". In: (2021).
- [139] Jella Pfeiffer et al. "Eye-tracking-based classification of information search behavior using machine learning: evidence from experiments in physical shops and virtual reality shopping environments". In: *Information Systems Research* 31.3 (2020), pp. 675–691.
- [140] Ken Pfeuffer et al. "Gaze+ pinch interaction in virtual reality". In: *Proceedings of the 5th symposium on spatial user interaction*. 2017, pp. 99–108.
- [141] Jingjing Zhang, Yi Li, Chao Mei, et al. "A methodology of eye gazing attention determination for VR training". In: *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. IEEE. 2020, pp. 138–141.
- [142] Tad T Brunyé et al. "A review of eye tracking for understanding and improving diagnostic interpretation". In: *Cognitive research: principles and implications* 4 (2019), pp. 1–16.
- [143] Baek-Hwan Cho et al. "Neurofeedback training with virtual reality for inattention and impulsiveness". In: *Cyberpsychology & Behavior* 7.5 (2004), pp. 519–526.
- [144] Gaetano Tieri et al. "Virtual reality in cognitive and motor rehabilitation: facts, fiction and fallacies". In: *Expert review of medical devices* 15.2 (2018), pp. 107–117.
- [145] Daniel Hufnal et al. "Impact of VR and desktop gaming on electroencephalogram (EEG) ratings". In: *2021 International Conference on Electrical, Communication, and Computer Engineering (ICECCE)*. IEEE. 2021, pp. 1–6.
- [146] Paul L Nunez. "The brain wave equation: a model for the EEG". In: *Mathematical Biosciences* 21.3-4 (1974), pp. 279–297.
- [147] Arne Dietrich and Riam Kanso. "A review of EEG, ERP, and neuroimaging studies of creativity and insight." In: *Psychological bulletin* 136.5 (2010), p. 822.
- [148] Herbert Ramoser, Johannes Muller-Gerking, and Gert Pfurtscheller. "Optimal spatial filtering of single trial EEG during imagined hand movement". In: *IEEE transactions on rehabilitation engineering* 8.4 (2000), pp. 441–446.

- [149] Ryan S McGinnis et al. "Quantifying the effects of load carriage and fatigue under load on sacral kinematics during countermovement vertical jump with IMU-based method". In: *Sports Engineering* 19 (2016), pp. 21–34.
- [150] A Nassef et al. "The assessment of heart rate variability (HRV) and task load index (TLI) as physiological markers for physical stress". In: *World Congress on Medical Physics and Biomedical Engineering, September 7-12, 2009, Munich, Germany: Vol. 25/13 Special Topics and Workshops*. Springer. 2010, pp. 146–149.
- [151] Weisheng Kong et al. "Development of a real-time IMU-based motion capture system for gait rehabilitation". In: *2013 IEEE International Conference on Robotics and Biomimetics (ROBIO)*. IEEE. 2013, pp. 2100–2105.
- [152] Steven N Blair and Jon C Connelly. "How much physical activity should we do? The case for moderate amounts and intensities of physical activity". In: *Research quarterly for exercise and sport* 67.2 (1996), pp. 193–205.
- [153] Léa Saunier, Marius Preda, and Catalin Fetita. "Hybrid Methodology Using Electroencephalogram and Eye-tracking for Virtual Reality Design and Optimization". In: *2024 23rd IEEE International Symposium on Mixed and Augmented Reality (ISMAR)* (2024). to be published, to appear.
- [154] Wolfgang Klimesch. "EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis". In: *Brain research reviews* 29.2-3 (1999), pp. 169–195.
- [155] Luis Cabañero et al. "Characterisation of mobile-device tasks by their associated cognitive load through EEG data processing". In: *Future Generation Computer Systems* 113 (2020), pp. 380–390.
- [156] Alexander J Casson et al. "Electroencephalogram". In: *Seamless healthcare monitoring: Advancements in wearable, attachable, and invisible devices* (2018), pp. 45–81.
- [157] Wayne K Kirchner. "Age differences in short-term retention of rapidly changing information." In: *Journal of experimental psychology* 55.4 (1958), p. 352.
- [158] Asma Islam et al. "Study and Analysis of Motion Artifacts for Ambulatory EEG". In: *Article in International Journal of Electrical and Computer Engineering* 9.4 (2021).
- [159] Zirui Hong et al. "Effect of virtual annotation on performance of construction equipment teleoperation under adverse visual conditions". In: *Automation in Construction* 118 (2020), p. 103296.
- [160] Irina Terentyeva et al. "The virtual construction site: Knowledge management in virtual environments". In: *International Journal of Emerging Technologies in Learning (iJET)* 15.13 (2020), pp. 81–95.
- [161] Colwyn B Trevarthen. "Two mechanisms of vision in primates". In: *Psychologische Forschung* 31.4 (1968), pp. 299–337.
- [162] Luis Cabañero Gómez et al. "Studying the generalisability of cognitive load measured with EEG". In: *Biomedical Signal Processing and Control* 70 (2021), p. 103032.
- [163] Tae Kyun Kim. "Understanding one-way ANOVA using conceptual figures". In: *Korean journal of anesthesiology* 70.1 (2017), pp. 22–26.
- [164] Thomas Potter, Zoran Cvetković, and Enzo De Sena. "On the relative importance of visual and spatial audio rendering on vr immersion". In: *Frontiers in Signal Processing* 2 (2022), p. 904866.
- [165] Hye-Geum Kim et al. "Stress and heart rate variability: a meta-analysis and review of the literature". In: *Psychiatry investigation* 15.3 (2018), p. 235.
- [166] David Bannach, Oliver Amft, and Paul Lukowicz. "Automatic event-based synchronization of multimodal data streams from wearable and ambient sensors". In: *Smart Sensing and Context: 4th European Conference, EuroSSC 2009, Guildford, UK, September 16-18, 2009. Proceedings 4*. Springer. 2009, pp. 135–148.
- [167] Robert JK Jacob. "The use of eye movements in human-computer interaction techniques: what you look at is what you get". In: *ACM Transactions on Information Systems (TOIS)* 9.2 (1991), pp. 152–169.
- [168] Ajoy S Fernandes, T Scott Murdison, and Michael J Proulx. "Leveling the playing field: A comparative reevaluation of unmodified eye tracking as an input and interaction modality for VR". In: *IEEE Transactions on Visualization and Computer Graphics* 29.5 (2023), pp. 2269–2279.
- [169] Xun Chen et al. "Removal of muscle artifacts from the EEG: A review and recommendations". In: *IEEE Sensors Journal* 19.14 (2019), pp. 5353–5368.
- [170] Olaf Dimigen et al. "Coregistration of eye movements and EEG in natural reading: analyses and review." In: *Journal of experimental psychology: General* 140.4 (2011), p. 552.

