



HAL
open science

Towards the Exploitation of Mental Workload in Virtual Reality

Tiffany Luong

► **To cite this version:**

Tiffany Luong. Towards the Exploitation of Mental Workload in Virtual Reality. Vision par ordinateur et reconnaissance de formes [cs.CV]. Université Rennes 1, 2021. Français. NNT : 2021REN1S069 . tel-03546229

HAL Id: tel-03546229

<https://theses.hal.science/tel-03546229>

Submitted on 27 Jan 2022

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.

THÈSE DE DOCTORAT DE

L'UNIVERSITÉ RENNES 1

ÉCOLE DOCTORALE N° 601
*Mathématiques et Sciences et Technologies
de l'Information et de la Communication*
Spécialité : *Informatique*

Par

Tiffany Luong

Towards the Exploitation of Mental Workload in Virtual Reality Systems

Thèse présentée et soutenue à Rennes, le 10 Février 2021

Unité de recherche : Inria

Thèse N° : 00000

Rapporteurs avant soutenance :

Daniel MESTRE
Anthony STEED

Directeur de Recherche, CNRS, France
Professeur, University College of London, United Kingdom

Composition du Jury :

Président : Catherine PELACHAUD
Examineurs : Mark BILLINGHURST
Catherine PELACHAUD
Dir. de thèse : Anatole LÉCUYER
Guillaume JEGOU
Encadr. de thèse : Ferran ARGELAGUET
Nicolas MARTIN

Directrice de Recherche, CNRS, France
Professeur, University of South Australia, Australia
Directrice de Recherche, CNRS, France
Directeur de Recherche, Inria, France
Responsable de Laboratoire, b<>com, France
Chargé de Recherche, Inria, France
Ingénieur de Recherche, b<>com, France

Attention, en cas d'absence d'un des membres du jury le jour de la soutenance, la composition du jury doit être revue pour s'assurer qu'elle est conforme et devra être répercutée sur la couverture de thèse

TOWARDS THE EXPLOITATION OF MENTAL WORKLOAD IN VIRTUAL
REALITY SYSTEMS

TIFFANY LUONG

Inria Rennes
Université de Rennes 1

*We have seen that computer programming is an art,
because it applies accumulated knowledge to the world,
because it requires skill and ingenuity, and especially
because it produces objects of beauty.*

— Donald E. Knuth [145]

ACKNOWLEDGEMENTS

Ces trois années sont passées à une vitesse folle, avec des moments forts qui resteront gravés en moi, notamment grâce à ce manuscrit. C'est l'occasion pour moi de remercier tous ceux qui ont, d'une manière ou d'une autre, contribué à ce résultat.

Je souhaite remercier en premier lieu l'ensemble des membres de mon jury : Dr. Catherine Pelachaud, qui m'a fait l'honneur de présider ma défense de thèse, Dr. Daniel Mestre et Pr. Mark Billinghurst. I would also like to thanks Pr. Anthony Steed, who kindly accepted to host me at UCL in the context of an international exchange, although unfortunately this did not happen due to the pandemic. It would have been an honor and a pleasure for me, and I hope to have the opportunity to visit the Virtual Environments and Computer Graphics group in the next few years. Tous ont accordé beaucoup de leur précieux temps à la lecture de mes travaux. J'ai eu l'honneur et le plaisir d'être évaluée par des chercheurs de grand calibre, et je leur suis particulièrement reconnaissante pour leurs commentaires pertinents, enrichissants et si constructifs.

Il va sans dire que mes prochains remerciement vont envers mes encadrants, sans qui rien de tout cela n'aurait été possible. J'ai eu la chance inestimable d'être entourée d'encadrants brillants, présents et bienveillants tout au long de ces trois années de thèse. Merci absolument pour tout, Anatole. Tes encouragements constants, ton expérience, et l'importance que tu portes au bien-être de tes doctorants et de ton équipe m'ont permis de tirer le meilleur de moi-même et d'effectuer cette thèse dans un cadre en or. Merci d'avoir toujours cru en moi et en mes compétences, de m'avoir fait confiance en toutes circonstances, d'avoir trouvé les mots justes, notamment dans les moments difficiles. Merci de m'avoir soutenu tout le long, de toujours t'être rendu disponible malgré un emploi du temps chargé, de m'avoir guidé tout en me laissant l'autonomie nécessaire pour m'épanouir dans mon travail, et de m'avoir encouragé dans tous mes projets, notamment créatifs, en m'intégrant à de nombreux comités et hackatons. Je pourrais sûrement continuer la liste indéfiniment, mais le fait est que je te suis infiniment reconnaissante. Ton éloquence et ton parcours sont une source d'inspiration pour moi. J'ai appris énormément à tes côtés, et je suis fière et honorée d'avoir travaillé sous ta direction pendant cette thèse !

Un très grand merci à mes encadrants Ferran Argelaguet et Nicolas Martin, qui ont été présents au quotidien au cours de ces trois années. Merci à

toi Ferran pour ta prévenance, tes conseils avisés, et pour avoir été un superviseur accessible. Il va sans dire que tes compétences m'ont grandement aidé, et je te remercie pour ta patience au vue de mes nombreuses questions et remises en question tout au long de ces trois années. Merci pour les « warnings » lorsque mes idées se démultipliaient, et je me souviendrai toujours de tes « cut » de slides et de paragraphes. J'espère appliquer tout ce que tu m'as appris encore longtemps ! Merci à toi également, Nicolas, pour ta patience et ton calme. Tu m'as beaucoup aidé à m'intégrer dans ce monde semi-industriel au cours de cette thèse. Merci de m'avoir aidé à gérer la charge de travail des deux côtés et d'avoir été là aux moments les plus importants de ma thèse. J'aimerais également te remercier, Jean-Marc. Malgré la distance, tu m'as introduite au domaine de la psychologie au début de ma thèse et m'a aidé à prendre en main les capteurs physiologiques, ce qui n'était pas une mince affaire. Merci aussi à toi Guillaume, pour m'avoir permise de faire cette thèse à l'IRT dans ces conditions. J'ai eu la chance d'avoir pu travailler sur un simulateur de vol en réalité virtuelle avec des capteurs physiologiques et des casques de réalité virtuelle dernier cri dans un espace de 100 m². Je suis consciente d'être particulièrement chanceuse.

Je tiens également à remercier ma famille. Merci papa, maman, de m'avoir soutenue dans ce que je voulais faire – même si cela vous semblait parfois obscure, d'être fiers de moi, et de me le rappeler dès que vous le pouvez. J'ai conscience que j'ai de la chance d'avoir pu mener mes études à bien, d'avoir pu choisir la voie et le métier que je voulais, ce qui n'était pas forcément donné aux femmes dans la culture dans laquelle vous avez grandi. Je n'ai manqué de rien à ce niveau-là, ai pu assouvir ma soif de connaissance grâce à tout ce que vous m'avez offert, et je vous en suis profondément reconnaissante. Le droit à l'éducation est une chance que je chérirai et porterai tout au long de ma vie. J'ai aussi la chance d'avoir des frères géniaux. Merci à tous les deux, Anthony et Christophe. On a souvent dû gérer des situations difficiles durant ces trois années, et je suis contente que vous soyez là. Malgré nos différences, vous avez toujours su garder l'esprit ouvert, notamment devant mon trop-plein de rationalité. Je sais que je pourrais toujours compter sur vous, et j'espère que vous savez que vous pouvez également compter sur moi.

Un remerciement particulier à toi, Mathieu. Merci d'avoir relu et corrigé mon manuscrit, de m'avoir aidé à répéter ma soutenance et corriger ma prononciation anglaise. Plus globalement, merci d'avoir été présent pour moi, d'avoir partagé mes moments de joie, de peine, et d'évasions pendant cette thèse et depuis plus longtemps encore. Merci de m'avoir encouragée, soutenue, écoutée râler, poussée en avant, et redonnée confiance en moi tout au long de mon parcours à chaque fois que je doutais. Je t'aime !

Je souhaite également remercier mes collègues et amis d'hybrid à l'Inria, en particulier Hakim Si Mohammed et Etienne Peillard pour avoir pris le temps de relire ma brique de thèse. Vos retours m'ont beaucoup aidé ! Je me suis faite de nombreux beaux souvenirs au cours des conférences,

sorties, midi SSB, séances de sport/dessin/peinture/cuisine, soirées (karao-ké, jeux-vidéos, jeux de plateau), évènements (barbecues Inria, JSM, séminaire au vert) et hackatons partagés ensemble dans l'équipe (on se souviendra des 20 km de marche en sortant de l'aéroport à Osaka à essayer des restaurants au hasard, Ferran, Hakim, Hugo, et du fameux 3DUI contest pyramidale avec des interfaces à retour haptique passives . . .). Merci aux actuels et anciens membres d'hybrid pour avoir été comme une deuxième grande famille pour moi pendant ces trois ans, ainsi qu'aux personnes rencontrées en cours de chemin à l'Inria : Adrien Reuzeau, Alexandre Audinot, Antoine Costes, Antonin Bernardin, Benoit Legouis, Bruno Arnaldi, Camille Jeunet, Carl Jorgensen, Caroline Lebrat, Diane Dewez, Emeric Goga, Fabien Lotte, Flavien Lécuyer, Florian Nouviale, François Lehericey, Gerard Gallagher, Giulia Lioi, Guillaume Bataille, Guillaume Costes, Guillaume Gicquel, Guillaume Moreau, Guillaume Vailland, Gwendal Fouché, Gwendal Lemoullec, Hugo Brument, Jean-Marie Normand, Jimmy Petit, Jussi Lindgren, Lorraine Perronnet, Ludovic Hoyet, Martin Guy, Mathis Fleury, Maud Marchal, Mélanie Cogné, Nami Ogawa, Nathalie Denis, Nicolas Olivier, Panagiotis Koutesis, Rebecca Fribourg, Romain Lagneau, Ronan Gaugne, Théo Perrin, Thierry Gaugry, Thomas Howard, Thomas Rinnert, Valérie Gouranton, Victor Mercado, Vincent Goupil, Xavier De Tinguy, et Yoren Gaffary. J'ai beaucoup partagé avec chacun d'entre vous, et j'aurai souhaité ajouter un petit mot spécial pour chacun, mais cette section risquerait de s'alourdir considérablement (et des dossiers risqueraient de sortir ahah). Vous êtes tous des gens géniaux ! J'ai passé beaucoup de temps à vos côtés, et je continuerai très certainement à garder contact avec une bonne partie d'entre vous dans le future.

Je tiens également à remercier mes collègues de b<>com ! Un grand merci, en particulier, à Ludovic Noblet pour m'avoir fait confiance et soutenue tout au long de ma thèse, Martin Ragot, pour m'avoir accueillie avec tant de bienveillance dans l'IRT (malgré quelques fourberies, on se souviendra du piège posé pendant mes deux premières semaines), Lucie Banatre, pour m'avoir accompagnée lors de cette première expérience de soumission de brevet, Pierre Le Gargasson, pour ton support technique et le partage de ta passion des simulateurs, ainsi que Agnès Vanlerenberghe, Albert Murienne, Amine Kacete, Anas El Rhammad, Antoine Turbin, Antonin Gilles, Aurélie Desmarest, Bertrand Guilbaud, Boris Labbe, Cambodge Bist, Cédric Morin, Charles Bonnineau, Delphine Monier, Emmanuelle Garnache, Fatemeh Nasiri, Guillaume Vallée, Glenn Herrou, Hadrien Gurnel, Jean-Michel Fazzari, Jérémy Lacoche, Jérôme Royan, Julia Jetin, Laurent Launay, Luc Le Magoarou, Lucie Petta, Mathieu Quentel, Maud Hervieux, Nam Duong Duong, Nico Pallamin, Nour Hoblosse, Patrick Gioia, Pauline Desnos, Phillipe Bertin, Romain Terrier, Sophie Clavier, Thomas Boggini, Vincent Savaux et Yann Cariou pour votre professionnalisme et tous les agréables moments partagés ensemble au cours de ces trois ans, notamment au cours des mythiques b<>com awards, séances de sport intensives, salons, interviews, midi « Zest » et sorties à « La Place ». Vous aurez chacun con-

tribué à égayer ma vie de doctorante dans cette fabuleuse région qu'est la Bretagne.

Merci à mes anciens camarades de classe et personnes rencontrées au cours de ma scolarité, notamment ceux de la prom's Me214 et ceux du Master Recherche MTI3D des Arts et Métiers. En particulier, merci à toi, Simon Richir pour ton enthousiasme, et à Geoffrey, Alex, Violaine, Edouard et Thibaud pour avoir contribué à maintenir mon émerveillement des technologies immersives qui m'a amenée à faire cette thèse.

Merci à mes amis pour m'avoir soutenue et pour m'avoir aidée à maintenir une vie sociale en dehors de la thèse. En particulier, merci à vous, Aude, Lisa, d'être là depuis des années, et à tous les gens du Thai Kung Fu, Pascal, Matthieu, Aurélien, Joy, Nelly, Céline, Angélique, Jordan, notamment pour avoir contribué à développer ma combattivité (cela m'a valu une sacrée réputation où que j'aille et le surnom de bulldozer) ! Je tiens également à remercier le reste de ma très grande famille, mes grands-parents, mes tatas et tontons en France et répartis dans le monde, et mes cousins, Julie, Elodie, Daniel, Camille, Kévin, Tom, Alexis, Axel, Célia, Cédric, Paul et Héloïse pour les cousinades, les barres de rire, les nouvel ans chinois à Angers, et les magnifiques souvenirs au mariage de Camille et Kévin ! Un grand merci aussi la famille Huot pour leur bienveillance, pour leur accueil chaleureux et pour m'avoir régaler avec leurs délicieux plats pendant toutes ces années.

Un remerciement particulier pour l'IRT b<>com et l'ANR, qui ont financé ma thèse.

A tous ceux que j'ai pu omettre de remercier, car j'ai eu la chance d'être entourée et de croiser le chemin de nombreuses personnes au cours de cette aventure qui ont pu m'apporter leur aide et leur soutien, sachez que je ne vous en suis pas moins reconnaissante !

Encore une fois, un immense **MERCI** à vous tous !

CONTENTS

INTRODUCTION	1
1 Context	1
2 Research Challenges	2
2.1 Induction of MW in VR	3
2.2 Recognition of MW in VR	4
2.3 Exploitation of MW in VR	4
3 Approach and Contributions	5
1 RELATED WORK ON COGNITIVE AND AFFECTIVE STATES IN VR	9
1 "Affective and Cognitive VR"	10
2 Cognitive and Affective States: Definitions and Models	12
2.1 Cognitive States	12
2.2 Affective States	15
2.3 Complementary Notions	17
3 Measuring Cognitive and Affective States	19
3.1 Self-Report Measures	20
3.2 Observational Measures	23
3.3 Task Performance Measures	25
3.4 Psychophysiological Measures	26
4 Overview of studies in Affective and Cognitive VR	32
4.1 Induction of Cognitive and Affective States in VR	32
4.2 Recognition of Cognitive and Affective States in VR	41
4.3 Exploitation of Cognitive and Affective States in VR	43
5 Mental Workload in VR Training	45
5.1 Designing Training Scenarios in VR	47
5.2 Using Mental Workload in VR Training Scenarios	48
6 Conclusion	50
2 STUDYING THE MENTAL EFFORT IN VIRTUAL VERSUS REAL ENVIRONMENTS	53
1 Experiment 1: Effect of VR HMD on mental effort	54
1.1 Apparatus	54
1.2 Participants	56
1.3 Experimental Design	57
1.4 Results	63
1.5 Discussion	65
2 Experiment 2: Accommodation to VR Effect	66
2.1 Apparatus	66
2.2 Participants	66
2.3 Experimental Design	67
2.4 Results	69
2.5 Discussion	71
3 General Discussion	72

3.1	Impact of Wearing an HMD on the User’s Mental Effort	72
3.2	Lessons Learned	73
4	Conclusion	74
3	A METHODOLOGICAL APPROACH TO INTRODUCE MENTAL WORKLOAD ASSESSMENT IN THE DESIGN OF VR TRAINING SCENARIOS	77
1	Methodological Approach	78
1.1	Tasks Identification and State Machine Generation .	78
1.2	Data Assessment	79
1.3	Scenario Generation	80
2	Validation of the Approach: Designing Training Scenarios Based On Mental Workload in a VR Flight Simulator	80
2.1	Tasks Design	80
2.2	State Machine Structure	84
2.3	Mental Workload Measures	85
2.4	Apparatus	85
2.5	Experiment 1: Data Assessment	86
2.6	Experiment 2: Scenarios Generation	92
3	General Discussion	96
4	Conclusion	98
4	TOWARDS REAL-TIME RECOGNITION OF USERS’ MENTAL WORK- LOADS USING INTEGRATED PHYSIOLOGICAL SENSORS INTO A VR HMD	99
1	All-in-one Approach to Assess Users’ MWs in VR in Real-Time	101
1.1	Integrated Hardware	101
1.2	Data Processing and MW Assessment	103
2	Data Acquisition Protocol	107
2.1	Tasks Design	108
2.2	Apparatus	109
2.3	Collected Data	110
2.4	Experimental Design	111
2.5	Participants	111
2.6	Experimental Procedure	112
2.7	Resulting Data	112
3	Results	113
3.1	Normalization of Physiological Features	113
3.2	CGS Sensors	114
3.3	Sensors Integrated Into the HMD	114
3.4	Mental Workload Recognition in Real-Time	117
4	Discussion	118
5	Conclusion	121
5	CONCLUSION	123
1	Contributions of This Thesis	123
2	Limitations and Future Work	125
2.1	Impact of VR on MW	125

2.2	Introducing MW assessment in the design of VR training scenarios	126
2.3	Towards the real-time recognition of users' MW in VR	127
3	Long-term perspectives	128
A	IMPACT OF DATA SPLITTING ON MENTAL WORKLOAD RECOGNITION ACCURACY	131
B	RÉSUMÉ LONG EN FRANÇAIS	133
	AUTHOR'S PUBLICATIONS	143
	BIBLIOGRAPHY	145

LIST OF FIGURES

Figure 1	The adaption of VR based on user’s MW. The adaptation is mainly composed of 3 blocks: the induction block, the recognition block, and the exploitation block.	3
Figure 2	Mapping the chapters of this thesis to the different research axes they targeted.	5
Figure 3	Illustration of the roadmap of this thesis.	6
Figure 4	Loop depicting an adaptation of VR content based on users’ CAS (Generalization of Figure 1). The 3 blocks represent the proposed categorization of “Affective and Cognitive VR ”studies. VR stimuli first <i>induce</i> a CS or an AS to users. Then, measures are assessed from users, which are used to <i>recognize</i> users’ CS or AS. Finally, the predicted CS or AS is <i>exploited</i> to adapt VR stimuli to drive users toward a targeted CS or AS.	11
Figure 5	The 4-D multiple resource model [304].	13
Figure 6	Schematic representation of the cognitive load construct [214].	14
Figure 7	The 3 types of cognitive load from the cognitive load theory [215].	14
Figure 8	Examples of representations of emotions from a discrete perspective.	15
Figure 9	Examples of representations of emotions: (a) from a dimensional perspective; (b) from a hybrid perspective.	16
Figure 10	Csikszentmihalyi’s Flow model [68]: mental state in terms of challenge level and skill level.	18
Figure 11	Some classical tools used for custom scales and questionnaires.	20
Figure 12	Example of self-reporting method in an IVE. Instantaneous Self-Assessment (ISA, measure of the user’s MW using 5 different ratings [287]) in a VR Flight Simulator [172].	21
Figure 13	The Self-Assessment Manikin (SAM) [39]. The first line of characters corresponds to the valence range, the second, to the arousal, and the third, to the dominance dimension.	22
Figure 14	Recognition of emotions based on facial expression recognition using photoreflective sensors on the VR HMD [283].	23

Figure 15	(a) Typical ECG waveform (image from [90]). (b) Typical PPG waveform (image from [222]).	26
Figure 16	Electrodermal Activity (EDA) signal decomposition into tonic and phasic components (image from [232]).	27
Figure 17	Constriction and dilation of the pupil (image from [234]).	28
Figure 18	The cortex subdivided into the frontal, temporal, parietal, and occipital lobes (image from [5]). . . .	29
Figure 19	Illustration of the path (in red) followed by the near-infrared photons from the light source to the detector through the different layers of the head using fNIRS systems. The penetration depth of the light is proportional to the source-detector distance (d_1 : deeper channel, d_2 : superficial channel) (image from [229]).	30
Figure 20	PhysioHMD [27]. Example of a module which integrates physiological sensors directly on the headset.	31
Figure 21	Examples of cognitive tasks performed in VR. . . .	34
Figure 22	Induction of 2 levels of arousal and 2 levels of valence using 4 architectural 360° panorama environments in a VR HMD [183]. Left: negative valence; right: positive valence; bottom: low arousal; top: high arousal.	35
Figure 23	Induction of anxiety in a competitive sports VR context (CAVE-like system) with animated virtual characters [11].	36
Figure 24	(a) The virtual pit experience developed by [191]. (b) Influence of avatars on cognitive performances and age-base discrimination (Einstein vs. gender-matched body) [17].	36
Figure 25	Example of implicit exploitation of users' emotions [27]. The particles size, density, brightness, and colour are adjusted to make avatars more emotionally expressive.	44
Figure 26	A Petri net enhanced with the event model, as used in #SEVEN [64].	48
Figure 27	Example of task difficulty adaptation. The distance of the object to grab (ball) in VR and the target (basket) was adapted based on users' performances [110]. The reach-to-grasp movements were supported by an exoskeleton because of the severity of the patients' impairments to provide passive gravity compensation.	49

Figure 28	Illustration of the first user study. The effect of wearing a VR HMD on users' mental efforts was measured using different metrics.	54
Figure 29	(top) Photo of the room in which the experiment took place. (bottom) Screenshot of the virtual environment representing the same room modelled in 3D for the VR conditions.	55
Figure 30	Photos of the experimental set-ups in the sitting condition (on the left) and in the walking condition (on the right). (1) HTC Vive HMD (VR) or SMI glasses (Real), (2) Micro-Headset, (3) Vive Controller with a Force Sensitive Resistance sensor on the trigger button, (4) Shimmer3 GSR+, (5) Vive tracker.	56
Figure 31	Experiment 1: Process of the experiment for one subject (excluding pre-, post-questionnaire, instructions, set-up, and calibrations).	58
Figure 32	Progress of one trial: (a) Visual display of the task difficulty on the (real or virtual) TV screen ; (b) Auditory announcement of the difficulty, animated GIF reminding the instruction for the difficulty, and initialization of the task (here, for the 1-back task) ; (c) Start of the task with a beep, followed by the sequence of audio letters (N+25 letters) ; (d) Rating Scale of Mental Effort (RSME) [318]. Steps (b) and (c) are triggered by the trigger button, and step (d) appears at the end of the task (step (c)).	62
Figure 33	Mean mental effort reported with the RSME and mean accuracy considering all independent variables.	64
Figure 34	Mean indicators of mental effort (response time, exerted pressure on the button, HR) in terms of the secondary task and difficulty. As the environment did not influence significantly these measures, only the effect of the task is displayed.	65
Figure 35	Experiment 2: Process of the experiment for one participant (excluding pre, post-questionnaire, instructions, set-up, and calibrations). The order of the exposure to each environment was counterbalanced (either Real-VR or VR-Real).	68
Figure 36	Mean indicators of mental effort (RSME, response time) depending on the Trial Number (= time) and environment (i.e., VR vs. Real).	70
Figure 37	Illustration of the methodological approach to introduce MW assessment for the design of VR training scenarios.	78

Figure 38	Virtual cockpit view. (1) Instantaneous Self Assessment (ISA) interface. (2) Resources management task interface (deactivated); when activated, the interface lit up (with a green outline). (3) Communication task interface (activated); when deactivated, the interface lit off (no green outline). (4) Informative panel which gives information about which task is activated or not at the current time. (5) Virtual representation of the joystick used to pilot the aircraft and of the right hand. The left hand is represented in the same way, but tracked by a Vive Controller and animated depending on the interaction.	81
Figure 39	The difficulty of the piloting task was modulated by changing the number of circles visible by users at a given time (i.e., $N_{visible}$) and the speed of the aircraft.	82
Figure 40	Resources management task interface. A and B are the main tanks; their fuel levels are indicated below the tanks. C and D are supply tanks; their fuel levels are indicated on their right side. E and F are supply tanks with unlimited capacities. The buttons numbered from 1 to 8 are pumps button. Pumps 4 and 6 are activated, pump 5 is failed, and all other grey pump buttons are deactivated.	83
Figure 41	State machine of the designed VR flight simulator. Each state is labelled using 3 digits. The first digit refers to the level of the piloting task (0-easy, 1-medium, 2-difficult), the second one to the level of the communication task (0-deactivated, 1-activated) and the third one to the level of the resources management task (0-deactivated, 1-activated). The colours of the nodes represent the mean ISA assessed during the first user study, which are as well indicated in the brackets. Only the transitions between consecutive states are depicted there.	84
Figure 42	Picture of the experimental set-up. The user is wearing a Vive Pro, and using a joystick and a Vive controller on the cockpit.	86
Figure 43	Progress of the tasks and of the ISA calls within one state in the first experiment. In this paper, 12 states were considered. The communication task and the resource management task were not always activated. The ISA is always present. Each state lasted at least 90 seconds (depending on the user's piloting path).	88

Figure 44	Mean ISA in terms of tasks levels and configurations. "Radio" refers to the communication task and "Fuel", to the resources management task.	89
Figure 45	All ISA values (subjective MW) obtained per state during the data assessment experiment (38 participants reported their MW 3 times per state).	91
Figure 46	Comparison of the users' subjective MW (mean ISA value) reported overtime for all 3 scenarios between the first and the second experiment. The first scenario was intended to induce a MW around $ISA \approx 3$ all long and the second one, to induce a low MW ($ISA \leq 3$) followed by a high MW at last. The last scenario objective was to induce a progressively increasing MW.	94
Figure 47	Results of the questionnaire asking users about their perceived progression of difficulty after each scenario. Corresponding "questions" and scenarios are highlighted in red.	95
Figure 48	Results of the mapping of scenarios descriptions to the experienced scenarios (after all 3 scenarios). D1: <i>The experiment induced a medium MW level all long.</i> D2: <i>The experiment induced a low MW level first, and a suddenly high MW level in the end.</i> D3: <i>The experiment induced a progressively increasing MW level throughout the experiment.</i> Corresponding descriptions and scenarios are highlighted in red.	95
Figure 49	Illustration of the all-in-one approach to do MW recognition in VR in real-time using physiological sensors integrated into a VR HMD. The MW recognition model was trained using the HMD sensors signal features labelled with users' self-reported MW level as input and the random forest algorithm.	101
Figure 50	Hardware solution. The sensors are placed on a Vive Pro Eye HMD, which has built-in eye-tracking. (1) PPG sensor, (2), electrodes to assess the EDA, (3) electronic card, (4) 3D printed case.	102
Figure 51	Software processing chain for the real-time recognition of users' MW. There are 2 main steps: the training of the recognition model and the real-time use of the model.	103
Figure 52	Illustration of a data windowing in the case of a physiological signal (in green). The time window (in red) is used to extract the physiological features based on given labels.	104

Figure 53	Interface for the MW recognition in real-time. It depicts on the left: the EDA, BVP, and pupil diameter signals evolution over time in a time window of 30 seconds; on the right: the predicted MW level (i.e., the probability that a user is at a particular MW level) using features extracted from the 30 seconds of data. The prediction and interface are updated every 1 second.	108
Figure 54	Multiple views of the experimental setup. The user is wearing a Vive Pro Eye equipped with sensors on the headset, a Shimmer3 GSR+ on his left hand, and using a joystick and a Vive Controller on the cockpit.	110
Figure 55	Confusion matrix (expected and predicted classes) in the "All physio+perf" using NSF normalization setup for (a) the "CGS sensors" and (b) the "HMD sensors".	115
Figure 56	Comparison of the "CGS sensors" and "HMD sensors" setups on the classification accuracy in the different measurement setups. Mean values are indicated and the error-bars correspond to the standard deviations considering the different normalization methods.	116
Figure 57	Comparison of the different normalization methods on the classification accuracy in the different measurement setups. Mean values are indicated and the error-bars correspond to the standard deviations considering the HMD and CGS sensors setups. . .	116
Figure 58	Simulation of the real-time prediction of MW using the data of one of the participants and the HMD sensors in the "All Physiological signals + task performance" setup with NSF normalization. The blue line represents the ground truth (i.e., the MW level reported by the participant); the red line represent the predicted MW using our trained recognition model.	117
Figure 59	Confusion matrix (5 classes) - <i>Accuracy</i> = 61.78%	131
Figure 60	Confusion matrix (3 classes) - <i>Accuracy</i> = 68.95%	132
Figure 61	Confusion matrix (2 classes) - <i>Accuracy</i> = 82.15%	132
Figure 62	Adaptation de la RV en fonction de la CMT des utilisateurs. L'adaptation est décomposée en principalement 3 blocs : l'induction, la reconnaissance et l'exploitation.	135
Figure 63	Correspondance entre les chapitres de la thèse et les différents axes de recherche ciblés.	138
Figure 64	Illustration de la feuille de route de cette thèse. . .	139

LIST OF TABLES

Table 1	The general movement protocols for the Big Six emotions recapitulated in Noroozi et al.'s survey [202].	24
Table 2	Summary of the 63 studies treated in the state of the art in the 3 proposed categories. In the "Induction of CAS" category, the works focused on the influence of variables on CAS. In "Recognition of CAS", studies tried to passively recognize CAS. In the "Exploitation of CAS" category, works focused on the adaptation of VR content and parameters based on CAS measures.	33
Table 3	Summary of works which induced Cognitive and Affective States (CAS) in Immersive Virtual Environments (IVEs). The CAS measures which have shown significant changes are indicated in bold in "Objective Measures" and "Subjective Measures" columns. Elements which had a significant effect on them are indicated in bold in the "Inducer" column. In the "Inducer" column, the emotional virtual environments (VEs), videos, and games are not detailed when several of them are used in the study.	38
Table 4	Summary of works studying the classification of Cognitive and Affective States (CAS) in IVEs.	43
Table 5	Summary of works which exploited Cognitive and Affective States (CAS) in IVEs. Among the different types of exploitation, Explicit refers to an adaptation of the IVE knowingly to users, and Implicit refers to an adaptation of the IVE unknowingly to users. BF : Biofeedback, it is a subcategory of "Explicit exploitation" and refers to studies which incite users to control their physiological state through feedback in the IVE. PF : Physiological Feedback, it is another subcategory of "Explicit exploitation" and refers to studies which directly display the users' physiological signals features in the virtual environment or to a third party interface.	46
Table 6	Experiment 1: independent and dependent variables.	57
Table 7	Experiment 2: independent and dependent variables.	67
Table 8	Data assessment experiment: independent and dependent variables.	87

Table 9	Scenarios generation: independent and dependent variables.	93
Table 10	Custom questionnaire on users' perceived progression of difficulty after each scenario. These questions were mixed with 3 other unrelated questions linked to users' absorption and sense of presence to mask the study objective. Users were asked to answer each affirmation using 7-points Likert-Scales.	95
Table 11	Target variables and extracted physiological and performance features used for the training of the recognition models.	106
Table 12	Contingency table of subjective responses.	113
Table 13	MW classification accuracy results (in %) using the Shimmer sensors (i.e., CGS sensors) in function of the normalization methods and of the type of measure. "Perf" corresponds to the task performance. Task performances and ocular activity are common to the CGS and HMD sensors.	115
Table 14	MW classification accuracy results (in %) using the sensors integrated into the VR HMD (i.e., HMD sensors) in function of the normalization methods and of the type of measure. "Perf" corresponds to the task performance. Task performances and ocular activity are common to the CGS and HMD sensors.	115

ACRONYMS

VR	Virtual Reality
MW	Mental Workload
HMD	Head-Mounted Display
IVE	Immersive Virtual Environment
ML	Machine Learning
ANS	Autonomic Nervous System
CNS	Central Nervous System
VE	Virtual Environment
CAS	Cognitive and Affective States
CS	Cognitive States
AS	Affective States
FACS	Facial Action Coding System
PAD	Pleasure, Arousal, Dominance
LS	Likert-Scale
VAS	Visual Analogue Scale
NASA-TLX	NASA Task Load Index
SWAT	Subjective Workload Assessment Technique
WP	Workload Profile
RSME	Rating Scale of Mental Effort
ISA	Instantaneous Self Assessment
SAM	Self-Assessment Manikin
PANAS	Positive and Negative Affect Schedule
STAI	State-Trait Anxiety Inventory
STAXI	State-Trait Anger Expression Inventory
EMG	Electromyography
ECG	Electrocardiography

PPG	Photoplethysmography
HR	Heart Rate
HRV	Heart Rate Variability
IBI	Interbeat Interval
BVP	Blood Volume Pulse
EDA	Electrodermal Activity
GSR	Galvanic Skin Response
SCL	Skin Conductance Level
SCR	Skin Conductance Response
Resp	Respiration
Temp	Skin Temperature
EOG	Electrooculography
EEG	Electroencephalography
fNIRS	functional Near-Infrared Spectroscopy
AI	Artificial Intelligence
IAPS	International Affective Picture System
IADS	International Affective Digital Sounds
SUDS	Subjective Units of Discomfort Scale
SVM	Support Vector Machines
LDA	Linear Discriminant Analysis
RF	Random Forest
PCA	Principal Component Analysis
BCI	Brain-Computer Interfaces
FSR	Force Sensitive Resistance
MATB-II	Multi-Attribute Task Battery II
CGS	Commercial Grade Systems
EMD	Empirical Mode Decomposition
RV	Réalité Virtuelle(FR)
CMT	Charge Mentale de Travail(FR)
EVI _s	Environnements Virtuels Immersifs(FR)
EV	Environnement Virtuel(FR)

INTRODUCTION

1 CONTEXT

A century ago, 80 percent of airplane crash accidents were caused by the machine, and 20 percent, by human error. Today the statistic has reversed, with approximately 80 percent of aviation crashes due to human error and 20 percent, due to machine failures [1, 305]. Hawkins depicted human factors as “*the Last Frontier of the airline safety problem*” [116]. Modern aircrafts are so complex today that they can disorient even the most experienced pilots. To palliate this issue, flight simulators have been developed in order to train aircrews to crisis situations. In particular, immersive technologies such as Virtual Reality (VR) are largely employed for training purposes in reason of their reduced cost compared to true-scale simulators.

In this training context, Mental Workload (MW) has long been a topic of interest for research [77, 114, 195, 205, 304, 312]. While there is no universal consensus over its definition, MW can be described as “*the relation between the (quantitative) demand for resources imposed by a task and the ability to supply those resources by the operator*” [303]. Its metrics remains popular today as systems have become more and more complex. This is not restricted to the aviation industry as a number of reports suggests that the incidence of human error has risen dramatically in recent years in a range of industries [121, 218]. Smartphones, media, and work tools are overflowed with information, and software and machines parameters options are increasing overtime. Multitasking is also prevalent as much in our working environment as in our daily lives. These situations can draw humans to be overloaded and to make mistakes. Different solutions were considered to answer this problem; among them: (1) training users to make them used to these ever-increasing cognitive demands, and (2) automatizing systems to make them adapted to users’ skills and psychological states.

To test these solutions and improve them, VR technologies are particularly versatile. On the one hand, VR allows to design training scenarios in order to improve operators’ skills and reactions in complex situations. Indeed, VR systems have the ability to simulate complex environments while immersing users into engaging experiences in a safe and controlled way. A wealth of research already demonstrated that immersive devices like Head-Mounted Displays (HMDs) and Cave-like systems were beneficial compared to more traditional supports to train users, for example, in the medical field [4, 112, 157, 265], in the educational field [99, 119, 194], and in the industrial field [24, 109, 159, 296]. On the other hand, VR technologies can be used to evaluate users’ behaviour, performances, and psychological states before systems manufacturing. Indeed, individuals are known

to react differently depending on factors such as their skills, knowledge, current mood, culture, and on the situation. Disregarding these factors in the design of systems can make the latter unfitted to the users and lead to human errors.

Nevertheless, the intersection between VR training and systems which are aware of users' MWs remains little explored. Current VR training scenarios mostly are predefined by sequences of events, interaction feedback, and performance measures. In this context, MW is more used as an off-line metric at the end of a chain to evaluate the user experience, than in the heart of an algorithm which adapts VR to users' mental workloads dynamically. We call "*adaptation of VR to users' mental workloads*" the dynamic process that fits VR parameters or content based on acquired information about users' MW. In computer science, such adaptive processes usually aim to maximize users' performances or users' pleasures [94, 204]. The overall objective is to provide a pleasant environment for the user by processing information about their psychological states, ideally in real-time [204]. Among the different benefits, considering MW assessment in the design of training applications could make the training more tailored to the individuals' available cognitive resources, engage users to the long-term, and enhance the learning and performance outcomes. If users are overloaded by the tasks they have to perform, the tasks difficulties can be decreased, and the other way around. Such a process can be generalized outside the training field for other psychological states [93, 106, 227]. For example, in an entertainment context, the stimuli could be intensified based on users' affective states to enhance users' emotional experiences. In the therapeutical field, progressive stimuli could be used to treat the patients' disorders. Overall, going toward adaptive systems based on users' psychological states could benefit user experience and improve systems robustness and adaptability.

2 RESEARCH CHALLENGES

In this PhD, we propose to decompose the "adaptation of VR based on users' MWs" into 3 research axes represented in Figure 1:

- The **induction of MW**: VR stimuli are presented to users, which triggers responses that can be assessed using different metrics relevant to users' MWs (e.g., self-reports, tasks performances, physiological measures, behavioural measures).
- The **recognition of MW**: The recognition model is feed with the previously collected metrics and predicts users' MW levels in real-time. Usually, the model is rule-based or trained using Machine Learning (ML) algorithms.
- The **exploitation of MW**: This step focuses on the feedback logic. A set of constraints and rules are defined to choose which VR stimuli

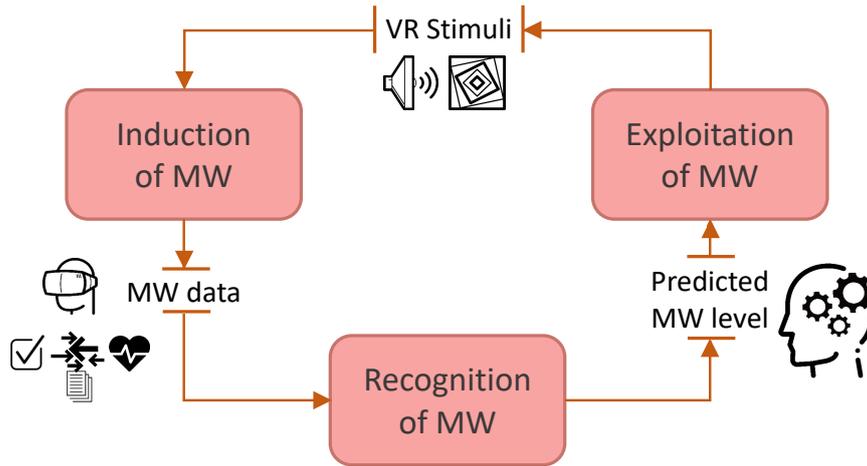


Figure 1: The adaptation of VR based on user's MW. The adaptation is mainly composed of 3 blocks: the induction block, the recognition block, and the exploitation block.

are going to be presented based on the predicted MW level to induce a targeted MW level.

These 3 categories each have their lots of challenges which are detailed hereafter.

2.1 Induction of MW in VR

Immersive Virtual Environments (IVEs) contain numerous information and parameters. This complexity makes it particularly difficult to identify which factors influence users' MWs in VR. Especially, the usage of VR devices implies a number of constraints (e.g., specific VR interaction, cybersickness) which can interact with the users' psychological state responses. VR paradigms such as redirected walking and the control of a self-avatar were already found to influence cognitive demands and users' behaviours [47, 277]. The influence of VR devices on users' MWs was, however, very little explored.

Another issue is the choice of stimuli. Those will mostly depend on the context of the application. To generalize the design of VR scenarios, the literature has presented different VR taxonomies and frameworks. One of their main challenges is to reach a balance between genericity and efficiency of the scenarios to meet the constraints and objectives of the applications. Most of these taxonomies proposed to design VR scenarios in a predefined way. Sequences of events are structured and fixed. They are triggered by users' choices and actions to meet narrative and interaction requirements. Currently, there are very few approaches that consider users' psychological states as a criterion to design VR scenarios while there are more and more applications that try to adopt a user-centred perspective. Some VR scenarios adapted the task difficulty level based on users' performances, but this was mostly done in a single task context or "by hand",

through subjective estimations. Identifying which stimuli can affect users' MWs in complex IVEs and how these stimuli interact with each other depending on the nature of the tasks could give more insight on how to structure complex IVEs in order to modulate users' MWs.

2.2 *Recognition of MW in VR*

The second research axis consists in achieving an automatic recognition of users' MWs. Recent breakthroughs in ML and deep learning made this theoretically possible [94, 227]. Especially, a wide range of research outside VR has focused on the recognition of users' psychological states based on physiological signals [204]. In the physiological computing field, physiological activities and changes are direct reflections of processes in the Autonomic Nervous System (ANS) and in the Central Nervous System (CNS) [93]. Since physiological responses are issued from psychological processes, it was hypothesized that it was possible to translate them to psychological states via the extraction of some specific features [228]. Many research in this field are showing promising results, even if the ability to make the recognition models generic to all users is still an issue [65, 182, 204]. Indeed, inferring users' subjective MW levels is intrinsically more complex than measuring the effect of parameters on MW. As a consequence, developing a recognition model necessitates to build large-sized datasets compared to usual user studies testing the effect of independent variables. There is also currently no consensus over which ML algorithms, sensors, and methodology in the signal treatment to adopt. Moreover, while the recognition of users' psychological states is an essential step to make VR applications more adapted to users, current setups are hardly adequate for daily usages. Monitoring users' physiological states usually requires a tedious process involving hardware setup, software initialization, signals monitoring, and constrained movements. These setups can be unadapted to the VR context where interactions can introduce motion artefacts to the signals. This can impede the signal treatment processing for MW recognition, especially if the objective is to perform the recognition in real-time.

2.3 *Exploitation of MW in VR*

The third research axis consists in the exploitation of MW in VR, or feedback logic using MW metrics. Different frameworks were proposed to provide feedbacks in Virtual Environments (VEs) based on users' physiological signals [29, 141] and affective states [33, 164]. However, aside from the clinical field [313], there are very little propositions of exploitation use cases supported by user studies in the VR literature, in particular, based on MW metrics. This can be explained by the challenges of the previous research axis: the recognition of users' MWs. For now, user studies mainly exploited users' psychological states measures in VR based on heuristic methods or by directly displaying the physiological signals changes in the

VE [2, 57, 78–80, 123]. The relation between psychological states and physiological signals is, however, quite complex as several physiological variables can be associated with several psychological or behavioural elements [93]. Therefore, using heuristic methods or directly physiological signals to recognize users' states can cause diagnosticity issues if the objective is to modulate users' psychological states and if the targeted psychological states can not be represented as a one-dimensional continuum [93]. On the other hand, current recognition models trained using ML algorithms allow to interpret complex physiological signals features changes but still lack of accuracy or are context-dependent, which limits the usage of the recognition outcomes to validate the proposed exploitation frameworks. Thus, there are very few studies which presented a logic of adaptation of VR stimuli and parameters based on users' MWs to draw them toward a targeted MW level. The impact of guiding users toward a targeted MW level or MW levels evolutions overtime is still little understood, as well as the effect of the modulation of the parameters used in the exploitation (e.g., changes timing).

3 APPROACH AND CONTRIBUTIONS

The main objective of this thesis is to study the exploitation of the user's mental workload in Virtual Reality systems, in particular, for VR training.

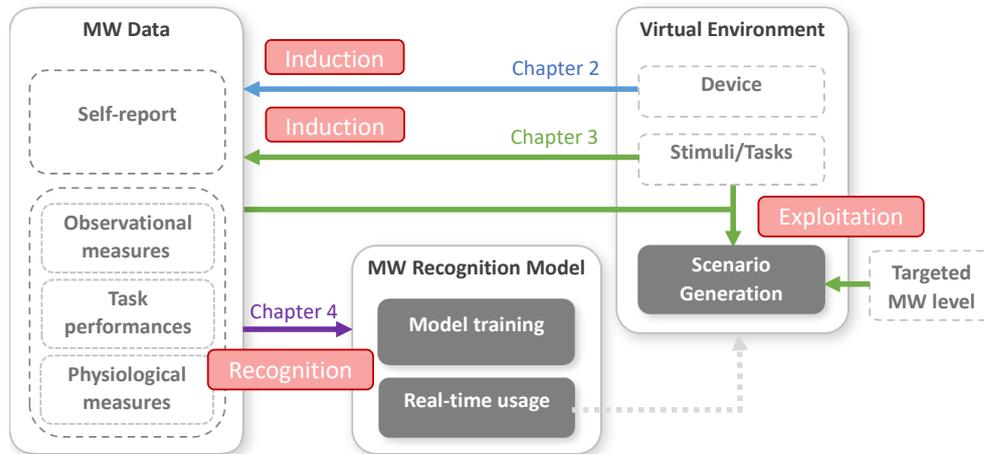


Figure 2: Mapping the chapters of this thesis to the different research axes they targeted.

Figure 2 presents the different contributions (presented in the chapters) mapped to the 3 research axes we proposed. Among the different challenges, for the induction, we chose to first control minimal effects. The effect of wearing a VR HMD on users' mental efforts was studied, as they are the most wide-spread fully-immersive VR systems [130]. In addition, the potential impacts of simple interactions like walking in IVEs and the effect of accommodation to VR were also investigated. Second, the induction and exploitation of MW parts were targeted. We focused on how to

introduce MW assessment in the design of VR training applications. The induction was tackled as the effect of different task difficulties and tasks interactions in a complex IVE was studied. For the exploitation, the VR training scenarios were designed based on a desired evolution of users' MWs overtime through the introduction of MW measures (see Fig. 2). Finally, the recognition of users' MWs in VR was explored. We proposed an all-in-one solution to recognize users' MWs in real-time using physiological sensors directly integrated into a VR HMD. Moreover, insights about the impact of sensors setups, type of measures, and signal normalization methods on the MW recognition accuracy were also investigated.

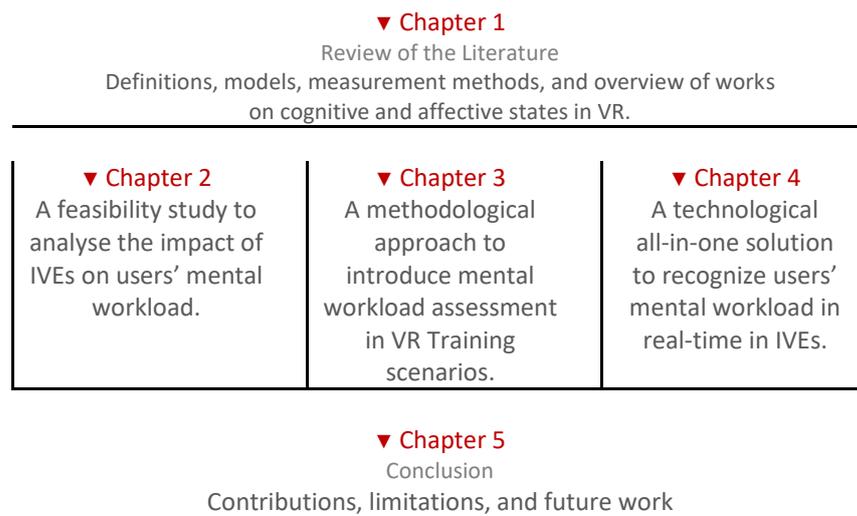


Figure 3: Illustration of the roadmap of this thesis.

Figure 3 presents a schematic representation of the roadmap of this thesis.

Chapter 1 presents a comprehensive state of the art of past works studying users' Cognitive and Affective States (CAS) in IVEs. MW is a concept which was given many definitions over the past years [312]. It is also strongly related to other psychological states such as emotional arousal, anxiety, and other concepts like stress, performances, and flow [69, 73, 196, 311]. MW should therefore be considered among them. Research studies associating VR and CAS often share the same scheme and common objectives. Therefore, we first propose a categorization of works studying CAS in VR and a definition of "Affective and Cognitive VR". Then, the definitions and models of the main CAS and psychological concepts studied in VR are clarified. Thereafter, the methods to measure them are presented, with a particular focus on their advantages and drawbacks in VR. We present an overview of past research studying CAS in IVEs in a non-clinical context. Finally, we focus more particularly on MW studies in VR training.

In **Chapter 2**, we study the impact of VR HMDs on users' mental effort. Because VR users feel "*there, in the virtual world*" [270], numerous studies

hypothesized that their responses in the IVEs were similar than in the real world. As such, it was shown that skills acquired in VR could be transferred in the real world [55, 246, 265, 300]. At the same time, users often continue to perceive and think that this world they visualize is not real. Such biased perception and the relative constraints of VR context (e.g., amount of interaction, cumbersomeness of VR equipment, lack of visual feedback of the real world, cybersickness) could require additional information processing from the users, which in turn might increase their MWs. Two user experiments were conducted to study the influence of wearing a VR HMD on users' mental efforts using an auditory cognitive task. Different metrics were assessed to study users' subjective mental efforts, task performances, behaviours, and physiological responses in VR (using an HMD) compared to in the real-world. In addition, the effect of natural walking (first experiment) and time during short exposures (second experiment) were also explored in both environments (i.e., virtual environment vs. real-world).

Chapter 3 proposes a methodological approach to introduce users' MWs measures in the design of VR training scenarios. Few VR applications were structured in order to modulate users' psychological states. Most VR scenarios are predefined or structured in order to trigger events based on users' choices and actions for narrative and interaction requirements. However, the same scenario and stimuli will induce different users' reactions based on, for example, their cognitive resources and their learning abilities. Therefore, these scenarios will not necessarily produce the expected outcomes and the best user experience depending on the users' profiles and states. In complex IVEs where there are multiple tasks and natures of stimuli, it can be particularly difficult to predict what effect a specific set of tasks will have on users. We proposed an approach to structure complex IVEs using a state machine based on tasks configurations. MW assessment is introduced in the state machine. Then, this information is used to design VR training scenarios which control MW modulation overtime. The proposed approach is illustrated by two user studies in a VR flight simulator.