- [171] Alexander V Kramarenko and Uner Tan. “Effects of high-frequency electromagnetic fields on human EEG: a brain mapping study”. In: *International Journal of Neuroscience* 113.7 (2003), pp. 1007–1019.
- [172] Christoph Tremmel, Christian Herff, and Dean J Krusienski. “EEG movement artifact suppression in interactive virtual reality”. In: *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE. 2019, pp. 4576–4579.
- [173] Christoph Tremmel, Christian Herff, and Dean J Krusienski. “EEG movement artifact suppression in interactive virtual reality”. In: *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE. 2019, pp. 4576–4579.
- [174] Ali Bulent Usakli. “Improvement of EEG signal acquisition: An electrical aspect for state of the art of front end”. In: *Computational intelligence and neuroscience* 2010.1 (2010), p. 630649.
- [175] Ricardo Vigário et al. “Independent component approach to the analysis of EEG and MEG recordings”. In: *IEEE transactions on biomedical engineering* 47.5 (2000), pp. 589–593.
- [176] V Krishnaveni et al. “Automatic identification and removal of ocular artifacts from EEG using wavelet transform”. In: *Measurement science review* 6.4 (2006), pp. 45–57.
- [177] Simon O’Regan and William Marnane. “Multimodal detection of head-movement artefacts in EEG”. In: *Journal of Neuroscience Methods* 218.1 (2013), pp. 110–120.
- [178] Christopher Beach et al. “Motion artefact removal in electroencephalography and electrocardiography by using multichannel inertial measurement units and adaptive filtering”. In: *Healthcare Technology Letters* 8.5 (2021), pp. 128–138.
- [179] Michael Plöchl, José P Ossandón, and Peter König. “Combining EEG and eye tracking: identification, characterization, and correction of eye movement artifacts in electroencephalographic data”. In: *Frontiers in human neuroscience* 6 (2012), p. 278.
- [180] João Jorge, Wietske Van der Zwaag, and Patrícia Figueiredo. “EEG–fMRI integration for the study of human brain function”. In: *Neuroimage* 102 (2014), pp. 24–34.
- [181] Gaetano Gargiulo et al. “A mobile EEG system with dry electrodes”. In: *2008 IEEE biomedical circuits and systems conference*. IEEE. 2008, pp. 273–276.
- [182] Asma Islam et al. “Study and analysis of motion artifacts for ambulatory eeg”. In: *Article in International Journal of Electrical and Computer Engineering* 9.4 (2021).
- [183] John Sweller. “Cognitive load during problem solving: Effects on learning”. In: *Cognitive science* 12.2 (1988), pp. 257–285.
- [184] Yunwang Li et al. “Navigation simulation of a Mecanum wheel mobile robot based on an improved A* algorithm in Unity3D”. In: *Sensors* 19.13 (2019), p. 2976.
- [185] Facundo Pessacg et al. “Simplifying UAV-based photogrammetry in forestry: How to generate accurate digital terrain model and assess flight mission settings”. In: *Forests* 13.2 (2022), p. 173.
- [186] Chi-Yuan Chang et al. “Evaluation of artifact subspace reconstruction for automatic artifact components removal in multi-channel EEG recordings”. In: *IEEE Transactions on Biomedical Engineering* 67.4 (2019), pp. 1114–1121.
- [187] Shohei Tsuchimoto et al. “Use of common average reference and large-Laplacian spatial-filters enhances EEG signal-to-noise ratios in intrinsic sensorimotor activity”. In: *Journal of neuroscience methods* 353 (2021), p. 109089.
- [188] Evelyn Karikari and Konstantin A Koshechkin. “Review on brain-computer interface technologies in healthcare”. In: *Biophysical Reviews* 15.5 (2023), pp. 1351–1358.
- [189] Jaime Riascos, Marta Molinas, and Fabien Lotte. “Machine Learning Methods for BCI: challenges, pitfalls and promises”. In: *ESANN 2024-European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*. 2024.
- [190] Immo Schuetz and Katja Fiehler. “Eye tracking in virtual reality: Vive pro eye spatial accuracy, precision, and calibration reliability”. In: *Journal of Eye Movement Research* 15.3 (2022).