Chapter 4 proposes an all-in-one approach to recognize users' MWs in real-time in VR. VR implies specific constraints and sensors setups adaptation. Few tools currently allow the monitoring of users' states in real-time in VR using ML algorithms. There is also currently no consensus over which ML algorithms, sensors, and methodology in the signal treatment to adopt. We proposed an all-in-one approach to recognize users' MWs in real-time in VR using physiological sensors directly integrated into the HMD. The approach presented in Chapter 3 was used to establish a data acquisition protocol to train MW recognition models using ML algorithms. The hardware setup is detailed, as well as the physiological signals processing pipeline to recognize users' MW levels in real-time. Sensors integrated into the VR HMD are compared to commercial-grade sensors in regards to the MW recognition accuracy. Different insights are also given about the effect of the different measures, sensors, and signals normalization treatments on the MW recognition performances.

Finally, **Chapter 5** summarizes our contributions, discuss the limits of our results and studies, and presents perspectives for future works.

RELATED WORK ON COGNITIVE AND AFFECTIVE STATES IN VR

SUMMARY This chapter aims to provide a state of the art of works studying Cognitive and Affective States (CAS) in VR. First, we propose a categorization of works studying CAS in VR, as well as a definition of “*Affective and Cognitive VR*”. Then, the definitions and models of some of the most studied CAS in VR are depicted. We present the different measurement methods to assess CAS, along with their advantages and drawbacks in VR. The main findings of research studying CAS in Immersive Virtual Environments (IVEs) in a non-clinical context are presented, as well as the different challenges and perspectives. Finally, we focus more particularly on related work on Mental Workload (MW) in VR training.

TABLE OF CONTENT

1	“Affective and Cognitive VR”	10
2	Cognitive and Affective States: Definitions and Models . . .	12
2.1	Cognitive States	12
2.2	Affective States	15
2.3	Complementary Notions	17
3	Measuring Cognitive and Affective States	19
3.1	Self-Report Measures	20
3.2	Observational Measures	23
3.3	Task Performance Measures	25
3.4	Psychophysiological Measures	26
4	Overview of studies in Affective and Cognitive VR	32
4.1	Induction of Cognitive and Affective States in VR . .	32
4.2	Recognition of Cognitive and Affective States in VR	41
4.3	Exploitation of Cognitive and Affective States in VR	43
5	Mental Workload in VR Training	45
5.1	Designing Training Scenarios in VR	47
5.2	Using Mental Workload in VR Training Scenarios . .	48
6	Conclusion	50

INTRODUCTION

The objective of this chapter is to provide a state of the art of works exploiting MW measures in VR. While MW concept seems intuitive, it remains surprisingly difficult to define [49, 312]. Past works have shown that it is strongly related to other psychological states, responses, and concepts such as emotions, anxiety, stress, task engagement, and flow [69, 73, 196, 311]. Similarly, it was often associated to other states or notions like cognitive load, mental effort, and cognitive performances. We can separate these notions into two categories: **Cognitive States (CS)**, which influence or are consequences of how information is processed (e.g., reasoning, deliberation, planning) [28] such as MW or cognitive performances, and **Affective States (AS)**, which refer to states like emotions, mood, feelings, and personality traits [258]. While there are some distinctions in their definitions and models, CS and AS have various similarities as multicomponent constructs [160, 215, 258] and in the methods to measure them [204].

There is a wealth of studies dealing with MW in training and ergonomics research [49, 312]. However, there are way fewer works studying MW in VR compared to works dealing with states such as emotions and anxiety in VR. CS and AS are studied in a very similar way in VR, which is why they will first be treated together in this chapter to reach a better understanding of how psychological states are processed in VR.

1 "AFFECTIVE AND COGNITIVE VR"

The interest in associating CAS and computers was raised early. In 1995, Picard et al. defined "*Affective Computing*" as a type of computing which "*relates to, arises from, or influences emotions*" [228]. She highlighted the fact that emotions can act as powerful motivators, influence perception, cognition, coping, and have an important role in creativity [228]. In that sense, giving computers the ability to recognize, express and have emotions could contribute to a richer quality of interactions, which is essential in VR. Similarly, taking into account users' cognitive abilities and limited cognitive resources [304] could make VR experiences more fitted to users. For these reasons, many researchers have shown their interest in associating VR and CAS by inducing, recognizing, and exploiting CAS in VR. The main ideas behind these studies were to understand human psychology and to go toward experiences which place users in the centre of the applications by customizing and adapting the content to each individual.

The adaptation of VR by taking into account users' CAS can be done following these steps (see Fig. 4): a stimulus induces a CS or an AS to a user. The CS or AS responses are measured via different metrics. These metrics can then be used to recognize the user's CS or AS using rule-based algorithms or models trained using ML algorithms. Finally, the recognized CS or AS can be exploited to modulate the VR content and parameters to conduct users toward a targeted CS or AS. Such a process is not new as sim-

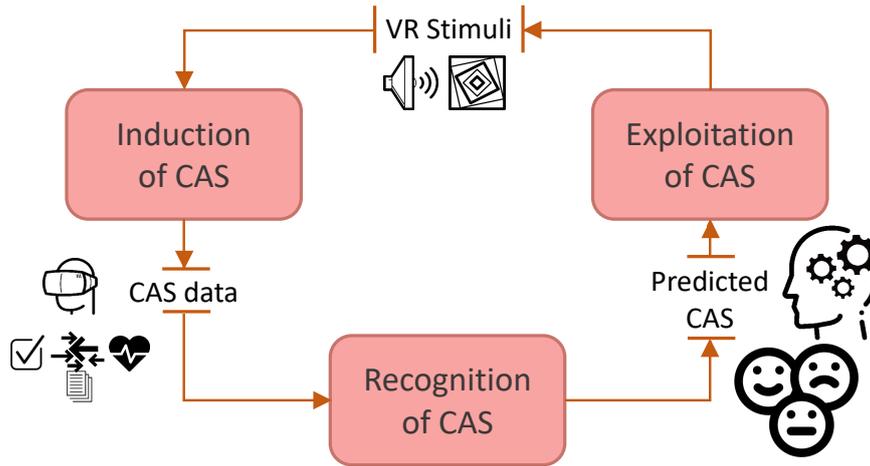


Figure 4: Loop depicting an adaptation of VR content based on users' CAS (Generalization of Figure 1). The 3 blocks represent the proposed categorization of "Affective and Cognitive VR" studies. VR stimuli first *induce* a CS or an AS to users. Then, measures are assessed from users, which are used to *recognize* users' CS or AS. Finally, the predicted CS or AS is *exploited* to adapt VR stimuli to drive users toward a targeted CS or AS.

ilar frameworks applied to emotions and games specifically [29, 164], outside VR [33, 141, 228], already exist. However, few studies went through the whole adaptation process in VR.

Following this scheme (see Fig. 4), we propose to categorize CAS VR studies following these 3 categories:

- The **induction of CAS** corresponds to the study, design, or development of content, parameters, or methods which can influence or induce changes in the users' CAS.
- The **recognition of CAS** corresponds to the study, design, or development of recognition models or methods to identify users' CAS.
- The **exploitation of CAS** corresponds to the study, design, or development of feedback or adaptation methods, logics, and parameters based on users' CAS measures.

Based on this categorization, we can define "Affective and Cognitive VR" to relate to works associating CAS and VR.

Definition:

Affective and Cognitive Virtual Reality is the study of virtual reality systems and applications that induce, recognize, or exploit affective and cognitive states.

In this definition, AS and CS are grouped because of the similarities they share in the way they are processed in VR studies. However, this can

be discussed as they can be differentiated in their concepts. In that case, works can speak of “*Cognitive VR*” when they relate to the study of CS, and “*Affective VR*” when they relate to the study of AS.

2 COGNITIVE AND AFFECTIVE STATES: DEFINITIONS AND MODELS

In this section, the focus is set on MW, cognitive load, emotions, and anxiety as they are the main CAS studied in the VR community. There is no real consensus over the definitions of psychological states, and CS and AS can overlap in their definitions and in the methods to measure them [204]. The definition of a CS or an AS is mainly chosen depending on the field, the context, and on the aim of the study [143, 312]. Nevertheless, the selection of an appropriate psychological model is crucial for the choice of the measurement method, and for all the other steps in a research dealing with CAS [204].

In this section, definitions and models of MW and cognitive load will first be addressed, followed by those of emotions and anxiety. Finally, complementary psychological notions related to Affective and Cognitive VR will be introduced.

2.1 Cognitive States

MW and cognitive load often designate the same concept in VR studies [235, 293, 317], while not necessarily referring to the same one in the psychological field [102]. They can sometimes be considered as even or distinct, and can especially be confounded with their components or other concepts such as mental load, mental effort, performance, or cognitive workload.

On the one hand, **Mental Workload (MW)** is a subcomponent of workload, which can be defined as the sum of the demands a task imposes on an individual. More precisely, Hart and Staveland [115] stated that workload “*emerges from the interaction between the requirements of a task, the circumstances under which it is performed, and the skills, behaviors, and perceptions of the operator*”. Among the numerous proposed definition, papers usually agree to determine MW as a multidimensional construct determined by characteristics of the task(s) (e.g., demands) and of the operator (e.g., skills) [195, 304, 312]. As such, Wickens described MW as “*the relation between the (quantitative) demand for resources imposed by a task and the ability to supply those resources by the operator*” [304]. Similarly, De Waard defined MW as “*the ratio of demand to allocated resources*” [72] in his paper on driver’s MW. Another way to consider MW is to describe it according to its different components. For example, Reid and Nygren described MW as a construct explained by 3 component factors: time load, mental effort load, and psychological stress load [240].

One theory is often associated with MW: Wickens [304]’s **multiple resources theory**. It proposes that human resources do not have only one

information processing, but multiple resources which can be exploited simultaneously or sequentially depending on their type. It is based on the multiple resource model (see Fig. 5), which is composed of 4 dimensions (i.e., *sensory modality*: auditory vs. visual, *types of memory code*: visual vs. spatial, *stages of information processing*: perceptual/cognitive vs. response, and *channels of visual information*: focal vs ambient). The theory also includes 3 components related to demands, resource overlap and allocation policy. It allows system designers to predict when certain tasks can be performed concurrently, when some tasks will interfere with each other, and what effect increasing the difficulty of a task will have on another task. For example, according to this theory, dual-task performance is more likely to be hampered by performing similar tasks than dissimilar tasks.

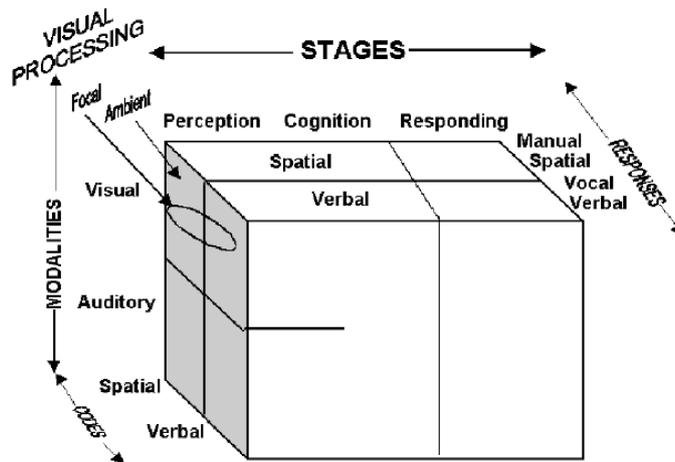


Figure 5: The 4-D multiple resource model [304].

On the other hand, Paas and Van Merriënboer described **cognitive load** as “a multidimensional construct representing the load that performing a particular task imposes on the learner’s cognitive system” [214] (note the singular “a particular task” and the employed verb “imposes”). They defined the cognitive load concept by taking into account the assessment factors AND the causal factors (see Fig. 6). The measurement dimensions of cognitive load are mental load, mental effort, and performance [214]. Mental load occurs before performing the task, to estimate the cognitive load at the current moment, mental effort can be measured while participants are working on a task, after the actual load has been allocated at a given time, and performance, while users are working on a task or thereafter, as a result of user’s cognitive capacity for the task [214] (see Fig. 6). The causal factors refer to characteristics of the task or of the environment (i.e., task complexity, time pressure, pacing of instruction, noises), of the learner (i.e., stable characteristics like cognitive capabilities, age, preferences, and prior knowledge), and of the interactions between both.

Based on this concept, the **cognitive load theory** “is concerned with the development of instructional methods that efficiently use people’s lim-

ited cognitive processing capacity to stimulate their ability to apply acquired knowledge and skills to new situations (i.e., transfer)" [215]. The theory distinguishes 3 types of cognitive load (see Fig. 7 [215]): intrinsic load, extraneous load, and germane load. The intrinsic load corresponds to the task complexity and is relative to the learner's knowledge. It can not be influenced by the instructional designer. The extraneous load is constituted of instructional features that are not beneficial for learning (e.g., unnecessary instructional design). The germane load is constituted of instructional features that are beneficial for learning (e.g., cognitive diagrams, patterns). To simplify, a good instruction should reduce extraneous cognitive load, and can increase germane cognitive load when intrinsic cognitive load is too low. The sum of the 3 cognitive loads should not exceed the learner's working memory limits.

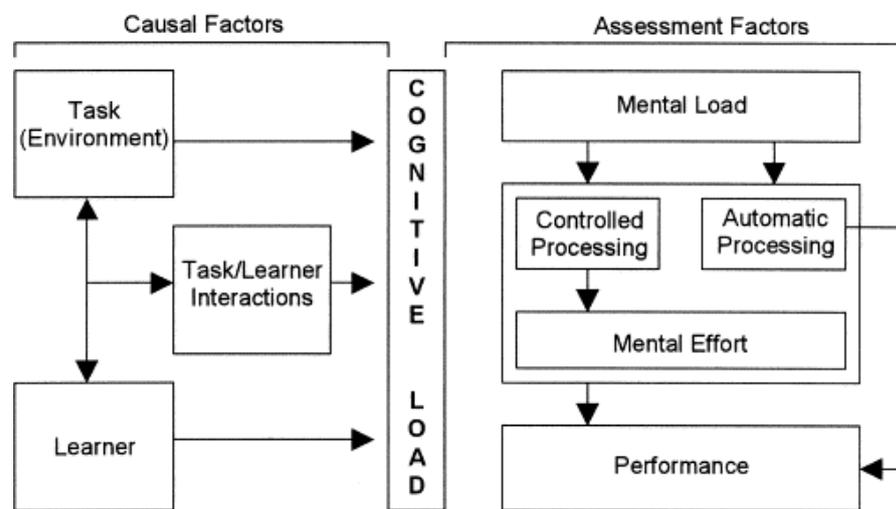


Figure 6: Schematic representation of the cognitive load construct [214].

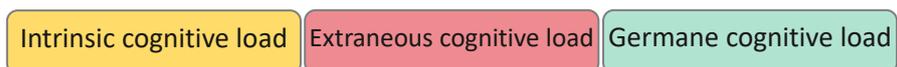


Figure 7: The 3 types of cognitive load from the cognitive load theory [215].

MW and cognitive load are very close concepts. However, their perspective approach is different. MW particularly focuses on the interaction between the nature of the task and the type of users' cognitive resources (modalities, demand, resources allocations), and cognitive load focuses more on the interaction between the instructional design (task intrinsic load and presentation) and users' cognitive process architecture. As a consequence, their measurement tools and induction methods also tend to differ. Indeed, MW is a notion which was mainly explored in aeronautics and ergonomics and target operators, while cognitive load was primarily explored in the educational field and target learners. These distinctions can

be interesting to understand to avoid confusion when transferring knowledge into VR studies.

2.2 Affective States

AS can be distinguished in terms of time (duration), intention (event focus), cause, and impacts on the behaviour and physiological responses among other criteria [22, 76, 258]. For example, Desmet explained how he distinguished emotions, mood, sentiments and, personality traits in terms of time and intention [76]. Emotions and moods are acute, meaning relatively short in duration, while sentiments and personality traits display themselves over longer periods of time. Emotions and sentiments are directed at something, while moods and personality traits are not directed at anything in particular. In VR, the most studied affects are emotions and anxiety, which is why we are going to focus more particularly on them.

A well-accepted definition of **emotion** is given by Scherer: “an episode of interrelated, synchronized changes in the states of all or most of the five organismic subsystems in response to the evaluation of an external or internal stimulus event as relevant to major concerns of the organism” [256, 257]. In his definition, the five organismic subsystems correspond to a categorization of the physiological responses. As for the emotional models, Kołakowska et al. [146] proposed to classify them into 3 categories following different perspectives:



Figure 8: Examples of representations of emotions from a discrete perspective.

- From a **discrete perspective**, each emotion is studied independently, as if all emotions were distinct (see Fig. 8a). The most popular discrete category of emotions is the Facial Action Coding System (FACS) [89] one for expressions recognition: the Big Six (Happiness, Sadness, Fear, Anger, Disgust and Surprise, see Fig. 8b). These fundamental emotions are considered as innate, basic, and determined in both psychological and biological sense from this perspective. To this basic set, some added the Neutral expression [124] or extended it with other emotions such as Amusement, Guilt, Pride and Shame, which have a more complex behavioural manifestation. Other sets of basic emotions are proposed and reviewed in [134, 211].

- Emotions, from a **dimensional perspective** are considered as continuous phenomena that can be represented in N -dimensional space, usually within 2 or 3 fundamental dimensions. Among the multiple models, Mehrabian and Russell [192] proposed a 3D environmental emotional scale of “*nearly independent bipolar dimensions*”. It is abbreviated as **PAD** for Pleasure, also referred as valence (from negative to positive), Arousal, which corresponds to the excitation, mental activation or level of alertness, and Dominance, which is the degree of feeling of having control over a specific situation. Another popular model is the Russel’s circumplex model of affect [248], which defines emotions based on 2 dimensions: the emotional Valence on the x-axis and the emotional Arousal, on the y-axis (see Fig. 9a). Besides the arousal, valence and dominance dimensions, Mendl, Burman and Paul included approach-avoidance motivations [193], which represent the tendencies to approach stimuli (reward acquisition) or to avoid them (punishment avoidance), and reflects behaviour activities of the affected person.
- Emotions can also be defined from a **hybrid perspective**, combining discrete and dimensional viewpoints. A good example is Plutchik’s model [230] which hierarchically arranges emotions (see Fig. 9b). In this model, complex emotions are combinations of pairs of more basic emotions, called dyads.

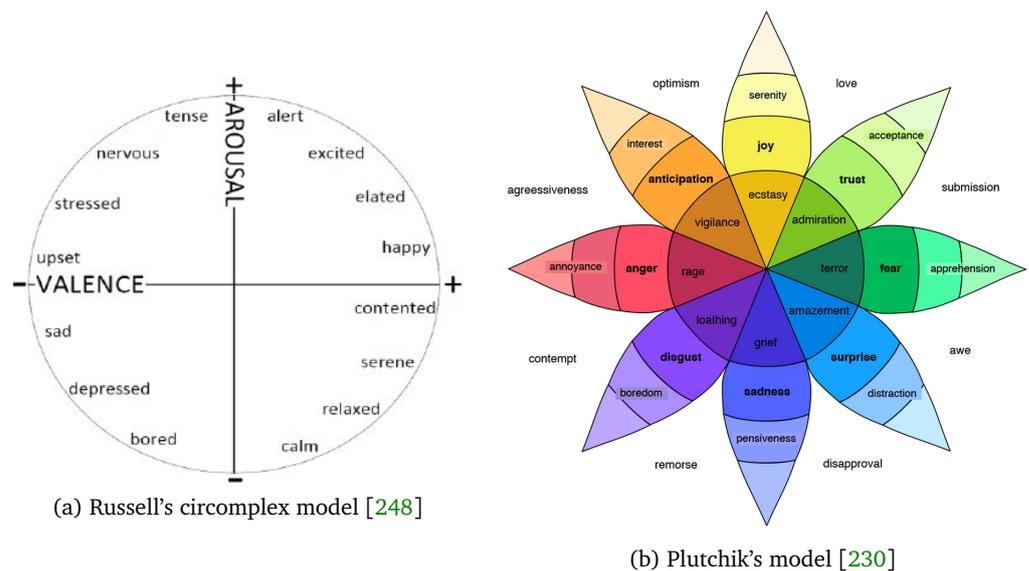


Figure 9: Examples of representations of emotions: (a) from a dimensional perspective; (b) from a hybrid perspective.

Studying emotions from a discrete perspective is at first very simple and intuitive. However, it describes emotions without any inter-relationship, which makes the recognition of non-basic affective states more difficult.

On the other hand, the dimensional perspective offers the possibility to study emotions as a function of 2 or 3 continuous and nearly independent variables. However, this model tends to be not expressive. The hybrid perspective combines both models advantages, but can be tedious to manipulate because of its large panel of emotions.

As for other AS, **anxiety** has been of particular interest in the VR community, even more in VR exposure therapy. Spielberger defined anxiety as “a complex emotional reaction or state that varies in intensity and fluctuates over time as a function of the intrapsychic or situational stresses that impinge upon an individual” [273]. He made a distinction between state and trait anxiety [273], and also between state and trait anger [274]. Upon his model, anxiety and anger can manifest themselves as personality traits, but also as emotions for shorter times. As such, state-anxiety corresponds to an emotional state that “consists of feelings of tension and apprehension and heightened autonomic nervous system activity”, and trait-anxiety corresponds to a personality trait that refers to “individual differences in anxiety-proneness” [273].

2.3 Complementary Notions

Global theories, definitions and models about the main CAS studied in IVEs were presented. VR technologies also involve several other affective and cognitive phenomena. Further notions have to be introduced to understand how some research treated in this survey undertook their studies.

2.3.1 Stress

The concept of **stress** was introduced by Selye et al. [263] in 1936. As such, it is “an organism’s response to an environmental situation or stimulus perceived negatively” (i.e., “stressor”), “which can be real or imagined” [128]. Depending on the type, level, and frequency of stress, it can have different impacts on individuals’ capacities as well as their mental and physical well-being. For example, people with chronic stress can be over-aroused, anxious and irritable.

Indeed, stress has a close relationship with MW, cognitive load, emotional arousal, and anxiety. It has similarities with them in its physiological responses. Some studies stated arousal was indicative of stress or mental illness and used arousal and valence stimuli to classify distress (which they defined as negative stress) versus calm conditions [314]. Moreover, some studies used stressors to induce high arousal or negative affect (e.g., in this state of the art [220, 262, 309]).

Dickerson and Kemeny [83] made the distinction between 3 different kinds of stress: physical stress, psychological stress and psychosocial stress. Physical stress can be induced by a lack of sleep or extreme temperature for example. Psychological stress is associated with difficult cognitive tasks, uncontrollability or negative emotions [83], and psychosocial stress

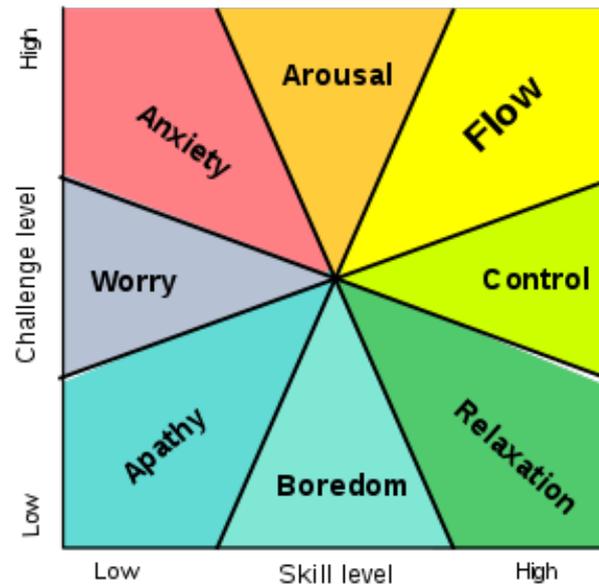


Figure 10: Csikszentmihalyi's Flow model [68]: mental state in terms of challenge level and skill level.

is triggered by a social evaluation threat (i.e., when a person's own estimated social value is likely to be degraded). These two last categories of stress are those which are the most studied in IVEs, mainly in the context of a cognitive challenge and of public speaking situations.

2.3.2 Flow

Particularly studied in the entertainment field, the **flow** is a concept developed by Csikszentmihalyi [69] and can sometimes be referred as zone [58] or stretch zone [122]. The latter associates the absorption and the engagement of the user, as well as positive and pleasurable feelings. The flow is defined by 8 components which are described in [68]. It corresponds to a balance between the "*inherent challenge of the activity*" and the "*player's ability*" required for the execution (see Fig. 10). Not respecting this balance is supposed to elicit worry, anxiety, and arousal, if the challenge surpasses the player's ability. On the contrary, if the challenge is too low compared to the player's ability, the user can feel control, relaxation, a lack of engagement, apathy, or boredom [68]. The flow concept is closely linked to happiness and the feeling of enjoyment [68]. It is also related to MW, cognitive load, and stress as seen in Lackey et al.'s study [154].

2.3.3 Presence and Cognitive and Affective States

VR technologies and IVEs were found to be efficient to induce and assess CAS, notably because they engage participants [244]. Users in VR are not just passive observers watching images on a screen, but active participants in a VE. From that aspect arises the sense of presence, an essential compon-

ent of VR defined by Slater, Usoh and Steed as the sense of “*being there*”, inside a virtual world [270]. It has also two other aspects: “*the extent to which the VE becomes the dominant one*”, and “*the extent to which participants, after the VE experience, remember it as having visited a ‘place’ rather than just having seen images generated by a computer*” [267]. Indeed, presence in VR is important to evaluate the quality of immersion experienced by users, and it can be treated as a neuropsychological phenomenon [243]. It should, therefore, be taken into account as it can interfere and play a great role when studying other CAS.

The link between presence and CAS has been widely studied in previous research. Riva et al. found a circular interaction between presence and emotions [244]. On one side, the feeling of presence is greater in “emotional” environments, and on the other side, the emotional state is influenced by the level of presence.

Indeed, most studies agree to say presence is influenced by the intensity of emotions [8, 9, 18, 74, 111, 180, 244, 282] and cognitive load [166]. Emotional content [18, 19] and stories which are emotionally powerful and richly narrated [282] could contribute more to presence than technological factors (i.e., the degree of immersion [282] and stereoscopy [18]). Furthermore, the nature of emotions could have an effect on presence [15, 209], even if some results do not support this statement [96]. Valence and arousal were found to have a positive relationship with presence [15, 209], and relaxation, a negative relationship with presence [209]. Some studies also found that presence is influenced by cognitive abilities [8, 74] and personality traits such as trait anxiety [8].

On the other hand, fewer results found an influence of presence on emotions. Studies usually agree to say this link exists but is more complex [18, 19, 95, 96, 111, 244]. For example, Gromer et al. found that presence due to the quality of the VE did not influence fear, but that presence due to individual variabilities predicted later fear responses [111]. As such, Felthofer et al. suggested that presence was a precondition for emotions to be felt, but that it did not influence the intensity of emotions [95].

In summary, presence, which is an important factor widely studied in IVEs, was found to have a circular relationship with CAS. On the one hand, the intensity of emotions and cognitive abilities influence presence. On the other hand, the effect of presence on CAS is more complex. In particular, it was suggested presence was a precondition for emotions to be felt in IVEs.

3 MEASURING COGNITIVE AND AFFECTIVE STATES

Various methods for measuring CAS have been proposed in the last decades. O’Donnell and Eggemeier classified the methods to measure MW in 3 groups [206]: subjective (or self-reports), physiological, and task performance measures. Similarly, Kaplan, Dalal and Luchman exposed 3 groups of strategies to measure emotions [132]: self-reports, observational meth-

ods, and psychophysiological measurements. We propose to merge their propositions as they have many common aspects by dividing the methods to measure CAS in 4 categories: self-reports, observational methods, task performances, and psychophysiological measures. Self-report methods are subjective. On the other hand, observational, performance, and psychophysiological methods are most of the time objective. We will expose each 4 measurement categories, present some of the most used methods, and discuss their advantages and drawbacks in an IVE context.

3.1 *Self-Report Measures*

Self-reports measures correspond to the methods in which individuals are asked to describe the CAS they experienced previously or at a given time. It usually involves the use of surveys or questionnaires, which can be paper-and-pencil, online or oral. It can also take other forms (e.g., interviews, self-confrontation).

In the following subsections, the standardized main tools used to measure CAS are going to be depicted. However, custom scales were also often used in the literature to measure CAS and their manifestations. Those usually takes the form of Likert-Scales (LSs), a symmetric agree-disagree scale in which users have to specify their level of agreement or disagreement for a series of statements (see Fig. 11a), or Visual Analogue Scales (VASs), which measure a characteristic or an attitude that is believed to range across a continuum of values and cannot easily be directly measured (see Fig. 21b).

This tool is easy to use.

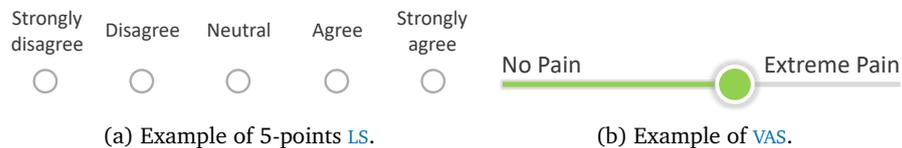


Figure 11: Some classical tools used for custom scales and questionnaires.

3.1.1 *Mental Workload Self-Report Measures*

Among CS, we are going to focus on **Mental Workload (MW)** as it is the cognitive concept which was the most developed in terms of measurement methods. Cain provided several recommendations to measure MW in his review [49]. The subjective methods can be divided into 2 categories: multidimensional and unidimensional scales.

The most popular scales are **multidimensional** ones: the NASA Task Load Index (NASA-TLX) [115], the Subjective Workload Assessment Technique (SWAT) [240], and the Workload Profile (WP) [294]. The NASA-TLX consists of 6 scales regarding workload: the mental demand, the physical demand, the temporal demand, the performance demand, the effort and

the frustration. Each of them is rated individually on a 0-100 scale. Then, they are weighted and aggregated for a global Workload Score. On the other hand, the SWAT considers the MW as composed of three components: the time load, the mental effort and the psychological stress load. It makes use of a method to take into account interval properties. Also, the WP is based on the assumption that MW can be defined by the dimensions outlined in Wickens's multiple resources theory [304]. Among **unidimensional** scales, the Rating Scale of Mental Effort (RSME) [318] evaluates the mental effort on a continuous vertical axis from 0 to 150, and the ISA evaluates MW using 5 different ratings [287] (see Fig. 12). The latter has been especially used during flight training [287]. Unidimensional scales are quicker to respond but they can have a smaller diagnosticity.



Figure 12: Example of self-reporting method in an IVE. Instantaneous Self-Assessment (ISA, measure of the user's MW using 5 different ratings [287]) in a VR Flight Simulator [172].

3.1.2 Affective States Self-Report Measures

For the measure of **emotions from a dimensional perspective**, Bradley and Lang developed the SAM [39], which depicts the three Pleasure, Arousal, Dominance (PAD) dimensions of emotions by representing each of them by five graphic characters along a nine-point scale (see Fig. 13). This method was successfully used for the assessment of players' emotions, especially the valence and the arousal dimensions. For the emotional valence dimension, the Positive and Negative Affect Schedule (PANAS) is a very popular method proposed by Watson, Clark and Tellegen [302]. This method consists of 2 scales to measure positive affect and negative affect using ten psychometric items each. A shorter form of the PANAS, the I-PANAS-SF, was also validated with international reliability concerning cross-cultural issues [290].

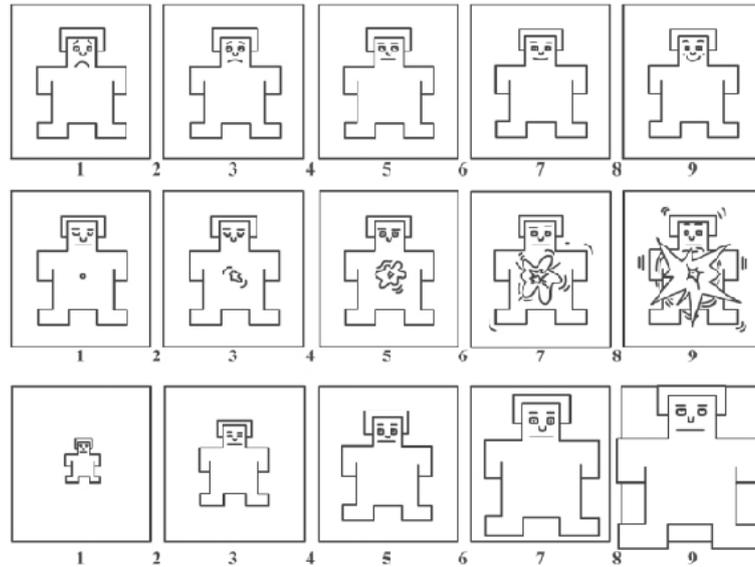


Figure 13: The Self-Assessment Manikin (SAM) [39]. The first line of characters corresponds to the valence range, the second, to the arousal, and the third, to the dominance dimension.

Self-reports is also often used to measure **affective states from a discrete perspective**. Among standardized questionnaires, the State-Trait Anxiety Inventory (STAI) is a popular form often used to measure anxiety [276]. This test is composed of two scales of 20 propositions each: the STAI-S (state), also referred as STAI form Y-A or Y1, and the STAI-T (trait), also referred as STAI form Y-B or Y2. The STAI-T is usually used at the beginning of a protocol to evaluate clinical anxiety (trait), allowing to recognize generally anxious people (who have higher scores). On the other side, the STAI-S score increases in a stressful situation. It is a good indicator of short-term anxiety, and it can be used after each condition of a protocol. Spielberger et al. also created a similar multi-scales for anger [275]: the State-Trait Anger Expression Inventory (STAXI), which evaluates trait anger and state anger.

Self-report methods are considered as cheap and convenient. It is a well-accepted measure of CAS thanks to their self-referential nature and high face validity [132]. One of the main issues concerns the ability of individuals to reveal their CAS and their interpretation of the latter. Results are restricted by human language and can be biased based on what individuals believed they felt, or on which CAS they wanted to report to be more socially desirable [169, 261]. Moreover, since real-time is an important aspect of VR, it is difficult to get a high-frequency measure of the CAS using self-reports. Depending on the way they are done, self-report measures can also break immersion in IVEs [268] and have an impact on the user's subjective impression, which adds difficulty in the care brought in the design of experimental protocols.

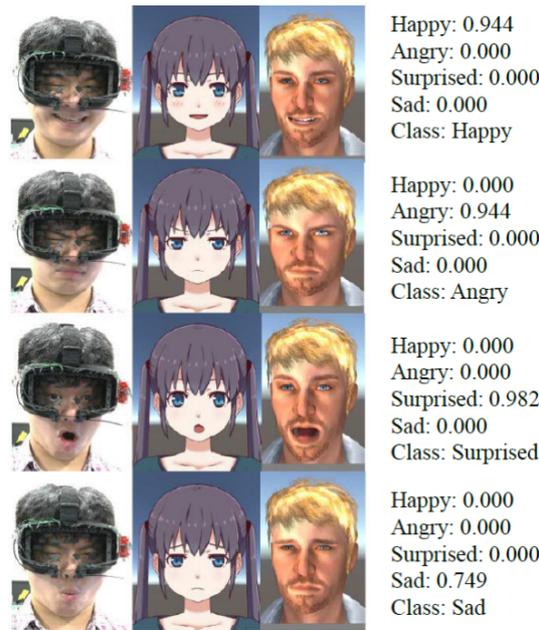


Figure 14: Recognition of emotions based on facial expression recognition using photoreflexive sensors on the VR HMD [283].

3.2 Observational Measures

Unlike self-reports, observational methods are based on an exterior point of view and evaluation of body or facial behaviours. Those behaviours are usually associated to body language and include nonverbal indicators such as facial expressions, body postures (e.g., positions of hands and legs, style of sitting, walking, standing, or lying), gestures and eyes movements (e.g., gaze direction), touch and the use of personal space [249]. One can identify patterns of behaviour, as well as their frequency of appearance and correlates them to CAS [202].

For the recognition of emotions, **facial expressions recognition** is from far the most popular observational method outside VR. Ekman, Sorenson and Friesen developed recognition methods for facial expressions of 6 primary universal emotions (i.e., the Big Six) [89]: the FACS. Based on this study, a detailed FACS [88] was elaborated as a taxonomy system presenting movements of human facial muscles by action units. It has proven to be a good standard for categorization and measurement of emotional expression but has some limitations in recognizing some expressions across different cultures due to facial deformations, and skin colour [125]. Moreover, it can be more challenging to do facial recognition in VR as parts of the face are most of the time occluded by the HMDs. Multiple research paths were explored to try answering this problem using Electromyography (EMG) [27, 185], photoreflexive sensors [133, 283] (see Fig. 14), proximity sensors [163], strain gauges to measure foam deformation on the HMD [162], and RGB-D camera on the mouth region [162, 210].

Emotion	Associated body language
Fear	Noticeably high heart beat-rate (visible on the neck). Legs and arms crossing and moving. Muscle tension: Hands or arms clenched, elbows dragged inward, bouncy movements, legs wrapped around objects. Breath held. Conservative body posture. Hyper-arousal body language.
Anger	Body spread. Hands on hips or waist. Closed hands or clenched fists. Palm-down posture. Lift the right or left hand up. Finger point with right or left hand. Finger or hand shaky. Arms crossing.
Sadness	Body dropped. Shrunk body. Bowed shoulders. Body shifted. Trunk leaning forward. The face covered with two hands. Self-touch (disbelief), body parts covered or arms around the body or shoulders. Body extended and hands over the head. Hands kept lower than their normal positions, hands closed or moving slowly. Two hands touching the head and moving slowly. One hand touching the neck. Hands closed together. Head bent.
Surprise	Abrupt backward movement. One hand or both of them moving toward the head. Moving one hand up. Both of the hands touching the head. One of the hands or both touching the face or mouth. Both of the hands over the head. One hand touching the face. Self-touch or both of the hands covering the cheeks or mouth. Head shaking. Body shift or backing.
Happiness	Arms open. Arms move. Legs open. Legs parallel. Legs may be stretched apart. Feet pointing something or someone of interest. Looking around. Eye contact relaxed and lengthened.
Disgust	Backing. Hands covering the neck. One hand on the mouth. One hand up. Hands close to the body. Body shifted. Orientation changed or moving to a side. Hands covering the head.

Table 1: The general movement protocols for the Big Six emotions recapitulated in Noroozi et al.'s survey [202].

On the other hand, VR often implies body engagement, which makes **body gesture** recognition particularly interesting to study CAS. As such, Noroozi et al. provided a survey on the emotional body gesture recognition, as well as a table with general movement protocols associated to the Big Six [202] (see Tab. 1). **Eye-related** measures has also shown a growing interest in recent years in VR studies [54, 56], with eye-tracking devices being easier and easier to equip if not directly integrated into the VR HMDs. Different observational measures can be extracted from them: gaze distribution, blink, fixation, and eyes saccades (i.e., the “*quick and simultaneous movement of both eyes between two or more phases of fixation in the same direction*” [52]) related measures are among the most popular ones [66]. Other methods such as **voice treatment** have been less explored in VR but could be interesting to research given the growing number of multi-user VR applications and its implication on oral interactions.

Interaction is one of the main components of VR [279], which makes users' behaviours particularly interesting to study in this context. In that sense, observational measures provide rich data, and can be useful for examining CAS in real-time [132]. However, there are some limitations. For instance, the behaviours of individuals greatly depend on factors such as their personalities, cultures and genders [184]. People can intentionally express, suppress or hide their psychological states during an observation. Their behaviours can be modulated by the environment (e.g., real or virtual) and the situation (e.g., experimental or entertainment). Moreover,

VR equipment can be cumbersome and sometimes make the behaviours in IVEs less natural. It is also worthwhile noting that most observational measures greatly depend on the context of the application and on the tasks users have to perform.

3.3 Task Performance Measures

Task performance measures have mainly been explored in works studying CS. O'Donnell distinguished 2 major types of performance-based measures [205]: primary task measures, which specify the adequacy of the operator to perform the principal task or system function of interest (e.g., “*the number of errors made by a pilot while flying an aircraft*”), and secondary task measures, which provide an index of primary task workload based on the operator's ability to perform an additional task (e.g., “*respond to a radio communications signal*”) concurrently with the primary task of interest (e.g., “*flying an aircraft*”) [205].

Primary task measures attempt to directly assess the user's performance on the task of interest. Typical measures of task performances include accuracy, reaction or response times, and error rates. They can be insensitive to users' MW if the variability in task demands are insufficient to result in observable changes as they do not give information on the remaining cognitive capacities of users.

On the contrary, secondary task measures provide an index of the remaining operator capacity while performing the primary tasks. There are two methodologies for secondary tasks: auxiliary task and loading task [49]. In auxiliary task methods, users are instructed to maintain consistent performances on the primary task regardless of the difficulty of the overall task. Therefore, the secondary task performances are an indirect indicator of the user's reserve capacity. In the loading task approach, the secondary task deliberately causes degradation of the primary task, which require consistent performances on the secondary task. The primary task performance measures are, in that case, more sensitive to the users' MW variations. Common secondary tasks include monitoring, tracking, memory tasks, mental arithmetics, or time estimation paradigms [207].

Task performance measures have great advantages as they are often easy to assess and can be continuous. However, similarly to observational measures, they are task and context-dependent. These measures often differ from one application to another and can hardly serve as a reference to compare different studies. While task performances were shown to have great correlations with MW, it should be noted that they can be modulated by users' engagements which can make them inaccurate for the measure of CAS [78, 294]. Many authors alerted that similar task performances should not be interpreted as an indication that the user's MW levels are equivalent [205, 294].

3.4 Psychophysiological Measures

The term psychophysiology refers to the physiological responding to psychological phenomena. In physiological computing [93], physiological activities and changes are direct reflections of processes in the Autonomic Nervous System (ANS) and in the Central Nervous System (CNS). The ANS is a system that acts unconsciously and regulates bodily functions such as heart rate and pupillary responses. The CNS comprises the brain and spinal cord. Since physiological responses are issued from psychological processes, it was hypothesized that it is possible to translate them to psychological states via the extraction of some specific features [228].

3.4.1 Measure of the ANS

The ANS changes are uncontrolled consequences of the CNS activity on physical processes. The two components of the ANS are the sympathetic division, which is associated with the “fight-or-flight response”, and the parasympathetic division, which is referred by the epithet of “rest and digest” [187]. The balance between these two systems is crucial for the homeostasis [187]. Kreibig provided a detailed review of the ANS activity when experiencing emotions [151]. Although there are still debates about the relation between ANS measures and psychological activities [184], the use of these measures to assess negative arousal is uncontested [132]. The most popular measures of the ANS activity for CAS include cardiovascular activity, electrodermal activity, and pupilometry indicators.

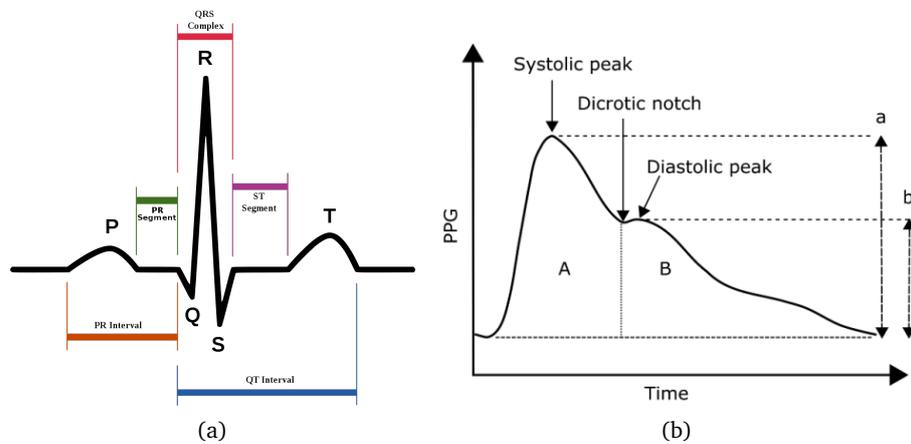


Figure 15: (a) Typical ECG waveform (image from [90]).
 (b) Typical PPG waveform (image from [222]).

Cardiovascular activity can be measured by means of Electrocardiography (ECG) and Photoplethysmography (PPG). On the one hand, ECG detects electrical changes that are the consequence of ventricles contracting and expelling blood. It has 3 main components: the P wave, which represents the depolarization of the atria, the QRS complex, which represents the depolarization of the ventricles, and the T wave, which represents the re-

polarization of the ventricles [190] (see Fig. 15a). ECG signal analysis are mainly based on RR interval, which is the time between successive R-peaks. For example, it can be used to extract the subject's Heart Rate (HR) and Heart Rate Variability (HRV), a recurrent cardiac measure that corresponds to the adaptation changes in the time intervals between two consecutive heartbeats [188]. On the other hand, PPG uses a small optical sensor in conjunction with a light source to measure changes in the skin light absorption as blood perfuses through the skin after each heartbeat. PPG signals depict pulse waves that exhibit systolic and diastolic peaks [189] (see Fig. 15b). Similarly to ECG, its signal analysis mainly relies on Interbeat Interval (IBI), also called beat-beat intervals. The main PPG features are Blood Volume Pulse (BVP), which represents the percentage of blood vessel pressure, HR, and HRV. It should be noted that if a person is strongly emotionally aroused, it will take more time for the BVP to return to its normal level [224]. PPG sensors are usually less obtrusive, but also less accurate than ECG methods.

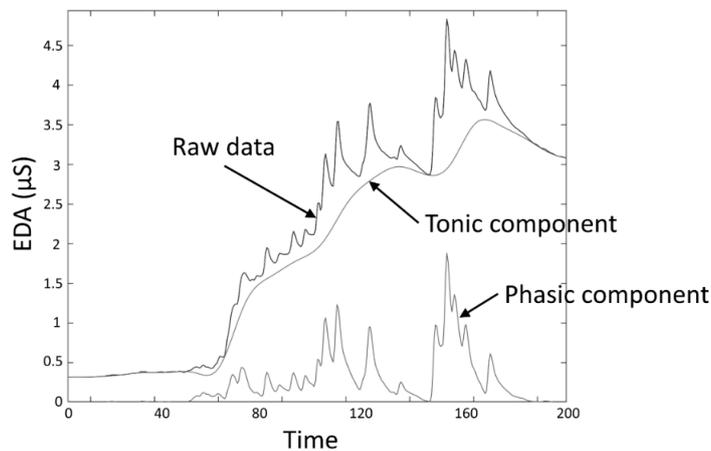


Figure 16: EDA signal decomposition into tonic and phasic components (image from [232]).

Electrodermal Activity (EDA), often referred to as Galvanic Skin Response (GSR), is the term used to define autonomic changes in the electrical properties of the skin [264]. The activity of the sympathetic nervous system is directly linked to the activity of the sweat glands, which is in turn related to the activity of the skin epidermis [36]. The EDA can be represented by its 2 main components: the tonic skin conductance and the phasic skin conductance (see Fig. 16). The tonic skin conductance corresponds to the normal conductance of an individual in the absence of any stimulus of discrete changes in the experimental environment and is related to the Skin Conductance Level (SCL). The phasic skin conductance happens in accordance to an affective event [168] and is related to the Skin Conductance Responses (SCRs). The EDA has proven strong content validity and is widely used to measure users' affective and cognitive changes [93].

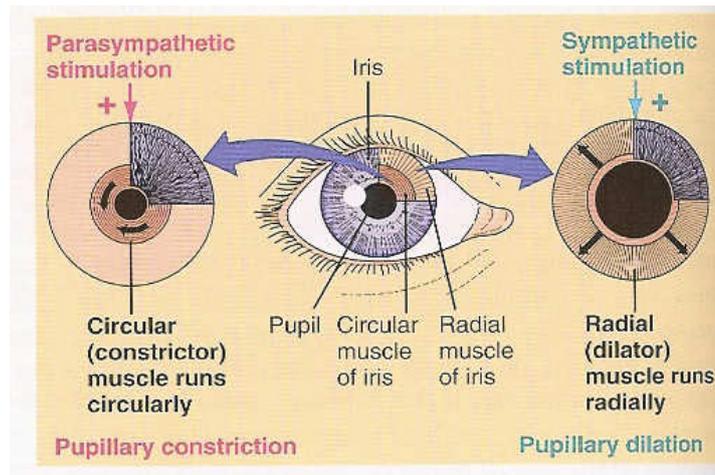


Figure 17: Constriction and dilation of the pupil (image from [234]).

Pupillometry measures can be assessed using eye-tracking devices. Changes in pupil size are caused by 2 antagonistic muscles: the dilator pupillae, which is located in the outer parts of the iris and dilates the pupil, and the sphincter pupillae, located in the central parts and constricting it [87] (see Fig. 17). The constricting sphincter muscle receives input from brain systems involved in the pupillary light reflex, and both muscles receive inputs from brain systems involved in cognitive and autonomic functions [87]. More precisely, the pupil dilatation is provoked by activity in the sympathetic pathway (fight-or-flight), and the constriction by activity in the parasympathetic division (rest-and-digest) [87]. Pupil dilation data was shown to be influenced by cognitive efforts [66, 87]. As a result, most popular pupil-related physiological features derive from pupil diameter, pupil dilatation, and pupil constriction (e.g., mean, max, min, standard-deviation).

Aside from these physiological measures, **Respiration (Resp)**, **Skin Temperature (Temp)**, **EMG**, and **Electrooculography (EOG)** measures can also be affected by CAS [93]. However, they will not be treated in-depth in this dissertation.

Good practices suggest placing ANS sensors at specific places [315]. In VR studies, ECG electrodes usually are placed on the user's torso (e.g., via a belt or using a set of electrodes), PPG sensors on the earlobes or on the wrist, and EDA sensors, on the fingers, wrist, or directly on the face via the headset. Eye-trackers systems are usually directly embedded inside the HMDs.

3.4.2 Measure of the CNS

Unlike the ANS methods, Electroencephalography (EEG) and other neuro-imaging methods measure directly the activity caused by the functioning of the CNS. Methods to measure the CNS are mostly known under the term Brain-Computer Interfaces (BCI). Among the different components of the

CNS, the cortex is the largest part of the human brain. It can be divided into the temporal, parietal, occipital, and frontal lobes (see Fig. 18) [5]. The temporal lobe is responsible for the senses of smell, sounds, and the processing of complex stimuli such as faces and scenes. The parietal lobe is responsible for integrating sensory information from various senses, as well as the manipulation of objects. The occipital lobe is responsible for the sense of sight. Finally, the frontal lobe is responsible for the conscious thought, which makes it the most impacted by CAS [5].

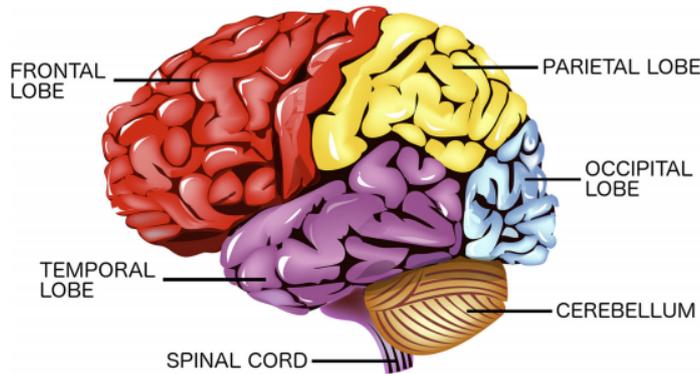


Figure 18: The cortex subdivided into the frontal, temporal, parietal, and occipital lobes (image from [5]).

Electroencephalography (EEG) measures the electrical activity generated by the brain cortex using electrodes placed on the scalp [201]. EEG signals can be measured using between 1 and 256 electrodes. Those are generally attached using an elastic cap, and the contact between the electrodes and the skins can be enhanced using a conductive gel or paste [201]. Their locations often follow standardized schemes, the most popular one being the international 10/20 system [127]. Because of weak amplitudes, EEG signals usually have to be amplified. Those are composed of different oscillations, namely “*rhythms*” or “*waves*”, that have distinct properties in terms of spatial and spectral localization. There are 6 classical rhythms [201]: δ (1-4 Hz), θ (4-7Hz), α (8-12Hz) (“idling” rhythm, default mode brain activity), μ (8-13Hz) (“sensorimotor” rhythm), β (13-30 Hz), and γ (30-100 Hz) rhythms. θ , α , and β waves have been found to be impacted by cognitive performance and mental effort (i.e., increasing working memory elicits increased activity in θ activity at frontal sites, suppression of α activity, and increased β) [94, 144, 239]. As for AS, Alarcao and Fonseca proposed a survey on emotions recognition using EEG signal [5]. Overall, the literature has corroborated that positive affects were shown to be associated with the frontal left hemisphere of the brain, and negative affects with the frontal right hemisphere of the brain [5, 129]. Compared to other neuroimaging methods, EEG is known to have a high temporal resolution but a low spatial resolution [71].

Another popular and affordable neuroimaging methods is the **functional Near-Infrared Spectroscopy (fNIRS)**, a multi-wavelength optical spectro-

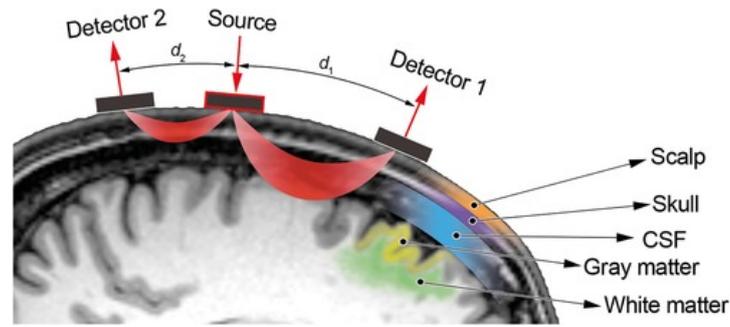


Figure 19: Illustration of the path (in red) followed by the near-infrared photons from the light source to the detector through the different layers of the head using fNIRS systems. The penetration depth of the light is proportional to the source-detector distance (d_1 : deeper channel, d_2 : superficial channel) (image from [229]).

scopy technique which measures hemodynamic brain responses in the cortex resulting from intentional sensory, motor, or cognitive activities [14, 97]. Similarly to PPGs, NIRS systems use skin tissue's particular absorption properties in the near-infrared range to measure and localize changes in oxygenated (i.e., HbO₂) and deoxygenated (i.e., HbR) hemoglobin concentration following neural activation [97]. Most fNIRS measure light attenuation due to tissue scattering at specific wavelengths based on the ratio of light intensity between the injected light (source/emitter) and output light (detector) (see Fig. 19). The difference in light attenuation between subsequent measures is used to derive the changes in concentration of HbO₂ and HbR. Each source-detector separation represents a measuring channel providing topographical representation of the distribution of the changes in concentration of HbO₂ and HbR over the cortical surface (see Fig. 19). Different commercial fNIRS systems are available and reviewed in [260]. There are many ways to analyse fNIRS data [229]. Most of these methods rely on *a-priori* knowledge of the timeline of the stimuli presentation [229], which makes it particularly challenging to use fNIRS in real-time outside laboratory conditions. Similarly to EEGs, fNIRS measures were shown to be impacted by individuals' CAS [229]. Overall, increasing task difficulty was found to increase HbO₂ and decrease HbR [229, 260]. AS results corroborate EEG findings on the fact positive and negative emotional valence respectively elicit of the left and right frontal cortex [16]. Compared to EEGs, fNIRS offer a less accurate temporal assessment of the brain activity (because of the intrinsically slow processes of hemodynamic changes) but a better spatial resolution. They are also less sensitive to movements [229], but can take time to prepare and install because of the emitters/detectors positioning.

Other technologies such as **magnetoencephalography (MEG)**, **functional Magnetic Resonance Imaging (fMRI)** and **Positron Emission Tomography (PET)** offer a more precise assessment of brain activity location and a good temporal resolution. However, they are more expensive and

can hardly be used with VR devices outside a clinical context as they are not mobile.

There is a rich literature demonstrating the efficiency of CNS measures to assess and classify CAS [5, 10, 13, 152, 165, 259], especially in the clinical field. However, the use of BCI can be particularly complex outside highly controlled laboratory environments. They can be obtrusive, sensitive to motions [101], have knowledge constraints for data-analysis [229], and thus be complex to use with a VR headset or in real-time.

Overall, psychophysiological measures have been widely studied in and outside VR. The measures of the ANS and CNS are both interesting since they provide high frequency and objective measures of the physiological activity influenced by the users' psychological state. One downside is that the sensors can be cumbersome, thus reduce immersion in VEs. For that reason, many laboratories and companies focused on introducing physiological sensors directly on VR headsets in recent years. Indeed, eye-trackers are more and more systematically embedded into the HMDs (e.g., FOVE, HTC Vive Pro Eye, Varjo VR-1, HP reverb G2 Omnicept) and add-on modules can be employed in other headsets to assess eye-related measures (e.g., Pupil Labs module). PhysioHMD introduces EEG, EMG, EOG, and EDA directly on the headset [27] (see Fig. 20), EmotionalBeasts study introduced EDA and PPG on the HMD [26], and LooxidVR [170] and Neurable [126], EEG on the HMD. Still, one matter of concerns is that physiological sensors are often sensitive to movements, which makes interactions and locomotion in VR sources of artefacts in the signals. Additionally, they can be greatly influenced by within and between-individual variabilities [288] and by cybersickness [75, 138]. Thus, physiological changes can be hard to interpret. Most studies use a baseline to take into account individual variability, to observe changes, and to normalize the obtained signals [204]. Usually, users are asked to relax or to be in a neutral state during this phase.



Figure 20: PhysioHMD [27]. Example of a module which integrates physiological sensors directly on the headset.

4 OVERVIEW OF STUDIES IN AFFECTIVE AND COGNITIVE VR

This section provides an overview of research in Affective and Cognitive VR in IVEs in a non-clinical context. This state of the art only considers HMDs and CAVE-like systems as IVEs [130]. It deals with MW, cognitive load, cognitive performances (i.e., CS); emotions, moods, personality traits (i.e., AS); and physiological responses to CAS stimuli. We focused on research which undertook experiments on healthy users. As such, we will not address VR studies which conducted experiments on pathological phobia. For reviews and meta-analysis interested readers can refer to [34, 221].

Combining CAS and VR has many advantages in several application fields. Among the 61 studies retained in this state of the art, most focused on **global VR** applications to understand how users felt in a VR context and how to enhance VR application using users' mental state (i.e., 36 studies) [8, 15, 17–20, 25, 26, 37, 47, 54, 56, 57, 78–80, 84, 95, 96, 105, 111, 147, 161, 166, 178–181, 191, 209, 212, 217, 225, 244, 269, 277]. 10 focused on **Artificial Intelligence (AI)** applications by training CAS recognition models in a VR context [3, 27, 31, 182, 183, 186, 198, 235, 245, 293]. 5 were done in a **gaming** context [2, 123, 164, 213, 216], 3 in a **military** context [219, 220, 309], 2 in a **sport** context [11, 280], 2 were applied in an **educational** context [65, 74], 1 was done to improve immersive **training** [61], 1 targeted an **aerospace** application [9], and 1 focused on the benefits of immersion for **journalism** applications [282].

We decided to treat the 63 studies following the 3 steps of adaptation of VR based on users' CAS (see Tab. 2 and Fig. 4). The first group of studies explored methods to **induce CAS**, as well as the influence of specific IVEs parameters on users' CAS (see Tab. 4.2). The second group deals with studies which developed models to **recognize CAS** using physiological signals and ML algorithms in IVEs (see Tab. 4). Finally, the last identified group target **the exploitation of CAS** to modulate IVEs content and parameters based on CAS measures to drive users toward a targeted CS or AS (see Tab. 5.1). Some studies were therefore classified in several of these categories.

4.1 *Induction of Cognitive and Affective States in VR*

One of the main advantages of VR in the study of CAS is its ability to create complex and controllable VEs. When designing an experimental protocol to study CAS, one of the first hypotheses is often that the stimuli will have an influence on a CS or an AS. The aim of this section is to depict the results of studies which found an influence of IVEs variables on users' CS or AS. Only user studies which measured users' CAS in IVEs in a non-clinical context were considered here. A summary regrouping the induction methods along with the measurement methods is given in Tab. 4.2). The studies can be divided in 2 sections depending on their purposes in regards to the induction of CS or AS: studies which aimed to induce a variety of different

	Induction	Recognition	Exploitation
CS	[17, 25, 47, 65, 78, 84, 181, 217, 235, 277, 293]	[65, 235, 293]	[65, 78, 105]
CAS	[61, 74, 178, 219, 309]	[309]	[220]
AS	[3, 8, 9, 11, 15, 18–20, 31, 37, 54, 56, 57, 79, 80, 95, 96, 111, 123, 147, 161, 164, 166, 179, 180, 182, 183, 186, 191, 209, 212, 213, 216, 225, 244, 245, 269, 280, 282]	[3, 182, 183, 186]	[2, 26, 27, 57, 79, 80, 123, 164, 198]

Table 2: Summary of the 63 studies treated in the state of the art in the 3 proposed categories. In the “Induction of CAS” category, the works focused on the influence of variables on CAS. In “Recognition of CAS”, studies tried to passively recognize CAS. In the “Exploitation of CAS” category, works focused on the adaptation of VR content and parameters based on CAS measures.

CAS or different levels of CAS, and studies which objective was to study the influence of specific parameters on a CS or an AS.

4.1.1 Induction of Several Cognitive and Affective States

Studies which intended to induce different CAS or different levels of CAS have mainly used cognitive tasks, emotional videos, emotional VEs, or emotional scenario in IVEs.

Cognitive tasks are the most employed methods to study CS. As such, the *N*-back task is a standardized cognitive task targeting working memory which can be declined in different difficulty levels and has been used extensively in VR to induce different levels of MW. It can be auditory [173] or visual, and the stimuli can take different forms such as color balls [235, 293], letters (for verbal task) or oriented symbols (for spatial task) [47]. Steed et al. also used a letter recall task, which targets users’ working memory, to test the influence of independent variables on cognitive task performances [217, 277]. For similar objectives, Bergström, Kilteni and Slater used a counting backward task [25], which target executive functioning. In the same line, Banakou, Kishore and Slater used a post-experiment Tower Of London task [17], which especially targets planning and problem-solving skills. On the mental rotation side, Collins et al. [65] used a spatial rotation task to induce different levels of cognitive load. Spatial rotation tasks were also used as distractors in [217, 277]. Dey, Chatourn and Billinghamurst [78] and Gerry et al. [105] made use of a visual searching task, where users had to find a target shape among several coloured visual distractors, to generate different task load levels. Recall of objects questionnaires are also often used to assess spatial awareness in VEs [84, 178, 181]. Cognitive tasks can also be used to induce arousal, such as arithmetic tasks (e.g., [9]), or the Virtual Reality Stroop Task used in [219, 309], ad-

adapted from the paper and pencil Stroop Test. Most of these tasks were used as single standardized tasks to induce different levels of CS using different difficulty levels. Parsons and Courtney also explored multitasking to modulate arousal level [219].

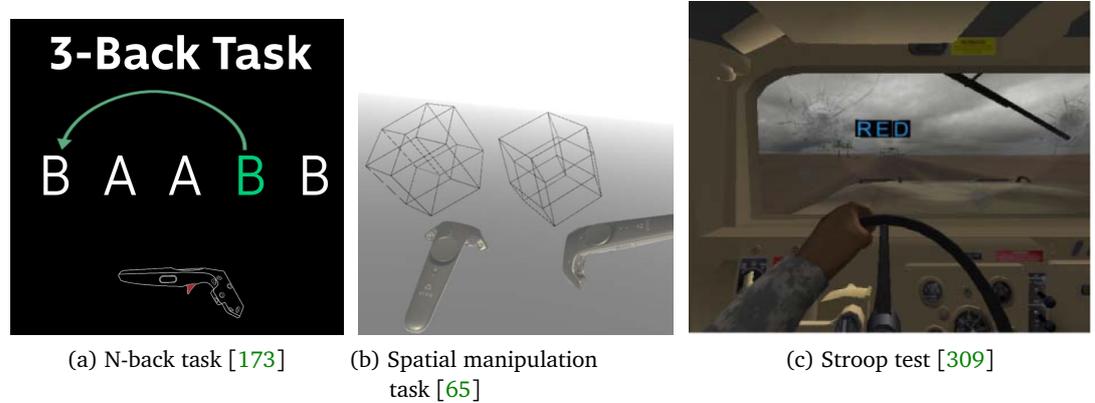


Figure 21: Examples of cognitive tasks performed in VR.

Emotional videos designate panoramic 360° videos displayed in a VR device. There is no 3D object nor interaction and users can only navigate in a limited space and/or move their head. As such, Li et al. developed a library of 73 immersive VR clips which were labelled using arousal and valence responses [161]. This was done similarly to previous works which developed databases of affective stimuli based on pictures (e.g., the International Affective Picture System (IAPS) [156]) or audio sources (e.g., the International Affective Digital Sounds (IADS) [40]) in VR. In another study, Mavridou et al. presented 20 non-panoramic emotional videos from an affective film library [251] along with 20 neutral videos in a VR cinema environment. The 20 emotional videos managed to induce balanced classes of emotions (i.e., 2 levels of arousal \times 2 levels of valence, 5 videos per class). In the same line, Marín-Morales et al. presented 4 architectural 360° panorama environments which succeeded to elicit 2 levels of arousal \times 2 levels of valence by varying colours and geometries [183] (see Fig. 22). On the other hand, Macedonio et al. presented 10 anger-provoking scenes via panoramic videos to users [179], which induced anger responses to users (see Table. 4.2). In the journalism field, Sundar, Kang and Oprean presented 2 stories in a VR HMD: a sad one about refugees, and a more positive one about dolphins [282]. Those induced the expected valence. It is also the case for Anderson et al.'s study, which presented a neutral 360° classroom class scene, and 2 natural panoramic views [9].

As for **emotional VEs**, those imply 3D objects but no other interactions than navigating in the VEs and moving the head. The Empathic Computing Lab developed 5 VE based on a jungle safari to induce happiness, anxiety, fear, disgust, and sadness [56, 57, 79]. Users are on a car which moves through the safari, and they are exposed to audio-visual stimuli. A pilot study was conducted to make sure the environments triggered the appro-



Figure 22: Induction of 2 levels of arousal and 2 levels of valence using 4 architectural 360° panorama environments in a VR HMD [183]. Left: negative valence; right: positive valence; bottom: low arousal; top: high arousal.

appropriate emotions [57, 79]. In the EMMA Project [6], a procedure was developed to induce different moods/emotions using a virtual park scene. The user explores a park, in which light and other audio-visual stimuli change depending on the targeted mood. The latter was shown to induce sadness, joy, anxiety, relaxation, and neutral states [18–20, 244]. In the same line, Felnhofer et al. developed 5 virtual parks which induced joy, sadness (no effect), boredom, anger, and anxiety by varying weather, sounds, light, and animated contents [96]. Robitaille and McGuffin also developed a virtual forest which could change from calm to stressful conditions by varying similar parameters [245]. Other studies made use of distinct VEs to elicit different emotions [54, 164, 209]. On the stressful and anxiety side, a famous study in VR consists in the virtual pit, which exposes a user to a fear of height situation [191] (see Fig. 24a). This was found to successfully induce presence and changes in physiological responses to users [111, 180, 191].

Furthermore, **emotional scenarios** are emotional VEs in which users can interact with virtual entities. Those can be horror games [80, 123, 216], games which are supposed to elicit sadness [213], joy [31, 80], neutral games [15], or serious games [61, 74]. VR games and scenarios are usually very engaging and were shown to influence emotions [61, 74, 80, 123, 213, 216], CS [61, 74], and flow [31]. Moreover, one large field of interest in VR concerns virtual public speaking situations [95, 147, 166, 225, 269]. Those were found to greatly influence anxiety, which can also be modulated by the virtual audience type [225, 269] and its reaction [147]. More particularly, some studies focused on the effect of threatening [219, 309] or emergency situations [3] on physiological signals and tasks performances. Others used everyday anxiety-provoking situations (e.g., a test day [8]), and VR sport to validate anxiety stimuli such as the crowd, and



Figure 23: Induction of anxiety in a competitive sports VR context (CAVE-like system) with animated virtual characters [11].

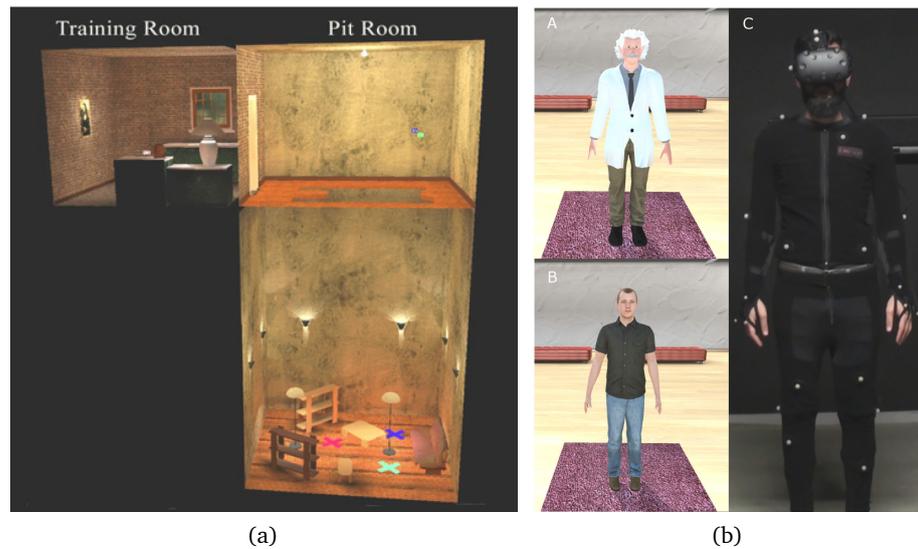


Figure 24: (a) The virtual pit experience developed by [191].
(b) Influence of avatars on cognitive performances and age-base discrimination (Einstein vs. gender-matched body) [17].

the presence and reaction of characters in a competitive situation [11, 280] (see Fig. 23).

4.1.2 Other Factors Influencing Cognitive and Affective States

In a more isolated way, factors such as the type of display, immersion levels, avatars, and user profiles were found to impact users' CAS.

The relation between the **display types**, **immersion levels** and users' CAS can be complex as some studies found an effect of IVEs on CAS [178, 182, 216], and others did not [147]. For example, Pallavicini et al. found no effect of VR vs. non-immersive console on anxiety. However, they found that VR provoked more happiness than non-IVEs [216]. These results go in the same line as other studies' findings [178, 182]. A supposition is that the novelty of VR could elicit this difference in emotional responses [182, 216]. Moreover, immersive display can induce more intense AS [61, 282], improve spatial awareness [178], and learning retention [61] compared to

non-immersive methods and display. Furthermore, simulation fidelity was found to influence sport anxiety [280] and memory of objects [181]. On the other hand, stereoscopy did not show an influence on emotions [18, 166].

Avatars were also found to influence CAS. For example, Osimo et al. proposed a self-therapy where the user is describing a problem and answering to themselves by embodying first the patient, and then the counsellor body [212]. The experiment, and the body ownership illusion were found to improve mood and emotions, especially when the user was embodied in Freud's body as the counsellor. Furthermore, embodying a virtual body associated with high cognitive abilities such as Einstein was found to result in better cognitive performances and a decrease of age-base discrimination in [17] (see Fig. 24b). Self-avatar with hand gestures allowance [277], as well as avatar types (full-body, real-body, hands-only, no body) [217] were also found to have significant effects on cognitive task performances. In the same line, being embodied in an uncomfortable posture resulted in changes in physiological signals and more mistakes in a cognitive task [25]. Bourdin et al. also discovered that a virtual out of body experience resulted in a lower fear of death [37].

Aside from avatars, **personality traits** such as trait anxiety can influence emotions [8, 95, 225]. **Other paradigms** like redirected walking were also found to impact cognitive task performances and walking behaviour [47]. On the interactions side, body participation and voice control improved emotions and enjoyment compared to more classical interactions [15, 213]. In another study, Dinh et al. found that multi-sensory feedbacks such as tactile, olfactory and auditory ones resulted in a greater memory for objects in the VEs [84]. In the same line, Chen et al. discovered HR feedbacks in emotional VEs were more enjoyable than no HR feedback, especially audio-haptic HR feedbacks [57]. They also found that altering the HR feedbacks could change participants' emotions but not their physiological signals themselves [79]. However, there was no effect of sharing HR to another player in a collaborative game on the observer's emotion and empathy [80]. In a horror game, Houzangbe et al. found that users who tried to influence their HR (biofeedback to control fear stimuli and stressors) experienced more fear than those who did not [123].

In conclusion, studies which intended to induce different CAS or different levels of CS or AS in IVEs have mainly used cognitive tasks, emotional videos, VEs, scenarios. Furthermore, the type of display and immersion level, avatars variables, user profiles, and other paradigms such as redirected walking and multi-sensory feedback were found to have an effect on CAS.

Table 3: Summary of works which induced Cognitive and Affective States (CAS) in VEs. The CAS measures which have shown significant changes are indicated in bold in "Objective Measures" and "Subjective Measures" columns. Elements which had a significant effect on them are indicated in bold in the "Inducer" column. In the "Inducer" column, the emotional virtual environments (VEs), videos, and games are not detailed when several of them are used in the study.

Ref.	CAS Measured	CAS	Subj.	Field	Inducer	Objective Measures	Subjective Measures
[65]	CS	Cognitive Load	24	Learning	Spatial rotation task (3 levels) for classification	Task perf., HR (PPG), EDA, Temp, and Accelerometer (filter)	Self-reported moments of insight, (Cognitive workload levels = cognitive task levels)
[235]	CS	Mental Workload	10	AI	N-back task (1, 2, 3)	fNIRS	(Mental workload levels = cognitive task levels)
[293]	CS	Cognitive Workload	15	AI	N-back task (0, 1, 2)	EEG	(Cognitive workload levels = cognitive task levels)
[17]	CS	Cognitive Perf.	30	VR	Virtual body (Einstein vs young male)	Task perf.	Implicit Association Test
[25]	CS	Cognitive Perf.	31	VR	Comfortable or not virtual body posture, body ownership illusion, cognitive task	Task perf., HR (ECG)	Autonomic Perception Questionnaire
[47]	CS	Cognitive Perf.	16	VR	Cognitive task (verbal vs spatial vs none), redirected walking curvature gains	Tasks perf., Behaviour	-
[217]	CS	Letter Recall Task perf.	32	VR	Avatar types (no body, hands-only representation, articulated body and real)	Task perf., Behaviour	-
[277]	CS	Letter Recall Task perf.	40	VR	Self-avatar presence, hands gestures allowed or not	Tasks perf., Behaviour	-
[181]	CS	Object Location Memory	30	VR	Rendering quality	Task perf.	-
[84]	CS	Object Location Memory	322	VR	Multi-sensory modalities (tactile, olfactory, auditory), visual details	Task perf.	-
[9]	CAS	Arousal (physiological), Valence	18	Aerospace, medicine	Emotional VEs, VE preference, cognitive task	HR (ECG), EDA	PANAS
[219]	CAS	Arousal (physiological)	50	Military	Threat situation levels, cognitive task presence	HR (ECG), EDA, Resp	(Cognitive workload levels = cognitive task levels)
[309]	CAS	Arousal (physiological), Cognitive perf.	18	Military and AI	Threat situation levels, cognitive task levels	Task perf., Behaviour, HR (ECG), EDA, EEG, Resp, Ver-tical EOG	-
[61]	CAS	Arousal (physiological), Fear	48	Transport	Aviation safety education method (card vs VR), aviation emergency scenario	Task perf. (knowledge retention), EDA, HR (PPG)	LS (7 points, 6 fear-related adjectives, 6 engagement-related statements)
[74]	CAS	Positive and Negative emotions	23	Learning	(Positive emotions, cognitive abilities, immersions) on presence, educational VEs	-	LS (6 points, positive emotions: enjoyment, hope, pride and positive deactivating emotion : relief)
[178]	CAS	Enjoyment, Fear, Cognitive Perf.	63	VR cinema	Emotional videos, display types	Tasks perf. (among which, spatial awareness)	LS (5 points - enjoyment ; fear, nervousness), Narrative Engagement Questionnaire
[209]	AS	Arousal, Valence, Dominance	146	VR	Emotional VEs, presence	-	SAM
[56]	AS	Arousal, Valence, Dominance	20	VR	Emotional VEs, HR sensory feedback	Behaviour, Pupil Dilatation	PANAS, SAM
[79]	AS	Arousal, Valence, Dominance	19	VR	Emotional VEs, HR feedback frequency	HR (ECG), EDA	PANAS, SAM
[161]	AS	Arousal, Valence	95	VR	Emotional videos	Behaviour	SAM (2D)

Continued on next page

Table 3 – Continued from previous page

Ref.	CAS	Measured CAS	Subj.	Field	Inducer	Objective Measures	Subjective Measures
[186]	AS	Arousal, Valence	11	AI	Emotional videos	HR (PPG and ECG)	Continuous Affect Self-Rating
[183]	AS	Arousal, Valence	38	AI	Emotional VEs	HR (ECG), EEG	SAM (2D)
[182]	AS	Arousal, Valence	60	AI	Emotional VEs, baseline with IAPS and emotional VEs	HR (ECG), EEG	SAM (2D)
[15]	AS	Arousal, Valence	56	VR	Presence, body participation in a pick and place game	-	SAM (2D)
[282]	AS	Arousal, Valence, enjoyment	129	Journalism	News, stories	-	Valence and emotional intensity scales, Empathy adjective scale
[212]	AS	Arousal, Valence, "global Mood"	22	VR	Personal self-therapy scenario, synchronic or not virtual body (Freud vs look-alike)	-	Profile Of Mood States, SAM (2D), SUDS
[245]	AS	Arousal	12	AI	Stressful situation	Behaviour, HR (ECG), EDA, SAM	SAM
[179]	AS	Arousal, Anger	41	VR	Emotional videos	HR (ECG and PPG), EDA, Resp, Temp	STAXI-2, PANAS, VAS (arousal), LS (10 points, Anger)
[80]	AS	Valence	26	VR	Emotional games, HR of a collaborator feedback	Eye Behaviour	PANAS, Inclusion of Other in Self scale
[20]	AS	Valence, Anxiety, Joy, Sadness	110	VR	Emotional VEs	-	LS (7 points - Sadness, joy, anxiety, relaxation), PANAS
[244]	AS	Valence, Anxiety, Anger, Disgust, Happiness, Sadness, Surprise	61	VR	Emotional VEs, presence	-	LS (10 points - Sad, Happy, Anxious, Relaxed), VAS (Happiness, Sadness, Anger, Surprise, Disgust, Anxiety, and Quietness), PANAS, STAI-S
[280]	AS	Anxiety	25	Sport	Anxiety triggers, field of regard, simulation fidelity, trait-anxiety, sport experience	Task perf., Behaviour, HR (ECG), EDA	State-Trait anxiety Inventory of Cognitive and Somatic Anxiety (state), Competitive state anxiety inventory-2 revised
[8]	AS	Anxiety	210	VR	Anxiety triggering situation, trait anxiety, personality, individual spatial abilities, computer experience	-	STAI-S, Subjective Units of Discomfort Scale (SUDS)
[11]	AS	Anxiety (Social Anxiety)	18	Sport	Sport competition scenario	Task perf., Behaviour, HR (ECG), EDA	LS (7 points - on mental state, motivation, perception of the VE during the competition)
[96]	AS	Anxiety, Boredom, Joy, Sadness	120	VR	Emotional VEs, presence	EDA	modified DES (Differential Emotions Scales)
[111]	AS	Anxiety, Discomfort	49	VR	Realism, presence, fear of height situation	HR (ECG), EDA	Acrophobia Questionnaire, STAI-S, SUDS
[166]	AS	Public Speaking Anxiety	86	VR	Stereoscopy, presence, virtual public speaking situation	HR (PPG)	MPRCS
[147]	AS	Public Speaking Anxiety/Fear (stress physiological response)	66	Therapy	Virtual/real, public speaking situation	HR (ECG), Salivary Cortisol	PRCS (Personal Report of Confidence as a Speaker), STAI-S
[225]	AS	Public Speaking Anxiety/Fear	40	VR	Virtual public speaking audience types, prior public-speaking anxiety, immersion	-	MPRCS (Modified Personal Report of Confidence as a Speaker), somatic response, self-rating of talk perf.
[269]	AS	Public Speaking Anxiety/Fear	41	VR/Therapy	Public speaking phobic participants, virtual public speaking situation	HR (ECG)	MPRCS, APQ

Continued on next page

Table 3 – Continued, from previous page

Ref.	CAS	Measured CAS	Subj.	Field	Inducer	Threatening situations (predicted or not)	Objective Measures	Subjective Measures
[180]	AS	Anxiety, Fear of Height (via HR and Discomfort)	41	VR			HR (ECG)	SUDS
[54]	AS	Anxiety, Happiness, Sadness, (Physiological signals)	20	VR	Emotional VEs, cybersickness inducer		HR (PPG), Eye-tracking data	LS (7 points - happy, sad, comfortable, anxious)
[18]	AS	Anxiety, Joy, Sadness, Valence	40	VR	Emotional VEs, stereoscopy		-	LS (10 points - Sadness, joy, anxiety, relaxation), PANAS
[216]	AS	Anxiety, Happiness, Surprise	26	Gaming	Horror game, immersion, trait anxiety		HR (ECG), EDA	VAS (Anxiety, Happiness, Surprise)
[95]	AS	Anxiety (social)	65	VR	Presence, social anxiety, virtual public speaking situation		HR (ECG)	STAI-S
[213]	AS	Enjoyment, Sadness	22	Gaming	Sad game, voice control interaction		-	LS (7 points - enjoyment, emotionally affected by the game)
[191]	AS	Fear of Height (via physiological signals)	259	VR	Fear of height situation, multiple exposures, passive haptics, frame rate, latency		HR (PPG and ECG), Temp	-
[37]	AS	Fear of Death	32	VR	Out of body experience types, visuotactile and visuomotor synchrony		-	Collett-Lester Fear of Death scale
[123]	AS	Fear	32	Gaming	Horror game, HR biofeedback		Task perf., HR	LS (5 points - fear (in engagement items))
[3]	AS	Excitement trend	20	AI	Emergency situation		EEG, Eye Behaviour	-
[31]	AS	Flow	36	AI	Shooting game		HR (ECG), Resp, EMG	Flow Short Scale
[164]	AS	(Physiological signals)	26	VR	Emotional VEs		EDA, HR (ECG)	-
[57]	AS	(Physiological signals)	7	VR	Emotional VEs, multi-sensory HR feedback		Eye-tracker data	-
[19]	AS	Emotional Engagement	60	VR	Emotional VEs, immersion		-	LS (10 points - emotional engagement and indifference)

4.2 Recognition of Cognitive and Affective States in VR

Recent studies tried to passively measure users' CAS in IVE by developing models of classification. This enables to assess users' psychological state without having to disrupt the immersive experience by asking users to report their state. However, classification models typically require the use of objective measures to be continuous. In IVE, studies mainly relied on physiological signals (see Tab. 4) as other observational and performance measures tend to be context and tasks dependent, which limits the genericity of the classification models. Due to the complexity of physiological data, Machine Learning methods are typically required in order to extract the users' CAS. These algorithms aim to infer the function between the input (e.g., physiological data) and the output data (e.g., the subjective CS or AS measure) using supervised techniques [148]. A challenge remains the training of ML models as they require the gathering of labelled data and specific induction protocols to ensure that different classes of CAS are generated. A summary of the different works which focused on developing a recognition model of users' CAS in an IVE is given in Table 4.

4.2.1 Classifiers

There are no consensus over which **ML algorithm** to use depending on the targeted CS or AS and the used sensors. While deep learning is making ground-breaking progress in several fields [131, 167, 301], supervised learning is by far the methods which was the most explored for CAS recognition [5, 316]. Support Vector Machines (**SVM**) and Linear Discriminant Analysis (**LDA**) are popular methods which have been used for years in the ML community. However, the Random Forest (**RF**) algorithm also has shown great results in recent years [65, 91]. For example, Collins et al. used Principal Component Analysis (**PCA**) and SVM with linear basis function, MultiLayer Perception, K-nearest neighbours (K=1), J48 Decision Tree, and RF (100 trees) to classify 3 cognitive load levels and 2 moments of insight [65]. They found RF to perform the best among the different classifiers. As for the model evaluation, it should be noted that most studies used a user-dependent approach (by using K-fold cross-validation, or by performing a leave-N out participant after having pooled all participants data together), meaning that these models require to be trained using the user's data to reach similar classification performances.

Model training using supervised learning methods requires to **label** the dataset for the classification of the users' CS or AS. Mavridou et al. [185] and Marín-Morales et al. [182, 183] used self-report measures to label their physiological dataset. However, the users' psychological state was sometimes inferred in other studies. For example, task difficulty levels were used to label the physiological dataset for cognitive load or MW recognition. It should be noted that this is more akin to measure the task intrinsic load rather than cognitive changes experienced by users (see Section 2.1). Tremmel et al. and Putze et al. both used the difficulty levels of the N-back task to label their dataset [235, 293]. In a similar way, Collins et al. extracted 3 difficulty levels of a spatial rotation task based on the users' overall performances. Using another approach, users' arousal level was identified based on users' performances (response time) and the Yerkes-

Dodson Law [311], which is an empirical law between the users' cognitive performances and their arousal level in Wu et al.'s work [309].

4.2.2 *Physiological input*

Concerning the sensors used in the classification of **AS**, cardiac activity sensors [182, 183, 185, 309], EDA sensors [309], RESP sensors [309] and EEG signals [182, 183, 309] have proved to contribute to the classification of arousal and valence. Wu et al. classified 3 arousal levels using EDA, RESP, EEG, ECG measures and SVM [309]. They ordered the physiological signals in terms of their importance in the classification performance. EDA and RESP features were found to be the most important, EEG features were moderately important (especially theta waves), followed by ECG features (IBI, then HR). Using EEG and ECG features, Marín-Morales et al. found that EEG features were more important than ECG features for both the classification of arousal and valence [183]. Their results are in agreement with Wu et al. [309]'s ones as they found that EEG theta waves played essentially a great role in the classification of emotions [183]. In another study ($N < 60$), Marín-Morales et al. also compared the classification of the 2 emotional dimensions in a virtual vs. in a physical museum environment and found out the classifier for the virtual environment needed less features than the classifier for the physical environment [182]. They also tested the classification of emotional valence and arousal using a "baseline" which induced a range of the emotions they targeted, using the IAPS, and the 4 emotional VEs used in the previous study [183] (see Fig. 22). By concatenating these baseline features to the tested condition features (i.e., features map), they improved the classification in both the physical (using the IAPS features) and virtual (using the 4 emotional VEs features) setups.

For **CS**, EEG [293], fNIRS [235], HR [65] and EDA [65] have been explored to classify different cognitive load levels in IVEs. In contrast to the recognition of emotions for which EEG θ waves were especially important [183, 309], Tremmel et al. found that the most consistent EEG features signals across participants to discriminate the cognitive load levels were the frontal β and γ signals [293]. Using the sensors separately, Collins et al. results showed greater cognitive load classification performances using HR than EDA features [65] (i.e., 91.71% vs. 50.83% in the best classification setup). Their classification of 2 moments of insight however reached a high accuracy using EDA features (i.e., 83.65% in the best setup).

To conclude, previous results in these studies are encouraging for the use of physiological computing in VR. However, most of them were presented using a user-dependent approach, which depicts a true challenge for the generalization of the classification of CAS using physiological signals between users. No consensus exists regarding which sensors to use for the classification of CAS, and on which classifier to use. The studies also used a variety of methods to label their dataset, from the users' CAS subjective

Ref.	CAS	Classification of:	Input	Algorithm	Subj.	Labeled Data	Model ation	Evalu-	Accuracy
[293]	CS	3 cognitive load levels (task levels)	EEG	Regularized LDA	15	3420	4-fold validation	cross-	63.90%
[235]	CS	3 cognitive load levels (task levels)	fNIRS	Shrinkage LDA	10	220	10-fold validation	cross-	42%
[65]	CS	3 cognitive load levels (task levels)	EDA	RF (best one)	24	345	10-fold validation	cross-	50.83%
		3 cognitive load levels	HR	RF	24	345	10-fold validation	cross-	91.71%
		2 moments of insight	EDA	RF	13	43	10-fold validation	cross-	83.65%
[309]	AS	3 arousal levels	EDA, RESP, EEG, ECG	SVM	18	2700	5-fold validation	cross-	96.90%
[183]	AS	2 arousal levels	EEG, ECG	SVM	38	152	leave-5 out participants	partici-	75%
		2 valence levels	EEG, ECG	SVM	38	152	leave-5 out participants	partici-	71.21%
[182]	AS	2 arousal levels	EEG, ECG	SVM	< 30	-	leave-1 out participant	partici-	75%
		2 valence levels	EEG, ECG	SVM	< 30	-	leave-1 out participant	partici-	71.08%
[186]	AS	2 arousal levels	PPG; ECG	SVM	11	10560	Bradley's test	-	-

Table 4: Summary of works studying the classification of Cognitive and Affective States (CAS) in IVEs.

measures to the task difficulty levels for CS. Few studies went further by exploiting their recognition model in real-time in IVEs.

4.3 Exploitation of Cognitive and Affective States in VR

The exploitation of CAS refers to the modulation of IVEs content based on users' CAS measure or recognition (see Fig. 1). Table 5.1 details studies exploiting CAS in real-time. Among the 13 studies which dealt with the exploitation of CAS in IVE (see Table 2), 2 studies presented a framework of adaptation of IVE based on users' CAS but did not give an example of possible adaptation [65, 164], which is why they are not presented in Table 5.1.

The methods to exploit CAS measures in real-time can be divided in 2 categories:

- **Explicit exploitation:** the users CAS measures are used to control virtual entities or parameters in the IVE, to raise awareness by the users of their own CAS.
- **Implicit exploitation:** the users CAS is measured and used independently from the users' knowledge to adapt features in the IVE.

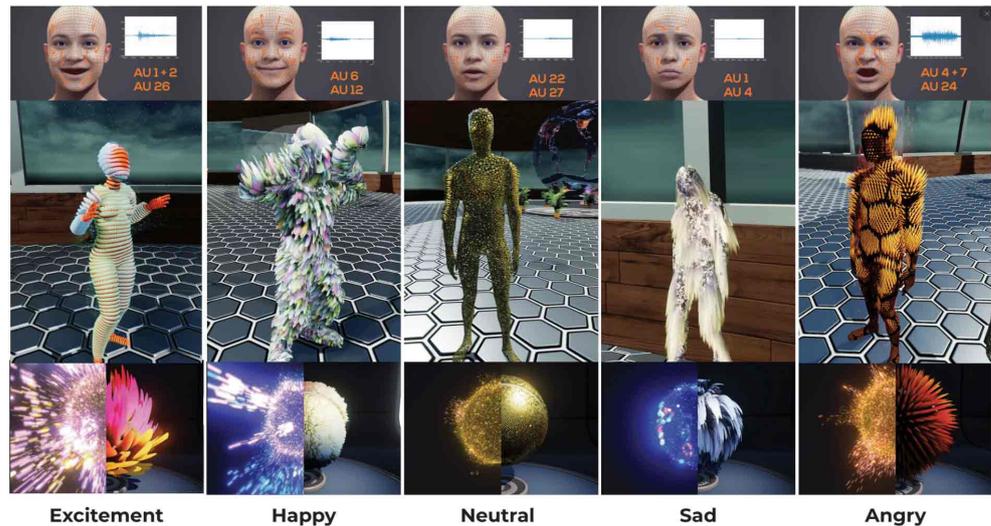


Figure 25: Example of implicit exploitation of users' emotions [27]. The particles size, density, brightness, and colour are adjusted to make avatars more emotionally expressive.

Explicit and implicit exploitations have several corresponding denominations in the literature. For example, Nacke et al. respectively used the terms "direct" and "indirect physiological control" in the game field [199].

Among explicit exploitations, a famous category of exploitation is **biofeedback**. In experiments relying on this type of feedback, users are given indications of their physiological states in the VE or via multi-sensory stimuli and they are encouraged to self-regulate their physiological states through these feedbacks. This is a type of exploitation which was particularly explored in VR therapy (e.g., [32]) and in 2 studies retained in this survey (i.e., [123, 198]). Another subcategory of explicit exploitations is **physiological feedback**. It corresponds to the direct display of users' physiological states to themselves in the IVE interface or to a third party. In this category, users are not particularly encouraged to control their physiological state and the feedback is mostly informative [57, 79, 80].

The **target of the adaptation** can take various forms, depending on the targeted CAS and the study objective. As such, task difficulty was often adapted to control cognitive load [78, 220], improve game experience [2], or to invite users to control their physiological signals in a horror game (i.e., a biofeedback case) [123]. Another classical category of feedback consists of audio-visual stimuli such as stressors (e.g., explosive device blasts [220], insects spawning [27], music box activation frequency [123], scene occlusion [27, 123]) to control users' arousal level and fear level. In Muñoz et al.'s study, they used auditory feedback to indicate the cat's location in a rescuing cat game in terms of the users' physiological signals to encourage users to control their physiological states [198]. The change of avatars appearances [26] and the direct display of the physiological signals [57, 79, 80] was also explored to communicate users' states to themselves or to a third party. Other types of feedback such as olfactory

ones or haptic ones were also used to enhance arousal stimuli [220] or as informative indications of the users' physiological state [57, 79].

As for the **events** which **trigger** changes in the IVEs features and content, CAS recognition was often set as a target method in previous frameworks [29, 33, 65, 164]. However, as seen in Table 5.1, only two protocols [2, 26] were proposed to adapt VR parameters based on the recognition of CAS using physiological signals and ML methods. Nonetheless, it was not validated through an experimental protocol in the case of Bernal and Maes [26]'s study, and the classification of emotions was not presented in the paper. Abdessalem, Boukadida and Frasson [2]'s used the excitement and frustration indexes from the Emotiv EEG device as a reference to trigger changes in the game difficulty. However, there is no published information about the algorithm used to extract these indexes. This lack of studies doing an adaptation based on CAS recognition can be explained by the difficulty to recognize users' CAS based on physiological signals, as underlined in Section. 4.2. To palliate this firsthand issue, other studies focused on determining thresholds based on physiological signals measured during the baseline [78, 123, 198], compared to a normative database [220], or in a custom way per participant [27] to trigger changes in the IVEs. On the other hand, some studies directly projected the users' physiological signals to them [57, 79] or to a collaborator in a VR game [80] in an explicit way. Bernal et al. also directly used users' EDA responses in real-time to continuously control the scene occlusion in a horror video [27].

To summarize, some studies adapted IVEs content based on CAS measures knowingly (i.e., explicit exploitation) or unknowingly to users (i.e., implicit exploitation). These adaptations take various forms depending on the targeted changes in users' CAS, from tasks difficulty, to haptic and olfactory stimuli. However, few of these studies used CAS recognition as a reference to perform the adaptation, while proposed in many frameworks [29, 33, 65, 164]. Most triggered changes in the IVEs based on physiological signals using rules-based methods (i.e., thresholds) or directly in real-time.

5 MENTAL WORKLOAD IN VR TRAINING

In this section, we are going to focus more especially on how MW is studied in VR training, in and outside the clinical field. First, the methods, frameworks, and models used to design and structure VR training scenarios are going to be depicted. Then, we are going to focus how on MW measures are exploited in training scenarios.

Table 5: Summary of works which exploited Cognitive and Affective States (CAS) in IVEs. Among the different types of exploitation, **Explicit** refers to an adaptation of the IVE knowingly to users, and **Implicit** refers to an adaptation of the IVE unknowingly to users. **BF**: Biofeedback, it is a subcategory of "Explicit exploitation" and refers to studies which incite users to control their physiological state through feedback in the IVE. **PF**: Physiological Feedback, it is another subcategory of "Explicit exploitation" and refers to studies which directly display the users' physiological signals features in the virtual environment or to a third party interface.

Ref.	ACS	Exploitation	Field	Targeted ACS	Feedback	Feedback Ref.	Feedback Triggering Event	Exp. validation
[78] [105]	CS	Implicit	VR	Cognitive Load	Visual search task difficulty (10 levels)	EEG	Threshold (based on EEG during baseline)	yes
[220]	ACS	Implicit	Military	Arousal (physiological), Cognitive Workload	Audio-visual-olfactive stressors for arousal, task difficulty for cognitive workload	ECG, EDA & EOG	Thresholds (based on a normative database)	no
[2]	AS	Implicit	Gaming	Excitement, Frustration	Obstacle game difficulty (2 levels)	EEG	EEG excitement and frustration indexes	yes
[26] [27]	AS	Implicit	AI	Anger, Stress, Neutral, Happiness (2 valence x 2 arousal)	Avatar appearance: Fur (2 happiness levels), particles brightness (2 stress levels) and colour (2 anger levels)	EDA, HR (PPG)	Emotion recognition	no
[27]	AS	Implicit	AI	Fear	360° horror video: Scene occlusion following user's gaze direction	EDA	Real-time	no
[27]	AS	Explicit	Therapy	Fear	Fearful stimuli for therapy: Insects spawn rate, speed, behaviour and size (3 arousal levels)	EDA	Thresholds (manually set for each participant)	no
[123]	AS	Explicit - BF	Gaming	Fear	Horror game difficulty (5 levels) via audio-visual fearful stimuli & field of view reduction	HR (PPG)	Thresholds (compared to HR during baseline)	yes
[198]	AS	Explicit - BF	Gaming	(Physiological Valence, Arousal, Dominance)	Rescuing a cat game: auditory indicators of cat's location	HR (PPG)	Threshold (at 95% of the average HR during baseline)	no
[79]	AS	Explicit - PF	VR	Valence, Arousal, Dominance	Altered audio-haptic HR feedback	HR (PPG)	Real-time	yes
[57]	AS	Explicit - PF	VR	Valence, Arousal, Dominance, preference	Multi-sensory HR feedback modalities	HR (PPG)	Real-time	yes
[80]	AS	Explicit - PF	VR	Valence	Visual feedback of collaborator's HR in a collaborative game	HR (PPG)	Real-time	yes

5.1 Designing Training Scenarios in VR

VR training often implies complex environments with several humanoid agents and interactable objects which each have a limited number of functions. For that reason, the approaches and models used to design VR scenarios often aim to simplify or to structure the creation of virtual scenarios. Those mostly are based on narrative or interaction requirements [104], and can be divided into two classes: predefined scenario models and emerging scenario models [64].

Predefined scenario models focus on the sequencing of the events. They orchestrate the events based on the users' actions and on the characteristics or attributes of the virtual objects. The models rely on diverse representations and are mostly based on automata. Among the different types of representation used in the models of scenarios in VEs, state machines are widely used (e.g., Story Nets [292], Hierarchical Concurrent State Machines (HCSM) [67] and Hierarchical Parallel Transition Systems++ (HPTS++) [155]). The events are defined in the states and triggered depending on different changes. For example, Story Nets will trigger changes in reactions and behaviours of virtual humans depending on the user's choices and actions [292]. On the other hand, HCSM and HPTS++ both represent the global virtual environment behaviour. They represent the different possible arrangements of events by structuring virtual object behaviours using hierarchical parallel automata. HPTS++ integrates in addition internal resources management. A change in the resources will trigger a respecification of the sub-scenarios. There are also other types of models which rely, for example, on Petri nets representation such as IVE [45] and #SEVEN [64] (see Fig. 26), graphcet-like representations such as LORA++ [104], or activity diagrams like HAVE [59]. These last models also consider collaborative interactions between users in the scenarios on top of arranging objects behaviours using parallel hierarchical automata. LORA++ and HAVE trigger sequences of predefined actions based on changes in the environments, and IVE, based on users' actions.

Emerging scenario models do not define precisely what events should occur and in which order. The scenario "emerge" during the simulation [63]. The simulations are driven based on a set of rules that constrains the behaviour of the VE and virtual agents. For example, IDTension [285] is based on a set of rules which define the actions that can be undertaken depending on the states of the virtual agents (e.g., "wish to realize an objective", "know an information", "can fulfil a task"). The agents' behaviour is driven by their personality (e.g., 'non-violent=0.8'), and the actions they are allowed to perform given the context and their objectives. Users are not constrained by sequences of actions, but they are not guided either in the VE. Similarly, EmoEmma [53] is based on a set of actions associated to each virtual agent. These actions are constrained by the state of the environment and by the states of the agents. Each action updates the state of the environment and the actions available to the agents. Another model, SELDON [50], was thought to adapt dynamically a scenario based on the user's actions. The user can interact freely and the application tries to reorient the scenario toward a specific path by launching events depending on a set of predefined constraints linked to pedagogical and narrative requirements.