Appendix A

Miscellaneous Figures

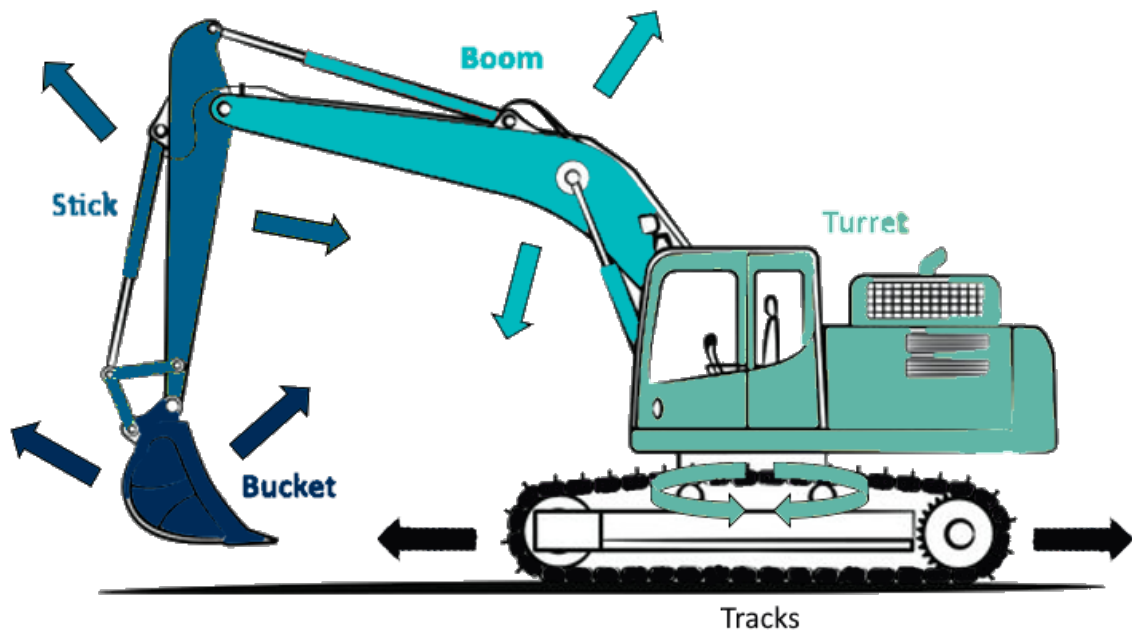


Figure A.1: Excavator Schematics



Figure A.2: Real (HERACLES Robotics') and Virtual (VR-SOLUS') Excavator

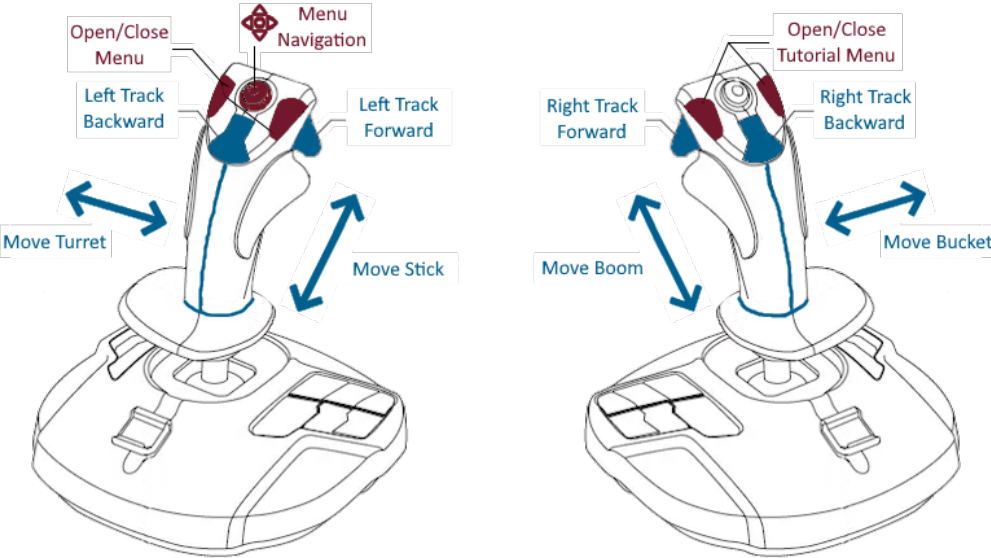
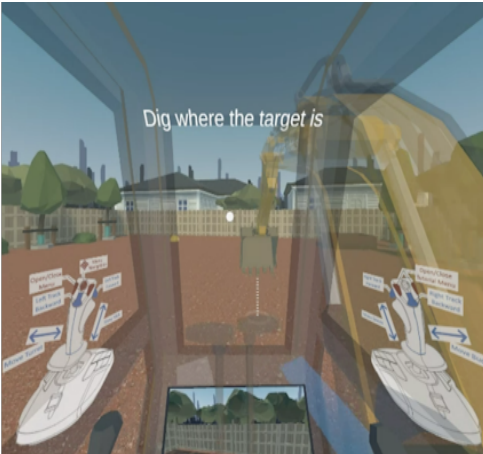


Figure A.3: Final Joysticks Mapping



USER SET INTERFACE



DATA-DRIVEN SET INTERFACE



HYBRID SET INTERFACE

Figure A.4: Virtual Annotations Study Interface

Appendix B

Experimental Surveys

This appendix contains the surveys templates (NASA Task-Load-Index and System-Usability-Scale-inspired) used to gather quantified user feedback.

ID _____

Name: _____ Interface: _____
 Date: _____

NASA Task Load Index

We are interested in the workload you experienced while completing this task. As workload can be caused by several different factors, we ask you to rate several of the factors individually on the scales provided.

Note: Performance goes from good on the left to bad on the right.

Mental Demand: How mentally demanding was the task?

Physical Demand: How physically demanding was the task?

Temporal Demand: How hurried or rushed was the pace of the task?

Performance: How successful were you in accomplishing what you were asked to do?

Effort: How hard did you have to work to accomplish your level of performance?

Frustration: How insecure, discouraged, irritated, stressed, and annoyed were you?

Figure B.1: NASA-Task Load Index Survey Template

Virtual Annotations User Survey (based on System Usability Scale SUS)

Participant ID:
 VA sets order:
 Environment order:

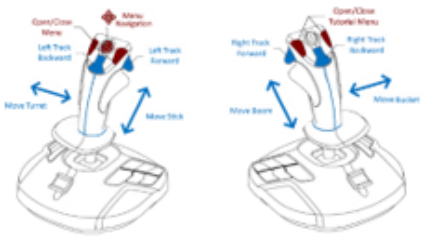
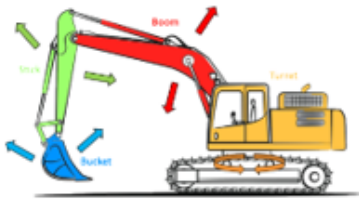
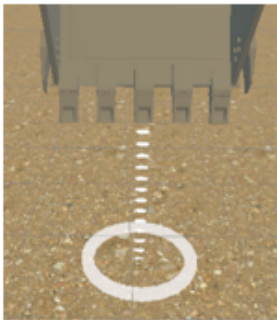
Virtual Annotation	User feedback																				
<p style="text-align: center;">Mappings</p> 	<p>This annotation is:</p> <p>Not useful Very helpful</p> <table border="1" style="width: 100%; height: 20px;"> <tr> <td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td> </tr> </table> <p>I hate it I love it</p> <p>This annotation should be:</p> <ul style="list-style-type: none"> <input type="checkbox"/> Deleted <input type="checkbox"/> In a menu <input type="checkbox"/> Activated by looking <input type="checkbox"/> Activated in special cases <input type="checkbox"/> Always visible 																				
<p style="text-align: center;">Tutorial</p> 	<p>This annotation is:</p> <p>Not useful Very helpful</p> <table border="1" style="width: 100%; height: 20px;"> <tr> <td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td> </tr> </table> <p>I hate it I love it</p> <p>This annotation should be:</p> <ul style="list-style-type: none"> <input type="checkbox"/> Deleted <input type="checkbox"/> In a menu <input type="checkbox"/> Activated by looking <input type="checkbox"/> Activated in special cases <input type="checkbox"/> Always visible 																				
<p style="text-align: center;">Reticle</p> 	<p>This annotation is:</p> <p>Not useful Very helpful</p> <table border="1" style="width: 100%; height: 20px;"> <tr> <td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td> </tr> </table> <p>I hate it I love it</p> <p>This annotation should be:</p> <ul style="list-style-type: none"> <input type="checkbox"/> Deleted <input type="checkbox"/> In a menu <input type="checkbox"/> Activated by looking <input type="checkbox"/> Activated in special cases <input type="checkbox"/> Always visible 																				

Figure B.2: System Usability Scale Inspired Survey Template (1st page)