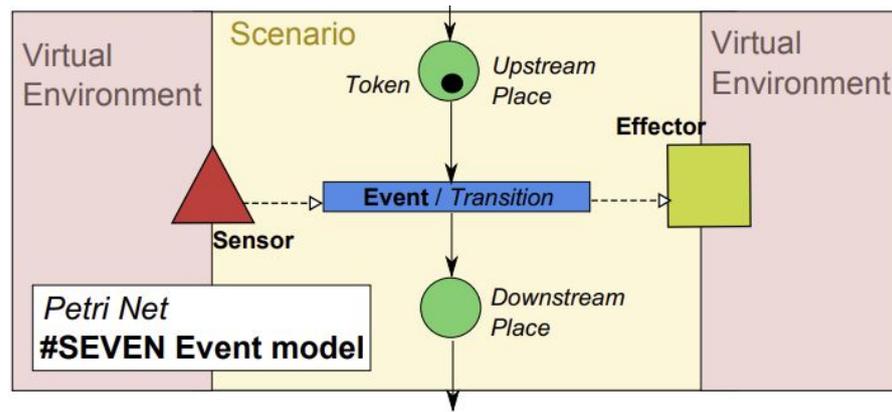


Figure 26: A Petri net enhanced with the event model, as used in #SEVEN [64].

The choice of a model depends on criteria such as the complexity of the application, the objective of the training, the flexibility of the scenario, and the collaborative aspect. In VR training, most studies focused on predefined scenario models as the objectives of the application often are defined in advance. On the one hand, predefined scenarios offer more control over the arrangement of the actions and can be represented graphically. However, they often lack of adaptability. On the other hand, emergent scenarios proposes a more natural approach as they let users interact freely, but they lack of guidance.

5.2 Using Mental Workload in VR Training Scenarios

There are mainly two usages of MW measures in VR training scenarios.

On the one hand, MW is majorly used as an **offline metric** in VR studies (see Section 4.1). Users are exposed to a VR training scenario once or multiple times, and their performances and MWs are being assessed [43, 159]. The purpose can be to show VR training can be as efficient or more efficient than more traditional methods [21, 38, 112, 200, 300] or to study the impact of other independent variables on users' MW [159]. Usually, the training scenarios are designed specifically for the study objective and can not be easily re-adapted without further coding and research.

On the other hand, MW can be implicitly **modulated by task difficulty** and assessed through task performances. For example, Parsons and Reinebold proposed a framework, which adapted the complexity of a Humvee follow task by varying the vehicle's acceleration and deceleration based on the distance between the user's vehicle and the following vehicle [220]. In the clinical field, adaptive VR training mainly targeted the rehabilitation of patients. For example, Grimm, Naros and Gharabaghi [110] adapted the level of difficulty of a reach-to-grasp task based on the performances of chronic stroke patients by adapting the distance between the object to grab in VR and the target where the patient had to release it (see Fig. 27). Kizony, Katz and Weiss developed 4 different clinical applications in VR,

each featuring one task which difficulty could be adapted based on the user's performance [142]. In the same line, Ma et al. developed adaptive VR games to encourage stroke patients with upper limb motor disorders to practice physical exercises [177]. One study adapted a visual search task difficulty depending on the EEG α band power [78] (see Fig. 25). The later was shown to be influenced by the task difficulty, and thus, the intrinsic cognitive load of the task [78]. They adapted the task difficulty based on the EEG alpha power value computed during a calibration phase before the experiment to adapt the experiment to each user.

MW was used either as an offline metric or through performance and EEG measures to adapt VR task difficulty. However, all studies which adapted task difficulty did it in a single task context. A training context generally implies complex environments with multiple tasks to perform, sometimes in parallel. In this case, the process of adaptation would be more difficult as an increase or a decrease in performances of one task might be due to the drop-out or the prioritization of another task.

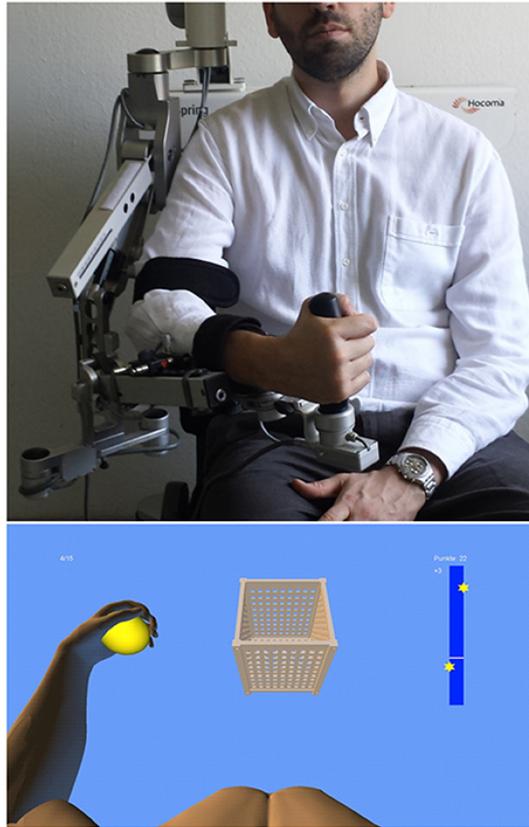


Figure 27: Example of task difficulty adaptation. The distance of the object to grab (ball) in VR and the target (basket) was adapted based on users' performances [110]. The reach-to-grasp movements were supported by an exoskeleton because of the severity of the patients' impairments to provide passive gravity compensation.

In summary, numerous models and approaches were proposed to design and control complex training scenarios. Nevertheless, as depicted, the use of cognitive states has been very little explored as a criterion to design VR training scenarios in those models. Some studies explored the adaptation of VR tasks difficulty based on task performances or EEG measures in a single-task context. While it is true performance is related to the user's MW [115], it is not a pure indicator of the user's cognitive state. For example, Tsang and Vidulich listed some conditions in which there was a dissociation between the performance measures and the subjective MW measures [295]. Users might show good results while getting bored or overloaded overtime, which can disengage them. In a context where most systems have interests in improving user experience and designing user-centred applications, considering users' perceived or experienced changes in CS is essential. Moreover, all of the presented studies which modulated task difficulty did it in a single task context. Combining the design of complex training scenarios while considering users' subjective psychological states to make VR training applications more aware of users' MWs is yet to be done.

6 CONCLUSION

In conclusion, this chapter provides theoretical background on CAS notions and measurement methods as well as an overview of related works on CAS in VR.

First, we proposed a categorization of works studying CAS in VR following: the induction, recognition, or exploitation of CAS in VR. We also proposed a definition of "**Affective and Cognitive VR**" to relate to works studying CAS in VR.

Then, psychological definitions and models of the most commonly studied CAS in VR were clarified, as well as their interaction with the notion of presence in VR.

The different methods to measure CAS were depicted along with their advantages and drawbacks in VR. To summarize, self-report assessment offers a high-face validity measure of users' subjective CAS but are invasive and non-continuous. Observational methods are particularly interesting in IVEs as interaction is an important component of VR. However, users' observational measures can be modulated by their personalities. Thus, they can be insensitive to users' CAS and unnatural in a VR context compared to in the real-world. Task performance measures offer easy-to-access continuous measures, but they are task and context-dependent and can be insensitive to CAS changes. Users' CAS can be measured through psychophysiological responses. Many of these physiological measures can be assessed at a high frequency and have shown to be correlated to users' CAS. As for the drawbacks, they necessitate sensors installation, calibration, complex signal processing methods, and can be hard to interpret.

Then, we presented an overview of Affective and Cognitive VR in IVEs in a non-clinical context following the 3 proposed categories: the induction, recognition, and exploitation of CAS in VR. Most studies focused on the induction of CAS in VR, with the objective to either present methods to induce specific or a wide range of CAS, or to study the influence of parameters and variables on users' CS or AS. Several stimuli were shown to impact users' CAS, but few standardized methods and ecological multi-tasking environments were employed to induce CAS. As for the recognition of CAS, studies used ML algorithms and physiological signals to passively identify users' CAS. Results are encouraging but the generalization of these algorithms to other users and other contexts than the ones in which they were trained is still an issue. Moreover, no work presented a real-time application and there is no consensus over the algorithms and sensors to adopt depending on the targeted CS or AS. Finally, the exploitation of CAS can be categorized in two groups, explicit and implicit exploitations, which occur respectively in the attentional foreground and background [93] (i.e., the changes are consciously registered by the users or not). Despite the proposition of many adaptation frameworks, few works presented an exploitation of IVEs based on users' CAS recognition supported by a user study, mainly because of the difficulty to recognize users' CAS in real-time.

In the last section, we focused more specifically on MW in VR training. Many frameworks and approaches were proposed to structure complex VEs for the design of VR applications. However, few of them considered MW as a criterion or parameter. Some works adapted the task difficulty based on users' task performances or EEG measures, but this was always done in a single-task context.

STUDYING THE MENTAL EFFORT IN VIRTUAL VERSUS REAL ENVIRONMENTS

SUMMARY Is there an effect of Virtual Reality (VR) Head-Mounted Displays (HMDs) on the user's mental effort? In this chapter, we compared the mental effort in VR versus in real environments. A first experiment ($N = 27$) was conducted to assess the effect of being immersed in a virtual environment (VE) using an HMD on the user's mental effort while performing a standardized cognitive task (the well-known N -back task, with three levels of difficulty, $N \in \{1, 2, 3\}$). In addition to test the effect of the environment (i.e., virtual versus real), we also explored the impact of performing a dual-task (i.e., sitting versus walking) in both environments on mental effort. The mental effort was assessed through self-report, task performance, behavioural and physiological measures. A second experiment ($N = 25$) was conducted to study a potential accommodation to VR exposure effect on users' mental efforts.

TABLE OF CONTENT

1	Experiment 1: Effect of VR HMD on mental effort	54
1.1	Apparatus	54
1.2	Participants	56
1.3	Experimental Design	57
1.4	Results	63
1.5	Discussion	65
2	Experiment 2: Accommodation to VR Effect	66
2.1	Apparatus	66
2.2	Participants	66
2.3	Experimental Design	67
2.4	Results	69
2.5	Discussion	71
3	General Discussion	72
3.1	Impact of Wearing an HMD on the User's Mental Effort	72
3.2	Lessons Learned	73
4	Conclusion	74

1 EXPERIMENT 1: EFFECT OF VR HMD ON MENTAL EFFORT

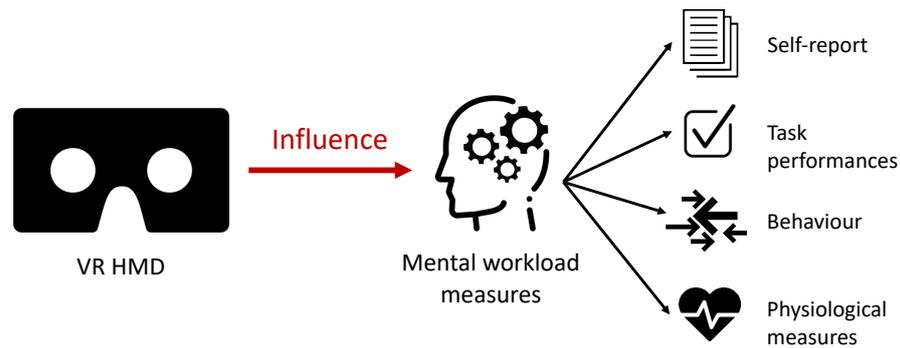


Figure 28: Illustration of the first user study. The effect of wearing a VR HMD on users' mental efforts was measured using different metrics.

This first experiment aimed to assess the effect of wearing a VR HMD on users' mental efforts using an auditory cognitive task with different levels of difficulty. In addition, the effect of a natural walking task on users' mental efforts was also explored.

1.1 Apparatus

The experiment was conducted in a room of 4.5m x 5m. The room was modelled in 3D using realistic textures, as illustrated in Figure 29. Special attention was given to the design of the virtual environment to make it correspond as much as possible to the real one. The light in the room was controlled using curtains and similar lights layout, intensity and colours. Small objects such as wires were not modelled. Otherwise, physical furniture and markers placed on the floor had their virtual counterparts in the virtual environment, which were mapped correspondingly. The orange rectangle drawn on the floor provided a visual indication to the user regarding the zone where he could walk freely without encountering any obstacle.

In the VR conditions, participants wore an HTC Vive HMD, which provides a resolution of 1200x1080 pixels per eye with a refresh rate of 90Hz. In the real conditions, all users wore the SMI eye-tracking glasses. Audio instructions were provided using a Sennheiser USB micro-headset. Some visual instructions, such as the announcement of the difficulty and a reminder of the task principle for each difficulty (see Fig. 32), were given on the TV screen. This was provided in the same way on the modelled TV in the VR conditions as on the real screen in the Real conditions.

The wires of the HTC Vive, the SMI glasses and the micro-headset were passed through a ring suspended from the ceiling. The length of the wires was adjusted each time before the walking conditions to match the user height. The cables were maintained in the user's back with a belt so they would not interfere with the user movement. A Vive Tracker was also at-

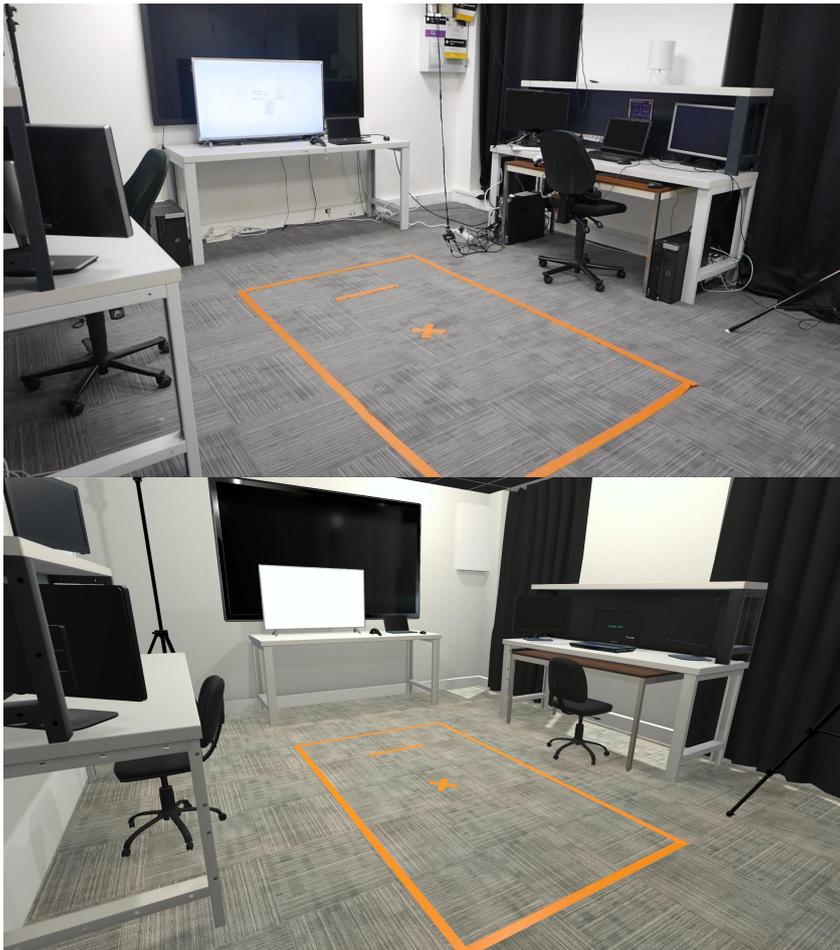


Figure 29: (top) Photo of the room in which the experiment took place. (bottom) Screenshot of the virtual environment representing the same room modelled in 3D for the VR conditions.

tached to the belt to track user's movements during the walking conditions (see Fig. 30). Only the trigger button of the Vive controller was used to interact with the support application throughout the experiment. In addition, the pressure exerted on the controller trigger button was recorded using a 1.8-cm diameter Force Sensitive Resistance (FSR) glued on the trigger button. A wireless module was placed on the Vive controller to allow the streaming of the FSR data on the computer (see Fig. 30). The total weight of the module was around 150g.

Two eye-tracking devices were used: the Pupil Labs Add-On which was integrated to the HMD (for the VR conditions), and the SMI eye-tracking glasses (for the real conditions). The participants' physiological responses were assessed using the Shimmer3 GSR+ [48]. In particular, BVP (using PPG signal) and EDA signals were recorded. The Shimmer module was placed on the forearm of the non-dominant hand (see Fig. 30). The EDA sensors of the Shimmer were placed on the middle phalanx of the non-dominant

hand two first fingers, and the BVP sensor, on the earlobe using the Shimmer ear clip. All data was recorded at 120Hz.

The support application was developed in Unity 3D, and run on a desktop computer equipped with an Intel Xeon CPU E5-1620 v4 processor, one Nvidia GeForce GTX 1080 graphic card, and 16Go Random-Access Memory.

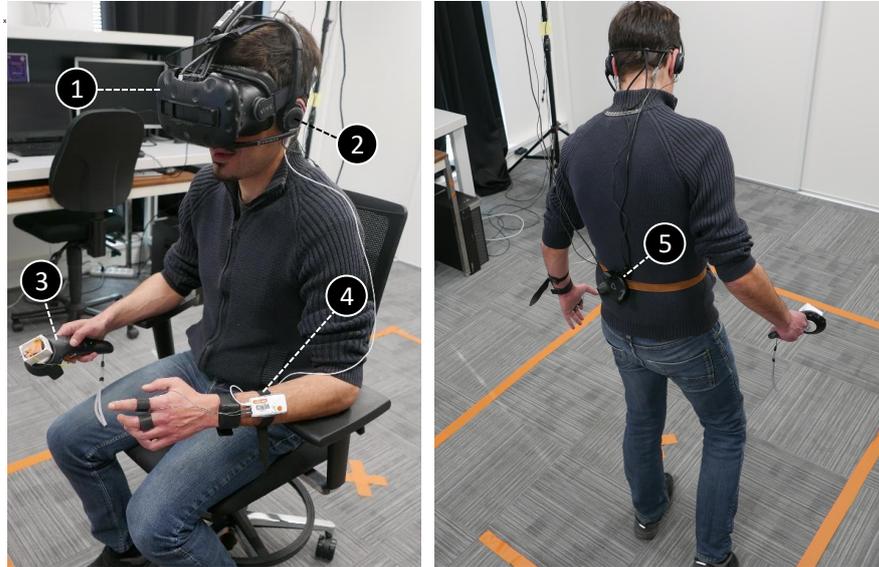


Figure 30: Photos of the experimental set-ups in the sitting condition (on the left) and in the walking condition (on the right). (1) HTC Vive HMD (VR) or SMI glasses (Real), (2) Micro-Headset, (3) Vive Controller with a Force Sensitive Resistance sensor on the trigger button, (4) Shimmer3 GSR+, (5) Vive tracker.

1.2 Participants

29 healthy participants from the b<>com research institute volunteered to take part in this study. One subject was excluded from the study due to technical issues, and another one, because of omissions in responses, resulting in a final sample of 27 participants (10 females, 17 males; ages 20-59, $M=32.8$). Two participants wore glasses, 6 participants wore contact lenses, and 1 participant reported a slight red-green weakness. One of the participants was left-handed, and all the others, right-handed. Three participants were at least regular users of HMD, and all others had no experience or a slight experience with HMD in the past. All participants were fluent in French and were naive to the experiment conditions and purpose. They all completed and signed an informed consent form before the start of the experiment. The total time per participant, including pre-questionnaires, instructions, training, experiment, breaks, post-questionnaire, and debriefing, ranged between 55 min and 1h 20 min. They wore the HMD for about 20 min and were allowed to take breaks at any time.

1.3 Experimental Design

As discussed above, exerting cognitive tasks in an immersive VE while wearing an HMD might require more cognitive resources than in the real world. The reported mental effort during natural walking in VR and in the real world might as well differ due to the cumbersomeness of the devices, the differences in perception between the real and the virtual environments, the lack of a virtual body [136] and other VR-related cognitive or perceptive phenomena.

The experiment followed a 3-factor within-subject design. The factors were the difficulty of the auditory cognitive task (3 levels), the environment (VR vs. Real), and the secondary task (Sitting vs. Walking). Each condition was repeated twice, resulting in a $3 \times 2 \times 2 \times 2$ within-subject experimental design (see Table 6). To minimize the learning effect, the order of the environment and task conditions were counterbalanced using a Latin square design, and the difficulty of the cognitive task conditions was randomized in each repetition (see Fig. 31).

Independent Variables		
<i>Participants</i>	27	(random variable)
<i>Environment</i>	2	Virtual, Real
<i>Secondary task</i>	2	Sitting, Walking
<i>Cognitive task</i>	3	1-back Task 2-back Task 3-back Task
<i>Repetition</i>	2	1, 2
Dependent Variables		
<i>Self-reports</i>	Rating Scale of Mental Effort [318]	
<i>Task performance</i>	Accuracy	
<i>Behavioural measure</i>	Response time Pressure exerted on the trigger button Walking trajectory and velocity	
<i>Physiological measure</i>	Heart Rate (HR) Skin Conductance Level (SCL)	

Table 6: Experiment 1: independent and dependent variables.

1.3.1 Independent Variables

In this experiment, three independent variables were considered: Difficulty of the Cognitive Task, Environment, and Secondary Task.

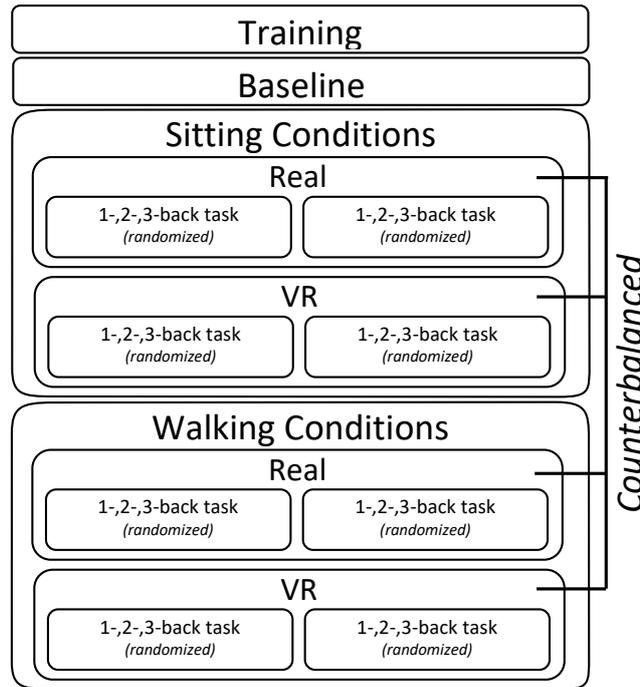


Figure 31: Experiment1: Process of the experiment for one subject (excluding pre-, post-questionnaire, instructions, set-up, and calibrations).

Cognitive Task (1- vs. 2- vs. 3-back Task)

An auditory N -back letter task was chosen to induce different levels of mental workload [12]. For each sequence of letters, the participants were instructed to press down the trigger button of the Vive controller, if (and only if) the current letter was the same as the N -th previous one and before the next letter presentation. The value of N determined the number of letters the participants had to remember, thus, the difficulty of the task [12]. Three values of N were chosen: 1, 2, and 3 as they are the most common difficulty levels in experiments relying on the N -back working memory paradigm [12]. We chose the 3-back task over the 0-back task since the latter mostly is used as a control condition [12], and may not elicit enough difference in mental workload with the 1-back task.

The sequences of letters were made of 3 distinct letters: "A", "B", and "C", which were spoken in French by a male synthetic voice¹. Each sequence of letters was $25 + N$ letters long. A letter could not appear more than $N + 2$ consecutive times as it would have resulted in a fluctuation of the difficulty. Out of the 25 letters which were relevant (the first N were irrelevant as there is no N -th previous one to compare to), 12 letters were the same as the N -th previous one. Consecutive letters were spaced in time by a constant interval of 2.0s (as in [118]), which resulted in a total duration ranged from 52s to 56s for each sequence of letters (depending on the difficulty).

¹ Text To Speech. <http://www.fromtexttospeech.com/>. Accessed: 2020-11-25

Environment (Virtual Reality vs. Real World)

The participants performed the cognitive tasks in both the Immersive VE (VR), wearing the Vive HMD, and in the real world (Real). A virtual representation of the room in which the experiment took place was given in the VR conditions (see Fig. 29). All visual stimuli were presented on the TV (real or virtual), and the sizes and positions of the virtual furniture and markers matched the real ones. In the sitting conditions, the participants were positioned in front of the TV at a distance of 1.5m.

Secondary Task (Sitting vs. Walking)

In order to see if a basic task had an impact on the reported mental effort in both environments, participants were asked to perform the *N*-back tasks while sitting and while walking.

In the sitting conditions, participants could blink, look down, fixate a point, but they were asked not to keep their eyes closed all trial long. Aside from this, they were not given any other specific instruction.

In the walking conditions, participants were asked to stay inside the orange rectangle drawn on the floor and to avoid walking on the markers. The participants were asked to start all trials (difficulty) behind the orange line, and to start walking forth in circles around the orange cross at the start of the cognitive task (when they heard the beep) until the end of the task. It was mentioned they could walk at the pace they felt natural and in the direction they wanted for the first difficulty of the condition (Walking-Real, Walking-VR). After having reported their mental effort, they were asked to alternate the direction in which they were walking between each trial to avoid ordering effects.

1.3.2 Collected Data

The dependent variables considered were: self-reports, which refer to the subjective mental effort reported by the participants, cognitive task performance, behavioural, and physiological measures.

Self-Reports: As this experiment includes many conditions, we decided to orient our choices toward a unidimensional scale: the Rating Scale of Mental Effort (RSME) [318], which is particularly sensitive over short periods of time [70, 298]. The latter evaluates the mental effort invested in a task on a continuous vertical axis from 0 to 150 units (see Fig. 32). Along the vertical axis, some landmarks are labelled with a verbal descriptor of effort, ranging from "absolutely no effort" to "extreme effort" (see Fig. 32).

Performance and Behavioural Measures: Since the primary task was the *N*-back task, performance indicators of the success of the cognitive task were recorded. We chose to focus on the accuracy value, which corresponds to the number of true responses (true positive and true negative), divided by the total number of responses (true and false values, here, 25).

For each trial, the participants should have pressed the trigger button 12 times, and not answered 13 times.

Aside from the performance linked to the cognitive task, we recorded the participants' click behaviour during the trials via the response time. The exerted pressure on the trigger button was also assessed, as it has been shown to be influenced by the arousal (which corresponds to the excitation, mental activation or level of alertness [192]) [255] and as MW and mental effort are related to the arousal [197].

Other behavioural measures were recorded during the walking conditions, such as the position and the velocity of the participants.

Physiological Measures: For all conditions, BVP and EDA data were recorded and specific physiological features were extracted. From the BVP signal, the Heart Rate (HR) was computed based on heartbeat detection. From the EDA signal, the Skin Conductance Level (SCL) was estimated using a low-pass filter (cutoff frequency = 0.05Hz, order = 3).

Since this study intends to compare VR and real conditions, we assumed that the use of two separated eye-tracking devices for each of these conditions would bias the analysis. Thus, eye-tracking data was not considered in this paper.

1.3.3 *Experimental Procedure*

The experiment had a total duration of around 1h10 and was subdivided into the following steps:

Written Consent and Instructions: The informed consent form was given in an email sent to recruit the participants. They had the possibility to read it prior to the experiment, and to bring it completed and signed at the start of the experiment. Otherwise, they were asked to read it, to complete it and to sign it. They were then instructed with the nature of the experiment, the equipment used, the data recorded (which was anonymized), the VR setup, the cognitive task instructions, and the proceedings of the experiment. Participants were also asked to fill a questionnaire (experience with VR, level of alertness, state of vision, demographic information) to gather information about their background.

Training: Users were then equipped with a Vive Controller and a microphone-headset. The RSME was displayed on the TV screen, and participants were asked to try to use the Vive controller to evaluate their level of mental effort. They were asked if they well understood the instruction and terms used on the scale. Once participants took control of the interaction, they could perform once every difficulty of the N -back task (1-, 2-, 3-back) to make sure they fully understood the instructions of the cognitive task. The training was performed in a Sitting configuration in the Real-World.

Equipment Set-Up and Baseline: At the end of the training, participants were equipped with physiological sensors, aside from the eye-tracking devices. They were then asked to remain seated and to relax while the TV screen was black for 1 minute, to record their physiological signals in a relaxed state.

Experiment: The experiment was divided in two parts: the Sitting condition and the Walking condition. Each of these two secondary task conditions (absence or presence of a secondary task) was associated with the environment conditions (Real or VR). The order of these conditions was counterbalanced using a Latin square design. Prior to the Walking group of conditions, the users were instructed with the Walking instructions (see 1.3.1). Each condition lasted around 8 minutes and was preceded by a 2 min calibration of the eye-tracking device (SMI glasses or Pupil Labs Add-on). Next, the participants performed the *N*-back task in two sets of randomized difficulties (6 trials per condition, see Fig. 31). They had the possibility to make a break at anytime and they were proposed to do it before each new condition. The participants were asked to do their best throughout the experiment.

A trial consisted of the following elements (see Fig. 32). It started with visual instructions on the TV screen, informing the participants which difficulty (1-, 2-, or 3-back) was about to start. After the user pressed the trigger button, an audio announcement of the difficulty was sent to the user, the sequence of letters was initialized, and a visual and animated reminder of the instructions for the current difficulty was displayed. Once the user pressed the trigger button again the task started with a beep followed after 1s by the stream of letters. By the end of the sequence the user was asked to answer the RSME by using the Vive controller. A pointer indicated the direction in which the Vive controller was pointing at on the TV, and the user could move a cursor on the vertical axis by dragging and dropping it. The "+" and "-" buttons could also be used to readjust the value of the reported mental effort. A click on the "Next" button marked the end of the trial, and the beginning of a new trial or the end of the current condition (Sitting-Real, Sitting-VR, Walking-Real, Walking-VR).

Debriefing: At the end of the experiment all sensors were removed from the participants. They were asked to fill a post-questionnaire regarding their subjective impressions. They were finally debriefed and invited to ask all questions they may have had.

1.3.4 Hypotheses

This study aims to evaluate the interaction between cognitive task, basic secondary task, and VR.

Our main hypothesis concerns the effect of the environment on mental effort. We hypothesize that being immersed in a VE through an HMD will impact mental effort [H1]. Moreover, since users are rarely static in VR the

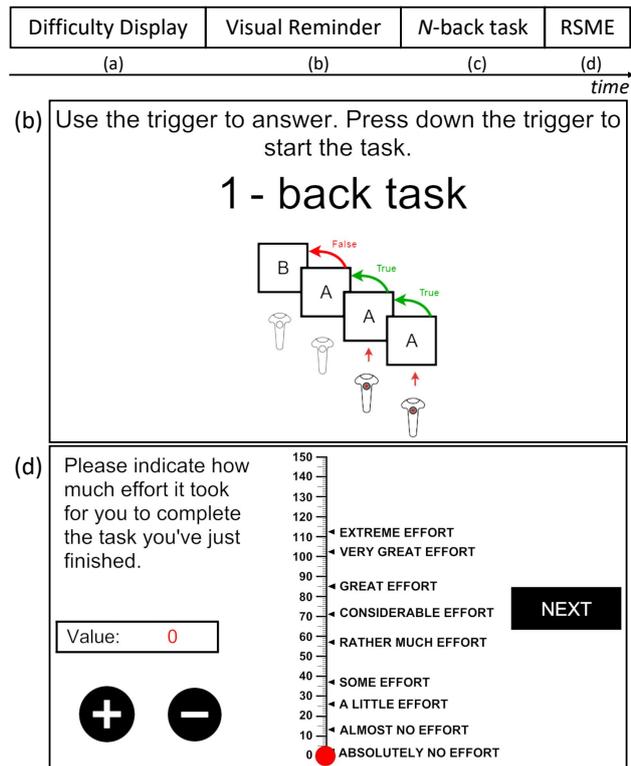


Figure 32: Progress of one trial: (a) Visual display of the task difficulty on the (real or virtual) TV screen ; (b) Auditory announcement of the difficulty, animated GIF reminding the instruction for the difficulty, and initialization of the task (here, for the 1-back task) ; (c) Start of the task with a beep, followed by the sequence of audio letters (N+25 letters) ; (d) Rating Scale of Mental Effort (RSME) [318]. Steps (b) and (c) are triggered by the trigger button, and step (d) appears at the end of the task (step (c)).

effect of basic and natural tasks such as walking may have an influence on the user's mental effort difference between the virtual and the real environments. This potential effect has to be investigated. It is supposed that doing a cognitive task while walking will significantly increase the reported mental effort compared to doing a cognitive task in a sitting condition [H2]. For both hypotheses it is supposed the results will be supported by the analysis of all the measured indicators of mental effort considered in this study. To summarize, our hypotheses are:

- **H1:** Users experience more mental effort in a VE while wearing a VR HMD than in the real world.
- **H2:** Users experience more mental effort while walking than while sitting.

1.4 Results

Generalized linear mixed model (GLMM) analysis was considered for all dependent variables (all parametric). For each variable, the user was considered as a random factor and all the independent variables as within-subject factors. When the equal variances assumption was violated, the degrees of freedom were correction using the Greenhouse-Geisser method. When needed, pairwise post-hoc tests (Bonferroni with adjustment) were performed, only significant differences ($p < 0.05$) are discussed. Order effects were also tested but there are only discussed if significant. The statistical analysis was performed using the R statistical software.

Figure 33 presents the results concerning the self-reported mental effort and the task performance, and Figure 34 presents part of the results on the behavioural and physiological data analysis.

1.4.1 Self-reports

RMSE: The GLMM showed a main effect of the Secondary Task $F_{1,26} = 20.81, p < 0.001, \eta_p^2 = 0.44$ and Difficulty $F_{1.76,45.70} = 127.50, p < 0.001, \eta_p^2 = 0.83$, but there was no effect of the Environment $F_{1,26} = 1.09, p = 0.31, \eta_p^2 = 0.04$. Post-hoc tests showed that as the level of difficulty increased, participants perceived the task as more mentally demanding. Moreover, participants perceived the Walking condition to be more mentally demanding (all $p < 0.05$). This results partially support [H2], but do not support [H1].

1.4.2 Task Performance

Accuracy: The GLMM showed a main effect of the Secondary Task $F_{1,26} = 7.14, p < 0.01, \eta_p^2 = 0.22$ and Difficulty $F_{1.80,46.84} = 208.14, p < 0.001, \eta_p^2 = 0.89$, in contrast there was no effect of the Environment $F_{1,26} = 1.26, p = 0.27, \eta_p^2 = 0.05$. The analysis did not show any significant interaction effect. Similarly as the RSME, post-hoc tests (all $p < 0.05$) showed that as the level of difficulty increased, the accuracy of participants significantly decreased. The analysis of the main effect on Secondary Task showed that the accuracy of participants was higher for the Sitting condition, yet a deeper analysis seems to suggest that this was mainly the case for the highest Difficulty ($N = 3$). This results partially support [H2], but do not support [H1].

1.4.3 Behavioural Measures

Response time: The GLMM analysis showed a main effect of the Difficulty $F_{1.75,45.41} = 94.83, p < 0.001, \eta_p^2 = 0.78$, a marginal effect of the Secondary Task $F_{1,26} = 3.93, p = 0.058, \eta_p^2 = 0.13$ and no effect of the Environment $F_{1,26} = 0.19, p = 0.66, \eta_p^2 = 0.01$. Post-hoc tests showed that the response time significantly increased as the difficulty increases (all $p < 0.05$). This results partially support [H2], but do not support [H1].

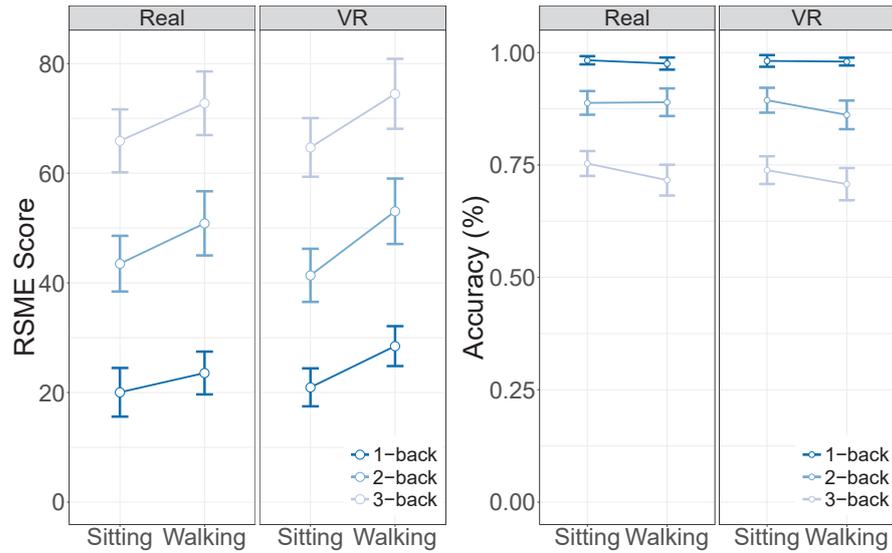


Figure 33: Mean mental effort reported with the RSME and mean accuracy considering all independent variables.

Exerted Pressure: Due to some measuring errors, only the data of 21 users could be analyzed. The GLMM analysis showed a main effect of the Secondary Task $F_{1,20} = 19.10, p < 0.001, \eta_p^2 = 0.49$ and Difficulty $F_{1.67,33.38} = 18.72, p < 0.001, \eta_p^2 = 0.48$, but no effect of the Environment $F_{1,20} = 0.48, p = 0.5, \eta_p^2 = 0.02$. Post-hoc tests showed that the exerted pressure was higher for the Walking condition, and that for the lowest Difficulty ($N = 1$) participants exerted the lowest force compared to the others ($N = 2$ and $N = 3$). This results partially support [H2], but do not support [H1].

Walking measures: No significant effect was found on the velocity and position data for any of the independent variables.

1.4.4 Physiological Measures

Heart Rate (HR): Due to some recording errors, the analysis was based on the data of 21 participants. The GLMM showed a main effect of the Secondary Task $F_{1,21} = 70.03, p < 0.0001, \eta_p^2 = 0.77$ and Difficulty $F_{1.75,36.67} = 11.67, p = 0.0002, \eta_p^2 = 0.36$. In contrast there was no effect of the Environment $F_{1,21} = 0.08, p = 0.77, \eta_p^2 = 0.004$. The analysis also revealed a significant interaction effect between Secondary Task and Difficulty $F_{1.56,32.72} = 5.40, p = 0.01, \eta_p^2 = 0.20$. Post-hoc tests (all $p < 0.05$) showed that the HR of participants significantly increased between the levels 1 and 2, but there was no significant difference between the levels 2 and 3. The analysis of the main effect of the Secondary Task showed that the HR of participants was higher for the Walking condition. The analysis of the interaction effect between Secondary Task and Difficulty showed a significant difference on HR between the Difficulty level in Sitting condition (all $p < 0.05$), but not in the Walking condition. This results partially support [H2], but do not support [H1].

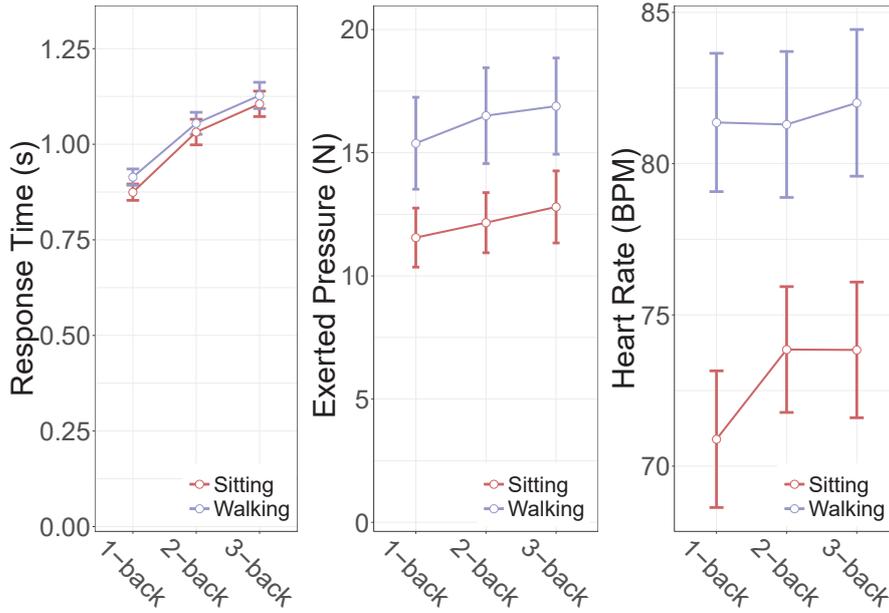


Figure 34: Mean indicators of mental effort (response time, exerted pressure on the button, HR) in terms of the secondary task and difficulty. As the environment did not influence significantly these measures, only the effect of the task is displayed.

Skin Conductance Level (SCL): The GLMM showed a main effect of the Secondary Task $F_{1,25} = 5.06$, $p = 0.03$, $\eta_p^2 = 0.17$ ($M_{Sitting} = 4.40$, $SD_{Sitting} = 3.39$; $M_{Walking} = 5.37$, $SD_{Walking} = 3.15$), and no effect of Difficulty $F_{1,98,49,51} = 0.13$, $p = 0.88$, $\eta_p^2 = 0.005$ or Environment $F_{1,25} = 0.28$, $p = 0.60$, $\eta_p^2 = 0.01$. The analysis did not show any significant interaction effect. This results partially support [H2], but do not support [H1].

1.5 Discussion

Results showed that being immersed in a VE while wearing a VR HMD did not differ from the Real World in terms of mental effort in this experiment. Other parameters such as the difficulty of the cognitive task and the presence or not of a basic secondary task were also tested to assess potential interaction effects. While the N -back task showed to be efficient to elicit different levels of mental effort (see Fig. 33), there was no significant difference in the responses reported by the participants in the real world and in the corresponding virtual environment. Moreover, we hypothesized that walking would amplify the potential difference in mental effort reported by the participants. However, in this experiment, no significant effect of the environment was found considering this variable, despite the absence of a virtual body. These results may be explained by the method used to induce mental effort. In this experiment, the auditory task was well suited as it could be done while walking. However, given the primarily visual nature of VR, it may not have been able to elicit enough difference in mental ef-

fort between VR and the Real World. As a consequence, the task difficulties and the sitting conditions did not differ much between the VR and the real conditions as users did not need to process visual information to answer the cognitive task. These conditions (tasks difficulties and sitting) are relevant as a first step to purely assess the effect of wearing a VR HMD on mental effort, without interactions linked to the nature of the interactions or factors such as realism. However, these conditions did not make full use of the immersion and interactions possibilities allowed by VR. Further work could help studying if stimuli necessitating more visual information processing could provoke different mental effort responses between the real world and the corresponding IVE.

In contrast, the results showed that the walking task had an effect on the mental effort and on the cognitive task accuracy. The effect of the walking condition was especially noticeable for the difficulty ($N = 3$), as seen in Fig. 33. The task was designed to be as natural as possible, considering the technical limitations. It was supposed not to require additional cognitive resources from the same pools than the ones affected by the mental effort elicited by the N -back task. However, it should be noted that the path constraint may have still influenced the latter. The fact that a simple walking task had an effect on the mental effort means the amount of interaction may need to be quantified in studies assessing similar cognitive states.

Further discussion about the metrics used to assess mental effort are given in the general discussion (see Section 3).

2 EXPERIMENT 2: ACCOMMODATION TO VR EFFECT

This second experiment aimed to assess the effect of time in VR on users' mental efforts during short VR exposures. More specifically, we wanted to study if what is commonly called the "awe" or "wow" effect [46, 120] or if the "adaptation/accommodation to VR" impacted users' mental efforts. In addition, this study was conducted to confront the first experiment results. We decreased the number of conditions that might have absorbed the potential effect of wearing a VR HMD on users' mental efforts. Only one difficulty of the auditory cognitive task (i.e., the 2-back task) and one secondary task condition (i.e., walking) were considered here.

2.1 Apparatus

The second experiment was conducted in the same conditions as in the first experiment (see Section 1.1), although physiological sensors (i.e., sensors from Shimmer bracelet) were not used.

2.2 Participants

25 participants (6 females, 19 males; ages 21-51, $M=31.3$) from the b<>com research institute, who did not take part in the first experiment, volun-

teered to participate in this experiment. 3 participants wore glasses, 4 participants wore contact lenses, and 2 participants reported a slight red-green weakness. 2 of the participants were left-handed, and all the others, right-handed. 2 participants were regular users of HMD, and all others had no experience or a slight experience with HMDs in the past. 2 subjects played games very frequently, 9 were regular players (once or twice a week), and all others played games occasionally. As in the first experiment, all participants were fluent in French and were naive to the experiment conditions and purpose. They all completed and signed an informed consent form before the start of the experiment.

2.3 Experimental Design

The purpose of this second experiment was to study the effect of “accommodation to VR” on the users’ mental efforts and performances.

Contrarily to the first experiment, only one difficulty (the 2-back task) and the walking condition were considered. The instructions for the tasks were similar, but the sequences of letters were made shorter (i.e., 15 + 2 here vs. 25 + N letters long in the first experiment).

This experiment followed a mixed design with 1 between-subject factor and 2 within-subject factors. The between-subject factor was the order of exposure to the environments (Real-VR vs. VR-Real) (see Fig. 35). The within-subject factors were the environment (VR vs. Real) and the trial number (i.e. representing the time: users had to do the 2-back task 3 times each time they were exposed to an environment). Each condition (Real-VR or VR-real) was repeated 3 times, resulting in a $2 \times 2 \times 3 \times 3$ mixed experimental design (see Table 7). The order of exposure to each environment was counterbalanced using a Latin square design.

Independent Variables		
<i>Participants</i>	25	(random variable)
<i>Order of exposure</i>	2	Virtual-Real, Real-Virtual (between-subject)
<i>Environment</i>	2	Virtual, Real
<i>Trial number (= time)</i>	3	1, 2, 3
<i>Repetition (= block)</i>	3	1, 2, 3
Dependent Variables		
<i>Self-reports</i>	Rating Scale of Mental Effort [318]	
<i>Task performance</i>	Accuracy	
<i>Behavioural measure</i>	Response time Pressure exerted on the trigger button Walking trajectory and velocity	

Table 7: Experiment 2: independent and dependent variables.

2.3.1 Independent Variables

In this experiment, three independent variables were manipulated: the Order of Exposure to the Environments, the Environment, and the Trial Number (i.e., time).

Order of Exposure to the Environments (Real-VR vs. VR-Real)

The experiment was divided in 3 identical blocks (see Fig. 35). In each block, users started either with the real environment followed by the virtual environment (Real-VR), or they started with the virtual environment followed by the real environment (VR-Real).

Environment (Virtual Reality vs. Real World)

The participants performed the cognitive tasks in both the IVE using the Vive HMD (VR) and in the real world (Real) wearing SMI glasses (see Fig. 29).

Trial Number

The Trial number was tested as an indicator of time. All participant did the cognitive task 3 times in each environment in each of the 3 blocks (i.e., they did the 2-back task 9 times in total in each environment).

2.3.2 Collected Data

In this experiment, physiological measures were not considered. Self-reports (RSME), task performance (accuracy), and behavioural measures (pressure exerted on the button, response time, and locomotion data) were collected.

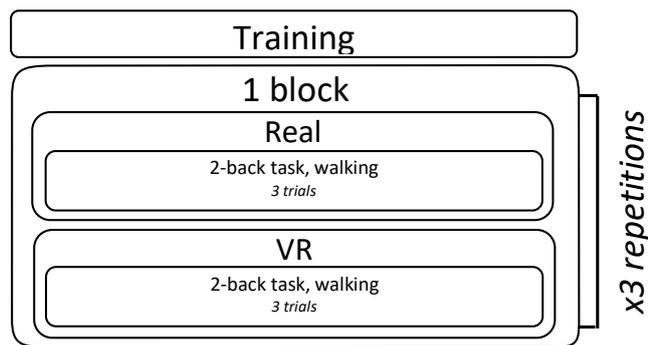


Figure 35: Experiment 2: Process of the experiment for one participant (excluding pre, post-questionnaire, instructions, set-up, and calibrations). The order of the exposure to each environment was counterbalanced (either Real-VR or VR-Real).

2.3.3 *Experimental Procedure*

The second experiment progress for one participant is given in Fig. 35. The procedure was very similar to the first experiment one (see Section. 2.5.3), except users only had to perform the 2-back task during the training. The total duration of this second experiment was around 45min.

2.3.4 *Hypotheses*

This study aims to study if there is an effect of acclimatization to IVEs on the users' mental effort responses.

The first hypothesis is the same as in the first experiment. Contrarily to in the first experiment which studied the effect of wearing a VR HMD on users' mental effort using different task difficulties and different secondary task conditions, we only considered the 2-back task difficulty and the walking condition. The results of the first experiment concerning the effect of VR were discussed as the different conditions (different auditory task difficulties and sitting condition) might not have sufficiently exploited the visual interactions allowed by VR technologies. Walking require the use of visual markers to move in the virtual or real environment, which is why we hypothesized again that there would be an effect of wearing a VR HMD on users' mental efforts [**H1**] (same H1 hypothesis than in the first experiment).

Our second hypothesis concerns the effect of wearing an HMD over-time on the mental effort. We hypothesize that users will experience an "adaptation effect" or a "wow effect" when exposed to the IVEs. Their mental effort in VR will be greater compared to in the real world at the beginning after having put on the headset and it will stabilize over-time. Therefore, we expect an interaction effect between the Trial Number (= time) and the Environment variables [**H3**].

As in the first experiment, it is supposed that the results will be supported by the analysis of the other indicators of the mental effort. To summarize, our hypotheses are:

- **H1**: Users experience more mental effort in a VE while wearing a VR HMD than in the real world.
- **H3**: Users will experience more mental effort during the first trial(s) compared to during the last trial(s) while wearing a VR HMD than in the real world.

2.4 *Results*

As in Section 1.4, GLMM analysis was considered for all dependent variables (all parametric). For each variable, the user was considered as a random factor. The order of exposure to each environment was considered

a between-subject factor and all other independent variables as within-subject factors (see Table 7). The statistical analysis was performed using the R statistical software.