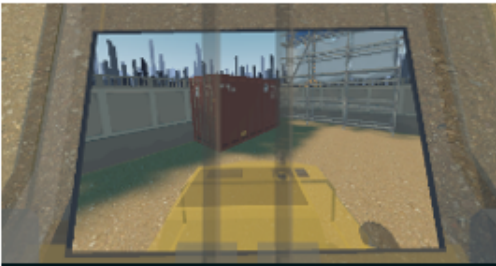



<p><u>Objectives in windshield</u></p> 	<p><u>This annotation is:</u> Not useful Very helpful</p> <table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 12.5%; height: 20px;"></td> <td style="width: 12.5%;"></td> <td style="width: 12.5%;"></td> <td style="width: 12.5%;"></td> <td style="width: 12.5%;"></td> <td style="width: 12.5%;"></td> <td style="width: 12.5%;"></td> <td style="width: 12.5%;"></td> <td style="width: 12.5%;"></td> <td style="width: 12.5%;"></td> </tr> <tr> <td style="text-align: left;">I hate it</td> <td colspan="8"></td> <td style="text-align: right;">I love it</td> </tr> </table> <p><u>This annotation should be:</u></p> <ul style="list-style-type: none"> <input type="checkbox"/> Deleted <input type="checkbox"/> In a menu <input type="checkbox"/> Activated by looking <input type="checkbox"/> Activated in special cases <input type="checkbox"/> Always visible 											I hate it									I love it
I hate it									I love it												
<p><u>Objective in the world</u></p> 	<p><u>This annotation is:</u> Not useful Very helpful</p> <table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 12.5%; height: 20px;"></td> <td style="width: 12.5%;"></td> <td style="width: 12.5%;"></td> <td style="width: 12.5%;"></td> <td style="width: 12.5%;"></td> <td style="width: 12.5%;"></td> <td style="width: 12.5%;"></td> <td style="width: 12.5%;"></td> <td style="width: 12.5%;"></td> <td style="width: 12.5%;"></td> </tr> <tr> <td style="text-align: left;">I hate it</td> <td colspan="8"></td> <td style="text-align: right;">I love it</td> </tr> </table> <p><u>This annotation should be:</u></p> <ul style="list-style-type: none"> <input type="checkbox"/> Deleted <input type="checkbox"/> In a menu <input type="checkbox"/> Activated by looking <input type="checkbox"/> Activated in special cases <input type="checkbox"/> Always visible 											I hate it									I love it
I hate it									I love it												
<p><u>Arrow to objective</u></p> 	<p><u>This annotation is:</u> Not useful Very helpful</p> <table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 12.5%; height: 20px;"></td> <td style="width: 12.5%;"></td> <td style="width: 12.5%;"></td> <td style="width: 12.5%;"></td> <td style="width: 12.5%;"></td> <td style="width: 12.5%;"></td> <td style="width: 12.5%;"></td> <td style="width: 12.5%;"></td> <td style="width: 12.5%;"></td> <td style="width: 12.5%;"></td> </tr> <tr> <td style="text-align: left;">I hate it</td> <td colspan="8"></td> <td style="text-align: right;">I love it</td> </tr> </table> <p><u>This annotation should be:</u></p> <ul style="list-style-type: none"> <input type="checkbox"/> Deleted <input type="checkbox"/> In a menu <input type="checkbox"/> Activated by looking <input type="checkbox"/> Activated in special cases <input type="checkbox"/> Always visible 											I hate it									I love it
I hate it									I love it												
<p><u>Rear view</u></p> 	<p><u>This annotation is:</u> Not useful Very helpful</p> <table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 12.5%; height: 20px;"></td> <td style="width: 12.5%;"></td> <td style="width: 12.5%;"></td> <td style="width: 12.5%;"></td> <td style="width: 12.5%;"></td> <td style="width: 12.5%;"></td> <td style="width: 12.5%;"></td> <td style="width: 12.5%;"></td> <td style="width: 12.5%;"></td> <td style="width: 12.5%;"></td> </tr> <tr> <td style="text-align: left;">I hate it</td> <td colspan="8"></td> <td style="text-align: right;">I love it</td> </tr> </table> <p><u>This annotation should be:</u></p> <ul style="list-style-type: none"> <input type="checkbox"/> Deleted <input type="checkbox"/> In a menu <input type="checkbox"/> Activated by looking <input type="checkbox"/> Activated in special cases <input type="checkbox"/> Always visible 											I hate it									I love it