Figure 36 presents the results concerning the self-reported mental effort and the response time.

2.4.1 Self-reports

RMSE: The GLMM showed a main effect of the Environment $F_{1,24} = 34.48$, $p < 0.001$, $\eta_p^2 = 0.59$, but no significant effect of the Trial Number and no interaction effect between the Trial Number and the Environment. These results support [H1] but do not support [H3].

2.4.2 Task Performance

Accuracy: The GLMM showed no significant effect of the Environment and of the Trial Number. The analysis did not show any significant interaction effect. These results do not support [H1] and [H3].

2.4.3 Behavioural Measures

Response time: Due to measuring errors, only the data of 22 users could be analyzed. The GLMM analysis showed a significant effect of the Environment $F_{1,22} = 10.00$, $p < 0.01$, $\eta_p^2 = 0.31$ and of the Trial Number $F_{1,88,41.29} = 9.66$, $p < 0.001$, $\eta_p^2 = 0.31$. The analysis did not show any significant interaction effect. Post-hoc tests show that the difference in mental effort is significant between the first and the third trials, and between the second and the third trials in a block. Users answered faster during the third trial

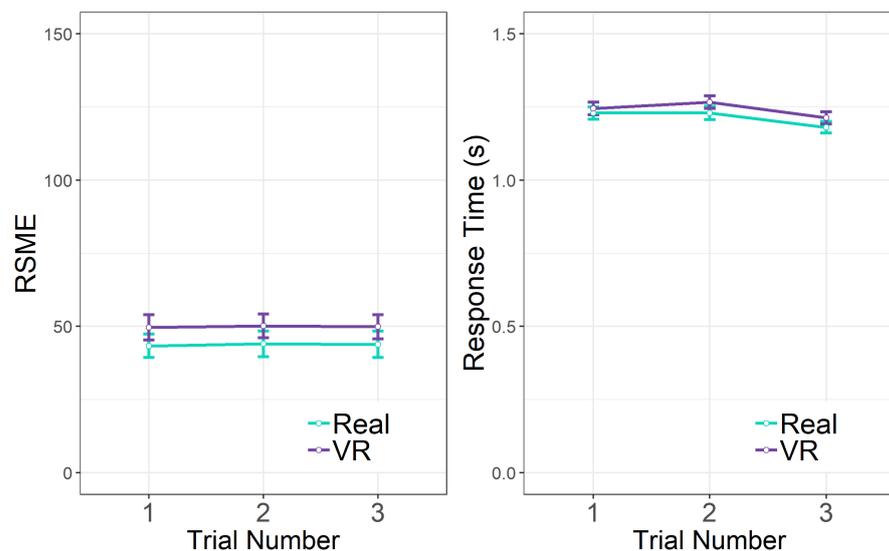


Figure 36: Mean indicators of mental effort (RSME, response time) depending on the Trial Number (= time) and environment (i.e., VR vs. Real).

compared to in the first and second ones. These results support [H1] and do not support [H3].

Exerted Pressure and Walking Measures: No significant effect was found on the pressure exerted on the button, and on the velocity and position data for any of the independent variables. These results do not support [H1] and [H3].

2.5 Discussion

In this experiment, we found a significant effect of the Environment on the subjective mental effort (RSME) and on the response time. This result was not supported by other behavioural measures and task accuracy. Therefore, the first hypothesis is partially supported. The effects on response time should be considered cautiously as the differences between the different conditions are relatively small (around 50 ms). In this experiment, users were walking in all conditions and did not control an avatar in the virtual environment. The absence of visual feedback of their feet and body movements might be at the origin of the effect of VR on the users' mental efforts. This result goes in the same line as previous studies that demonstrated that the control and type of self-avatar could impact users' cognitive performances [217, 277]. Another result can be highlighted. While there is a significant effect of the environment on the RSME, there is no effect of the environment on the accuracy (see Fig. 36). Participants would perceive the task as more mentally demanding in the VE than in the real environment while demonstrating similar levels of performance. This result could reflect the fact that there is indeed a difference between the demands imposed on users' cognitive resources and the tasks performances, or it could reflect a perceptual bias linked to the exposure to IVEs.

As for the adaptation effect, there is no effect of time (i.e., trial number) on users' mental effort responses, except for the response time. No interaction effect was found between time and environment on the users' mental effort responses. The outcomes considering the participants' profiles (VR and gaming experience) and first repetition (i.e., first block) only were also explored and demonstrated similar results. Therefore, the third hypothesis was not validated by this experiment. Users did not experience more mental effort during the first trial compared to during the second and third trials of each block in VR compared to in the real world. This result can be discussed. The literature shows that users can easily be impressed by the immersion allowed by VR technologies [60]. While this "wow effect" or "adaptation to IVEs effect" did not show an effect on users' mental efforts in this experiment, it could impact other psychological processes. Moreover, we tested only 3 time conditions every 34s (duration of a 2-back task trial) during each exposure. The conditions might not have been correctly scaled in time to catch the "wow effect." The learning effect should also be taken into consideration as it is a major phenomenon studied in the training and education fields. Parting the effects linked to the improvement of the skills

after repetitions of the task from the effects linked to the accommodation to IVEs can be complex. Further works could help to determine at which time scale they intervene. Nevertheless, the fact no effect of time during short exposures to VR on mental effort was found is encouraging as it supports the fact that mental effort in VR might quickly stabilize and is not impacted by the accommodation to VR effect.

3 GENERAL DISCUSSION

3.1 *Impact of Wearing an HMD on the User's Mental Effort*

In the first experiment, 3 levels of difficulty for the cognitive task and 2 secondary task conditions were considered. In the second one, only the 2-back task and the walking condition were taken into account. We found contradictory results concerning the effect of wearing a VR HMD on the users' mental effort. No effect of wearing the VR HMD was found in the first experiment as opposed to in the second experiment. This can be explained by the fact the VR conditions were very similar to the real conditions in the first experiment. The virtual environment was designed to be highly faithful to the real one. Moreover, the interaction in the sitting condition did not require any visual information processing and were identical in both environments. On the contrary, there was a difference of body representation and control in the second experiment as only the walking condition was considered. Users did not have visual feedback of their body in the VE, contrarily to in the real world. We pushed the analysis further considering only the walking conditions in the first experiment and indeed found a slight effect of the Environment on the RSME ($F_{1,26} = 4.32$, $p = 0.05$, $\eta_p^2 = 0.14$), but no effect of the Environment on the response time ($F_{1,26} = 0.42$, $p = 0.52$, $\eta_p^2 = 0.02$). Moreover, there was no interaction effect between the Secondary Task (Sitting vs. Walking) and the Environment (VR vs. Real) on both measures. This small difference in results might be due to the difference in statistical power between the first and second experiments for the walking conditions. There was an inferior number of data considered for the walking conditions in the first experiment compared to in the second one (i.e., experiment 1: 2 repetitions \times 3 difficulties vs. experiment 2: 3 repetitions \times 3 trials in each environment), and the task difficulty conditions might have absorbed the effect of the environment. To summarize, these results support that wearing a VR HMD has a limited impact on users' mental efforts compared to in the real world. The type of tasks, nature of stimuli, and VR interactions can impact mental effort in VR compared to in the real world. In particular, walking in VR with no avatar increased mental effort compared to walking in the real world.

In addition, walking significantly increased mental effort in both environments compared to the sitting condition. This result was supported by all objective indicators. Therefore, interactions need to be quantified in studies assessing similar cognitive states.

The accommodation to VR effect did not have a significant impact the users' mental efforts compared to in the real world during short environment exposures. This supports the idea that mental effort measures were quickly stabilized following the IVE exposure.

3.2 *Lessons Learned*

Besides giving insights about the effect of wearing a VR HMD on users' mental efforts, this study provides lessons on how to measure mental effort in VR in particular about what worked and what did not.

First, the mental effort reported via the RSME was consistent with the subject's accuracy while performing different levels of difficulty of the N -back task in the first experiment (see Fig. 33). RSME scores were however almost constant in a block between two consecutive 2-back task trials in the second experiment. Thus, RSME can be an appropriate tool in order to measure the level of mental effort when using clearly different tasks difficulties. However, finer effects might be more difficult to assess using this measurement method. While this scale can be used between different conditions rather quickly, reporting users' mental effort using the latter still is impracticable at a higher frequency in VR.

Second, the results showed that the level of difficulty had an effect on the HR, the response time, and on the pressure exerted on the trigger button. Since the mental effort was found to be influenced by the difficulty of the cognitive task (see Section 1.4.1), these findings are consistent with the literature [72, 255]. Nevertheless, depending on the context, not all of the indicators were efficient to measure the mental effort. On the one hand, the response time was able to discriminate between the three levels of difficulty in the presence and in the absence of the secondary task. On the other hand in the sitting condition both HR and exerted pressure measures were able to differentiate the 1-back task from the 2-back task, but not the 2-back task from the 3-back task (see Fig. 34). This can be explained by the between-individual variability. Most people consider the 1-back task as an easy task, and the 3-back task as a difficult task. However, the perceived difficulty tend to vary concerning the 2-back task, which may cause more subtle changes in the physiological and behavioural measures. In the walking condition, the exerted pressure could still discriminate between the 1-back task and the 2-back task, but the HR was unable to discriminate between any of the difficulties. These outcomes are consistent with the literature since movements tend to add noises in the physiological signals and to decrease the distinction between signal features [284]. Fine changes in physiological responses linked to psychological processes can be absorbed by more impactful factors such as physical movements.

Concerning other manipulated variables, the response time results were consistent with the RSME scores. There was no effect of the Environment on the response time in the first experiment, contrarily to in the second

one. On the other hand, HR, SCL, accuracy, and pressure exerted on the trigger button were not affected by the Environment. Results also showed a clear influence of the secondary task conditions on the HR, SCL, response time, and on the pressure exerted on the trigger button. The latter should, however, be taken with caution since they may be explained by the way users held controllers in the sitting condition versus in the walking condition. While the effect observed on the SCL and the HR may partially be explained by changes in mental effort, those were probably mainly affected by the physical activity caused by the walking conditions, compared to the sitting conditions [281].

Finally, while users' body behaviour and gesture showed to be influenced by the cognitive load in past studies, the analysis of the walking behaviour was not successful to show an effect in these two experiments. This might be explained by the fact the space in which the subjects could walk in VR and in the real world was restrained. Bigger spaces and no path constraint may have lead to different outcomes. Also, the analysis of the SCL was not able to discriminate between the different independent variables in the first experiment, aside from the secondary task condition. An explanation could be that the cognitive task required the subject to be focused, and was not emotionally arousing.

In summary, in our experiment, among behavioural and physiological measures, the response time was the best indicator of mental effort, followed by the exerted pressure, then, by the HR. Cautions should however be taken when using physiological sensors and behavioural measures in interactive environments, as their viability tend to greatly decrease with the involvement of movements [284].

4 CONCLUSION

In this chapter, we studied the impact of being immersed in a VE using a VR HMD on the user's mental effort. We conducted two experiments. In the first one, participants had to perform a standardized cognitive task with various levels of difficulty in two different environments (real vs. virtual), and in the presence or not of a secondary task (sitting vs. walking). In the second one, users had to perform a single cognitive task difficulty while walking and were exposed multiple times to both environments (real vs. virtual) in the same order. Their mental efforts were assessed multiple times during each exposure. The results suggest that wearing a VR HMD has a limited impact on the users' mental efforts. Given similar interaction and tasks conditions, there was no effect of wearing a VR HMD on the users' mental efforts. Moreover, the experiments did not show an effect of accommodation to VR on users' mental effort compared to in the real world. However, walking in VR with no avatar increased users' mental efforts compared to walking in the real world. In addition, walking significantly increased mental effort and decreased task performance in both environments compared to being seated. These findings are grounded on

various kinds of measurement: self-report, task accuracy, mean heart-rate, response time, and pressure exerted on the answering button, which validities are discussed in the different contexts.

Taken together, our results support the view that being immersed in a VE while wearing an HMD has a limited impact on users' mental efforts. In particular, interactions and stimuli should carefully be controlled, especially if they differ between the virtual and the real environments when VR is used to simulate the real world.

A METHODOLOGICAL APPROACH TO INTRODUCE MENTAL WORKLOAD ASSESSMENT IN THE DESIGN OF VR TRAINING SCENARIOS

SUMMARY In this chapter, we propose to consider MW assessment for the design of complex training scenarios involving multiple parallel tasks in VR. The proposed approach is based on the assessment of the MW elicited by each potential task configuration in the training application. Following the assessment, the resulting model is then used to create training scenarios able to modulate the users' MWs over time. This approach is illustrated by a VR flight training simulator based on the Multi-Attribute Task Battery II, a computer application developed by NASA which solicits different cognitive resources. A first user study ($N = 38$) was conducted to assess the MW in each task configuration using self-reports and performance measures. This assessment was then used to generate three training scenarios in order to induce different levels of MW over time. A second user study ($N = 14$) confirmed that the proposed approach was able to induce the expected MW over time for each training scenario.

TABLE OF CONTENT

1	Methodological Approach	78
1.1	Tasks Identification and State Machine Generation	78
1.2	Data Assessment	79
1.3	Scenario Generation	80
2	Validation of the Approach: Designing Training Scenarios Based On Mental Workload in a VR Flight Simulator	80
2.1	Tasks Design	80
2.2	State Machine Structure	84
2.3	Mental Workload Measures	85
2.4	Apparatus	85
2.5	Experiment 1: Data Assessment	86
2.6	Experiment 2: Scenarios Generation	92
3	General Discussion	96
4	Conclusion	98

1 METHODOLOGICAL APPROACH

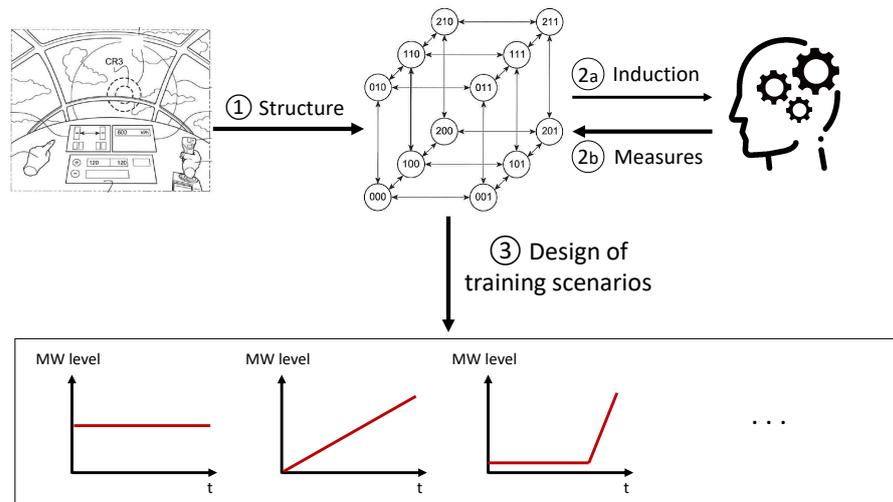


Figure 37: Illustration of the methodological approach to introduce MW assessment for the design of VR training scenarios.

In this section, we propose an approach to introduce MW assessment in the design of VR training scenarios to make them more adapted to users' cognitive states [51]. The goal is to modulate the user's MW over time in complex contexts involving multiple parallel tasks.

Our approach is based on the model of the states of the training application using the different allowed task configurations. State machines are convenient in our proposed approach as it can be used to generate scenarios based on the arrangement of tasks configurations (i.e., nodes or states) over time. Each state contains data about the tasks setting and the users' MWs, and the transitions help to constrain the sequencing of the tasks.

This methodology is divided into the following steps. First, tasks are selected based on the objective and on the context of the training study. Then, each selected task is subdivided into different difficulty levels. These task levels are combined to form the states of a state machine. The state machine transitions are defined to allow users to navigate from a task configuration (i.e., a state) to another. Following these steps, the experimenter chooses which MW data to collect. The MW is thereafter assessed and will determine the attributes of the states. These measures can finally be used to design training scenarios by defining paths inside the state machine.

1.1 Tasks Identification and State Machine Generation

The objective of this first step is to design the states and the transitions of the state machine. The states contain data about the tasks configurations and the MWs they induce to users. The transitions (i.e., a vector with a

start and an end states) determine if users can transit from one state to another or not.

Starting with the selection of the tasks, those should be defined based on the training context. For example, in a car manufacturing training scenario, the main task can be the assembly of car pieces and in a fire safety application, the main task can be the extinction of a fire situation. The idea of this first step is to extract all tasks relevant to a training purpose in the imposed context. Once the tasks are selected, those can be decomposed into multiple discrete intrinsic levels of difficulty, including their presence (activation) or absence (deactivation), if relevant. Each task level should be combined to form the "states", except for the tasks levels which can not be associated given the training environment. In any case, all tasks should be represented in each state with a specific level of difficulty.

In the same way, the transitions should be created depending on the constraints of the training context. By default, those can be defined so only one task at a time can be upgraded or downgraded in its difficulty level, or so all states are linked together. However, if one task level can only come after another specific task level, this should be taken into account. The condition to trigger a transition can be of multiple types (e.g., a time limit, once all the tasks of a state have been fulfilled, after users have reported their MW).

1.2 Data Assessment

Following the identification of the tasks levels, the trainer has to choose which MW data to collect (see Section. 3). Then, the objective is to assess a maximum of data to have an estimation of the MW induced by each state.

For a clean estimation and to avoid order effects, the experimenter can make all participants traverse the states in a randomized or counterbalanced order so each state is explored in a balanced way. In the case the training context would be too complex and the states too numerous, only certain states can be used to assess data. For example, for tasks with more than 2 levels of difficulty, the intermediary states considering a variation of difficulty of one task can be ignored during the assessment and estimated afterwards. Nevertheless, this will draw to less accuracy about the MW induced by these states.

Once the data is measured, it can be assigned to the states as attributes. At this point, the states should contain data about the task configuration (i.e., "task1 difficulty level", "task2 difficulty level", ...) and about the mean MW induced by the task configuration to users (i.e., "MW measure 1", "MW measure 2", ...). The measured data can be used to weight the transitions (e.g., the difference between the subjective mental workload of the end state and of the start state) for the scenarios design purposes. Also, a confidence value can be attributed based on the MW data distribution inside a state.

1.3 Scenario Generation

The MW measures can now be used to design scenarios by defining the order of the task configurations (i.e., states) through time.

From this point, the keys of the design belong to the trainer. Depending on the motive of the training, they have to define the scenario depending on the MW they want to induce to the trainee. The trainer can as well make a list of constraints linked to the duration of the training, the number of simultaneous changes of tasks for the transitions, and the appearance or not of a task level. The computation of the paths can be done using the transitions weights (which can be defined as the difference of MW data between the end and the start state). For example, if a scenario should maintain the user's MW at the same level all long, the path can be computed by choosing the transitions with the cheapest costs (i.e., lowest absolute differences of MW).

This last step should result in the design of a training scenario which modulates the progress of MW over time, as defined by the trainer.

2 VALIDATION OF THE APPROACH: DESIGNING TRAINING SCENARIOS BASED ON MENTAL WORKLOAD IN A VR FLIGHT SIMULATOR

To illustrate the proposed methodological approach, we designed a VR application based on a flight simulator. This section is structured following the proposed methodology in section 1. First, tasks and their difficulty levels were defined following the context of the training. Those were combined to form states, which were used to structure a state machine. Then, the MW measures to collect were identified. Two experiments ($N1 = 38$, $N2 = 14$) were conducted to validate the proposed approach. The objective of the first experiment was to collect MW data in the defined task configurations. The objectives of the second experiment was to use these measures to create scenarios based on training objectives and to compare the MW results to the expected outcomes.

2.1 Tasks Design

The illustrative application was designed based on a VR flight simulator. It is inspired by the second version of the Multi-Attribute Task Battery II (MATB-II) [253], a computer-based application designed to induce and evaluate an operator's performance and workload developed by NASA. This battery of tasks has been widely used to study multitasking and the use of automation [253]. The original application comprises a monitoring task, a tracking task, a schedule window, a communication task, and a resources management task. Those were intended to be presented on a single computer window and are analogous to tasks that aircraft crew-members perform in flight, while being accessible to non-pilot subjects.

In the current studies, three tasks were selected and adapted to a VR environment: the piloting task, which would be an analogy to the tracking

task, the communication task, and the resources management task (see Fig. 38).



Figure 38: Virtual cockpit view. (1) Instantaneous Self Assessment (ISA) interface. (2) Resources management task interface (deactivated); when activated, the interface lit up (with a green outline). (3) Communication task interface (activated); when deactivated, the interface lit off (no green outline). (4) Informative panel which gives information about which task is activated or not at the current time. (5) Virtual representation of the joystick used to pilot the aircraft and of the right hand. The left hand is represented in the same way, but tracked by a Vive Controller and animated depending on the interaction.

Piloting Task

The tracking task of the MATB-II required the user to keep a target in the centre of a square controlled by a joystick. In the following experiments, the task was adapted to fully exploit the immersion permitted by VR. Participants were piloting an aircraft in the first-person perspective using a joystick and could see the environment being refreshed in real-time depending on their actions. The speed of the piloting task was imposed on the users (they could not accelerate nor decelerate). They could, however, orientate the aircraft using a joystick. Preliminary tests were done to tweak the sensitivity and control of the interaction. Since the participants were non-pilot subjects, only two degrees of freedom of the aircraft were retained: the yaw (i.e., rotation upon the vertical axis) and the pitch (i.e., rotation to go up and down). The roll degree (i.e., rotation upon the forward vector of the aircraft) was not included. The objective of this task was to follow as closely as possible the green line which passed through all circles centres.

Three levels of difficulty were proposed: easy piloting (0), medium piloting (1), and hard piloting (2). The task difficulty was modified by adapting the speed of the aircraft and the number of visible circles at a time (see Fig. 39). In the easy difficulty, users could see 3 circles at a time and the air-

craft was advancing at a slow pace (about 20 seconds between 2 circles). In the medium difficulty, this was changed to 2 circles and a speed which was multiplied by about 2, and in the hard difficulty, to one circle and an original speed which was multiplied by about 3.

Each circle was separated by a constant distance. Only the easiest task configuration was set to have a straight alignment of circles. Otherwise, the trajectory was randomized for each participant so the horizontal and vertical distances between two circles could not exceed an imposed value.

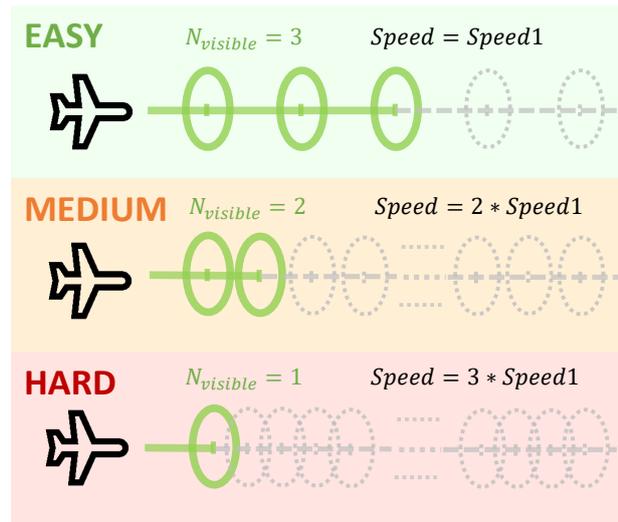


Figure 39: The difficulty of the piloting task was modulated by changing the number of circles visible by users at a given time (i.e., $N_{visible}$) and the speed of the aircraft.

Communication Task

The communication task was designed similarly to the one from the MATB-II, but without the radio channel selection. This task was designed to have two levels: it was either disabled (0) or enabled (1). When the task was enabled, audio messages were sent to the participant's audio headset, and the interface of interaction lit up (see Fig. 38).

The operators were asked to answer when the message was intended for their aircraft. In the case of our experiment, the ID of the operator's aircraft was "HDG219". When the message was directed to the aircraft, the participant was required to change the frequency of the radio in accordance with the message, by pushing the "plus" and "minus" buttons, which changed the left screen (see Fig. 38). Once users finished inputting the frequency, they were asked to click on the validate button, which changed the right screen so it matched the left screen (see Fig. 38). No action was expected when the message was directed to another aircraft. In the studies, half of the calls in each state with the communication task were set to target the user.

The target frequency was an integer number of 3 digits with no decimal part. It was computed so the user was asked to click on the plus or minus

button (see Fig. 38) a random number of times between 2 and 5 included. When the message was not directed to the user, the requested aircraft was a random one between 4 different IDs.

Resources Management Task

This task was designed in accordance with the one from the MATB-II. It depicts a generalized fuel management system. This task had two levels: disabled (0) and enabled (1). When the task was enabled, the interface of interaction of the task was lit up (see Fig. 40 and Fig. 38).

The interface displayed 8 different pumps numbered from 1 to 8, and 6 tanks labelled from A to F (see Fig. 40). When the task was enabled, the fuel of the tanks A and B started to decrease. The objective of the user was to keep the levels of the tanks A and B in the blue zone displayed on the two tanks sides. Those indicated the critical levels of fuel for those tanks. Tanks D and F had unlimited capacities. Not transferring fuel to tanks A and B resulted in empty tanks after some time, while tanks C and D only lost fuel if they were transferring fuel to another tank.

Users could use the 8 pumps at their disposal to maintain the two tanks levels in the appropriate zone. There were 3 possible states for each pump button: grey, which meant the pump was deactivated, green, which meant the pump was activated and red, which meant the pump was failed and could not be used. An activated pump button (green) meant fluid was circulating in the direction indicated by the button arrow. Clicking on a grey button would turn the button green, and clicking on it again would turn the button back in grey, like a switch button. The red state was activated at predefined moments.

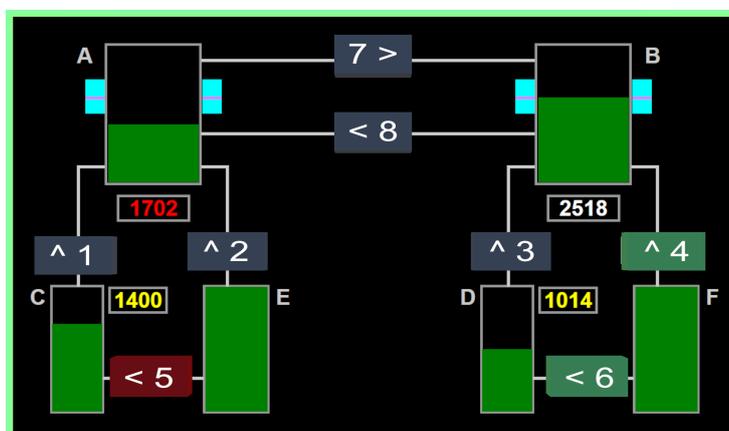


Figure 40: Resources management task interface. A and B are the main tanks; their fuel levels are indicated below the tanks. C and D are supply tanks; their fuel levels are indicated on their right side. E and F are supply tanks with unlimited capacities. The buttons numbered from 1 to 8 are pumps button. Pumps 4 and 6 are activated, pump 5 is failed, and all other grey pump buttons are deactivated.

2.2 State Machine Structure

As seen in Section 2.1, 3 tasks were considered, each of them with their own difficulty levels. This gives us: 3 piloting task (0-easy, 1-medium, 3-hard) \times 2 radio task (0-deactivated, 1-activated) \times 2 resources management task (0-deactivated, 1-activated). All these levels were combined to form 12 states ($3 \times 2 \times 2$). To ease the understanding, each state was labelled using 3 digits, one for each task. The first digit represents the difficulty of the piloting task, the second one, the level of the communication task, and the third one, the level of the resources management task. For example, in the state "201", the difficulty of the piloting task is set to "2-hard", the radio task is "0-disabled" and the resources management task is "1-enabled". Following this labelling system, the resulting state machine is depicted in Figure 41. For convenience purposes, we will call "consecutive states" two states which differ in only one task of one difficulty level. All the states could transit from one to another (i.e., there were transitions between all states). The transitions were triggered once users passed the last circle of the state they were in.

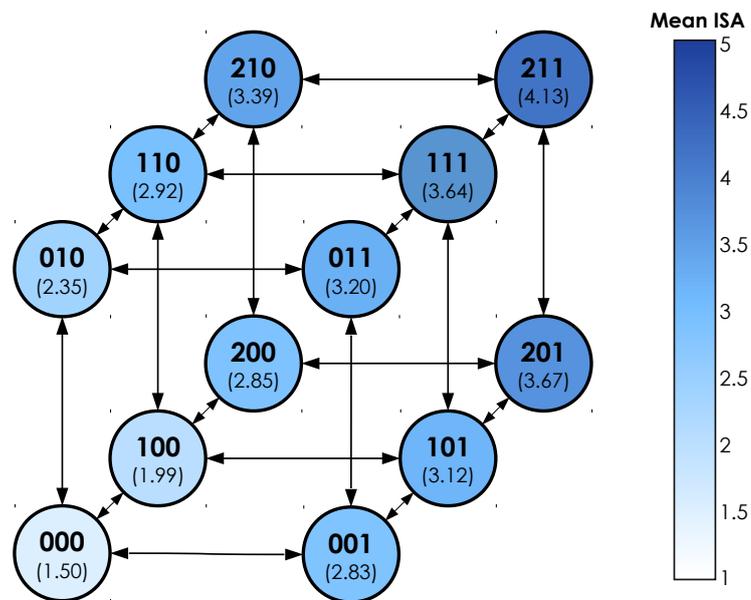


Figure 41: State machine of the designed VR flight simulator. Each state is labelled using 3 digits. The first digit refers to the level of the piloting task (0-easy, 1-medium, 2-difficult), the second one to the level of the communication task (0-deactivated, 1-activated) and the third one to the level of the resources management task (0-deactivated, 1-activated). The colours of the nodes represent the mean ISA assessed during the first user study, which are as well indicated in the brackets. Only the transitions between consecutive states are depicted there.

2.3 Mental Workload Measures

The dependent variables considered were: self-reports (i.e., the subjective MW reported by the participants) and task performances.

Self-Reports: Originally, the MATB-II was designed so the MW could be assessed using the NASA-TLX [115] after the experiment. However, the latter is a multidimensional scale, which is not adapted to the several ratings of the MW throughout a scenario. Therefore, the focus was set on the ISA [287], which rates the MW using five different ratings (under-utilized, relaxed, comfortable, high, excessive) and has been especially used during air traffic control tasks [287]. The meaning of each rating levels of the ISA was made explicit to each subject using the description given by Kirwan et al. [140] before each experiment.

During the experiments, users were asked to report their MW when a screen appeared in front of them with 5 buttons (see Fig. 38). They were asked to push the button corresponding to their MW level and then, to click on the validate button on the same screen to make it disappear.

Performance Measures: Concerning the piloting task, the distance to each circle centre was recorded throughout the experiment. The participants were given indication before the experiment of how to align the plane with the green line optimally.

As for the communication task, performance indicators of the success of the task were recorded: true positive and negative, accuracy, difference between the input and target frequencies, reaction and response time, global success for each communication call compared to the expected reaction.

Lastly, the global success time ratio of the resources management task was recorded as well as events corresponding to a success or a fail of the task.

2.4 Apparatus

Each participant was installed on a cockpit assembled for the experiment (see Fig. 42). The virtual environment was modelled in 3D so the virtual cockpit matched the real one in position and size.

As for the interactions, users piloted the aircraft using a Logitech 3M X52 PRO with the right hand, and they interacted with the virtual interfaces using a Vive controller with the left hand. Both hands were represented by transparent virtual hands. The right hand was placed on the virtual joystick (see Fig. 38), which moved when the user was interacting with the real one. The left hand was tracked with the Vive controller. As seen in Section 2.1, all interactions linked to the tasks are buttons based. The users did not need to use any of the buttons on the Vive Controller or on the joystick to do the tasks. Pushing a virtual button would be processed as follows: the

user approaches her.his left hand to the button. The interactable object highlights and the hand animation changes to a pointing index. The user pushes further the button as s.he would have done with her.his real finger. This triggers a small haptic pulse on the Vive Controller, which gives a feedback that the action has been well resolved. The environment updates according to the button action.

The VR headset used during the experiments was an HTC Vive Pro. Audio instructions were provided using the audio headset supplied with the HMD. The support application was developed in Unity 3D, and run on a laptop computer equipped with an Intel(R) Core(TM) i7-6820HK CPU (2.7 GHz), one Nvidia GeForce GTX 1070 graphic card, and 16 Go Random-Access Memory.



Figure 42: Picture of the experimental set-up. The user is wearing a Vive Pro, and using a joystick and a Vive controller on the cockpit.

2.5 *Experiment 1: Data Assessment*

The objective of this first experiment was to collect data on the users' MW in each state (i.e., tasks levels combination) to get an overview of their effect on the MW (see Fig. 41).

2.5.1 *Participants*

39 healthy participants from our research institute volunteered to take part in this study. One subject was excluded from the study because of motion sickness issues, resulting in a final sample of 38 participants (10 females, 28 males; ages 21-62, $M=36.97$). Four participants reported having a small experience with aircraft piloting, and all others never had any experience with it. All participants were fluent in French and were naive to the experiment conditions and purpose. They all completed and signed an informed consent form before the start of the experiment.

Independent Variables		
<i>Participants</i>	38	(random variable)
<i>States</i>	12	12 tasks configurations
<i>Piloting task level</i>	3	0-Easy, 1-Medium, 2-Difficult
<i>Communication tasks level</i>	2	0-Deactivated 1-Activated
<i>Ressources management tasks level</i>	2	0-Deactivated 1-Activated
Dependent Variables		
<i>Self-reports</i>	ISA [287]	
<i>Piloting task performance</i>	Distance to the centres of the circles	
<i>Communication task performance</i>	Global success, accuracy, reaction and response time	
<i>Resources management task performance</i>	Success time ratio	

Table 8: Data assessment experiment: independent and dependent variables.

2.5.2 Experimental Design

The goal of this experiment was to measure the users' MW induced by each state of the graph. Thus, each user went by all the states. Only transitions between consecutive states were considered in this study. To minimize the learning effect, the order of the states was randomized for each participant. However, they always started with the state "000" (easy piloting, no communication task, no resources management task).

The objective was to make sure that the designed states could induce different levels of MW. Therefore, we hypothesized that:

- **H1:** Increasing the level of difficulty for each task will increase the users' MW.
- **H2:** Additional task will increase the users' MW.
- **H3:** The subjective MW will have an effect on the tasks performances.

2.5.3 Experimental Procedure

The experiment had a total duration of around 1h and was subdivided into the following steps:

Written Consent and Instructions: Users completed a consent form, prior to the experiment. They were then instructed with the nature of the experiment, the equipment used, the data recorded (which was anonymized), and the tasks instructions. Participants were also asked to fill a questionnaire (experience with VR, video games, and piloting an aircraft,

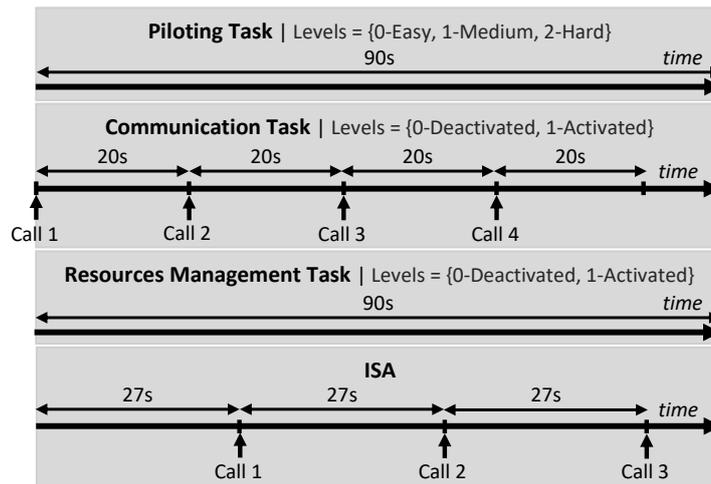


Figure 43: Progress of the tasks and of the ISA calls within one state in the first experiment. In this paper, 12 states were considered. The communication task and the resource management task were not always activated. The ISA is always present. Each state lasted at least 90 seconds (depending on the user's piloting path).

dominant hand, level of alertness, state of vision, demographic information, simulator sickness questionnaire (SSQ) [135]) to gather information about their background and their state before the start of the experiment.

Training: Users were then equipped with a Vive Controller and the HMD. They were first asked to interact with the buttons of the tasks interfaces to familiarized themselves with the interactions. Then, they travelled the states following this path: "000 – 010 – 011 – 001 – 101 – 201 – 211" (see Fig. 41), which gave them a good overview of each task and their levels.

The training part was followed by a 2-minutes pause where users did not wear the HMD and were invited to ask any question they may have had.

Experiment: In this experiment part, users travelled all the states in a randomized order, starting with the state "000". Only consecutive states were considered for the transitions. The states were set to last 90 seconds with 4 communication calls and 3 ISA calls. The progress of the tasks within a state is depicted in Fig. 43.

Debriefing: At the end of the experiment, they were asked to fill the SSQ again, debriefed and invited to ask questions.

2.5.4 Results

Generalized linear mixed model (GLMM) analysis was considered for all dependent variables (all parametric). For each variable, the user was considered as a random factor and all the independent variables as within-subject factors. When the equal variances assumption was violated, the degrees of freedom were corrected using the Greenhouse-Geisser method. When needed, pairwise post-hoc tests (Bonferroni with adjustment) were

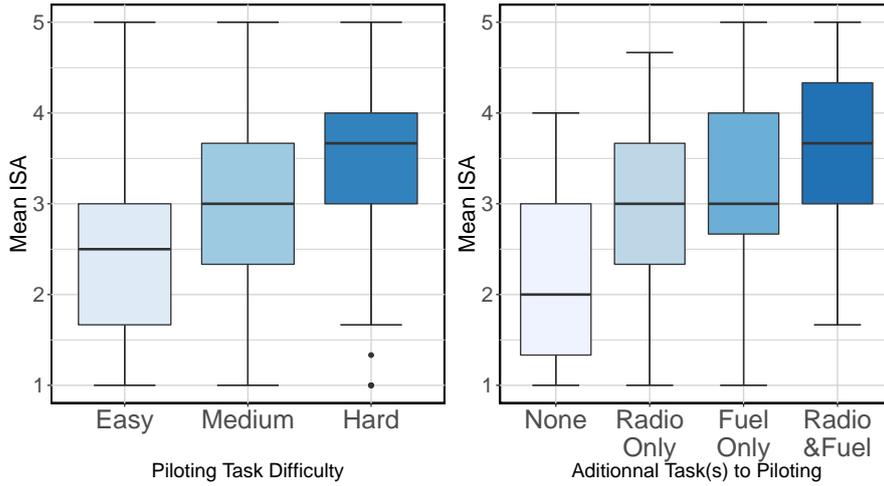


Figure 44: Mean ISA in terms of tasks levels and configurations. “Radio” refers to the communication task and “Fuel”, to the resources management task.

performed. Only significant differences ($p < 0.05$) are discussed. The statistical analysis was performed using the R statistical software. Figure 44 presents the effect of tasks difficulty levels and secondary tasks interactions on the mean ISA value reported by users, and Figure 45 depicts all ISA values reported by users in each state.

ISA: The GLMM showed a main effect of the piloting difficulty $F_{1,83,67.62} = 100.77$, $p < 0.001$, $\eta_p^2 = 0.73$, communication task $F_{1,37} = 85.23$, $p < 0.001$, $\eta_p^2 = 0.70$, and resources management task $F_{1,37} = 154.68$, $p < 0.001$, $\eta_p^2 = 0.81$. It also showed an interaction effect between the communication task and the resources management (fuel) task $F_{1,37} = 12.81$, $p = 0.001$, $\eta_p^2 = 0.26$, as well as an interaction effect between the piloting difficulty and the fuel task $F_{1,81,67.39} = 3.72$, $p < 0.05$, $\eta_p^2 = 0.09$. Post-hoc tests showed that as the level of difficulty increased, participants perceived the task as more mentally demanding (see Fig. 44) (all $p < 0.001$). The tasks were also perceived as more demanding when a task was activated (see Fig. 44). The resources management task was perceived as more difficult than the communication task ($p < 0.01$; all $p < 0.001$ for other tasks combinations effects otherwise; see Fig. 44). These results support [H1] and [H2].

Piloting task performance: The considered variable is the distance to circles centres. The GLMM showed a main effect of the piloting difficulty $F_{1,06,39.10} = 16.07$, $p < 0.001$, $\eta_p^2 = 0.20$, communication task $F_{1,37} = 16.33$, $p < 0.001$, $\eta_p^2 = 0.31$, and resources management task $F_{1,37} = 19.95$, $p < 0.001$, $\eta_p^2 = 0.35$. It also showed a main effect of the mean ISA value (mean of the 3 ISA values per state) $F_{1,447.80} = 46.30$, $p < 0.001$, $\eta_p^2 = 0.08$. Post-hoc tests showed that as the difficulty level increased, the participants tended to go further away from the circles centres (all $p < 0.01$ except between easy and medium difficulties). In the same way, the activation of the communication task or of the resources management task increased the distance

to the circles centres. Overall, the more the MW increased (ISA value), the greater the distance to the circles centres was. These results support [H3].

Resources management task performance: The considered variable is the success time ratio of the task. The GLMM showed a main effect of the activation of the communication task $F_{1,37} = 7.52, p < 0.01, \eta_p^2 = 0.17$, and of the mean ISA value (mean of the 3 ISA values per state) $F_{1,219.64} = 31.15, p < 0.001, \eta_p^2 = 0.10$. Overall, the activation of the communication task decreased the success time ratio of the resource management task. Moreover, as the MW increased, the success time ratio of the resources management task decreased. These results support [H3].

No significant effect was found on the communication task performance indicators, which does not support [H3].

2.5.5 Discussion

The levels of the tasks were able to induce different levels of MW (see Fig. 44). Figure 41 depict a global map of the mean MW assessed during this first experiment and Figure 45 presents all ISA reported in each state. Overall, the ISA values follow a normal distribution in each state. Increasing task difficulty and additional tasks increased users' subjective MW. Both [H1] and [H2] were supported in our experiment. The resources management task induced a higher MW than the communication task. Different explanations can be provided. This effect could be explained by the fact it stimulated similar pools of cognitive resources to the ones stimulated by the primary piloting task (visuomotor) compared to the ones stimulated by the communication task (mainly auditory-motor). This would also explain the interaction effect between the resources management task and the piloting difficulty on the reported MW. These results go in the same line than Wicken's Multiple Resources Theory [304]). The resources management task could also directly be considered as intrinsically more difficult than the communication task. Users did, however, not report having perceived one of these two tasks as being more difficult after the experiment.

The subjective results were supported by piloting and resources management tasks performances, which supports [H3]. However, piloting task performances were less steady in the states compared to the self-report measures (e.g., no effect between easy and medium piloting difficulty). No significant effect was found for the communication task as it was mainly used to distract users from the other tasks and to induce time load. Participants managed to complete this task most of the time.

Overall, the designed tasks configurations were successful into inducing different levels of MW to the users. Moreover, the subjective MW and the states had an effect on the piloting task performance (distance to the circles centres) and on the resources management task performance (success time ratio). However, the ISA value was more reliable to differentiate the different levels of mental workload in the states than the performance measures considering the effect of the states and the data distribution. Finally, each state was characterized by the following attributes: (*piloting task level, com-*

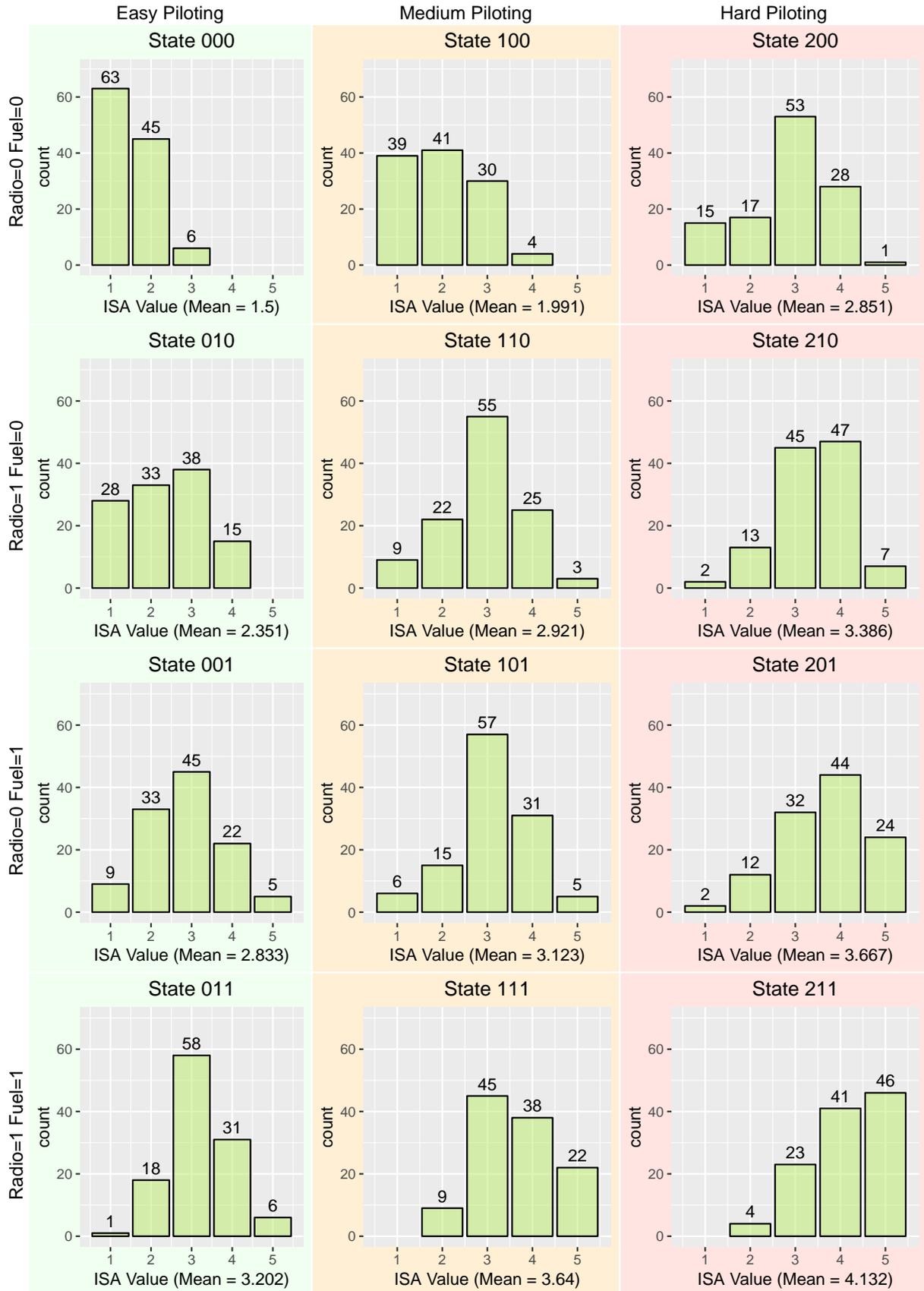


Figure 45: All ISA values (subjective MW) obtained per state during the data assessment experiment (38 participants reported their MW 3 times per state).

munication task level, resources management task level); (*ISA value, piloting task performance, resource management task performance*). As [H3] was partially supported, we only used the *ISA value* among the MW measures in the subsequent study.

2.6 Experiment 2: Scenarios Generation

The objective of this second experiment was to build scenarios based on the states defined in Section 2.5 and to validate the proposed approach.

2.6.1 Participants

14 healthy people from our research institute, who did not participate in the first experiment, volunteered to take part in this study (2 females, 12 males; ages 21-52, $M=32.21$). None of the participants had an experience with piloting an aircraft in the past. As in the first experiment, all participants were fluent in French and were naive to the experiment conditions and purpose. They all completed and signed an informed consent form before the start of the experiment.

2.6.2 Experimental Design

We chose to focus on 3 different scenarios using a fixed number of 5 states. They were all designed based on the subjective MW (ISA).

The **scenario 1** *induces a medium MW level all long*. This scenario is a classical use case where a trainer wants to keep the users' MW at a medium level during the whole training, while varying the experimental conditions. The goal can be to train users while keeping them engaged, by not overloading nor underloading them, and without using repetitive tasks. This scenario was designed by choosing the 5 states with the MW the closest to $ISA = 3$. The transitions between each of these states were computed so there was a prioritization of transitions between consecutive nodes. The orientation of the given path was chosen randomly. The resulting path is: 200 - 101 - 001 - 011 - 110 (see Fig. 41).

The **scenario 2** *induces a low MW level first, and a suddenly high MW level in the end*. This scenario is a typical scenario where the reactivity of the pilot is studied in an urgency or surprising situation. The 4 first states were chosen as the 4 consecutive states with the lowest ISA. The first state was chosen by prioritizing the state "000", and the orientation of the path was chosen randomly. The last state was chosen as the one with the highest ISA value. The resulting sequence is: 000 - 010 - 110 - 100 - 211.

The **scenario 3** *induces a progressively increasing MW level throughout the experiment*. It can be used to train progressively pilots to different levels of difficulty they might experience. The scenario was defined so the first state was the one with the lowest ISA, and the last state, the one with the highest ISA. All paths of 5 consecutive states going from the first state to the last one where computed, as well as the sum of the transitions weights for each path. The final path was selected as the one the lowest sum of

transitions weights: 000 - 100 - 110 - 210 - 211.

This experiment followed a one-factor (scenario) within-subject design. All participants experienced all 3 scenarios. The order of the scenarios was counterbalanced using a Latin square design, except for the 2 last participants who did the 3 scenarios in random order.

Concerning the hypotheses, they are defined by the scenarios depicted above. We expect no significant effect of the experiment (first and second one) on the mean ISA of the states.

2.6.3 Experimental Procedure

The procedure is almost the same as the one defined in Section 2.5.3, except the SSQ was not included this time. Still, all participants were asked how they felt after the scenarios.

The experiment was divided into 3 blocks: one for each scenario. During the scenarios, users answered the ISA twice per state. After each scenario, users were asked to answer a set of custom questions on a 7-points Likert-Scale (1: fully disagree, 7: fully agree) to define their perceived difficulty during the experiment (see Table 10). The 3 questions respectively transcribe the progress of the scenarios 1, 2, and 3, from MW to perceived difficulty. They were mixed with 3 other unrelated questions linked to the users' absorption and sense of presence [270] to mask the objective of the study to participants. At the end of the 3 scenarios, participants were asked to map the descriptions of the 3 scenarios to the 3 experiences they did.