I hate it									I love it												

Figure B.3: System Usability Scale Inspired Survey Template (2nd page)

<p style="text-align: center;"><u>Tracks direction</u></p> 	<p><u>This annotation is:</u></p> <p>Not useful Very helpful</p> <table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 12.5%; height: 20px;"></td> <td style="width: 12.5%;"></td> <td style="width: 12.5%;"></td> <td style="width: 12.5%;"></td> <td style="width: 12.5%;"></td> <td style="width: 12.5%;"></td> <td style="width: 12.5%;"></td> <td style="width: 12.5%;"></td> <td style="width: 12.5%;"></td> <td style="width: 12.5%;"></td> </tr> <tr> <td style="width: 12.5%;"></td> <td style="width: 12.5%;"></td> <td style="width: 12.5%;"></td> <td style="width: 12.5%;"></td> <td style="width: 12.5%;"></td> <td style="width: 12.5%;"></td> <td style="width: 12.5%;"></td> <td style="width: 12.5%;"></td> <td style="width: 12.5%;"></td> <td style="width: 12.5%;"></td> </tr> </table> <p>I hate it I love it</p> <p><u>This annotation should be:</u></p> <ul style="list-style-type: none"> <input type="checkbox"/> Deleted <input type="checkbox"/> In a menu <input type="checkbox"/> Activated by looking <input type="checkbox"/> Activated in special cases <input type="checkbox"/> Always visible 																				
<p style="text-align: center;"><u>Engine sound</u></p> 	<p><u>This annotation is:</u></p> <p>Not useful Very helpful</p> <table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 12.5%; height: 20px;"></td> <td style="width: 12.5%;"></td> <td style="width: 12.5%;"></td> <td style="width: 12.5%;"></td> <td style="width: 12.5%;"></td> <td style="width: 12.5%;"></td> <td style="width: 12.5%;"></td> <td style="width: 12.5%;"></td> <td style="width: 12.5%;"></td> <td style="width: 12.5%;"></td> </tr> <tr> <td style="width: 12.5%;"></td> <td style="width: 12.5%;"></td> <td style="width: 12.5%;"></td> <td style="width: 12.5%;"></td> <td style="width: 12.5%;"></td> <td style="width: 12.5%;"></td> <td style="width: 12.5%;"></td> <td style="width: 12.5%;"></td> <td style="width: 12.5%;"></td> <td style="width: 12.5%;"></td> </tr> </table> <p>I hate it I love it</p> <p><u>This annotation should be:</u></p> <ul style="list-style-type: none"> <input type="checkbox"/> Deleted <input type="checkbox"/> In a menu <input type="checkbox"/> Activated in special cases <input type="checkbox"/> Always audible 																				
<p style="text-align: center;"><u>Backing sound (beeping)</u></p> 	<p><u>This annotation is:</u></p> <p>Not useful Very helpful</p> <table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 12.5%; height: 20px;"></td> <td style="width: 12.5%;"></td> <td style="width: 12.5%;"></td> <td style="width: 12.5%;"></td> <td style="width: 12.5%;"></td> <td style="width: 12.5%;"></td> <td style="width: 12.5%;"></td> <td style="width: 12.5%;"></td> <td style="width: 12.5%;"></td> <td style="width: 12.5%;"></td> </tr> <tr> <td style="width: 12.5%;"></td> <td style="width: 12.5%;"></td> <td style="width: 12.5%;"></td> <td style="width: 12.5%;"></td> <td style="width: 12.5%;"></td> <td style="width: 12.5%;"></td> <td style="width: 12.5%;"></td> <td style="width: 12.5%;"></td> <td style="width: 12.5%;"></td> <td style="width: 12.5%;"></td> </tr> </table> <p>I hate it I love it</p> <p><u>This annotation should be:</u></p> <ul style="list-style-type: none"> <input type="checkbox"/> Deleted <input type="checkbox"/> In a menu <input type="checkbox"/> Activated in special cases <input type="checkbox"/> Always audible 																				