The parameters used in this experiment were set for each state to last 70 seconds, with 3 communications calls (2 which targeted the user) and 2 ISA calls.

Independent Variables		
<i>Participants</i>	38 + 14	(random variable)
<i>Experiment</i>	2	Experiment 1 ($N = 38$), Experiment 2 ($N = 14$)
<i>Scenario 1 States</i>	5	200 - 101 - 001 - 011 - 110
<i>Scenario 2 States</i>	5	000 - 010 - 110 - 100 - 211
<i>Scenario 3 States</i>	5	000 - 100 - 110 - 210 - 211
Dependent Variables		
<i>Self-reports</i>	ISA (3 times per state) [287]	
<i>Piloting task performance</i>	Distance to the centres of the circles	
<i>Communication task performance</i>	Global success, accuracy, reaction and response time	
<i>Resources management task performance</i>	Success time ratio	

Table 9: Scenarios generation: independent and dependent variables.

2.6.4 Results

Evaluation of the MW overtime

ANOVAs were run to analyse the results. To ensure the replication of the first experiment results, the outcomes between experiment 1 and 2 (between-subject factor “Experiment” with N=35 in Experiment 1 and N=14 in Experiment 2) were compared. The results are presented in Figure 46. For each scenario, no significant effect was found on the ISA and tasks performances. As well, each state in each scenario was also tested separately considering the experiment factor, and it did not show any significant effect on the ISA and tasks performances.

Perception of the progression of scenarios A Posteriori

The perceived progression of difficulty was assessed after each scenario using the questionnaire presented in Table 10. Results are presented in Figure 47. For the first scenario, the mean ratings were (Q1 : 3.36), (Q2 : 3.14), and (Q3 : 2.64). For the second scenario, the results gave (Q1 : 2.43), (Q2 : 5.86), and (Q3 : 4.14). As for the third scenario, the average scores were (Q1 : 1.64), (Q2 : 3.79), and (Q3 : 5.71). For each scenario, the related question was noted the best among the 3 in average. The second and the third scenario were globally successfully noted based on the MW they were supposed to induce, but the first scenario did not induce the right perceived progression of the difficulty.

As for the mapping of the descriptions of the scenarios to the appropriate experiments after all 3 scenarios, the results are presented in Figure 48. A confusion was observed mainly between the second and the third scenarios.

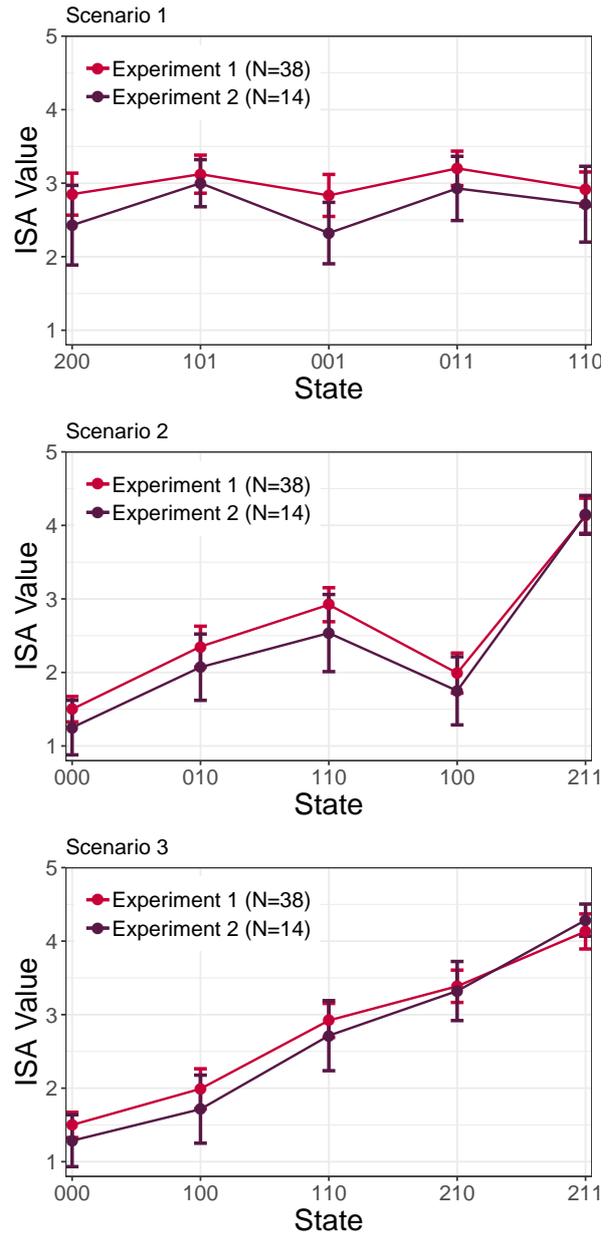


Figure 46: Comparison of the users’ subjective MW (mean ISA value) reported overtime for all 3 scenarios between the first and the second experiment. The first scenario was intended to induce a MW around $ISA \approx 3$ all long and the second one, to induce a low MW ($ISA \leq 3$) followed by a high MW at last. The last scenario objective was to induce a progressively increasing MW.

Question	Corresponding Scenario	Affirmation
Q1	Scenario 1	I felt that the difficulty was approximately the same during the scenario.
Q2	Scenario 2	I felt the difficulty suddenly increased at a given time.
Q3	Scenario 3	I felt the difficulty increased progressively.

Table 10: Custom questionnaire on users’ perceived progression of difficulty after each scenario. These questions were mixed with 3 other unrelated questions linked to users’ absorption and sense of presence to mask the study objective. Users were asked to answer each affirmation using 7-points Likert-Scales.

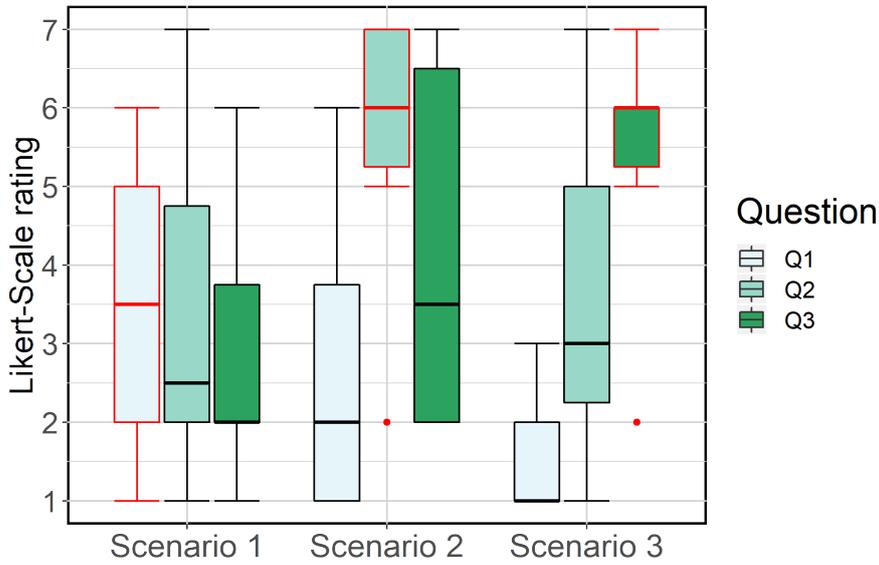


Figure 47: Results of the questionnaire asking users about their perceived progression of difficulty after each scenario. Corresponding “questions” and scenarios are highlighted in red.

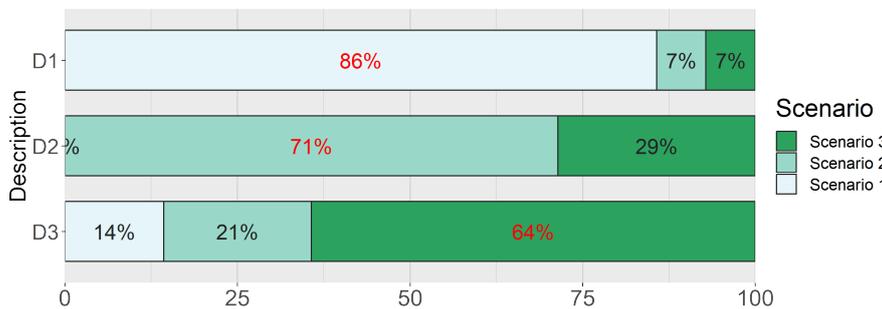


Figure 48: Results of the mapping of scenarios descriptions to the experienced scenarios (after all 3 scenarios). D1: *The experiment induced a medium MW level all long.* D2: *The experiment induced a low MW level first, and a suddenly high MW level in the end.* D3: *The experiment induced a progressively increasing MW level throughout the experimly.* Corresponding descriptions and scenarios are highlighted in red.

2.6.5 Discussion

The ISA scores and the task performances followed the expected results for scenarios 1, 2, and 3, which were designed based on the first experiment measures (see Fig. 46). The slight variations between the two experiments can be explained by the fact there might have been a small order effect as the travel of the states was randomized in the first experiment and not in the second one. However, as tested, those are not significant. As well, the subjective measures were once again supported by the piloting and the resources management tasks performances.

The perceived difficulty was found to lack of accuracy compared to the expected outcomes, on the contrary to the ISA values. This can be due to the fact these questions were non-standardized and to the difference in the rating delays [306]. Some users were found to answer in unexpected ways compared to their ISA score, even just after having completed the scenario. Because the tasks configurations changed through time for the first scenario, users did not feel that the difficulty was approximately the same during the experiment. As for the second and third scenarios, users had issues to determine if the task difficulty suddenly or progressively increased. As well, the fact there was a confusion between the mapping of the second and the third scenario at the end of the experiment might be explained by the fact users mainly remembered the first and the last states of the scenarios, which appear to be the same in the two scenarios. Therefore, subjective ratings have to be performed cautiously and delays in rating, considered carefully, as mentioned in previous studies [306]. The subjective MW rating throughout the scenarios was more accurate than the post-experiment self-report questions and the performance measures.

3 GENERAL DISCUSSION

In this chapter, we proposed an approach to create scenarios in order to induce different levels of MW during a VR training routine. It differs from the literature as it focuses directly on the user's subjective cognitive state and targets multiple parallel tasks environments (see Section 5). The methodology is divided into the following parts: the tasks levels identification and association, the MW assessment, and the scenarios design. Two studies based on a VR flight simulator were performed to test the method. The first study results support that the designed states (i.e., tasks configurations), which are based on the combination of multiple tasks with varying levels of difficulty, were able to induce different levels of MW. This is consistent with previous findings [55, 113, 159, 304]. They also support that using two tasks stimulating the same pools of cognitive resources will further increase MW compared to two tasks stimulating different pools of cognitive resources, which goes in the same line as Wickens's Multiple Resources Theory [304]. Both subjective and task performance measures were influenced by the states, but the subjective MW measures appeared

to be more accurate, as outlined in [49, 51, 310]. From the subjective MW measures gathered and the designed states, 3 training scenarios inducing different progressions of MW through time were generated in a second study. Those were able to reproduce the same MW profiles than in the first experiment. However, some users did not perceive the entire scenario difficulty progression after the experiment as accurately as their subjective MW was predicted over time, which can be explained by the difference of delays in rating [306]. Overall, the approach successfully designed training scenarios which were able to induce the expected MW through time. This supports the introduction of MW assessment in training scenarios, as expressed by other studies [51, 55, 159], and encourages the use of MW in the design of VR training scenarios to make them more fitted to users' cognitive resources in a controlled, reproducible, and safe environment. On a side note, in the two presented studies, most users reported having enjoyed the training and being engaged throughout the experiment, which encourages the use of VR in the training field and is in line with previous studies [21, 30, 38, 55, 112, 200, 233, 300].

In the second study, the 3 scenarios were designed based on the mean subjective MW assessed in the first experiment and considering realistic training purposes. However, the scenarios could also have been constrained by other criteria such as task performance measures, the presence and the absence of one task, or the probability to induce a subjective MW rating level for example. Given the designed states with just subjective and task performance measures, there are already numerous ways to generate scenarios. As such, if the first scenario ($ISA \approx 3$ during the whole scenario) appears to be too easy or too difficult for some outliers trainees, a similar design process can easily be replicated to generate a new scenario which induces a higher or a lower MW all long. Also, we can fully imagine combining our approach to other tools developed to simplify the design of scenarios in VEs which focus more on narrative and interaction requirements, by introducing more complex mechanisms, labels and constraints for example (see Section 5).

Concerning the limits of the approach, the assessment part can be time-consuming depending on the complexity and the number of tasks, even if it helps accelerating the design of a wide variety of training scenarios afterward. Some simplification can be made by assessing the MW only in strategical nodes, and by inferring the others (see Section 3). However, this will draw to less accuracy in the prediction of MW throughout the scenarios. Also, in the first study, the transitions were only set between consecutive nodes, which was not the case in the second study. It did not have an effect on the MW measures. Yet, the transition effects might have to be taken into account in some contexts. Finally, the self-report request (ISA, see Section 2.3) could be considered as a fourth task as users were required to answer each time they saw the screen appearing. This probably similarly increased their MW for each state as tests did not show an effect of the tasks levels on the ISA response time. While this approach was

successful in inducing the expected MW progression overtime, Figure 45 clearly depict individual differences as the subjective MW data distributions follow a normal law in each state. A further step would be to extend this methodology to the adaptation of VR training scenarios in real-time based on users' estimated MWs.

4 CONCLUSION

This chapter proposed an approach to introduce MW assessment in the design of complex training scenarios involving multitasking in VR. First, tasks levels are identified and associated. Then, MW is assessed inside each task configuration. Finally, the training scenarios can be designed based on the MW measures. This approach allows the generation of different training scenarios based on the progression of MW over time using different task levels combinations.

Two studies based on a VR flight simulator were performed to test the approach. Taken together, the results support the idea that the approach was successful into designing training scenarios which induced the expected progression of MW through time. These results pave the way to further studies exploring how MW modulation can be used to improve VR training applications.

TOWARDS REAL-TIME RECOGNITION OF USERS' MENTAL WORKLOADS USING INTEGRATED PHYSIOLOGICAL SENSORS INTO A VR HMD

SUMMARY This chapter describes an “all-in-one” solution for the real-time recognition of users’ MWs in VR through the customization of a commercial HMD with physiological sensors. First, we describe the hardware and software solution employed to build the system. Second, we detail the machine learning methods used for the automatic recognition of the users’ MWs, which are based on the well-known Random Forest (RF) algorithm. In order to gather data to train the system, we conducted an extensive user study with 75 participants using a VR flight simulator to induce different levels of MW. In contrast to previous works which label the data based on a standardized task (e.g., *N*-back task) or on a pre-defined task-difficulty, participants were asked about their perceived MW level along the experiment. With the data collected, we were able to train the system in order to classify four different levels of MW with an accuracy up to 65%. In addition, we evaluated the approach using HMD sensors in comparison to Commercial Grade Systems (CGS) sensors and propose an original evaluation based on classification performances. We also discuss the role of the signal normalization procedures, and compare the contribution of the different physiological signals on the recognition accuracy.

TABLE OF CONTENT

1	All-in-one Approach to Assess Users’ MWs in VR in Real-Time	101
1.1	Integrated Hardware	101
1.2	Data Processing and MW Assessment	103
2	Data Acquisition Protocol	107
2.1	Tasks Design	108
2.2	Apparatus	109
2.3	Collected Data	110
2.4	Experimental Design	111
2.5	Participants	111
2.6	Experimental Procedure	112
2.7	Resulting Data	112
3	Results	113
3.1	Normalization of Physiological Features	113
3.2	CGS Sensors	114

3.3	Sensors Integrated Into the HMD	114
3.4	Mental Workload Recognition in Real-Time	117
4	Discussion	118
5	Conclusion	121

1 OUR "ALL-IN-ONE" APPROACH TO ASSESS USERS' MENTAL WORKLOADS IN VR IN REAL-TIME

In this section, we present our all-in-one solution to assess MW in VR in real-time. First, the hardware components concerning the sensor integration in the HMD will be addressed. Next, the software components presenting the solution proposed for the data synchronization and the real-time recognition pipeline will be presented.

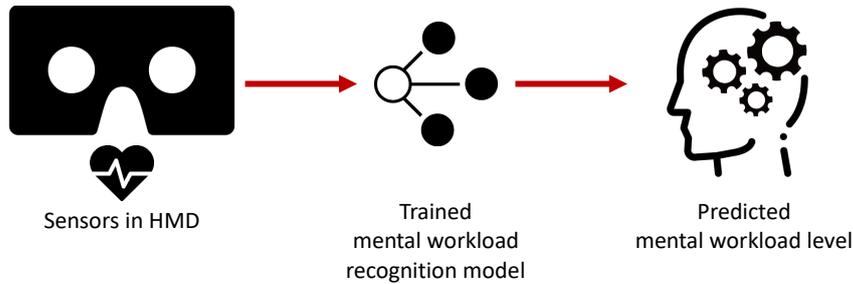


Figure 49: Illustration of the all-in-one approach to do MW recognition in VR in real-time using physiological sensors integrated into a VR HMD. The MW recognition model was trained using the HMD sensors signal features labelled with users' self-reported MW level as input and the random forest algorithm.

1.1 Integrated Hardware

Several physiological dimensions relevant to MW measurement were mentioned in Section 3. We chose to focus on cardiac activity, EDA, and oculomotor activity as they were shown to be influenced by MW [150], are non-intrusive, and can easily be positioned on and inside VR HMDs. The main efforts focused on the integration of the cardiac and EDA sensors into the VR HMD.

The cardiac activity was monitored via a PPG sensor: the Maxim MAX-30102, which was fixed on a small clip to assess data on one of the user's earlobes (see Fig. 50). This location was chosen following the recommendation of the literature for PPG sensors, as blood vessels are close to the surface of the skin and light can readily be detected [286, 315].

For the EDA, pairs of electrodes were made out of a flexible printed circuit board. The latter was chosen based on previous works (e.g., [27]) and for its various qualities. It is robust enough to weld electrical wires on it to make the connection to the electronic card. As for the conductive material, gold was chosen as it is stainless and biocompatible. The prefrontal area has been found to be relevant in order to measure EDA [86, 315], so we chose to place one pair of electrodes on the foam in contact with the forehead part of the headset (see Fig. 50). The sizes of the electrodes were chosen to be thin (i.e., 100 μm) to not mark the skin, and large enough to palliate the reduced presence of sweat glands in the prefrontal area.

Those were spaced a few inches apart to let a weak current flow through the skin, and the EDA was given as the difference in potential between the 2 electrodes.

All these sensors were plug in a custom-made electronic card (see Fig. 50), based on SOM Variscite i.MX8M Mini, which was powered by a 5V power bank. The later was placed in a designed 3D printed case, which was positioned in the back of the HMD using the vertical strap of the headset (see Fig. 50). The electronic card main features are that: (i) it can collect physiological data from multiple sensors, (ii) it enables the aggregation and time stamping of all samples, (iii) it has the capacity to process AI models and algorithms (not currently considered), and (iv) it can transmit the data to a computer either by a wired medium, using ethernet, or in a wireless way, via wifi connection. Additional slots were available on the electronic card to plug in more sensors if necessary.

As for the ocular activity assessment, several HMDs already integrate eye-tracking solutions (e.g., HTC Vive Pro Eye, FOVE [149], HP Reverb G2 Omnicept). Pupil Labs also offers add-on solutions for virtual and augmented reality headsets. We chose to use the Vive Pro Eye HMD.

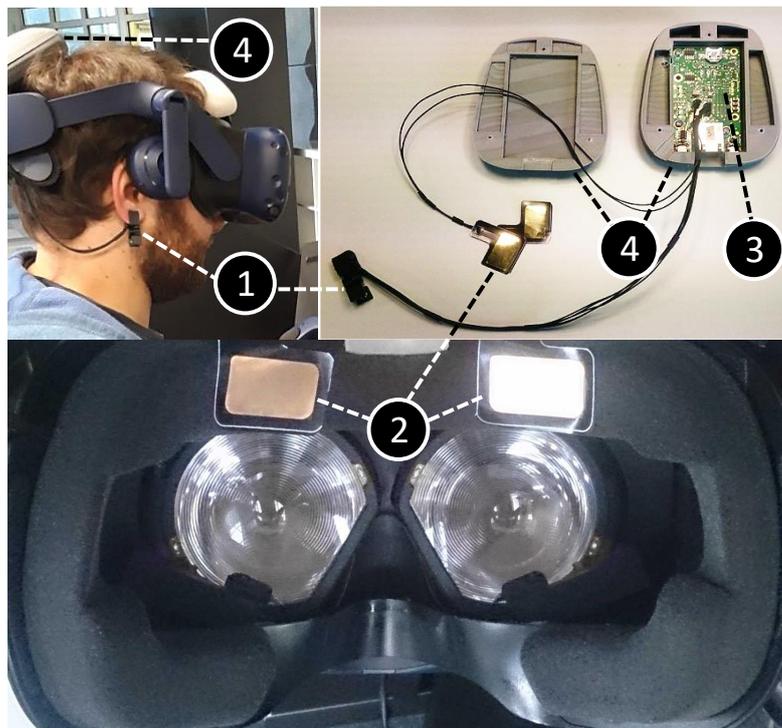


Figure 50: Hardware solution. The sensors are placed on a Vive Pro Eye HMD, which has built-in eye-tracking. (1) PPG sensor, (2), electrodes to assess the EDA, (3) electronic card, (4) 3D printed case.

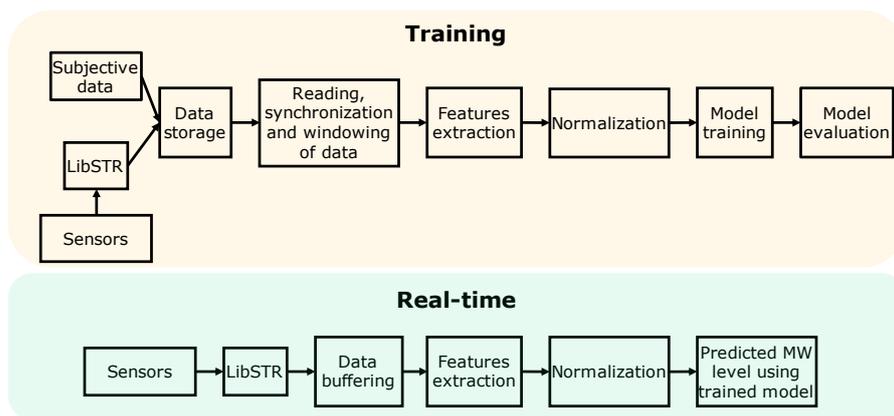


Figure 51: Software processing chain for the real-time recognition of users' MW. There are 2 main steps: the training of the recognition model and the real-time use of the model.

1.2 Data Processing and MW Assessment

The software part of the proposed approach aims at recognizing users' MW levels in real-time using multiple sources of measurements. It is composed of several steps depicted in the Figure 51 which can be divided in 2 parts: the training phase and the real-time phase.

1.2.1 Data collection

The recording of data coming from various sensors may be challenging especially due to issues like time synchronization, or data format. Moreover, sensors usually come with their own software, and handling all of them to record participants' signals can be tedious. For these reasons, a middleware, called LibSTR, has been developed for the collection of physiological and behavioural data. It enables to collect, synchronize and distribute the data coming from various sensors in real-time by abstracting the capture of data. It is composed of three components: hub, sensors wrappers and listeners. The hub collects and synchronizes the data from the sensors wrappers, and exposes them to the listeners. Developed in C and without software dependency, it is compatible with multiple platforms.

1.2.2 Data windowing

Supervised learning model training requires labelled data [203] (i.e., the labels correspond to the subjective measures of the user's state and the baseline in our case). For this purpose, a fixed-size window of the collected data (e.g., blood volume pulse (BVP), performance) are extracted before each label (see Fig. 52). Based on previous studies [272], a window size of 30 seconds was selected as it seems to be an appropriate compromise between performance and real-time use. Thus, for each label, the 30

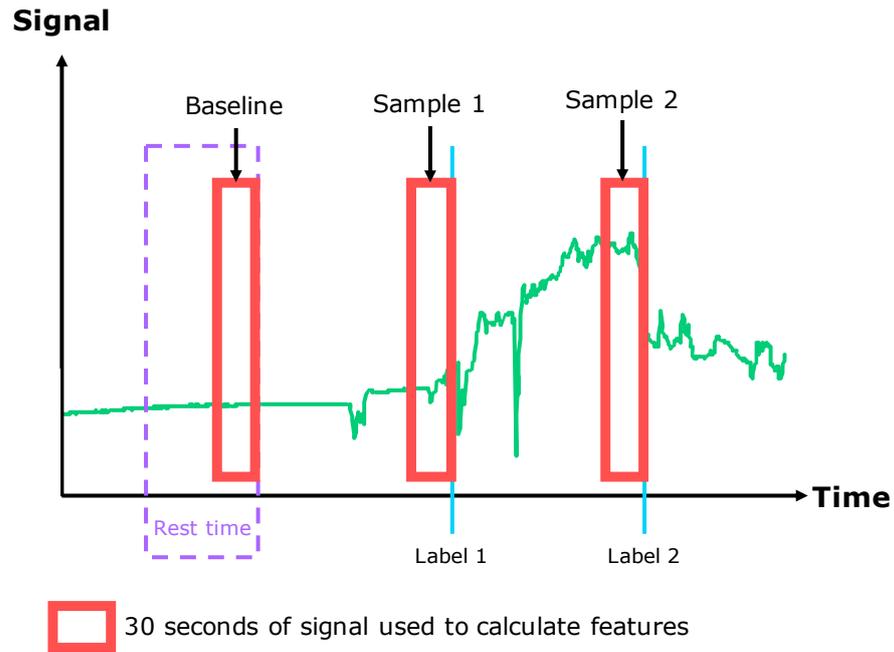


Figure 52: Illustration of a data windowing in the case of a physiological signal (in green). The time window (in red) is used to extract the physiological features based on given labels.

seconds of data preceding the label timestamp are used to calculate the said-label dataset features [98] (see Fig. 52).

1.2.3 Features extraction

The exploitation of physiological signals requires the extraction of specific physiological features [7]. Based on the window of signals described above, common features from the relevant state of the art were used. It should be noted that performance measures depend on the context of the application and on the tasks users have to perform. A summary of all the features extracted using our setup can be found in Table. 1.2.3. The windowing of data as well as the extraction of features was implemented in Python using numpy [208], pandas [319] and scipy [299].

Cardiac activity. Six time-domain and 5 frequency-domain features related to the HRV were extracted [152]. These features are based on the IBI. For this purpose, a bandpass filter was first applied (cutoff frequency = [0.66; 3.33] Hz, order = 3). It enables the reduction of noise (e.g., motions artefacts) [7] and the restriction of the cardiac activity assessment between 40 and 200 beats per minute. Then, the peaks on the BVP signal were detected using a threshold (arbitrarily set) and an estimation of local minima/maxima [291]. This enabled the estimation of IBIs to calculate related features (see Table 1.2.3).

Electrodermal activity. The EDA signal is composed of 2 components: the phasic part and the tonic part [35]. The phasic part (i.e., SCL) corres-

ponds to slow changes in the EDA while the tonic part (i.e., *SCR*) corresponds to the rapid physiological responses to a stimulus. The extraction of those two components from the raw signal is composed of several steps. First, a low-pass filter (cutoff frequency = 1Hz, order = 3) is applied to reduce noise in the signal [35]. Second, a low-pass filter (cutoff frequency = 0.05Hz, order = 3) is applied on previously filtered signals to extract the tonic part of the EDA [41]. Lastly, the phasic part is obtained by subtracting the tonic signal to the filtered signal. For the phasic part, 9 features were extracted based on the estimated EDA peaks¹. For the tonic part, 29 features were calculated (see Table 1.2.3). Inspired by research from other domains, some are computed based on the shape of the signal [171], others from the data-driven signal decomposition (i.e., Empirical Mode Decomposition - *EMD*) [117] and finally, some are EDA components from the frequency-domain [252, 266].

Ocular activity. 17 features were extracted from the pupil diameter, the dynamic of pupil diameter, and the dilatation and constriction of the pupil (see Table 1.2.3). The signal was cleaned by taking into account the data only when the pupil was detected.

1.2.4 Normalization of physiological features

Considering the inter-individual variability is a key point in research dealing with physiological data [23, 98, 204], several methods have been proposed in the literature to reduce its influence on the recognition accuracy and to make data comparable across participants. One of the most common approaches is to collect data during a rest time (i.e., the baseline) and to subtract the mean value of the data collected during this rest time from the whole signal [41]. This approach can be effective on non-periodic signals such as the EDA signal, but is not compatible with periodic signals (e.g., BVP²). As such, the methods of normalization at feature level seem interesting (e.g., subtraction of feature values during rest time from other feature values) as they can be applied to all signals. However, as no *de facto* normalization method exists, the most common approaches used in the literature will be evaluated.

1.2.5 Model training using Machine Learning

Models are trained, using supervised ML algorithms, to recognize MW based on the extracted physiological features, task performance measures, and related subjective responses (i.e., the subjective measures of users' state). In this way, the function between the input data (e.g., extracted

1 An EDA peak is characterized by the amplitude (the height of the peak) and the recovery time (time to return to the level of EDA before the peak) [35, 271].

2 Contrarily to the EDA signal, the BVP signal is periodic. Therefore, subtracting the mean value of the BVP signal collected during the baseline to the signal will only bring the BVP signal to the same amplitude level (roughly the same mean). However, the features related to the BVP signal are time-based, which makes such normalization not relevant.

Target Variables		
<i>Subjective MW level</i>	4	1, 2, 3, 4
Predictor Variables		
<i>Cardiac activity features</i>	11	HR Average of NN intervals Standard deviation of NN intervals (SDNN) Root mean square of successive differences between NN intervals (RMSQ) Number of interval differences of successive NN intervals greater than 50 ms Percentage of interval differences of successive NN intervals greater than 50 ms Very low frequency (0.003 to 0.004 Hz) Low frequency (0.04 to 0.15 Hz) High frequency (0.15 to 0.4 Hz) Ratio of low-frequency and high-frequency Total spectral power
<i>Tonic EDA features</i>	29	Max, range, inter-quartile range, root mean square error, mean, SD, skewness and kurtosis of the signal Mean absolute value of 1st differences and mean absolute value of 2nd differences of the signal Mean absolute value of the 1st differences and mean absolute value of the 2nd differences of the standardized signal Mean, SD, min and max of 3 Intrinsic Mode Functions Very low frequency (0 to 0.1 Hz) Low frequency (0.1 to 0.2 Hz) Middle frequency (0.2 to 0.3 Hz) High frequency (0.3 to 0.4 Hz) Very high frequency (0.4 to 0.5 Hz)
<i>Phasic EDA features</i>	9	Number of peaks Mean, SD, min and max of peak amplitude Mean, SD, min and max of half of recovery time of peaks
<i>Ocular activity features</i>	17	Number of constrictions Min, max, range, mean and SD on pupil diameter Min, max, range, mean and SD of the pupil amplitude Min, max, mean and SD of pupil constriction and dilation speed
<i>Task performance features</i>	5	Min, max, range, mean and SD on distance to the centre of the circle

Table 11: Target variables and extracted physiological and performance features used for the training of the recognition models.

physiological and performance features) and output data (i.e., subjective responses) is automatically inferred [203]. The objective is to enable the detection of users' subjective MW levels using only objective measures (i.e., physiological responses and/or task performance) without requesting any evaluation from the users, which could disrupt their experiences [236]. Random Forest (RF) [42] was selected as it presented the best performance in similar contexts (e.g., [65, 91]). The number of trees as well as the number of randomly selected predictors at each cut in the tree were tuned during the training. As the evaluation of feature selection (i.e., Principal Component Analysis (PCA)) showed no improvement of the recognition accuracy and as the tree-based models are generally robust against unhelpful features, all the features per sensor were considered. All trainings were realized using R [238] and the caret library [153].

1.2.6 Real-time recognition

The real-time pipeline adopts a similar processing chain as the training part although it includes some adaptations.

Data buffering. As the data is progressively captured, it is necessary to store it in a buffer. Indeed, 30 seconds windows of data are required to calculate the features. The buffer starts empty and is progressively filled with available data. When the 30 seconds of data are reached, the data are fed to the rest of the processing chain. Then, whenever new data is available, it is added to the buffer by pushing the oldest data at a 1-second step.

Normalization. The recording of data during a baseline is required to normalize the calculated features, regardless of the method used. The normalization should be done in the same way as in the training part.

Real-time prediction. The predicted MW level and related physiological signals are displayed in an interface (see Fig. 53). Wrappers were written to communicate the estimated MW level output from Python to other environments, such as C# for Unity3D VR environments.

In order to have a unified processing chain, the best ML configuration was implemented in Python using scikit-learn [223] for real-time purpose.

2 DATA ACQUISITION PROTOCOL

An experiment was conducted to test the viability of our solution to classify MW levels. The objective was to create a dataset to train the recognition model. Three major constraints were enforced. First, a real-life task which required the users to perform different sub-tasks. Second, the possibility to modulate the task difficulty ensuring that a wide range of MW levels could be induced. Third, enforce a good balance between the different levels of MW induced in order to ensure an optimal classification model. The VR application which was previously developed to support the methodological approach contribution (see Chapter 3) meets these requirements. In ad-

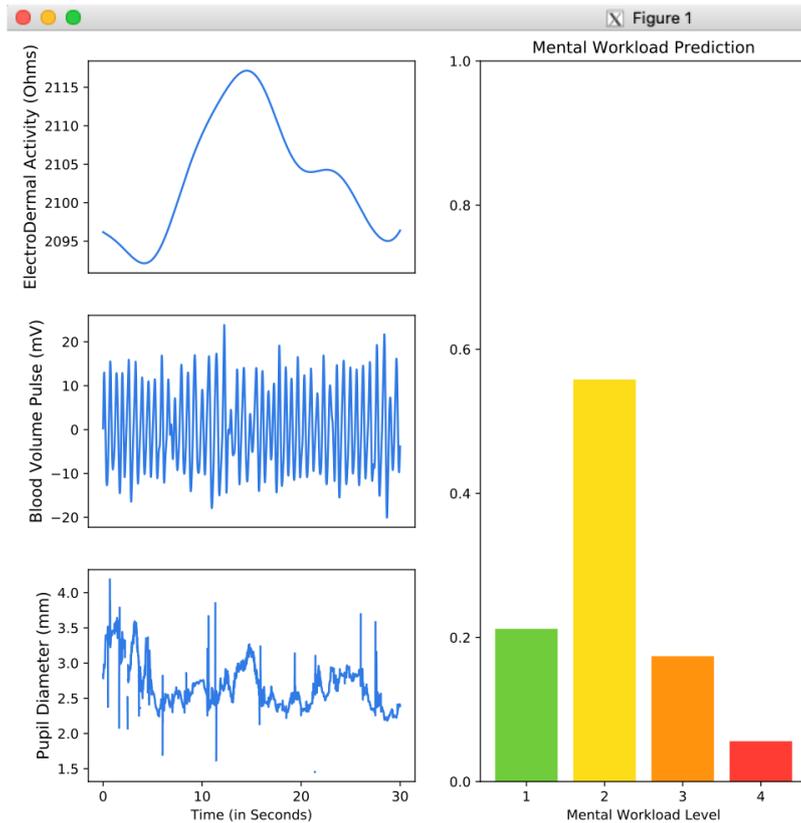


Figure 53: Interface for the MW recognition in real-time. It depicts on the left: the EDA, BVP, and pupil diameter signals evolution over time in a time window of 30 seconds; on the right: the predicted MW level (i.e., the probability that a user is at a particular MW level) using features extracted from the 30 seconds of data. The prediction and interface are updated every 1 second.

dition, we compared our solution to CGS sensors, in regards to the MW classification performances.

2.1 Tasks Design

The illustrative application which was used to demonstrate our approach to introduce MW measures in the design VR training scenario is based on the [MATB-II](#). The MATB-II was originally developed to study human performances in a multi-task context [253]. It has been used extensively to study MW and to train users in situations where they might be overloaded. This makes it a very suitable tool to induce different MW levels to users. All tasks and tasks configurations are presented in Section 2. To summarize, three tasks of the MATB-II [253] were adapted in VR to induce different levels of subjective MW: the tracking task (into a piloting task), the communication task, and the resources management task.

More details about the tasks instructions can be found in Section 2 and [253].

All these tasks levels could be associated to form 12 tasks configurations (3 piloting levels \times 2 communication levels \times 2 resources management levels), which formed the states of a state machine (see Section 2.2). The states were labelled as following. The first, second, and third digits represent respectively the piloting task difficulty level (0-easy, 1-medium, 2-hard), the radio task level (0-deactivated, 1-activated), and the resources management task level (0-deactivated, 1-activated). For instance, in the state "110", the difficulty of the piloting task is set to "1-medium", the radio task is "1-activated", and the resources management task is "0-deactivated". Each state was set to last 110s in average.

2.2 Apparatus

The participants were installed on a cockpit, which was assembled for the experiment (see Fig. 54). The virtual cockpit was modelled in 3D and calibrated so it matched the real one in position and size.

Users were equipped with the customized Vive Pro Eye (see Section. 1.1) and the Shimmer3 GSR+ [48] sensors (i.e., the CGS sensors) (see Fig. 54). The Shimmer wristband was disposed on users' left wrist. Its EDA sensors were placed on the middle phalanx of the users' left ring and middle fingers, and the PPG sensor ear clip, on users' left earlobe. The data were collected using our libSTR middleware (see Section 1.2.1) with homogeneous timestamp formats at different frequencies (Vive Pro Eye-tracker: 250Hz, HMD-PPG: 100Hz, HMD-EDA: 50Hz, Shimmer sensors: 60Hz).

Users could pilot the virtual aircraft using the Logitech 3M X52 Pro joystick with their right hand. They were also equipped with a Vive Controller with their left hand to interact with the virtual interfaces. In the VE, both hands were represented by transparent virtual hands. The right hand was placed on the virtual joystick (see Fig. 38), which moved when the user was interacting with the real one, and the left hand was tracked with the Vive Controller. The users did not need to use any of the buttons on the Vive Controller or on the joystick to perform the tasks. The interactive virtual objects were highlighted when the participants advanced their virtual hands in their direction, and the animation turned into a pointing index upon approaching. Haptic pulse feedbacks on the Vive controller informed the participants that their action had been carried out. Audio instructions were provided using the audio headset supplied with the Vive Pro Eye Headset during the flight simulation.

The support application was developed in Unity 3D, and run with the recording on a computer equipped with an Intel(R) Core(TM) i9-7900X CPU, a Nvidia Titan V graphic card, and 16 GB Random-Access Memory.



Figure 54: Multiple views of the experimental setup. The user is wearing a Vive Pro Eye equipped with sensors on the headset, a Shimmer3 GSR+ on his left hand, and using a joystick and a Vive Controller on the cockpit.

2.3 Collected Data

The assessed data were: self-report (i.e., the subjective MW reported by participants), physiological, and tasks performance measures.

2.3.1 Self-Report

As in Chapter 3, we wanted to assess users' MWs while they were performing other tasks. Therefore, the focus was set on the ISA [287], which rates the MW level using 5 different ratings (1-underutilized, 2-relaxed, 3-comfortable, 4-high, 5-excessive) and has been especially used during air traffic control tasks [287]. The meaning of each rating of the ISA is described in [140] and was explained to each participant before the experiment. They were asked to report the MW level they experienced during the last 30s, when a screen appeared in front of them with the 5 buttons corresponding to the ISA ratings (see Fig. 38). First, they were asked to push the button corresponding to their MW level, and then, to click on the validate button on the same screen to make it disappear.

2.3.2 Physiological Measures

Multiple types of physiological signals were assessed during the experiment: ocular activity via the eye-tracking cameras present in the HMD, and cardiac activity and EDA via our hardware solution integrating sensors into the HMD (see Section 1.1) and via the Shimmer3 GSR+ (i.e., the CGS sensors). The features extracted from these sensors are depicted in Table 1.2.3.

2.3.3 Task Performance Measures

Some performance indicators of all 3 tasks were assessed throughout the experiment. However, among the 3 tasks used in the experiment, only one was always present across the different states: the piloting task. Thus, only this measure is considered as a performance measure.

Users were given indications on how to align the aircraft with the green line optimally before the experiment, and the distance to each circle centre when they passed it was recorded throughout the experiment.

2.4 Experimental Design

In a previous study [172] (see Section 2.5), 38 participants traversed the 12 states in a row in a randomized order and reported their subjective MW level using the ISA scale [287], 3 times per state. Since unbalanced datasets can lead to poor recognition performance for minority classes [85], the experimental protocol was designed to induce the highest number of different subjective MW levels in the most balanced way possible.

The constraints were the following: a state could not appear more than 3 times in total, and the total duration of the experiment was set not to last more than 25 min, conducting to the choice of a number of 10 states. Following these constraints and based on the ISA responses reported during the previous study [172], the best subjective MW level distribution was given for the following 10 states: “000 – 000 – 000 – 100 – 111 – 111 – 201 – 211 – 211 – 211”. Those were selected for the new experimental study, which results and analysis are depicted and discussed in this chapter.

2.5 Participants

77 healthy participants, who were completely naive to the experiment, were recruited through an external cabinet. They were paid 30€ for their participation to the study. Two of the users were excluded from the study due to motion sickness, resulting in a final sample of 75 participants with ages ranging from 18 to 64 (38 females, 37 males; $M = 38.69$, $SD = 13.54$).

There were some inclusion criteria: the participants had to be fluent in french. They should not have taken any medication that could influence their physiological responses. They were also asked not to consume coffee and/or tea in the 2 hours preceding the experiment. Moreover, variables such as their experience in VR, games, flight simulator, aircraft piloting, vision state, and dominant hand were controlled. One participant reported having a great experience in VR, 4 reported having few experiences in VR, and all others (i.e., 70 participants), none. As for the gaming experience, 64% of the participants were novice, 20% played occasionally, and 16% regularly. All of them reported having no experience in flight simulator and aircraft piloting.

In accordance with ethical principles, participants were required to complete an informed consent form, advising them of their right to withdraw at any time from the study, of the preservation of their anonymity and about the potential side effects of VR.

2.6 *Experimental Procedure*

The experiment lasted around 1h and was subdivided into the following steps:

Written Consent and Instructions: Users completed a consent form, prior to the experiment. They were then instructed with the nature of the experiment, the equipment used, the data recorded (which was anonymized), and the tasks instructions. Participants were also asked to fill a questionnaire (experience with VR, video games, and piloting an aircraft, dominant hand, level of alertness, state of vision, demographic information, simulator sickness questionnaire (SSQ) [135]) to gather information about their background and their state before the start of the experiment.

Training: Users were then equipped with the Shimmer sensors, a Vive Controller and the Vive Pro Eye HMD with the sensors. The eye-tracker of the Headset was first calibrated following the instructions given in the headset. They were then immersed in the virtual cockpit environment. Once they got used to the VE, they were asked to breath normally and to remain still for 1 minute, to record their physiological signals in a neutral state (i.e., for the physiological baseline). Users were then asked to interact with the buttons of the tasks interfaces to familiarized themselves with the interactions. Then, they travelled the states following this path: “000 – 010 – 011 – 001 – 101 – 201 – 211”, which gave them a good overview of each task and their levels. Users were then invited to ask any question they may have had.

Experiment: In the experiment part, users were first asked to do a 1-min baseline again. Then, they travelled the 10 states mentioned in Section. 2.4 (i.e., “000 – 000 – 000 – 100 – 111 – 111 – 201 – 211 – 211 – 211”) in a pseudo-randomized order (two identical states could not appear twice in a row). The states were set to last 110 seconds with 4 communication calls and 3 ISA calls (spaced in time of 30s).

Debriefing: At the end of the experiment, they were asked to fill the SSQ again, debriefed and invited to ask questions.

2.7 *Resulting Data*

3265 subjective responses were collected. As unbalanced datasets can lead to poor recognition performance for minority classes [85], we chose to re-

tain 4 classes of subjective MW level, using the following data distribution (see Table. 12):

MW level	1	2	3	4
ISA value	1	2	3	4 & 5
N	624	822	1143	676
Perc.	19 %	25 %	35 %	21 %

Table 12: Contingency table of subjective responses.

The results for other data splittings (2 classes, 3 classes, and 5 classes) based on the reported ISA levels are presented in Appendix A.

3 RESULTS

This section presents an overview of the classification accuracy of the proposed system, considering the 4 levels of MW chosen (see Table 12). Table 13 presents the classification results using the EDA and PPG from the CGS sensors, and Table 14 presents the classification results using the EDA and PPG from our custom sensors integrated into the VR HMD (see Section 1.1). Ocular and task performance measures are the same in these 2 groups. The normalization method, as well as the different combinations of sensors using either the Shimmer sensors (i.e., CGS sensors) or our setup (i.e., HMD sensors) are provided. In order to facilitate the interpretation of results, the performance of a naive model was calculated (i.e., model always predicting the most represented class in the training dataset).

Classification accuracy was computed using a 10-fold cross-validation method [203]. To reduce the potential problems linked to random splitting with cross-validation, 10 independent 10-fold cross-validations were performed for each trained model.

3.1 Normalization of Physiological Features

As previously explained, the physiological signals are subject to inter-individual variability, which can affect recognition performance. Thus, three types of normalization approaches have been evaluated in the context of MW detection:

- **Normalization by subtraction of features (NSF):** For each participant, subtracting the features estimated for each subjective response with the features estimated during the baseline [62, 137, 278];
- **Normalization by adding features (NAF):** For each participant, adding the features estimated during the baseline to the dataset (i.e., the feature space) [44, 226].

- **Normalization by Min-Max:** For each participant, each feature is normalized using the following formula:

$$\frac{X - \min}{\max - \min}$$

Data about participants were looped over to find the first occurrence of their highest ISA score and the first occurrence about their minimum ISA score. Their features values were respectively associated with the max and min values in the formula.

3.2 CGS Sensors

The recognition performances based on data collected with the Shimmer sensors are presented in Table 13. According to the results of the training, the best accuracy using the CGS sensors is 64.2%. It is obtained with a combination of performance and all physiological signals, using either NSF or NAF normalization.

Taking sensors and task performance measures individually, the best classification performance is achieved either with the ocular activity (in the cases of no-normalization and NAF normalization) or the EDA (in the cases of NSF and min-max normalizations), followed by the cardiac activity, and the task performance measures.

There is a mean drop in accuracy of only 1.33% when considering only the EDA and ocular signals instead of all physiological signals and performance measure, using any normalization method. A confusion matrix using all data and the NSF normalization is given in Fig. 55.

3.3 Sensors Integrated Into the HMD

The recognition performances based on data collected with the sensors integrated into the HMD are presented in Table 14. Similarly to results with CGS sensors, the best accuracy is 65%. It was obtained when all data are considered, and with NSF normalization.

Taking sensors and task performance measures individually, the ocular activity lead to the best classification performance, followed by the EDA, the cardiac activity, then the task performance measure.

Similarly to results with CGS sensors, the combination of EDA and ocular signals lead to a mean drop of only 2.1% compared to the best configuration requiring more sensors.

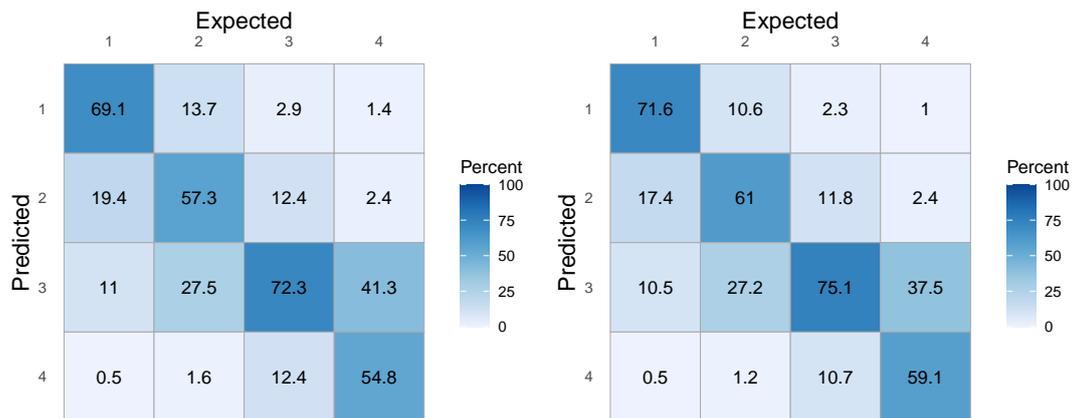
A confusion matrix using all data and the NSF normalization is given in Fig. 55.

Normalization	Input								
	Perf	Ocular	Shimmer Sensors (CGS)						
			Cardiac	EDA	Cardiac + EDA	Cardiac + Ocular	EDA + Ocular	All physio	Physio + Perf
<i>Naive model</i>	19.1								
No Normalization	41.4	44.4	36.7	41.1	44.8	50.5	49.9	52.5	57.6
NSF	41.5	52.3	42.9	54.2	56.4	57.3	62.2	62.5	64.2
NAF	45.8	60.5	50.8	56.9	56.3	61.1	62.1	62.3	64.2
Min-Max	45.3	53.6	45.8	56.6	58.2	57.8	63.3	62.5	63.2

Table 13: MW classification accuracy results (in %) using the Shimmer sensors (i.e., CGS sensors) in function of the normalization methods and of the type of measure. “Perf” corresponds to the task performance. Task performances and ocular activity are common to the CGS and HMD sensors.

Normalization	Input								
	Perf	Ocular	HMD Sensors						
			Cardiac	EDA	Cardiac + EDA	Cardiac + Ocular	EDA + Ocular	All physio	Physio + Perf
<i>Naive model</i>	19.1								
No Normalization	41.4	44.4	36.5	44.1	49.7	51.0	53.6	57.7	61.2
NSF	41.5	52.3	44.9	50.7	54.5	59.1	62.0	62.8	65.0
NAF	45.8	60.5	49.9	52.9	56.3	60.3	62.6	62.1	63.8
Min-Max	45.3	53.6	45.6	51.6	54.5	56.9	61.3	61.8	63.4

Table 14: MW classification accuracy results (in %) using the sensors integrated into the VR HMD (i.e., HMD sensors) in function of the normalization methods and of the type of measure. “Perf” corresponds to the task performance. Task performances and ocular activity are common to the CGS and HMD sensors.