Rank the VA sets (set1, set2, set3) in your preferred order:

Figure B.4: System Usability Scale Inspired Survey Template (3rd page)

Appendix C

Supplementary Data Analysis

A one-way ANOVA (Analysis of Variance) is a statistical test used to determine whether there are significant differences between the means of three or more independent groups. It compares the variability within each group to the variability between the groups, helping to assess whether any observed differences in group means are due to chance or a specific factor. This test is useful when comparing different conditions or treatments to see if they affect the outcome in a meaningful way.

Below is the complete statistical analysis (one-way ANOVA) of our VR-SOLUS methodology's measures across our different studies. Measures evaluated:

- **Time:** completion time of tasks/exercises;
- **Errors:** number of inputs errors;
- **MI%:** percentage of multiple inputs simultaneous usage;
- **Eye-tracking:** tracked gaze fixations;
- **TAR:** Theta-Alpha Ratio, cognitive load measure;
- **IMU:** Inertial Measurement Unit, accelerometers and gyroscopes, physical load measure.

ANOVA statistics analysed:

- **Degrees of Freedom (df):** number of independent values that can vary;
- **Sum of Squares (SoS):** total variation;
- **Mean Square (MS):** sum of squares divided by the respective degrees of freedom;
- **F-Statistic (F):** ratio of the Mean Square Between Groups to the Mean Square Within Groups. Used to determine the p-value and assess the statistical significance of the differences between group means;
- **P-Value:** probability that the observed differences among group means occurred by chance;
- **Effect Size (η^2):** provides information on the proportion of variance explained by the independent variable. Useful for understanding the practical significance of the results.

Overall, the statistical power of each study is relatively weak, indicating that a larger participant group is needed to generalise the findings of this thesis. Nevertheless, despite this limitation, clear trends emerge that can be used to formulate hypotheses and demonstrate the value of the various metrics discussed. The Interface Comparison study was excluded from the following table because it involves only two groups: the virtual reality group and the real machine group.

Study	Statistics	Measures					
		Time	Errors	MI%	Eye-tracking	TAR	IMU
Mappings Evaluation 3 Controllers Comparison	df	2	2	2	-	2	-
	SoS	11054.7	274.7	2050.9	-	0.03	-
	MS	5527.3	137.3	1025.5	-	0.01	-
	F	0.29	2.76	0.62	-	1.86	-
	p-value	0.75	0.07	0.54	-	0.16	-
	η^2	0.016	0.13	0.033	-	0.094	-
Virtual Annotations Design Phase 3 Empirical VA Sets Comparison	df	2	2	2	7	2	2
	SoS	4787.3	5.4	33.9	93880.7	0.005	8.6
	MS	2393.6	2.7	16.9	13411.5	0.002	4.3
	F	0.28	0.32	0.007	3.44	0.42	0.06
	p-value	0.75	0.73	0.99	0.002	0.66	0.94
	η^2	0.012	0.014	0.00032	0.17	0.018	0.0026
Virtual Annotations Evaluation Phase 3 Methodical VA Sets Comparison	df	2	2	2	7	2	2
	SoS	22318.3	37.8	1553.6	121372.1	0.006	6.6
	MS	11159.1	18.9	776.8	17338.9	0.003	3.3
	F	1.17	0.48	0.34	10.5	0.41	0.04
	p-value	0.32	0.62	0.71	3E-10	0.66	0.96
	η^2	0.049	0.021	0.015	0.38	0.018	0.0016