(a) CGS Sensors - Accuracy = 64.2%

(b) HMD Sensors - Accuracy = 65%

Figure 55: Confusion matrix (expected and predicted classes) in the “All physio+perf” using NSF normalization setup for (a) the “CGS sensors” and (b) the “HMD sensors”.

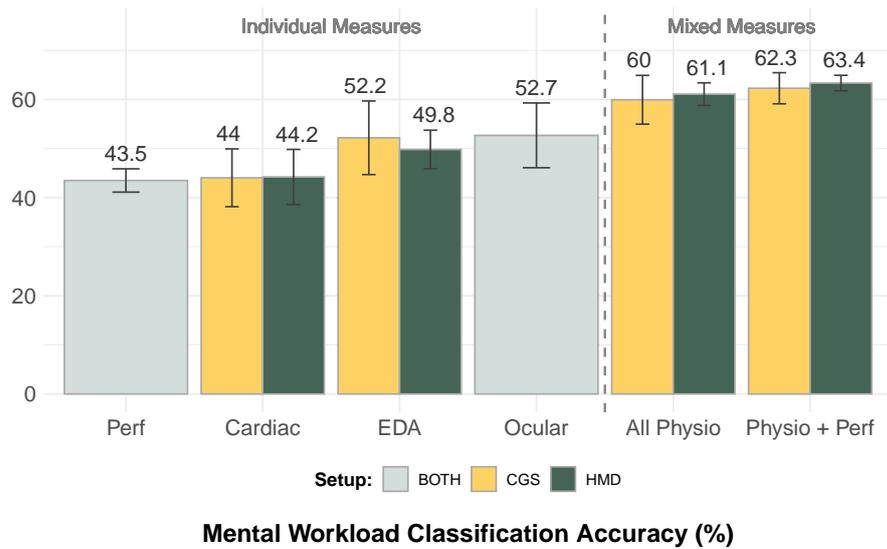


Figure 56: Comparison of the “CGS sensors” and “HMD sensors” setups on the classification accuracy in the different measurement setups. Mean values are indicated and the error-bars correspond to the standard deviations considering the different normalization methods.

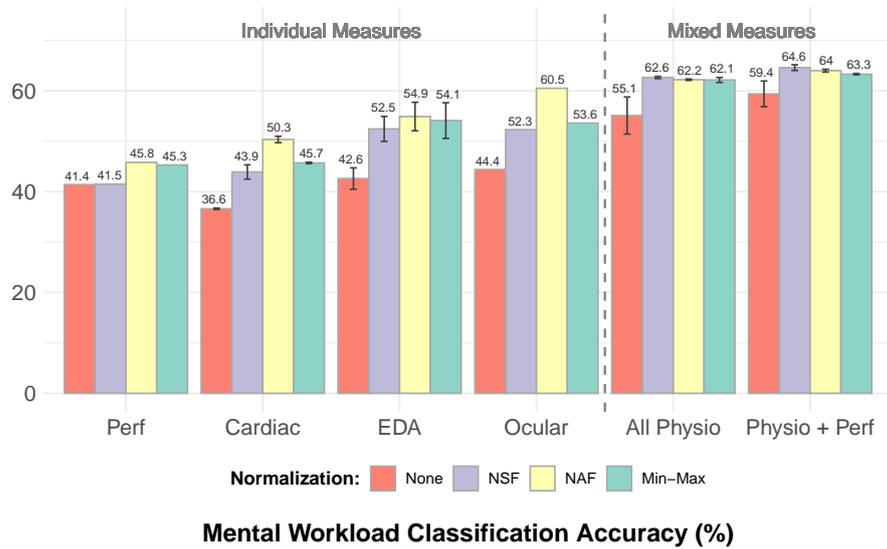


Figure 57: Comparison of the different normalization methods on the classification accuracy in the different measurement setups. Mean values are indicated and the error-bars correspond to the standard deviations considering the HMD and CGS sensors setups.

3.4 Mental Workload Recognition in Real-Time

The recognition in real-time was simulated using the collected data and the configuration offering the best performance (i.e., input: all physiological sensors and performance, NSF normalization, HMD sensors; see Tables 13 & 14).

To simulate real-time, a sliding window of 30s with 1s step was applied to the physiological data. Then, these windows of data were fed every second to the recognition processing chain until the last available data (see Fig. 50).

For illustrative purposes, the real-time was simulated on one participant (the stages of the processing chain are identical between participants). The predicted MW as well as the ground truth (i.e., true MW level, see Table 12) are depicted in Fig. 58. As subjective responses are only provided punctually, the MW level was linearly interpolated between the different subjective MW responses in order to have a continuous ground truth. A moving average (on 10s) was also used to smooth the prediction and limit the effect of misclassified MW level. The prediction latency did not exceed 0.20s.

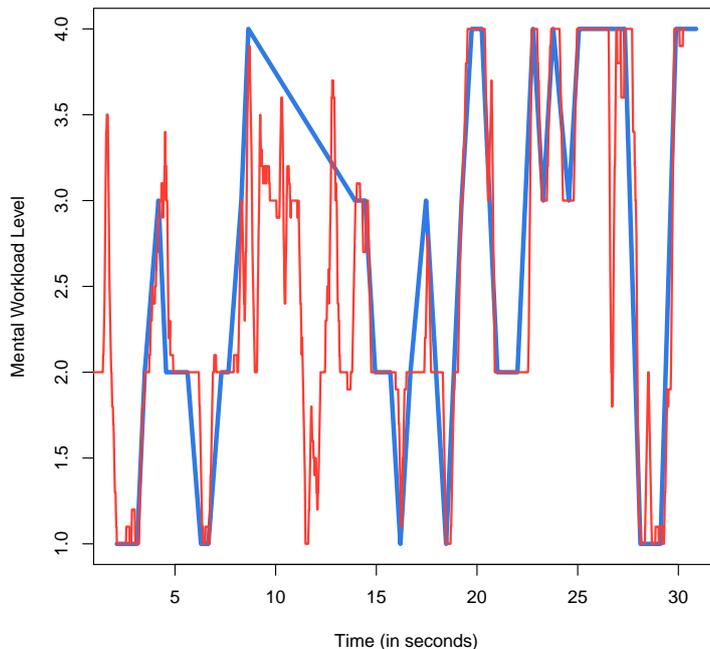


Figure 58: Simulation of the real-time prediction of MW using the data of one of the participants and the HMD sensors in the “All Physiological signals + task performance” setup with NSF normalization. The blue line represents the ground truth (i.e., the MW level reported by the participant); the red line represent the predicted MW using our trained recognition model.

4 DISCUSSION

In this chapter, the recognition of MW using physiological signals and task performance data was explored. Based on data from 75 participants, the trained models were able to classify 4 levels of MW with an accuracy up to 65%. Yet, it should be noted that the misclassified levels are mostly contained in the adjacent classes, as shown in Figure 55. As such, the classification accuracy reaches 95.5% using NSF normalization when considering the classes which are adjacent to the predicted mental workload level (see Fig. 55). This highlights the good performance of our approach. Moreover, the recognition accuracy considering CGS sensors and sensors integrated in the HMD were compared. The normalization method as well as each sensor and combination of sensors were also tested in regards to the classification accuracy.

Unlike previous research exploring the recognition of MW in VR [65, 235, 293], the data were labelled using subjective responses. This approach is novel in VR, as previous work labelled their physiological data based on the task difficulty levels [235, 293] or based on task performances [65]. While it is true that task difficulty was shown to be correlated to users' MW [12], it does not take into account the users' subjective impressions. The same task can induce different MW levels to users because of individual differences [297]. In the same way, while performance measures have shown to be correlated to users' MW in some studies [206], some users might show similar performances while experiencing different levels of mental workload [78, 295]. In addition, relying on task difficulty or task performance for the labelling could result in a recognition based on features linked to the task specificities, and not linked to the users' psychological state. Labelling data based on users' subjective MW levels via self-report responses provides an efficient way to capture the actual state of the user as they show greater face validity [203]. Furthermore, subjective measures are task agnostic, as opposed to task performance measures and task difficulty. This enables to use models trained using subjective measures in different contexts than those in which they were trained.

Special attention was given to the data acquisition protocol. It differs from previous work as it was done in a multitask context. Usual protocols, which gather data for classification purposes, focus on a single standardized task which stimulates limited pools of cognitive resources depending on the nature of the stimuli (e.g., visual, auditory) and information processing (e.g., perception, action) [304]. This has for consequence to influence physiological signals in a way that might prevent the generalization of the recognition models. Knowing MW is mostly studied in complex contexts where users have to perform multiple tasks in parallel, we chose to assess MW data in a VR flight simulator, with different tasks and stimuli natures.

Although the integration of sensors in HMDs was already explored in the literature [26, 27, 126], to the best of our knowledge, the recognition

of MW based on ML using multiple physiological sensors integrated into a VR HMD was yet not investigated. Figure 56 compares CGS and HMD sensors in different configurations for MW recognition. The current results showed overall similar recognition performances between CGS sensors and integrated sensors in an HMD (see Fig. 56). Individually, the CGS EDA sensor worked better than the HMD EDA sensor to recognize users' MW. This result could be explained by the fact that sweat glands density in the forehead is lower than on the fingers [250]. However, when combining sensors and measures, the HMD sensors worked slightly better than the CGS sensors (see Fig. 56). In particular, the best classification performance was achieved with the HMD sensors when using NSF normalization (i.e., +0.8% compared to CGS sensors with NSF). These differences are settled but could support that sensors in the HMD, together with ocular activity measures and/or task performance measures would explain more the users' MW changes than CGS sensors. Given the reduced cumbersomeness of HMD sensors compared to CGS sensors, this encourages the integration of sensors into VR HMDs to monitor users' psychological states in VR.

Moreover, multiple types of data were gathered to classify MW level: cardiac, electrodermal, and ocular activities, as well as task performance measures. The contribution of each signal was explored. When taking sensors individually, ocular activity leads overall to the highest classification accuracy (except using CGS sensors and NSF or Min-Max normalization, where it goes second to EDA), followed by EDA, cardiac activity, and finally performance measures. Otherwise, the combination of sensors improved the recognition accuracy with maximum performance when all physiological sensors are included as well as performance data. Performance measures have been used extensively in the literature as an indicator of users' MW level [206]. While it has the disadvantage of being task-dependent and not generic, its impact was expected to be major in the classification accuracy. Therefore, it is encouraging to observe that taken individually, physiological signals, which are less task-dependent, contributed more than performance features to the classification accuracy of MW levels in our study. With a normalization, the mean gain of the "All Physio + Perf" configuration compared to the "All physio" configuration is low (i.e., +1.6%) (see Tables 13 & 14), which is why we would advise not to use performance measures to make the recognition model more generic to other contexts.

From a features perspective, 71 were calculated on the cardiac, EDA and ocular signals, ranging from conventional ones such as heart rate or mean pupil diameter to less explored ones such as EMD ones [117] (see Table 1.2.3). In particular, the introduction of frequency-domain EDA measurements features [171, 252, 266] appears promising. Additional analysis showed that they strongly contributed to the MW recognition ac-

ording to the estimated features importance ³. Four (CGS sensors) and two (integrated sensors) of the five EDA frequency features appeared among the 20 most important variables.

The exploitation of physiological signals commonly lead to the well-know problem of inter-individual variability [289]. To minimize it, 3 normalization methods have been evaluated: subtraction of features calculated during baseline, adding features calculated during baseline, and min-max normalization (see Fig. 57). Overall, normalizing the data improved the classification accuracy compared to no normalization (i.e., with a mean gain of +8.2%). When taking sensors or performance individually, the best normalization method was found to be NAF with a mean gain of +11.64% compared to the no-normalization case. As for the min-max method, the results are inconsistent between the different sensors associations. It has the disadvantage to require the recording of the users' data when they are experiencing the lowest and highest mental workload possible prior to the training of the classification model, which makes it a bad normalization candidate for real-time use. In the "All Physio" and "Physio + Perf" configurations, the approach by subtraction was found to be the most effective (see Fig. 57). In particular, there is a mean gain of +5.2% in recognition accuracy using the NSF normalization compared to the no-normalization case in the "Physio+Perf" configuration (i.e., the one which had the best results). Besides, it has the advantage to be compatible with real-time applications, to be easy to implement, and low-cost in signal processing.

A preliminary study allowed to demonstrate the ability of our processing chain to classify MW in real-time through a simulation. It showed that the whole pipeline using multiple sensors is compatible with real-time usages, which is presented for the first time in the context of MW recognition in VR. This result paves the way to new HMDs with integrated sensors facilitating the real-time adaptation of VR environments based on detected MW levels.

As for the limits, while efforts have been made to try to balance the 4 MW level classes, our model tends to predict more often the class 3 due to its over-representation compared to the other classes (see Table 12). The classification performance of subject-independent models (i.e., generalisation of recognition on unseen participants) was not explored in this experiment. The current approaches (i.e., feature extraction followed by model training) does not seem to be adapted to this rarely explored problem [103]. Recent advances in deep learning seem to provide a solution and offer unmatched performances in various fields [158]. Nevertheless, this promising approach usually requires particularly large datasets [108]. Thus, even if the number of participants in the current paper is similar to comparable studies (e.g., [100, 272]), the number of labelled data is small compared to datasets in some other areas (e.g., ImageNet for object recog-

³ The importance of features was estimated based on the impurity decrease. It corresponds to the mean decrease in impurity averaged over all nodes where that feature was used to split the node [237].

dition [247]). Collecting such large physiological datasets is very complex, in particular due to the cost of data collection.

5 CONCLUSION

In this chapter, we proposed an all-in-one approach to assess users' MW in VR in real-time, using physiological sensors directly embedded into the headset and the Random Forest algorithm. The hardware and software solutions employed to build the system were depicted, and a user study with 75 participants was conducted to train the system to recognize 4 MW levels using physiological and performance measures. Contrarily to previous work which focused on single standardized tasks to elicit MW, users performed different tasks on a VR flight simulator and reported their subjective MW level during the experiment, which was then used to label the dataset. Moreover, the contribution of different normalization procedures, and of different types of measures and sensors, considering our solution integrating sensors into the HMD and CGS, are compared with regards to the recognition accuracy. Results showed similar recognition performances between the HMD sensors and the CGS sensors with an accuracy up to 65%. This encourages the use of sensors integrated into HMD as they are less cumbersome than CGS sensors in VR. Normalizing the dataset features also greatly improved the classification performance. As for the type of measures, ocular activity features were found to be especially important, followed by EDA, cardiac activity, and task performance features. Preliminary results demonstrate the ability of our pipeline to recognize MW in real-time. Taken together, the results support that our all-in-one approach is promising for real-time recognition of MW in VR.

CONCLUSION

1 CONTRIBUTIONS OF THIS THESIS

In this manuscript, entitled "**Towards the Exploitation of the User's Mental Workload in VR**", we studied the adaptation of Immersive Virtual Environments (IVEs) using Mental Workload (MW). We targeted 3 research axes: the induction, recognition, and exploitation of MW in VR. The induction focuses on the impact of VR parameters and stimuli on users' MWs. The recognition focuses on developing a model able to infer users' subjective MW levels using objective metrics. The exploitation focuses on guiding users towards a targeted MW level using MW measures.

First, we focused on the induction of MW. More precisely, we studied the impact of wearing a VR HMD on users' mental efforts. The influence of simple interactions like walking in IVEs and the effect of accommodation to VR were analyzed. Then, we targeted the induction and exploitation of MW in VR. We proposed a methodological approach to introduce MW assessment in the design of VR training scenarios. The induction was tackled as the effects of different tasks difficulties and tasks interactions in a complex IVE were studied. The exploitation was dealt with as the methodology enables to control MW modulation overtime through the introduction of MW measures in VR training scenarios. Finally, we proposed an all-in-one approach to recognize users' MW levels in real-time using physiological sensors directly embedded into a VR HMD. Insights about the impact of the sensors setups, the type of measures and signal normalization methods on MW recognition accuracy were also investigated.

Chapter 1 first aimed to present theoretical background and a comprehensive state-of-the-art of past contributions dealing with Cognitive and Affective States (CAS) in VR. We proposed a categorization of works studying CAS in VR following: the induction, recognition, or exploitation of CAS in VR. We also proposed a definition of "Affective and Cognitive VR" to relate to works associating CAS and VR. The definitions and models of MW, cognitive load, and emotions were depicted, as well as the different methods to measure them along with their advantages and drawbacks in VR. These assessments can overall be categorized in self-report, observational, task performance, and psychophysiological methods. Then, an overview of "Affective and Cognitive VR" in IVEs in a non-clinical context was presented. The main results following the induction, recognition, and exploitation of CAS in VR were summarized. Finally, a state of the art on MW in VR training was presented with a particular focus on the methodology and frameworks proposed to design VR training scenarios and the usage of MW assessment in VR training. This first chapter showed that few

works exploited CAS in VR, mainly because of the difficulty to perform an adaptive loop as the current recognition models tend to lack of accuracy or are user-dependent. Studies used physiological measurements and task performance measures as a way to provide feedback in the IVEs. However, changes in physiological signals can be difficult to interpret because of their complex relationships with psychological states which can bring diagnosticity issues [93]. Task performances on the other hand can be insensitive to users' CAS changes [294]. Thus, most works focused on the impact of factors on users' CAS and MW assessment was rarely considered as a criterion in the design of VR training scenarios. Multitasking was also little explored when MW modulation was targeted.

In **Chapter 2**, we studied the effect of wearing a VR HMD on users' mental efforts [130]. Users performed a standardized cognitive task in both the real environment and in a realistic virtual representation of this real environment in VR. In addition, the potential impacts of simple interactions like walking in IVEs and the effect of accommodation to VR were also investigated. The objective was to control minimal effects for the induction of MW in VR. The results support that wearing a VR HMD has a limited impact on the users' subjective mental efforts. In similar conditions (i.e., sitting), no effect was found. The accommodation to VR effect also did not impact users' mental efforts. However, walking in VR significantly increased users' mental efforts compared to walking in the real world. Moreover, natural walking increased users' mental efforts in both conditions. These results were supported by various kind of measurements. They support that secondary interactions such as walking and differences in representations between the real and virtual world can increase users' mental efforts and have to be controlled.

In **Chapter 3**, we proposed a methodological approach to introduce MW assessment in the design of VR training applications in 3 steps. First, the different tasks configurations considering each task and their intrinsic levels of difficulty are used to generate a state machine. Then, MW measures are assessed inside each state of the state machine to build a dataset of *A-priori* knowledge about the induction of each state on users' MW levels. Finally, the MW assessment and the state machine are used to generate VR training scenarios able to modulate users MW overtime. This contribution targeted the induction and exploitation research axes (see Fig. 1). More particularly, we wanted to study tasks interactions for the induction of users' MWs, and the approach allows to exploit MW measures to design VR training applications given a targeted progression of users' MW overtime. An illustrative application, inspired by the [MATB-II](#), was developed to validate the approach. The two experiments showed that the scenarios were able to induce the expected MW progression through time. Besides, two tasks stimulating similar pools of cognitive resources induced more MW than two tasks stimulating different pools of cognitive resources. In addition, traversing the state machine in a randomized order vs. in a predefined order did not significantly impact users' MW levels for the same tasks config-

urations, which supports that the relative MW levels differences between the tasks configurations were stable.

Finally, in **Chapter 4**, we proposed an all-in-one approach to recognize users' MW levels in real-time in VR using physiological sensors directly integrated into a VR headset. Moreover, the approach integrating sensors in an HMD was compared to the use of Commercial Grade Systems (CGS) sensors in terms of MW recognition accuracy. Insights about the impact of the different types of measures, sensors, and signal normalization methods were also explored. Overall, the recognition performance reached 65% of accuracy for 4 MW levels using the HMD sensors and by normalizing the physiological features by subtracting the physiological features assessed during the baseline to the dataset. These results support the integration of sensors into VR HMDs as there were similar recognition accuracies between the HMD sensors and the CGS sensors. Pupil measures of users were the most important in the recognition of users' MW levels, followed closely by the EDA measures, then the cardiac measures, and finally, the performance measures. The normalization of physiological signals overall greatly improved MW recognition accuracy. Especially, adding baseline features to the feature map worked the best for individual sensors and subtracting baseline features to the data worked well when combining sensors and/or task performances measures.

2 LIMITATIONS AND FUTURE WORK

This section addresses the current limitations of the contributions that were presented in this manuscript and proposes future works that could be investigated in the short/middle-term.

2.1 *Impact of VR on MW*

- **Effect of the nature of the tasks in VR:** In the first contribution, the results supported that wearing a VR HMD had a limited impact on users' mental efforts. In similar conditions (i.e., sitting), using an auditory task, no effect of the environment (i.e., VR vs. Real) was found. However, in the walking conditions, which is a task requiring visual cues to navigate, an effect of the environment was found. These preliminary results open further questions considering the nature of the tasks. Further works could consider a primary cognitive task relying on different pools of cognitive resources [304] (i.e. spatial or motor skills) or on different kind of stimuli (i.e. number, shape, haptics) to help understanding which sensory information and stage of information processing import the most on users' subjective MWs in VR.
- **Effect of differences in VR representation:** An effect of the environment was found in the walking conditions, where users did

not have a feedback of their body in VR, contrarily to in the real world. This difference in representation might be the cause of this increase in cognitive demands, as seen in previous studies modulating the angle of redirected walking [47] and hand control of a self-avatar [277]. Further works degrading the representations and controls in VR could help determining the minimum requirements not to trigger a difference in MW with the real world. Factors such as the interactions control, avatar's control and visual feedback, framerate, field of view, screens resolution, scene lighting, and scene realism could be manipulated with varying levels of differences with the real-world.

- **Effect of VR devices:** Only the effect of wearing a VR HMD in an IVE was tested in the first contribution. However, there are also other VR devices allowing different degrees of immersions such as CAVE-like systems or computers [130]. Physical workload can also impact MW [81, 82], so other parameters linked to the ergonomics of VR equipment such as the weight of the VR HMD, the obtrusiveness of VR equipment (e.g., presence of wires), and screen brightness could be tested on users' MWs.

2.2 Introducing MW assessment in the design of VR training scenarios

- **Updating A-priori knowledge:** The methodological approach we proposed in Chapter 3 requires to assess the users' MW levels in each state to be accurate. This process can be long and time-consuming if many tasks and tasks levels exist in the training environment. Suggestions were proposed to only measure MW in strategical nodes (e.g., min and max difficulty levels of the tasks) and to interpolate the induced MW in-between. A-priori knowledge of the mean MW level distributions in each state can also be arbitrarily predefined and updated following users' MW assessments iteratively.
- **Continuous tasks difficulty:** Only discrete task difficulty levels were considered here. Supplementary work could be done to investigate the possibility to extend the approach to continuous task difficulties. Interpolations could be considered depending on the relation between MW levels and tasks difficulty levels.
- **Other stimuli and collaborative aspects:** The goal of the methodological approach we proposed targeted users' MWs in VR training. We focused on tasks levels and tasks interactions as they were the most obvious factors that could impact users' MW in a VR training environment. However, other stimuli such as stressors, virtual agents' behaviours, IVEs visual load (e.g., amount of objects and distractors), can also impact users' MW and could be considered for the modulation of users' MW levels. Moreover, the approach is single-user at the moment. Further works could be done to merge the introduction of

MW assessment to existing frameworks which consider, for example, collaborative context [59, 64], virtual agents [53, 285, 292], and guidance elements [50].

2.3 Towards the real-time recognition of users' MW in VR

- **Central Nervous System (CNS) Measures:** Our all-in-one solution allows to estimate MW in real-time using physiological sensors directly integrated into a VR HMD. Only the Autonomic Nervous System (ANS) and performance measures were considered and labelled using users' self-reported MW levels. CNS measures have demonstrated impressive results to measure users' MWs and cognitive loads in past studies [144, 229, 239, 260]. Frontal measures were shown to be directly implicated in the working memory process of the users [94, 144, 239]. This type of measure would be very valuable to qualify the nature of the detected state in our approach. In particular, EEG and fNIRS are the most popular portable methods to measure users' CNS activity (see Section 3). However, they can impede the all-in-one approach we proposed to recognize users' MWs in real-time because of the complexity in the signal treatment in real-time for the fNIRS [229], and because of the tediousness of the installation and sensitivity to motion artefacts for the EEG [101]. In typical cases where users would not move, measuring users' EEG signals would indeed be a more direct measure of users' cognitive activity. Nevertheless, in our applications where users had visuomotor tasks to perform, EEG signals would have been contaminated by users' muscle activities, motions, and blinks. Technological improvements still are made in the community to use CNS measures in ecological situations. Thus, future work could focus on integrating this type of sensors in a similar setup than the one we proposed. CNS measures could also be compared to other sensors in terms of subjective MW recognition accuracy.
- **Robustness of the HMD sensors:** In the same veins as the previous point, the robustness of physiological sensors to motion artefacts could be tested in environments where users are required to move. The methodology was demonstrated in a situation where users were seated. We could further extend the work by studying the robustness of integrating sensors in VR compared to CGS setup to motion artefacts in a context requiring full-body motions. The impact of different interactions depending on sensors placements on signals features and MW recognition could also be interesting to analyze for signal treatment.
- **Genericity of the recognition model:** Our methodology relies on subjective assessment, which makes it application-agnostic when not taking into account task performance measures. A further direction

would be to test if the recognition of MW can be shared between different applications (e.g., training the model using a standardized task, and using it in another application). Moreover, the results of the second user study in Chapter 3 demonstrate that users' perceived progression of their own MW overtime could be twisted after each scenario (which lasted around 10 min). Further work on self-reports validity could be done, as well as studying the relation between perceived MW and users' brain working memory using BCI systems. Works focusing on approaches to normalize self-report scales could also be considered to address users' scale perceptions and to consider relative between-individual differences based on their data distribution and profiles.

3 LONG-TERM PERSPECTIVES

Regarding long-term perspectives, the increasing number of autonomous and adaptive systems comes with several questions and paths for improvements [231, 254, 308].

The approaches we proposed focused on MW measures. However, similar approaches could be imagined for other Cognitive States (CS), fatigue, stress, or Affective States (AS) in other fields and for other use cases. For example, scenarios could be built based on the evolution of users' psychological or physiological states over time. Moreover, because the approach we proposed in Chapter 3 allows to modulate MW overtime, a whole new set of works could be dedicated to understanding the effect of MW variation overtime on users' states, performances, and experiences (e.g., stress, time indicators, engagement) in VR training.

As underlined in the state-of-the-art, few works explored the exploitation of users' CAS and supported their work with a user-study in VR, especially for MW. Future works could focus on studying the effect of VR automation based on users' CAS measures on users' experiences and performances. For example, autonomous cars and AI are not currently fully trusted by users because of a lack of understanding of AI and because of the feeling of not being in control [307]. Similar behaviours could be expected in VR experiences and the acceptance of such systems could be studied. The balance between control and automation could be investigated based on, for example, the invasiveness of the stimuli and parameters. Moreover, we focused on CAS that are short in duration. Focusing on adapting VR experiences based on users' profiles or personality traits could also be relevant as these characteristics tend to be more stable for each individual. Another theory in affective games also proposed an interesting categorization of adaptive responses based on users' preferences [106]. Gilleade, Dix and Allanson proposed to adapt the responses of the system to an undesirable user state upon three broad categories: assist me, challenge me, and emote me [106]. These different paths could be interesting to approach the exploitation of users' CAS in VR.

As for the improvement of CAS recognition algorithms, efforts are made in the Affective Computing and Physiological Computing fields considering the recent breakthroughs in ML [93, 204, 227, 231, 254, 308]. Still, while many frameworks propose to aim for the dynamic adaptation of systems taking into account users' CAS [94, 106, 204], more work remains to perform the recognition in real-time independently from the user and the context. Compared to other fields, the availability of labelled data is also very poor (e.g., ImageNet, for object recognition, contains more than 14 millions labelled images [247]). Efforts have been made to build a dataset for biomedical research with PhysioNet that contains different banks of labelled and unlabelled physiological datasets using different sensors setups [107]. However, these data can hardly be exploited in the absence of the used sensors and if the labels do not fit the targeted users' states. Approaches such as semi-supervised learning with deep generative models could provide a way to decrease the number of labelled data needed to recognize users' CAS [139, 242].

Ethical issues should also be further considered as such recognition models and adaptive systems might be subverted to undesirable ends [93, 228, 241]. Fairclough [93] and Reynolds and Picard [241] addressed some of the issues when dealing with physiological signals and proposed directions to reduce the potential for abuses. As the goal is to propose a unique and tailored experience for each individual, data protection and the invasion of privacy can be of crucial concerns for the users [93, 228, 241]. Granting the user full control over the data collection and transparency are essential to reinforce the trust of such systems, especially when the data can emphasize medical conditions. Moreover, feedbacks which explicitly depict the users' states can embarrass users in presence of other individuals. Thus, discretion should be taken into account when using adaptive applications in public spaces, and users should be able to disable this feedbacks [93]. Adaptive systems based on users' CAS intrinsically are designed to guide users towards a targeted CS or AS. This can be interpreted as "manipulating" the users' psychological states [93]. For the explicit exploitation, users have control over their CAS, as when they are listening to music, doing meditation, or taking psychoactive agents (e.g., caffeine, alcohol) [93]. However, when the exploitation is implicit, users do not retain control over the process of adaptation. In that case, Fairclough suggests that users need to at least understand the range of manipulations they may be subjected to [93]. While it is impossible to safeguard any new technology, we propose that systems integrating an adaptive loop based on users' CAS always clearly depict the goal states, objectives of the adaptation process, and enable users to have control over it. More work could be dedicated to characterizing the ethical issues and solutions that are specific to VR adaptive systems and affective experiences.

To conclude, we believe that making VR applications aware of the users' states should enable the emergence of 'smart' and more robust technolo-

gies that could adapt to a “*specific individual in a defined place at a precise time*” [93]. Especially, because users feel “*there, in the virtual world*” [270], VR technologies are particularly efficient to engage participants into virtual experiences and to provide impactful stimuli [243] through the wide spectrum of parameters, interactions, and sensory feedback they can provide. In this manuscript, we proposed different research axes to build adaptive VR systems, conducted studies and proposed approaches to go toward the adaptation of VR applications based on MW in real-time. As such, we hope that this thesis work is now going to inspire further research and development of novel VR systems that could adapt based on users’ affective and cognitive states.

IMPACT OF DATA SPLITTING ON MENTAL WORKLOAD RECOGNITION ACCURACY

Several data splitting have been evaluated on mental workload recognition accuracy based on the reported ISA levels (1, 2, 3, 4 and 5). Figures 59, 60, and 61 present the recognition accuracy obtained for respectively 5, 3, and 2 classes of mental workload. They were obtained using the CGS sensors and NSF normalization in the "All physio + Perf" configuration. The following confusion matrixes have been normalized to obtain a total mental workload recognition accuracy of 100% per expected class.

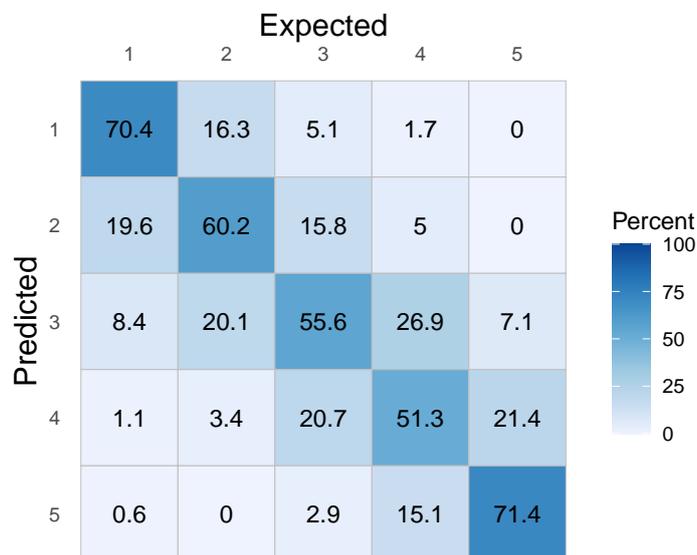


Figure 59: Confusion matrix (5 classes) - Accuracy = 61.78%

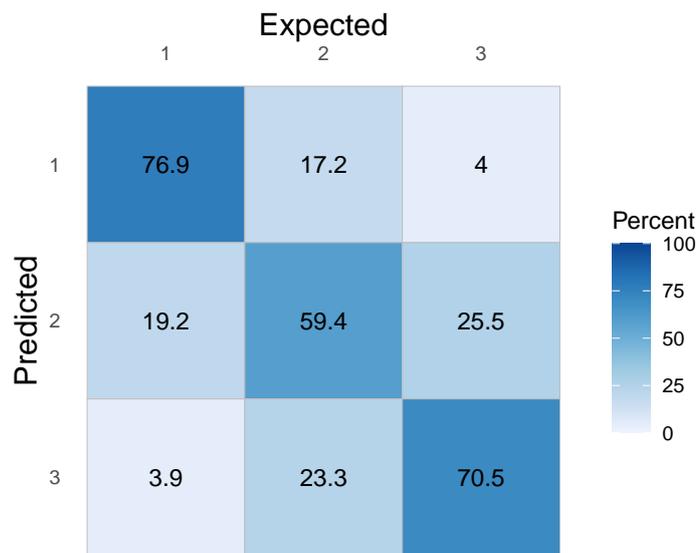


Figure 60: Confusion matrix (3 classes) - Accuracy = 68.95%

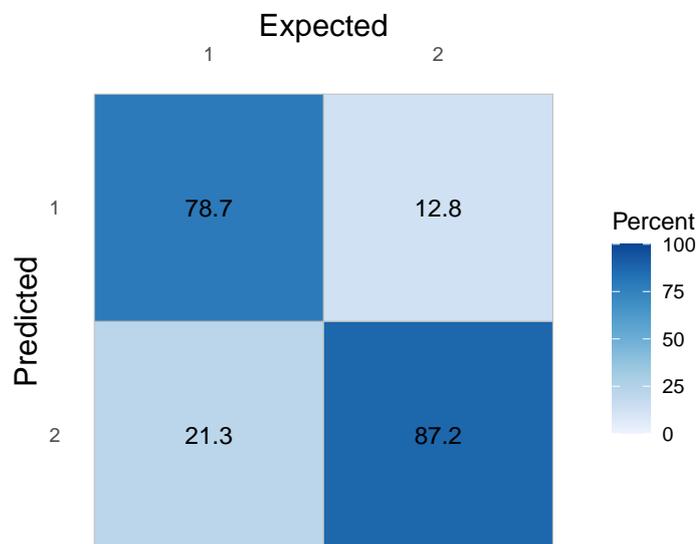


Figure 61: Confusion matrix (2 classes) - Accuracy = 82.15%

1 CONTEXTE

Il y a un siècle de cela, 80% des accidents d'avion étaient causés par des machines, et 20% des accidents d'avion, par des erreurs humaines. La tendance s'est aujourd'hui inversée. 80% des accidents d'avion sont maintenant dus à des erreurs humaines et 20%, à des défaillances liées aux machines [1, 305]. Hawkins a décrit les facteurs humains comme étant "*la dernière frontière contre les problèmes de sécurité aérienne*" [116]. En effet, les avions modernes sont aujourd'hui si complexes qu'ils peuvent désorienter les pilotes les plus expérimentés. Pour pallier ce problème, des simulateurs de vol ont été développés afin de former les équipages aux situations de crise. En particulier, les technologies immersives telles que la Réalité Virtuelle (RV) sont aujourd'hui largement employées à des fins de formation en raison de leur coût réduit par rapport aux simulateurs taille réelle.

Dans ce contexte de formation, la Charge Mentale de Travail (CMT) est depuis longtemps un sujet d'intérêt pour la recherche [77, 114, 195, 205, 304, 312]. Bien qu'il n'y ait pas de consensus universel sur sa définition, la CMT peut être décrite comme "*le rapport entre la demande (quantitative) de ressources imposée par une tâche et la capacité de l'opérateur à fournir ces ressources*" [303]. La CMT est un sujet d'intérêt populaire en raison de la complexité grandissante des systèmes. Cela ne se limite pas à l'industrie aéronautique. En effet, un certain nombre de rapports suggèrent que les accidents liés aux erreurs humaines ont augmenté de façon spectaculaire ces dernières années dans un grand nombre d'industries [121, 218]. Les smartphones, les supports de médias et les outils de travail débordent d'informations, et les options de paramétrage des logiciels et des machines ne cessent d'augmenter. Faire du multitâche est également devenu répandu partout, que ce soit dans un environnement de travail ou dans la vie quotidienne. Ces situations peuvent amener les humains à être *surchargés* et à faire des erreurs. Différentes solutions ont été envisagées afin de répondre à ce problème ; parmi elles : (1) former les utilisateurs afin de les habituer à des demandes cognitives toujours plus importantes, et (2) automatiser les systèmes afin de les rendre plus adaptés aux compétences et aux états psychologiques des utilisateurs.

Afin d'évaluer ces solutions et de les améliorer, la RV est particulièrement adaptée. D'une part, la RV permet de concevoir des scénarios de formation afin d'améliorer les compétences et les réactions des opérateurs dans des situations difficiles où ils pourraient être surchargés mentalement. En effet, les systèmes de RV ont la capacité de simuler des environnements com-

plexes et réalistes tout en plongeant les utilisateurs dans des expériences engageantes de manière sûre et contrôlée. De nombreuses recherches ont déjà démontré que les dispositifs immersifs (par ex., casques de RV, CAVEs) étaient bénéfiques par rapport aux supports plus traditionnels afin de former les utilisateurs, par exemple dans le domaine médical [4, 112, 157, 265], dans le domaine éducatif [99, 119, 194], et dans le domaine industriel [24, 109, 159, 296]. D'autre part, les technologies de RV peuvent être utilisées pour évaluer l'état mental, le comportement, et les performances des utilisateurs lors de l'usage d'un système avant sa fabrication. En effet, on sait que les individus réagissent différemment en fonction de leurs compétences, leurs connaissances, leur humeur, leur culture, et en fonction de la situation. Ne pas tenir compte de ces facteurs dans la conception d'un système peut rendre ce dernier inadapté aux utilisateurs et conduire à des erreurs humaines.

Néanmoins, l'intersection entre la formation dans la RV et les systèmes intelligents, capables de prendre en compte l'état psychologique des utilisateurs, reste peu explorée. Les scénarios actuels de formation en RV sont pour la plupart prédéfinis par des séquences d'événements, et régis par l'action des utilisateurs et des mesures de performance. Dans ce contexte, la CMT est davantage utilisée comme une finalité, une mesure à la fin du processus afin d'évaluer l'expérience utilisateur, qu'au cœur d'un algorithme qui adapte la RV à la CMT des utilisateurs de manière dynamique. Nous appelons "*adaptation de la RV à la CMT des utilisateurs*" le processus dynamique qui adapte les paramètres ou le contenu de la RV en fonction des informations acquises sur la CMT des utilisateurs. En informatique, de tels processus adaptatifs visent généralement à maximiser la performance des utilisateurs ou leur expérience utilisateur [94, 204]. L'objectif est généralement de rendre l'expérience agréable à l'utilisateur en traitant des informations sur son état psychologique, idéalement en temps réel [204]. Parmi les différents avantages d'un tel traitement, la prise en compte de la CMT des utilisateurs dans la conception des applications de formation pourrait rendre ces applications plus adaptées aux individus, engager les utilisateurs sur le long terme dans la formation et améliorer l'apprentissage de ces derniers. Si les utilisateurs sont surchargés mentalement, la difficulté des tâches à effectuer peut être réduite, et inversement. Un tel processus peut être généralisé en dehors du domaine de la formation pour d'autres états psychologiques [93, 106, 227]. Par exemple, dans un contexte de divertissement, les stimuli pourraient être modulés et intensifiés en fonction des états affectifs des utilisateurs afin d'améliorer leur expérience utilisateur. Dans le domaine thérapeutique, les stimuli pourraient être utilisés de façon progressive afin de traiter divers troubles ressentis par les patients. Dans l'ensemble, évoluer vers des systèmes adaptatifs basés sur les états psychologiques des utilisateurs peut améliorer l'expérience utilisateur ainsi que la robustesse et la flexibilité des systèmes.

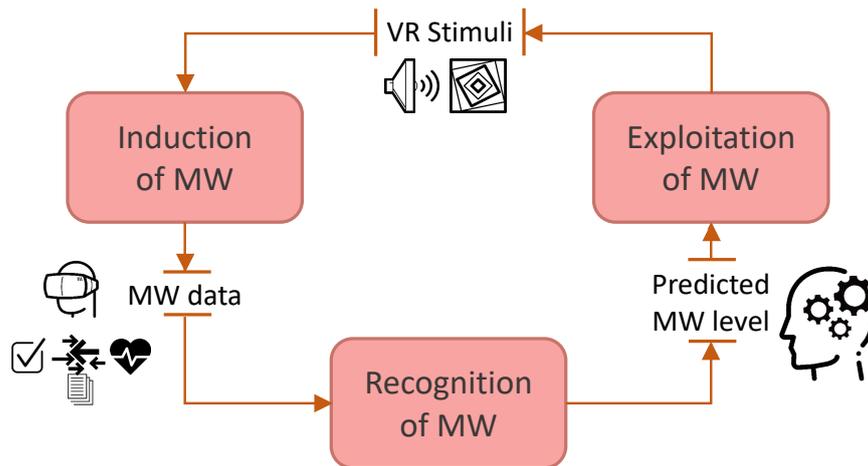


Figure 62: Adaptation de la RV en fonction de la CMT des utilisateurs. L'adaptation est décomposée en principalement 3 blocs : l'induction, la reconnaissance et l'exploitation.

2 VERROUS SCIENTIFIQUES

Durant cette thèse, nous proposons de décomposer l'adaptation de la RV basée sur la CMT des utilisateurs en 3 axes de recherche, représentés dans la Figure 62 :

- **L'induction de la CMT** : Des stimuli sont présentés aux utilisateurs dans la RV, déclenchant des réponses qui peuvent être évaluées à l'aide de différentes mesures pertinentes pour la CMT des utilisateurs (par exemple, des questionnaires, des mesures de performances aux tâches, des mesures physiologiques, des mesures comportementales).
- **La reconnaissance de la CMT** : Le modèle de reconnaissance utilise les mesures collectées précédemment et prédit le niveau de CMT des utilisateurs en temps réel. Généralement, le modèle est basé sur des règles heuristiques ou entraîné à partir d'algorithmes de [ML](#).
- **L'exploitation de la CMT** : Cette étape se concentre sur la logique rétroactive. Un ensemble de contraintes et de règles sont définies afin de choisir comment moduler les stimuli de RV en fonction du niveau de CMT prédit afin d'induire un niveau de CMT visé.

Chacune de ces 3 catégories possèdent des verrous, détaillés ci-dessous.

2.1 Induction de la CMT dans la RV

Les Environnements Virtuels Immersifs ([EVIs](#)) contiennent de nombreuses informations et paramètres. Cette complexité rend particulièrement difficile l'identification des facteurs influençant la CMT des utilisateurs dans la RV. En particulier, l'utilisation de dispositifs de RV implique un certain

nombre de contraintes (par exemple, des interactions spécifiques à la RV, l'induction de *cybersickness*) qui peuvent interagir avec les réponses liées à l'état psychologique des utilisateurs. Des études passées ont déjà constaté que certains paradigmes spécifiques à la RV tels que la marche redirigée et le contrôle d'un avatar influençaient la charge cognitive et le comportement des utilisateurs [47, 277]. L'influence des dispositifs de RV sur les la CMT des utilisateurs a cependant été très peu explorée.

Un autre verrou concerne le choix des stimuli. Ceux-ci dépendent principalement du contexte de l'application. Pour généraliser la conception de scénarios dans la RV, la littérature a présenté différentes taxonomies et frameworks. L'un des principaux problèmes est de trouver un équilibre entre la généricité et l'efficacité des scénarios pour répondre aux contraintes et aux objectifs des applications. La plupart de ces taxonomies proposent de concevoir des scénarios de RV de façon prédéfinie. Nous entendons par là que les séquences d'événements dans l'application de RV sont structurées de manière fixe. Elles sont déclenchées dans un certain ordre par les choix et les actions des utilisateurs pour répondre à des exigences de narration et d'interaction et le nombre de possibilités est limité. Actuellement, très peu d'approches considèrent les états psychologiques des utilisateurs comme références dans la conception des scénarios de RV malgré le fait que de plus en plus d'applications essaient d'adopter une perspective centrée utilisateur. Certains scénarios de RV adaptent le niveau de difficulté des tâches en fonction de la performance des utilisateurs, mais cela est fait le plus souvent dans un contexte uni-tâche et/ou "à la main", par l'expérimentateur. L'identification des stimuli pouvant affecter la CMT des utilisateurs dans des EVIs complexes, ainsi que la manière dont ces stimuli interagissent les uns avec les autres en fonction de la nature des tâches permettraient de mieux comprendre comment structurer les EVIs complexes afin de contrôler ou moduler le niveau de CMT des utilisateurs.

2.2 Reconnaissance de la CMT dans la RV

Le deuxième axe de recherche consiste à reconnaître le niveau de CMT des utilisateurs de façon automatique. Les récentes avancées en matière de ML et de *Deep Learning* rendent cela théoriquement possible [94, 227]. En particulier, un large éventail de recherches en dehors de la RV porte sur la reconnaissance des états psychologiques des utilisateurs à partir de signaux physiologiques [204]. Dans le domaine de l'informatique physiologique, les activités et les changements physiologiques sont le reflet direct de processus dans le système nerveux autonome et le système nerveux central [93]. Puisque les réponses physiologiques sont issues de processus psychologiques, il devrait être en théorie possible de traduire ces réponses en états psychologiques via l'extraction de caractéristiques physiologiques spécifiques [228]. De nombreuses recherches dans ce domaine montrent des résultats prometteurs, même si la capacité à rendre les modèles de reconnaissance génériques à tous les utilisateurs reste un problème ma-

jeur [65, 182, 204]. En effet, inférer le niveau de CMT perçu par les utilisateurs est intrinsèquement plus complexe que de mesurer l'effet d'une variable indépendante sur la CMT des utilisateurs. Par conséquent, le développement d'un modèle de reconnaissance nécessite de rassembler un nombre de données généralement plus important que celui habituellement rassemblé pour des études utilisateur dans le domaine des interactions homme-machine. Il n'existe pas non plus de consensus sur les algorithmes, les capteurs et les méthodologies de traitement du signal à adopter. De plus, si la reconnaissance des états psychologiques des utilisateurs est une étape essentielle afin de rendre les applications de RV plus adaptées aux utilisateurs, les dispositifs actuels de reconnaissance basés sur les signaux physiologiques sont difficilement appropriés à des usages quotidiens. L'acquisition des signaux physiologiques des utilisateurs consiste généralement en un processus fastidieux impliquant l'installation du matériel, l'initialisation du logiciel, la surveillance rigoureuse des signaux par un expert et la contrainte de limiter les mouvements des utilisateurs. Ces configurations peuvent être inadaptées au contexte de la RV où les interactions peuvent introduire des artefacts de mouvement dans les signaux physiologiques. Cela peut entraver le traitement du signal pour la reconnaissance de la CMT, en particulier si l'objectif est de reconnaître le niveau de CMT des utilisateurs en temps réel.

2.3 *Exploitation de la CMT dans la RV*

Le troisième axe de recherche consiste à exploiter la CMT dans la RV, c'est à dire la logique de rétroaction qui adapte l'Environnement Virtuel (EV) à partir des mesures de CMT. Différents frameworks ont été proposés afin de proposer des logiques de rétroaction dans les EVs basées sur les signaux physiologiques des utilisateurs [29, 141] et leurs états affectifs prédits [33, 164]. Cependant, en dehors du domaine clinique [313], il n'existe que très peu d'exemples de cas d'exploitation basée sur l'état psychologique des utilisateurs soutenus par des études utilisateurs dans la RV dans la littérature scientifique, d'autant plus avec la CMT comme référence. Cela peut être expliqué par les verrous de l'axe de recherche précédent : la reconnaissance de la CMT des utilisateurs. Jusqu'à présent, les études d'utilisateurs ont principalement exploité les mesures d'états psychologiques des utilisateurs dans la RV en se basant sur des méthodes heuristiques ou en affichant les changements physiologiques directement dans les EVs [2, 57, 78–80, 123]. La relation entre les états psychologiques et les signaux physiologiques est cependant relativement complexe, les variables physiologiques pouvant être associées à plusieurs éléments psychologiques ou comportementaux [93]. Par conséquent, l'utilisation de méthodes heuristiques pour directement reconnaître les états psychologiques des utilisateurs à partir de signaux physiologiques peut poser des problèmes de diagnostique, notamment si les états psychologiques ciblés ne peuvent pas être représentés dans un continuum unidimensionnel [93]. D'autre part, les modèles de

reconnaissance actuels entraînés à l'aide d'algorithmes de ML permettent d'interpréter des changements physiologiques complexes, mais manquent actuellement souvent de précision ou sont dépendants du contexte, ce qui limite leur utilisation afin de valider les frameworks d'exploitation proposés. Ainsi, très peu d'études ont proposé une logique d'adaptation des stimuli et des paramètres de RV basée sur la CMT des utilisateurs afin de les amener vers un niveau de CMT désiré. Par conséquent, l'impact du guidage des utilisateurs vers un niveau ou une évolution du niveau de CMT désiré est encore peu compris, de même que l'effet de la modulation des paramètres utilisés dans l'exploitation (par exemple, le moment d'enclenchement).

3 APPROCHE ET CONTRIBUTIONS

L'objectif principal de cette thèse est d'étudier l'exploitation de la charge mentale de l'utilisateur dans des systèmes de RV, en particulier pour la formation en RV.