Table C.1: Statistical Analysis of the VR-SOLUS Measures Across the Different Studies (One-way ANOVA)

Titre : Conception pilotée par les données en réalité virtuelle : Combiner la perception de l'utilisateur avec des mesures opérationnelles et physiologiques

Mots clés : réalité virtuelle, interfaces homme-machine, interactions homme-machine, conception, capteurs physiologiques

Résumé : Cette thèse explore l'intégration des méthodologies hybrides pour l'optimisation de l'expérience utilisateur et des interactions homme-machine en réalité virtuelle (RV). Alors que les méthodes traditionnelles de conception reposent largement sur les retours des utilisateurs, ces approches, bien que précieuses, présentent des limites en raison de leur subjectivité. En parallèle, les méthodes basées sur les données, qui utilisent des mesures opérationnelles et physiologiques, offrent une perspective plus objective mais demeurent imparfaites, souvent insuffisantes pour capturer la complexité de l'expérience utilisateur en RV.

Pour surmonter ces limitations, nous avons développé et testé la méthodologie VR-SOLUS (Méthode de synergie des données sensorielles, opérationnelles et de l'utilisateur pour la conception de la RV), qui combine des mesures opérationnelles (performances, efficacité), physiologiques (suivi oculaire, électroencéphalogramme, fréquence cardiaque) et les retours des utilisateurs. Cette approche a été appliquée à deux études de cas distinctes : la téléopération de machines de terrassement et la

thérapie pour la dyslexie. À travers une série d'expériences impliquant la validation d'interfaces, la sélection de contrôleurs et l'optimisation d'annotations virtuelles, nos résultats démontrent que la méthodologie hybride offre une analyse plus complète et nuancée que les approches purement subjectives ou objectives.

Les résultats révèlent également les défis techniques et méthodologiques liés à l'intégration des biocapteurs dans les environnements VR, notamment en termes de synchronisation des données, de filtrage des signaux EEG et de la fidélité des simulations physiques. Malgré ces défis, cette thèse propose des pistes pour améliorer l'analyse des données en VR, notamment par l'intégration de nouveaux capteurs et l'utilisation de techniques avancées de traitement du signal.

En conclusion, cette recherche contribue à l'avancement des méthodologies de conception en RV, en démontrant l'efficacité de l'approche hybride pour optimiser l'expérience utilisateur et en offrant des perspectives pour de futures recherches dans ce domaine en pleine évolution.

Title : Data-driven Design in Virtual Reality: Combining User Perception with Operational and Physiological Measures

Keywords : virtual reality, human-machine interfaces, human-machine interactions, design, bio-sensors

Abstract : This thesis explores the integration of hybrid methodologies for optimising user experience and human-machine interactions in virtual reality (VR). Traditional design methods rely heavily on user feedback, which, while valuable, are limited by their subjectivity. In parallel, data-driven methods, which use operational and physiological measures, offer a more objective perspective but often fall short in capturing the full complexity of the VR user experience.

To address these limitations, we developed and tested the VR-SOLUS methodology (Sensory, Operational and User data Synthesis for Virtual Reality), which combines operational measures (performance, efficiency), physiological data (eye-tracking, electroencephalogram, heart rate), and user feedback. This approach was applied to two distinct case studies: earthwork machinery teleoperation and dyslexia therapy. Through a series of experiments involving inter-

face validation, controller selection, and virtual annotation optimisation, our results demonstrate that the hybrid methodology provides a more comprehensive and nuanced analysis than purely subjective or objective approaches.

The results also highlight the technical and methodological challenges associated with integrating biosensors into VR environments, particularly regarding data synchronization, EEG signal filtering, and the fidelity of physical simulations. Despite these challenges, this thesis proposes ways to enhance data analysis in VR, including the integration of new sensors and the use of advanced signal processing techniques.

In conclusion, this research contributes to the advancement of VR design methodologies by demonstrating the effectiveness of the hybrid approach in optimising user experience and providing insights for future research in this rapidly evolving field.