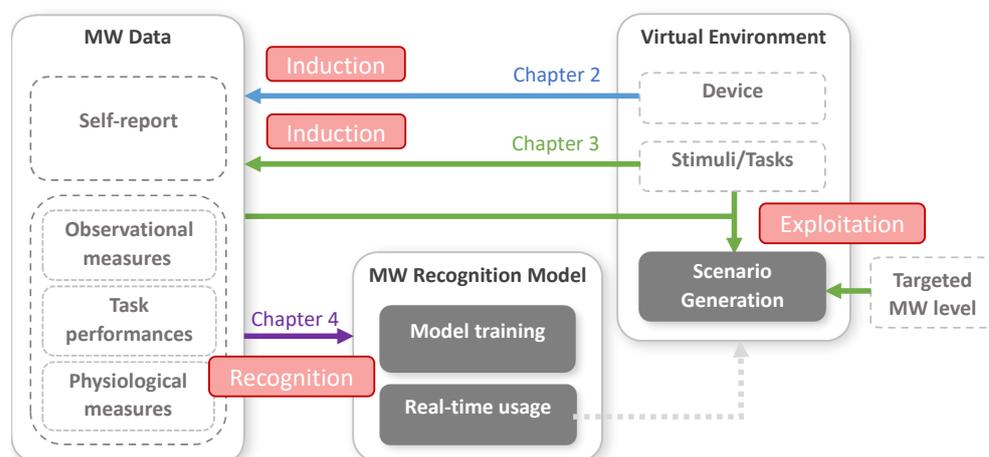


Figure 63: Correspondance entre les chapitres de la thèse et les différents axes de recherche cibles.

La figure 63 présente la correspondance entre les différentes contributions (présentées dans les chapitres) et les 3 axes de recherche proposés. Parmi les différents verrous mentionnés pour l'induction, nous avons tout d'abord choisi de contrôler les effets externes liés à l'utilisation de la RV. Nous avons étudié l'effet du port d'un casque de RV sur l'effort mental des utilisateurs, notamment car il s'agit des systèmes de RV immersifs les plus répandus sur le marché [130]. Nous avons également étudié l'impact potentiel d'interactions simples comme la marche dans les EVIs et l'effet d'accommodation à la RV sur l'effort mental. En second lieu, nous avons ciblé l'induction et l'exploitation de la CMT. Nous avons proposé d'introduire les mesures de CMT dans la conception d'applications de formation en RV. L'induction de la CMT a été abordée en étudiant l'effet de tâches de différentes natures, de leurs difficultés, et de leurs interactions sur la

CMT des utilisateurs dans un EVI complexe. Pour l'exploitation, des scénarios de formation en RV ont été conçus à partir de niveaux de CMT introduits dans l'EV mesurés à priori afin d'obtenir une évolution souhaitée du niveau de CMT des utilisateurs au cours du temps (voir Fig. 2). Enfin, nous avons exploré la reconnaissance de la CMT des utilisateurs dans la RV. Nous avons proposé une solution tout-en-un afin de reconnaître le niveau de CMT des utilisateurs en temps réel dans la RV en utilisant des capteurs physiologiques directement intégrés dans un casque de RV. En outre, nous avons étudié l'impact du dispositif de captation, du type de mesures et des méthodes de normalisation des signaux physiologiques sur la précision de la reconnaissance de la CMT.

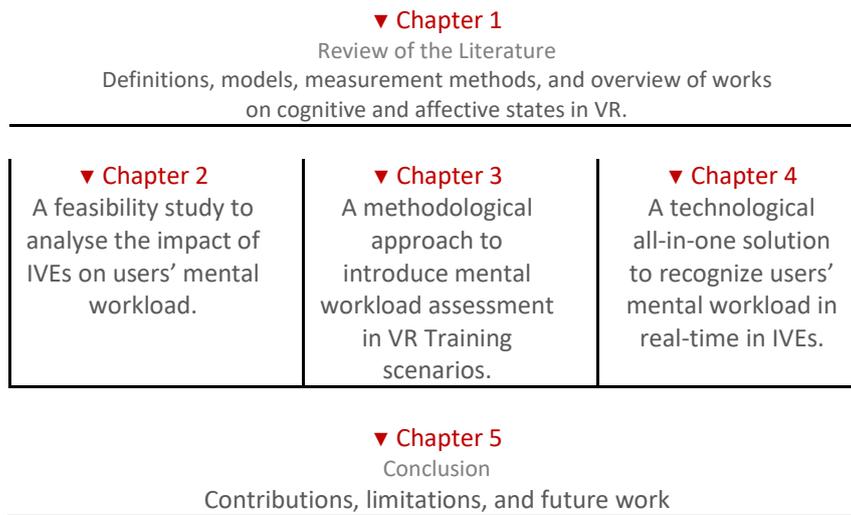


Figure 64: Illustration de la feuille de route de cette thèse.

La figure 64 présente une représentation schématique de la feuille de route de cette thèse.

Le **chapitre 1** présente un état de l'art exhaustif des travaux passés portant sur l'étude des états cognitifs et affectifs dans des EVIs. La CMT est un concept qui a reçu de nombreuses définitions au cours des dernières années [312]. Elle est également fortement liée à d'autres états psychologiques tels que l'éveil émotionnel, l'anxiété, et d'autres concepts comme le stress, les performances cognitives et le "flow" [69, 73, 196, 311]. C'est pourquoi la CMT doit donc être considérée parmi ces différentes notions. Les recherches associant la RV et les états cognitifs et affectifs suivent souvent le même schéma et des objectifs communs. Par conséquent, nous proposons tout d'abord une catégorisation des travaux portant sur l'étude des états cognitifs et affectifs dans la RV, ainsi qu'une définition de la "Réalité Virtuelle Affective et Cognitive". Dans un second temps, nous clarifions les définitions et les modèles des principaux d'états cognitifs et affectifs et concepts psychologiques étudiés dans la RV. Ensuite, les méthodes pour mesurer ces états sont présentées, ainsi que leurs avantages et inconvéni-

ents pour la RV. Nous présentons une vue d'ensemble des recherches antérieures ayant étudié les états cognitifs et affectifs dans un EVI dans un contexte non clinique. Enfin, nous nous concentrons plus particulièrement sur les études portant sur la CMT dans le milieu de la formation en RV.

Dans le **chapitre 2**, nous étudions l'impact des casques de RV sur l'effort mental des utilisateurs. Parce que les utilisateurs se sentent "*là, dans le monde virtuel*" [270] dans la RV, de nombreuses études ont émis l'hypothèse que les réponses physiologiques et comportementales des utilisateurs dans les EVIs étaient similaires à celles dans le monde réel. Ainsi, il a été démontré que les compétences acquises dans la RV pouvaient être transférées dans le monde réel [55, 246, 265, 300]. Cependant, les utilisateurs continuent malgré tout à percevoir et à penser que le monde qu'ils visualisent dans la RV n'est pas réel. Une telle perception biaisée et les contraintes relatives au contexte de la RV (par exemple, les interactions en environnement immersif, l'encombrement des dispositifs de RV, le manque de visibilité de ce qui entoure l'utilisateur dans le monde réel, la *cybersickness*) pourraient exiger un traitement d'information supplémentaire de la part des utilisateurs, ce qui pourrait augmenter leur CMT. Deux expériences d'utilisateurs ont été menées afin d'étudier l'influence du port du casque de RV sur l'effort mental des utilisateurs lors d'une tâche cognitive auditive. Différentes mesures ont été récoltées afin d'étudier l'effort mental des utilisateurs dans la RV (en utilisant un casque de RV) par rapport à la réalité : des mesures de performances, comportementales, et des mesures physiologiques. En outre, l'effet de la marche naturelle dans la RV (première étude utilisateur) et la durée d'une exposition courte dans la RV (deuxième étude utilisateur) sur l'effort mental ont également été explorés. .

Le **chapitre 3** propose une approche méthodologique afin d'introduire les mesures de CMT des utilisateurs dans la conception de scénarios de formation en RV. Peu d'applications de RV ont été construites avec en tête de moduler les états psychologiques des utilisateurs. La plupart des scénarios de RV sont prédéfinis ou structurés afin de déclencher des événements basés sur les choix et les actions des utilisateurs suivant des exigences de narration et d'interaction. Cependant, un même scénario et des mêmes stimuli induiront des réactions différentes de la part des utilisateurs en fonction, par exemple, de leurs ressources cognitives et de leurs capacités d'apprentissage. Par conséquent, ces scénarios ne produiront pas nécessairement les résultats attendus et la meilleure expérience utilisateur selon le profil et l'état psychologique des utilisateurs. Dans des EVI complexes où les tâches et les natures des stimuli sont multiples, il peut être particulièrement difficile de prévoir l'effet qu'un ensemble spécifique de tâches aura sur les utilisateurs. Nous proposons donc une approche afin de structurer des EVI complexes en utilisant une machine à états basée sur les configurations de tâches. Des mesures de la CMT déterminées de façon empirique sont introduites dans la machine à états. Ensuite, ces informations sont utilisées afin de concevoir des scénarios de formation en RV qui permettent de moduler la CMT des utilisateurs au cours du temps.

L'approche proposée est illustrée par deux études d'utilisateurs dans un simulateur de vol en RV.

Le **chapitre 4** propose une approche tout-en-un afin de reconnaître le niveau de CMT des utilisateurs en temps réel dans la RV. La RV implique des contraintes spécifiques et des dispositifs de captation particuliers. Peu de dispositifs permettent actuellement le suivi de l'état psychologique des utilisateurs en temps réel dans la RV en utilisant des algorithmes de ML. Il n'existe pas non plus actuellement de consensus sur quel algorithme de ML, quels capteurs et quelle méthodologie adopter pour le traitement des signaux physiologiques en RV. Nous proposons ici une approche tout-en-un afin de reconnaître la CMT des utilisateurs en temps réel dans la RV en utilisant des capteurs physiologiques directement intégrés dans un casque de RV. L'approche présentée au chapitre 3 est utilisée ici afin d'établir un protocole d'acquisition de données afin d'entraîner plusieurs modèles de reconnaissance de CMT à l'aide d'algorithmes de ML. La configuration matérielle est présentée, ainsi que le processus de traitement des signaux physiologiques afin de reconnaître le niveau de CMT des utilisateurs en temps réel. Les capteurs intégrés dans le casque de RV sont comparés à des capteurs trouvables dans le commerce en termes de précision des modèles de reconnaissance de la CMT. Différents résultats sont également détaillés vis à vis de l'effet des différentes mesures, capteurs et traitement de normalisation des signaux sur les performances de reconnaissance de la CMT.

Enfin, le **chapitre 5** résume les contributions, discute les limites des résultats et études délivrés au cours de cette thèse, et présente quelque perspectives pour les travaux futurs.

AUTHOR'S PUBLICATIONS

INTERNATIONAL CONFERENCES

- *T. Luong*, N. Martin, A. Raison, F. Argelaguet, J.-M. Diverrez and A. Lécuyer. 'Towards Real-Time Recognition of Users' Mental Workload Using Integrated Physiological Sensors Into a VR HMD'. In: *IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. Virtual, Brazil, Nov. 2020, pp. 1–13. URL: <https://hal.archives-ouvertes.fr/hal-03008918>
- *T. Luong*, F. Argelaguet, N. Martin and A. Lécuyer. 'Introducing Mental Workload Assessment for the Design of Virtual Reality Training Scenarios'. In: *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 2020, pp. 662–671. DOI: [10.1109/VR46266.2020.00089](https://doi.org/10.1109/VR46266.2020.00089)
- *T. Luong*, N. Martin, F. Argelaguet and A. Lécuyer. 'Studying the Mental Effort in Virtual Versus Real Environments'. In: *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 2019, pp. 809–816. DOI: [10.1109/VR.2019.8798029](https://doi.org/10.1109/VR.2019.8798029)

PATENTS

- *T. Luong*, F. Argelaguet, N. Martin and A. Lécuyer. 'French Patent no FR2000704. Issued on 2020-01-24.' Pat.

SUBMITTED PAPERS

- *T. Luong*, A. Lécuyer, N. Martin and F. Argelaguet. 'A Survey on Affective and Cognitive VR'. In: *Submitted to IEEE Transactions on Visualization and Computer Graphics*, pp. 1–22

a

- [1] AERO - MEDA Investigation Process. URL: https://www.boeing.com/commercial/aeromagazine/articles/qtr_2_07/article_03_2.html (visited on 22/09/2020).
- [2] H. B. Abdessalem, M. Boukadida and C. Frasson. 'Virtual reality game adaptation using neurofeedback'. In: *The thirty-first international flairs conference*. 2018.
- [3] H. B. Abdessalem, M. Chaouachi, M. Boukadida and C. Frasson. 'Toward Real-Time System Adaptation Using Excitement Detection from Eye Tracking'. In: *International Conference on Intelligent Tutoring Systems*. Springer. 2019, pp. 214–223.
- [4] G. Ahlberg, L. Enochsson, A. G. Gallagher, L. Hedman, C. Hogman, D. A. McClusky III, S. Ramel, C. D. Smith and D. Arvidsson. 'Proficiency-based virtual reality training significantly reduces the error rate for residents during their first 10 laparoscopic cholecystectomies'. In: *The American journal of surgery* 193.6 (2007), pp. 797–804.
- [5] S. M. Alarcao and M. J. Fonseca. 'Emotions recognition using EEG signals: A survey'. In: *IEEE Transactions on Affective Computing* 10.3 (2017), pp. 374–393.
- [6] M. Alcañiz, R. Baños, C. Botella and B. Rey. 'The EMMA project: emotions as a determinant of presence'. In: *PsychNology Journal* 1.2 (2003), pp. 141–150.
- [7] M. Ali, A. H. Mosa, F. A. Machot and K. Kyamakya. 'Emotion Recognition Involving Physiological and Speech Signals: A Comprehensive Review'. In: *Recent Advances in Nonlinear Dynamics and Synchronization*. Ed. by K. Kyamakya, W. Mathis, R. Stoop, J. C. Chedjou and Z. Li. Vol. 109. Springer International Publishing, 2018, pp. 287–302. ISBN: 978-3-319-58995-4. DOI: [10.1007/978-3-319-58996-1_13](https://doi.org/10.1007/978-3-319-58996-1_13). URL: http://link.springer.com/10.1007/978-3-319-58996-1_13.
- [8] I. Alsina-Jurnet and J. Gutiérrez-Maldonado. 'Influence of personality and individual abilities on the sense of presence experienced in anxiety triggering virtual environments'. In: *International journal of human-computer studies* 68.10 (2010), pp. 788–801.

- [9] A. Anderson, M. Mayer, A. Fellows, D. Cowan, M. Hegel and J. Buckey. 'Relaxation with Immersive Natural Scenes Presented Using Virtual Reality'. In: *Aerospace Medicine and Human Performance* 88 (June 2017), pp. 520–526. DOI: [10.3357/AMHP.4747.2017](https://doi.org/10.3357/AMHP.4747.2017).
- [10] P. Antonenko, F. Paas, R. Grabner and T. Van Gog. 'Using electroencephalography to measure cognitive load'. In: *Educational Psychology Review* 22.4 (2010), pp. 425–438.
- [11] F. Argelaguet Sanz, F. Multon and A. Lécuyer. 'A methodology for introducing competitive anxiety and pressure in VR sports training'. In: *Frontiers in Robotics and AI* 2 (2015), p. 10.
- [12] J. Au, E. Sheehan, N. Tsai, G. J. Duncan, M. Buschkuhl and S. M. Jaeggi. 'Improving fluid intelligence with training on working memory: a meta-analysis'. In: *Psychonomic bulletin & review* 22.2 (2015), pp. 366–377.
- [13] H. Ayaz, P. A. Shewokis, S. Bunce, K. Izzetoglu, B. Willems and B. Onaral. 'Optical brain monitoring for operator training and mental workload assessment'. In: *Neuroimage* 59.1 (2012), pp. 36–47.
- [14] H. Ayaz, P. A. Shewokis, S. Bunce, M. Schultheis and B. Onaral. 'Assessment of cognitive neural correlates for a functional near infrared-based brain computer interface system'. In: *International Conference on Foundations of Augmented Cognition*. Springer. 2009, pp. 699–708.
- [15] L. Aymerich-Franch. 'Presence and emotions in playing a group game in a virtual environment: the influence of body participation'. In: *Cyberpsychology, Behavior, and Social Networking* 13.6 (2010), pp. 649–654.

b

- [16] M. Balconi, E. Grippa and M. E. Vanutelli. 'What hemodynamic (fNIRS), electrophysiological (EEG) and autonomic integrated measures can tell us about emotional processing'. In: *Brain and cognition* 95 (2015), pp. 67–76.
- [17] D. Banakou, S. Kishore and M. Slater. 'Virtually being einstein results in an improvement in cognitive task performance and a decrease in age bias'. In: *Frontiers in psychology* 9 (2018), p. 917.
- [18] R. M. Baños, C. Botella, I. Rubió, S. Quero, A. García-Palacios and M. Alcañiz. 'Presence and emotions in virtual environments: The influence of stereoscopy'. In: *CyberPsychology & Behavior* 11.1 (2008), pp. 1–8.

- [19] R. M. Baños, C. Botella, M. Alcañiz, V. Liaño, B. Guerrero and B. Rey. 'Immersion and emotion: their impact on the sense of presence'. In: *Cyberpsychology & behavior* 7.6 (2004), pp. 734–741.
- [20] R. M. Baños, V. Liaño, C. Botella, M. Alcañiz, B. Guerrero and B. Rey. 'Changing induced moods via virtual reality'. In: *International Conference on Persuasive Technology*. Springer. 2006, pp. 7–15.
- [21] W. Barfield, C. Hendrix and K.-E. Bystrom. 'Effects of stereopsis and head tracking on performance using desktop virtual environment displays'. In: *Presence: Teleoperators & Virtual Environments* 8.2 (1999), pp. 237–240.
- [22] C. Beedie, P. Terry and A. Lane. 'Distinctions between emotion and mood'. In: *Cognition & Emotion* 19.6 (2005), pp. 847–878.
- [23] M. Benedek and C. Kaernbach. 'A continuous measure of phasic electrodermal activity'. In: *Journal of Neuroscience Methods* 190.1 (June 2010), pp. 80–91. ISSN: 1872-678X. DOI: [10.1016/j.jneumeth.2010.04.028](https://doi.org/10.1016/j.jneumeth.2010.04.028).
- [24] L. P. Berg and J. M. Vance. 'Industry use of virtual reality in product design and manufacturing: a survey'. In: *Virtual reality* 21.1 (2017), pp. 1–17.
- [25] I. Bergström, K. Kilteni and M. Slater. 'First-person perspective virtual body posture influences stress: a virtual reality body ownership study'. In: *PloS one* 11.2 (2016).
- [26] G. Bernal and P. Maes. 'Emotional Beasts: Visually Expressing Emotions through Avatars in VR'. In: *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. ACM. 2017, pp. 2395–2402.
- [27] G. Bernal, T. Yang, A. Jain and P. Maes. '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.
- [28] K. A. Bernhardt, D. Poltavski, T. Petros, F. R. Ferraro, T. Jorgenson, C. Carlson, P. Drechsel and C. Iseminger. 'The effects of dynamic workload and experience on commercially available EEG cognitive state metrics in a high-fidelity air traffic control environment'. In: *Applied ergonomics* 77 (2019), pp. 83–91.
- [29] D. Bersak, G. McDarby, N. Augenblick, P. McDarby, D. McDonnell, B. McDonald and R. Karkun. 'Intelligent biofeedback using an immersive competitive environment'. In: Paper at the Designing Ubiquitous Computing Games Workshop at UbiComp. 2001.
- [30] K. K. Bhagat, W.-K. Liou and C.-Y. Chang. 'A cost-effective interactive 3D virtual reality system applied to military live firing training'. In: *Virtual Reality* 20.2 (2016), pp. 127–140.

- [31] Y. Bian, C. Yang, F. Gao, H. Li, S. Zhou, H. Li, X. Sun and X. Meng. 'A framework for physiological indicators of flow in VR games: construction and preliminary evaluation'. In: *Personal and Ubiquitous Computing* 20.5 (2016), pp. 821–832.
- [32] E. Bisson, B. Contant, H. Sveistrup and Y. Lajoie. 'Functional balance and dual-task reaction times in older adults are improved by virtual reality and biofeedback training'. In: *Cyberpsychology & behavior* 10.1 (2007), pp. 16–23.
- [33] B. Bontchev. 'Adaptation in affective video games: A literature review'. In: *Cybernetics and Information Technologies* 16.3 (2016), pp. 3–34.
- [34] C. Botella, J. Fernández-Álvarez, V. Guillén, A. García-Palacios and R. Baños. 'Recent progress in virtual reality exposure therapy for phobias: a systematic review'. In: *Current psychiatry reports* 19.7 (2017), p. 42.
- [35] W. Boucsein. *Electrodermal Activity*. 2nd ed. Springer-Verlag New York Inc., Sept. 2011. ISBN: 978-1-4614-1125-3.
- [36] W. Boucsein. *Electrodermal activity*. Springer Science & Business Media, 2012.
- [37] P. Bourdin, I. Barberia, R. Oliva and M. Slater. 'A virtual out-of-body experience reduces fear of death'. In: *PloS one* 12.1 (2017).
- [38] D. A. Bowman and R. P. McMahan. 'Virtual reality: how much immersion is enough?' In: *Computer* 40.7 (2007), pp. 36–43.
- [39] M. M. Bradley and P. J. Lang. 'Measuring emotion: the self-assessment manikin and the semantic differential'. In: *Journal of behavior therapy and experimental psychiatry* 25.1 (1994), pp. 49–59.
- [40] M. Bradley and P. J. Lang. *The International affective digitized sounds (IADS): stimuli, instruction manual and affective ratings*. NIMH Center for the Study of Emotion and Attention, 1999.
- [41] J. J. Braithwaite, D. G. Watson, R. Jones and M. Rowe. *A Guide for Analysing Electrodermal Activity (EDA) & Skin Conductance Responses (SCRs) for Psychological Experiments*. 2015.
- [42] L. Breiman. 'Random Forests'. In: *Machine Learning* 45.1 (Oct. 2001), pp. 5–32. ISSN: 1573-0565. DOI: [10.1023/A:1010933404324](https://doi.org/10.1023/A:1010933404324).
- [43] J. Bric, M. Connolly, A. Kastenmeier, M. Goldblatt and J. C. Gould. 'Proficiency training on a virtual reality robotic surgical skills curriculum'. In: *Surgical endoscopy* 28.12 (2014), pp. 3343–3348.

- [44] E. L. van den Broek, V. Lisý, J. H. Janssen, J. H. D. M. Westerink, M. H. Schut and K. Tuinenbreijer. ‘Affective Man-Machine Interface: Unveiling Human Emotions through Biosignals’. en. In: *Biomedical Engineering Systems and Technologies*. Ed. by A. Fred, J. Filipe and H. Gamboa. Communications in Computer and Information Science. Springer Berlin Heidelberg, 2010, pp. 21–47. ISBN: 978-3-642-11721-3.
- [45] C. Brom and A. Abonyi. ‘Petri-nets for game plot’. In: *Proceedings of AISB artificial intelligence and simulation behaviour convention, Bristol*. Vol. 3. 2006, pp. 6–13.
- [46] M. Bruce and H. Regenbrecht. ‘A virtual reality claustrophobia therapy system-implementation and test’. In: *2009 IEEE Virtual Reality Conference*. IEEE. 2009, pp. 179–182.
- [47] G. Bruder, P. Lubas and F. Steinicke. ‘Cognitive resource demands of redirected walking’. In: *IEEE Transactions on Visualization and Computer Graphics* 21.4 (2015), pp. 539–544.
- [48] A. Burns, B. R. Greene, M. J. McGrath, T. J. O’Shea, B. Kuris, S. M. Ayer, F. Stroiescu and V. Cionca. ‘SHIMMER™—A wireless sensor platform for noninvasive biomedical research’. In: *IEEE Sensors Journal* 10.9 (2010), pp. 1527–1534.

C

- [49] B. Cain. *A review of the mental workload literature*. Tech. rep. Defence Research And Development Toronto (Canada), 2007.
- [50] K. Carpentier, D. Lourdeaux and I. M. Thouvenin. ‘Dynamic Selection of Learning Situations in Virtual Environment.’ In: *ICAART* (2). 2013, pp. 101–110.
- [51] C. M. Carswell, D. Clarke and W. B. Seales. ‘Assessing mental workload during laparoscopic surgery’. In: *Surgical innovation* 12.1 (2005), pp. 80–90.
- [52] B. Cassin, M. L. Rubin and S. Solomon. *Dictionary of eye terminology*. Vol. 10. Triad Publishing Company Gainesville, 1984.
- [53] M. Cavazza, J.-L. Lugrin, D. Pizzi and F. Charles. ‘Madame bovary on the holodeck: immersive interactive storytelling’. In: *Proceedings of the 15th ACM international conference on Multimedia*. ACM. 2007, pp. 651–660.
- [54] B. Cebeci, U. Celikcan and T. K. Capin. ‘A comprehensive study of the affective and physiological responses induced by dynamic virtual reality environments’. In: *Computer Animation and Virtual Worlds* 30.3-4 (2019), e1893.

- [55] C.-J. Chao, S.-Y. Wu, Y.-J. Yau, W.-Y. Feng and F.-Y. Tseng. 'Effects of three-dimensional virtual reality and traditional training methods on mental workload and training performance'. In: *Human Factors and Ergonomics in Manufacturing & Service Industries* 27.4 (2017), pp. 187–196.
- [56] H. Chen, A. Dey, M. Billinghamurst and R. W. Lindeman. 'Exploring pupil dilation in emotional virtual reality environments'. In: (2017).
- [57] H. Chen, A. Dey, M. Billinghamurst and R. W. Lindeman. 'Exploring the design space for multi-sensory heart rate feedback in immersive virtual reality'. In: *Proceedings of the 29th Australian Conference on Computer-Human Interaction*. ACM. 2017, pp. 108–116.
- [58] J. Chen. 'Flow in games (and everything else)'. In: *Communications of the ACM* 50.4 (2007), pp. 31–34.
- [59] P. Chevaillier, T.-H. Trinh, M. Barange, P. De Loor, F. Devillers, J. Soler and R. Querrec. 'Semantic modeling of virtual environments using mascaret'. In: *2012 5th Workshop on Software Engineering and Architectures for Realtime Interactive Systems (SEARIS)*. IEEE. 2012, pp. 1–8.
- [60] A. Chirico, P. Cipresso, D. B. Yaden, F. Biassoni, G. Riva and A. Gaggioli. 'Effectiveness of immersive videos in inducing awe: an experimental study'. In: *Scientific Reports* 7.1 (2017), pp. 1–11.
- [61] L. Chittaro and F. Buttussi. 'Assessing knowledge retention of an immersive serious game vs. a traditional education method in aviation safety'. In: *IEEE transactions on visualization and computer graphics* 21.4 (2015), pp. 529–538.
- [62] I. C. Christie and B. H. Friedman. 'Autonomic specificity of discrete emotion and dimensions of affective space: a multivariate approach'. In: *International Journal of Psychophysiology* 51.2 (Jan. 2004), pp. 143–153. ISSN: 01678760. DOI: [10.1016/j.ijpsycho.2003.08.002](https://doi.org/10.1016/j.ijpsycho.2003.08.002).
- [63] G. Claude. 'Séquencement d'actions en environnement virtuel collaboratif'. PhD thesis. Rennes, INSA, 2016.
- [64] G. Claude, V. Gouranton, R. B. Berthelot and B. Arnaldi. 'Short paper: # SEVEN, a sensor effector based scenarios model for driving collaborative virtual environment'. In: 2014.
- [65] J. Collins, H. Regenbrecht, T. Langlotz, Y. S. Can, C. Ersoy and R. Butson. 'Measuring Cognitive Load and Insight: A Methodology Exemplified in a Virtual Reality Learning Context'. In: *2019 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE. 2019, pp. 351–362.
- [66] M. P. Coral. 'Analyzing cognitive workload through eye-related measurements: A meta-analysis'. In: (2016).

- [67] J. Cremer, J. Kearney and Y. Papelis. 'HCSM: a framework for behavior and scenario control in virtual environments'. In: *ACM Transactions on Modeling and Computer Simulation (TOMACS)* 5.3 (1995), pp. 242–267.
- [68] M. Csikszentmihalyi. 'Flow and the psychology of discovery and invention'. In: *HarperPerennial, New York* 39 (1997).
- [69] M. Csikszentmihalyi. 'Toward a psychology of optimal experience'. In: *Flow and the foundations of positive psychology*. Springer, 2014, pp. 209–226.

d

- [70] F. P. Da Silva. 'Mental workload, task demand and driving performance: what relation?' In: *Procedia-Social and Behavioral Sciences* 162 (2014), pp. 310–319.
- [71] A. M. Dale and M. I. Sereno. 'Improved localizadon of cortical activity by combining EEG and MEG with MRI cortical surface reconstruction: a linear approach'. In: *Journal of cognitive neuroscience* 5.2 (1993), pp. 162–176.
- [72] D. De Waard. *The measurement of drivers' mental workload*. Groningen University, Traffic Research Center Netherlands, 1996.
- [73] F. Dehais, A. Lafont, R. Roy and S. Fairclough. 'A Neuroergonomics Approach to Mental Workload, Engagement and Human Performance'. In: *Frontiers in Neuroscience* 14 (2020), p. 268.
- [74] A. Dengel and J. Mägdefrau. 'Presence Is the Key to Understanding Immersive Learning'. In: *International Conference on Immersive Learning*. Springer. 2019, pp. 185–198.
- [75] M. S. Dennison, A. Z. Wisti and M. D'Zmura. 'Use of physiological signals to predict cybersickness'. In: *Displays* 44 (2016), pp. 42–52.
- [76] P. Desmet. *Designing emotions*. Delft University of Technology, Department of Industrial Design, 2002.
- [77] P. A. Desmond and P. A. Hancock. *Stress, workload, and fatigue*. L. Erlbaum Associates, 2001.
- [78] A. Dey, A. Chatourn and M. Billinghamurst. 'Exploration of an EEG-Based Cognitively Adaptive Training System in Virtual Reality'. In: *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE. 2019, pp. 220–226.

- [79] A. Dey, H. Chen, M. Billinghamurst and R. W. Lindeman. 'Effects of manipulating physiological feedback in immersive virtual environments'. In: *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play*. 2018, pp. 101–111.
- [80] A. Dey, T. Piumsomboon, Y. Lee and M. Billinghamurst. 'Effects of Sharing Physiological States of Players in a Collaborative Virtual Reality Gameplay'. In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM. 2017, pp. 4045–4056.
- [81] A. DiDomenico and M. A. Nussbaum. 'Interactive effects of physical and mental workload on subjective workload assessment'. In: *International journal of industrial ergonomics* 38.11-12 (2008), pp. 977–983.
- [82] A. DiDomenico and M. A. Nussbaum. 'Effects of different physical workload parameters on mental workload and performance'. In: *International Journal of Industrial Ergonomics* 41.3 (2011), pp. 255–260.
- [83] S. S. Dickerson and M. E. Kemeny. 'Acute stressors and cortisol responses: a theoretical integration and synthesis of laboratory research.' In: *Psychological bulletin* 130.3 (2004), p. 355.
- [84] H. Q. Dinh, N. Walker, L. F. Hodges, C. Song and A. Kobayashi. 'Evaluating the importance of multi-sensory input on memory and the sense of presence in virtual environments'. In: *Proceedings IEEE Virtual Reality (Cat. No. 99CB36316)*. IEEE. 1999, pp. 222–228.
- [85] Q. Dong, S. Gong and X. Zhu. 'Imbalanced Deep Learning by Minority Class Incremental Rectification'. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2018), pp. 1–1. ISSN: 0162-8828. DOI: [10.1109/TPAMI.2018.2832629](https://doi.org/10.1109/TPAMI.2018.2832629).
- [86] M. van Dooren, J. H. Janssen et al. 'Emotional sweating across the body: Comparing 16 different skin conductance measurement locations'. In: *Physiology & behavior* 106.2 (2012), pp. 298–304.

e

- [87] M. K. Eckstein, B. Guerra-Carrillo, A. T. M. Singley and S. A. Bunge. 'Beyond eye gaze: What else can eyetracking reveal about cognition and cognitive development?' In: *Developmental cognitive neuroscience* 25 (2017), pp. 69–91.
- [88] P. Ekman. 'Facial action coding system (FACS)'. In: *A human face* (2002).
- [89] P. Ekman, E. R. Sorenson and W. V. Friesen. 'Pan-cultural elements in facial displays of emotion'. In: *Science* 164.3875 (1969), pp. 86–88.

- [90] *Electrocardiography*. <https://en.wikipedia.org/wiki/Electrocardiography>. Accessed: 2020-11-04.
- [91] C. Elkin and V. Devabhaktuni. 'Comparative Analysis of Machine Learning Techniques in Assessing Cognitive Workload'. en. In: *Advances in Neuroergonomics and Cognitive Engineering*. Ed. by H. Ayaz. Advances in Intelligent Systems and Computing. Cham: Springer International Publishing, 2020, pp. 185–195. ISBN: 978-3-030-20473-0. DOI: [10.1007/978-3-030-20473-0_19](https://doi.org/10.1007/978-3-030-20473-0_19).
- [92] *Ex Machina*. 2015.

f

- [93] S. H. Fairclough. 'Fundamentals of physiological computing'. In: *Interacting with computers* 21.1-2 (2008), pp. 133–145.
- [94] S. H. Fairclough, L. Venables and A. Tattersall. 'The influence of task demand and learning on the psychophysiological response'. In: *International Journal of Psychophysiology* 56.2 (2005), pp. 171–184.
- [95] A. Felnhöfer, O. D. Kothgassner, T. Hetterle, L. Beutl, H. Hlavacs and I. Kryspin-Exner. 'Afraid to be there? Evaluating the relation between presence, self-reported anxiety, and heart rate in a virtual public speaking task'. In: *Cyberpsychology, Behavior, and Social Networking* 17.5 (2014), pp. 310–316.
- [96] A. Felnhöfer, O. D. Kothgassner, M. Schmidt, A.-K. Heinzle, L. Beutl, H. Hlavacs and I. Kryspin-Exner. 'Is virtual reality emotionally arousing? Investigating five emotion inducing virtual park scenarios'. In: *International journal of human-computer studies* 82 (2015), pp. 48–56.
- [97] M. Ferrari, L. Mottola and V. Quaresima. 'Principles, techniques, and limitations of near infrared spectroscopy'. In: *Canadian journal of applied physiology* 29.4 (2004), pp. 463–487.
- [98] A. Fortin-Côté, C. Chamberland, M. Parent, S. Tremblay, N. Beaudoin-Gagnon, A. Campeau-Lecours, J. Bergeron-Boucher and L. Lefebvre. 'Predicting Video Game Players' Fun from Physiological and Behavioural Data'. In: *Advances in Information and Communication Networks*. Ed. by K. Arai, S. Kapoor and R. Bhatia. Advances in Intelligent Systems and Computing. Springer International Publishing, 2018, pp. 479–495.
- [99] L. Freina and M. Ott. 'A literature review on immersive virtual reality in education: state of the art and perspectives'. In: *The International Scientific Conference eLearning and Software for Education*. Vol. 1. " Carol I" National Defence University. 2015, p. 133.

- [100] L. Fridman, B. Reimer, B. Mehler and W. T. Freeman. ‘Cognitive Load Estimation in the Wild’. In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. CHI ’18. event-place: Montreal QC, Canada. New York, NY, USA: ACM, 2018, 652:1–652:9. ISBN: 978-1-4503-5620-6. DOI: [10.1145/3173574.3174226](https://doi.org/10.1145/3173574.3174226). URL: <http://doi.acm.org/10.1145/3173574.3174226>.



- [101] P. Gajbhiye, R. K. Tripathy, A. Bhattacharyya and R. B. Pachori. ‘Novel approaches for the removal of motion artifact from EEG recordings’. In: *IEEE Sensors Journal* 19.22 (2019), pp. 10600–10608.
- [102] E. Galy, M. Cariou and C. Mélan. ‘What is the relationship between mental workload factors and cognitive load types?’ In: *International Journal of Psychophysiology* 83.3 (2012), pp. 269–275.
- [103] P. van Gent, T. Melman, H. Farah, N. van Nes and B. van Arem. ‘Multi-Level Driver Workload Prediction using Machine Learning and Off-the-Shelf Sensors’. en. In: *Transportation Research Record: Journal of the Transportation Research Board* 2672.37 (Dec. 2018), pp. 141–152. ISSN: 0361-1981, 2169-4052. DOI: [10.1177/0361198118790372](https://doi.org/10.1177/0361198118790372). (Visited on 25/02/2020).
- [104] S. Gerbaud, N. Mollet and B. Arnaldi. ‘Virtual environments for training: from individual learning to collaboration with humanoids’. In: *International Conference on Technologies for E-Learning and Digital Entertainment*. Springer. 2007, pp. 116–127.
- [105] L. Gerry, B. Ens, A. Drogemuller, B. Thomas and M. Billingham. ‘Levity: A virtual reality system that responds to cognitive load’. In: *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*. 2018, pp. 1–6.
- [106] K. Gilleade, A. Dix and J. Allanson. ‘Affective videogames and modes of affective gaming: assist me, challenge me, emote me’. In: *DiGRA 2005: Changing Views—Worlds in Play*. (2005).
- [107] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng and H. E. Stanley. ‘PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals’. In: *circulation* 101.23 (2000), e215–e220.
- [108] I. Goodfellow, Y. Bengio and A. Courville. *Deep Learning*. Anglais. The MIT Press, Nov. 2016.

- [109] J. Goulding, W. Nadim, P. Petridis and M. Alshawi. ‘Construction industry offsite production: A virtual reality interactive training environment prototype’. In: *Advanced Engineering Informatics* 26.1 (2012), pp. 103–116.
- [110] F. Grimm, G. Naros and A. Gharabaghi. ‘Closed-loop task difficulty adaptation during virtual reality reach-to-grasp training assisted with an exoskeleton for stroke rehabilitation’. In: *Frontiers in neuroscience* 10 (2016), p. 518.
- [111] D. Gromer, M. Reinke, I. Christner and P. Pauli. ‘Causal interactive links between presence and fear in virtual reality height exposure’. In: *Frontiers in psychology* 10 (2019), p. 141.
- [112] K. S. Gurusamy, R. Aggarwal, L. Palanivelu and B. R. Davidson. ‘Virtual reality training for surgical trainees in laparoscopic surgery’. In: *Cochrane database of systematic reviews* 1 (2009).

h

- [113] S. Haga, H. Shinoda and M. Kokubun. ‘Effects of task difficulty and time-on-task on mental workload’. In: *Japanese Psychological Research* 44.3 (2002), pp. 134–143.
- [114] P. A. Hancock and N. Meshkati. *Human mental workload*. North-Holland Amsterdam, 1988.
- [115] S. G. Hart and L. 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.
- [116] F. H. Hawkins. *Human factors in flight*. Routledge, 2017.
- [117] L. He, D. Jiang, L. Yang, E. Pei, P. Wu and H. Sahli. ‘Multimodal Affective Dimension Prediction Using Deep Bidirectional Long Short-Term Memory Recurrent Neural Networks’. In: *Proceedings of the 5th International Workshop on Audio/Visual Emotion Challenge*. AVEC ’15. ACM, 2015, pp. 73–80. ISBN: 978-1-4503-3743-4. DOI: [10.1145/2808196.2811641](https://doi.org/10.1145/2808196.2811641). URL: <http://doi.acm.org/10.1145/2808196.2811641>.
- [118] C. Herff, D. Heger, O. Fortmann, J. Hennrich, F. Putze and T. Schultz. ‘Mental workload during n-back task—quantified in the prefrontal cortex using fNIRS’. In: *Frontiers in human neuroscience* 7 (2014), p. 935.
- [119] K. F. Hew and W. S. Cheung. ‘Use of three-dimensional (3-D) immersive virtual worlds in K-12 and higher education settings: A review of the research’. In: *British journal of educational technology* 41.1 (2010), pp. 33–55.

- [120] C. Hinsch, R. Felix and P. A. Rauschnabel. 'Nostalgia beats the wow-effect: Inspiration, awe and meaningful associations in augmented reality marketing'. In: *Journal of Retailing and Consumer Services* 53 (2020), p. 101987.
- [121] A. Hobbs. 'Human factors: the last frontier of aviation safety?' In: *The International Journal of Aviation Psychology* 14.4 (2004), pp. 331–345.
- [122] K. Höök. 'Affective loop experiences—what are they?' In: *International Conference on Persuasive Technology*. Springer. 2008, pp. 1–12.
- [123] S. Houzangbe, O. Christmann, G. Gorisse and S. Richir. 'Fear as a biofeedback game mechanic in virtual reality: Effects on engagement and perceived usability'. In: *Proceedings of the 13th International Conference on the Foundations of Digital Games*. 2018, pp. 1–6.
- [124] I. Hupont, S. Baldassarri and E. Cerezo. 'Facial emotional classification: from a discrete perspective to a continuous emotional space'. In: *Pattern Analysis and Applications* 16.1 (2013), pp. 41–54.

j

- [125] R. E. Jack. 'Culture and facial expressions of emotion'. In: *Visual Cognition* 21.9-10 (2013), pp. 1248–1286.
- [126] J. Jantz, A. Molnar and R. Alcaide. 'A brain-computer interface for extended reality interfaces'. In: *ACM SIGGRAPH 2017 VR Village*. 2017, pp. 1–2.
- [127] H. Japer. 'Ten-twenty electrode system of the international federation'. In: *Electroencephalogr Clin Neurophysiol* 10 (1958), pp. 371–375.
- [128] C. Jeunet, C. Mühl and F. Lotte. 'Design and Validation of a Mental and Social Stress Induction Protocol Towards Load-Invariant Physiology-Based Detection'. In: *International Conference on Physiological Computing Systems*. 2014.
- [129] N. A. Jones and N. A. Fox. 'Electroencephalogram asymmetry during emotionally evocative films and its relation to positive and negative affectivity'. In: *Brain and cognition* 20.2 (1992), pp. 280–299.

k

- [130] R. S. Kalawsky. 'Exploiting virtual reality techniques in education and training: Technological issues'. In: <http://www.man.ac.uk/MVC/SIMA/vrtech/title.html> (1996).
- [131] A. Kamilaris and F. X. Prenafeta-Boldú. 'Deep learning in agriculture: A survey'. In: *Computers and electronics in agriculture* 147 (2018), pp. 70–90.
- [132] S. Kaplan, R. Dalal and J. Luchman. 'Measurement of emotions.' In: (Jan. 2013), pp. 61–75.
- [133] K. Kawahara, M. Sakashita, A. Koike, I. Suzuki, K. Suzuki and Y. Ochiai. 'Transformed Human Presence for Puppetry'. In: *Proceedings of the 13th International Conference on Advances in Computer Entertainment Technology*. 2016, pp. 1–6.
- [134] T. D. Kemper. 'How many emotions are there? Wedding the social and the autonomic components'. In: *American journal of Sociology* 93.2 (1987), pp. 263–289.
- [135] R. S. Kennedy, N. E. Lane, K. S. Berbaum and M. G. Lilienthal. 'Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness'. In: *The international journal of aviation psychology* 3.3 (1993), pp. 203–220.
- [136] K. Kilteni, R. Groten and M. Slater. 'The sense of embodiment in virtual reality'. In: *Presence: Teleoperators and Virtual Environments* 21.4 (2012), pp. 373–387.
- [137] K. H. Kim, S. W. Bang and S. R. Kim. 'Emotion recognition system using short-term monitoring of physiological signals'. In: *Medical and Biological Engineering and Computing* 42.3 (May 2004), pp. 419–427. ISSN: 1741-0444. DOI: [10.1007/BF02344719](https://doi.org/10.1007/BF02344719).
- [138] Y. Y. Kim, H. J. Kim, E. N. Kim, H. D. Ko and H. T. Kim. 'Characteristic changes in the physiological components of cybersickness'. In: *Psychophysiology* 42.5 (2005), pp. 616–625.
- [139] D. P. Kingma, S. Mohamed, D. J. Rezende and M. Welling. 'Semi-supervised learning with deep generative models'. In: *Advances in neural information processing systems*. 2014, pp. 3581–3589.
- [140] B. Kirwan, A. Evans, L. Donohoe, A. Kilner, T. Lamoureux, T. Atkinson and H. MacKendrick. 'Human factors in the ATM system design life cycle'. In: *FAA/Eurocontrol ATM R&D Seminar, Paris, France*. 1997, pp. 16–20.
- [141] J. M. Kivikangas, G. Chanel, B. Cowley, I. Ekman, M. Salminen, S. Järvelä and N. Ravaja. 'A review of the use of psychophysiological methods in game research'. In: *journal of gaming & virtual worlds* 3.3 (2011), pp. 181–199.

- [142] R. Kizony, N. Katz and P. L. Weiss. 'Adapting an immersive virtual reality system for rehabilitation'. In: *The Journal of Visualization and Computer Animation* 14.5 (2003), pp. 261–268.
- [143] P. R. Kleinginna and A. M. Kleinginna. 'A categorized list of emotion definitions, with suggestions for a consensual definition'. In: *Motivation and emotion* 5.4 (1981), pp. 345–379.
- [144] W. 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.
- [145] D. E. Knuth. 'Computer Programming as an Art'. In: *Communications of the ACM* 17.12 (1974), pp. 667–673.
- [146] A. Kołakowska, A. Landowska, M. Szwoch, W. Szwoch and M. R. Wróbel. 'Modeling emotions for affect-aware applications'. In: *Cover and title page designed by ESENCJA Sp. z oo* 55 (2015).
- [147] O. D. Kothgassner, A. Felnhöfer, H. Hlavacs, L. Beutl, R. Palme, I. Kryspin-Exner and L. M. Glenk. 'Salivary cortisol and cardiovascular reactivity to a public speaking task in a virtual and real-life environment'. In: *Computers in human behavior* 62 (2016), pp. 124–135.
- [148] S. B. Kotsiantis. 'Supervised Machine Learning: A Review of Classification Techniques'. In: *Proceedings of the 2007 Conference on Emerging Artificial Intelligence Applications in Computer Engineering: Real World AI Systems with Applications in eHealth, HCI, Information Retrieval and Pervasive Technologies*. Amsterdam, The Netherlands, The Netherlands: IOS Press, 2007, pp. 3–24. ISBN: 978-1-58603-780-2. (Visited on 04/02/2016).
- [149] G. A. Koulouris, K. Akşit, M. Stengel, R. K. Mantiuk, K. Mania and C. Richardt. 'Near-Eye Display and Tracking Technologies for Virtual and Augmented Reality'. In: *Computer Graphics Forum*. Vol. 38. 2. Wiley Online Library. 2019, pp. 493–519.
- [150] A. F. Kramer. 'Physiological metrics of mental workload: A review of recent progress'. In: *Multiple-task performance* (1991), pp. 279–328.
- [151] S. D. Kreibig. 'Autonomic nervous system activity in emotion: A review'. In: *Biological Psychology* 84.3 (2010). The biopsychology of emotion: Current theoretical and empirical perspectives, pp. 394–421. ISSN: 0301-0511. DOI: <https://doi.org/10.1016/j.biopsycho.2010.03.010>. URL: <http://www.sciencedirect.com/science/article/pii/S0301051110000827>.
- [152] S. D. Kreibig. 'Autonomic nervous system activity in emotion: a review'. In: *Biological Psychology* 84.3 (July 2010), pp. 394–421. ISSN: 1873-6246. DOI: [10.1016/j.biopsycho.2010.03.010](https://doi.org/10.1016/j.biopsycho.2010.03.010).
- [153] M. Kuhn et al. *caret: Classification and Regression Training*. 2018. URL: <https://CRAN.R-project.org/package=caret>.

- [154] S. Lackey, J. Salcedo, J. Szalma and P. Hancock. ‘The stress and workload of virtual reality training: the effects of presence, immersion and flow’. In: *Ergonomics* 59.8 (2016), pp. 1060–1072.
- [155] F. Lamarche and S. Donikian. ‘Automatic orchestration of behaviours through the management of resources and priority levels’. In: *Proceedings of the first international joint conference on Autonomous agents and multiagent systems: part 3*. 2002, pp. 1309–1316.
- [156] P. J. Lang, M. M. Bradley and B. N. Cuthbert. ‘International affective picture system (IAPS): Technical manual and affective ratings’. In: *NIMH Center for the Study of Emotion and Attention* (1997), pp. 39–58.
- [157] C. R. Larsen, J. L. Soerensen, T. P. Grantcharov, T. Dalsgaard, L. Schouenborg, C. Ottosen, T. V. Schroeder and B. S. Ottesen. ‘Effect of virtual reality training on laparoscopic surgery: randomised controlled trial’. In: *Bmj* 338 (2009), b1802.
- [158] Y. LeCun, Y. Bengio and G. Hinton. ‘Deep learning’. en. In: *Nature* 521.7553 (May 2015), pp. 436–444. ISSN: 1476-4687. DOI: [10.1038/nature14539](https://doi.org/10.1038/nature14539). URL: <https://www.nature.com/articles/nature14539> (visited on 15/06/2018).
- [159] G. T. Leung, G. Yucel and V. G. Duffy. ‘The effects of virtual industrial training on mental workload during task performance’. In: *Human Factors and Ergonomics in Manufacturing & Service Industries* 20.6 (2010), pp. 567–578.
- [160] H. Leventhal and K. Scherer. ‘The relationship of emotion to cognition: A functional approach to a semantic controversy’. In: *Cognition and emotion* 1.1 (1987), pp. 3–28.
- [161] B. J. Li, J. N. Bailenson, A. Pines, W. J. Greenleaf and L. M. Williams. ‘A public database of immersive VR videos with corresponding ratings of arousal, valence, and correlations between head movements and self report measures’. In: *Frontiers in psychology* 8 (2017), p. 2116.
- [162] H. Li, L. Trutoiu, K. Olszewski, L. Wei, T. Trutna, P.-L. Hsieh, A. Nicholls and C. Ma. ‘Facial performance sensing head-mounted display’. In: *ACM Transactions on Graphics (ToG)* 34.4 (2015), pp. 1–9.
- [163] R. Li and G. Reyes. ‘Buccal: low-cost cheek sensing for inferring continuous jaw motion in mobile virtual reality’. In: *Proceedings of the 2018 ACM International Symposium on Wearable Computers*. 2018, pp. 180–183.

- [164] Y. Li, A. S. Elmaghraby, A. El-Baz and E. M. Sokhadze. 'Using physiological signal analysis to design affective VR games'. In: *2015 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*. IEEE. 2015, pp. 57–62.
- [165] Y.-P. Lin, C.-H. Wang, T.-P. Jung, T.-L. Wu, S.-K. Jeng, J.-R. Duann and J.-H. Chen. 'EEG-based emotion recognition in music listening'. In: *IEEE Transactions on Biomedical Engineering* 57.7 (2010), pp. 1798–1806.
- [166] Y. Ling, W.-P. Brinkman, H. T. Nefs, C. Qu and I. Heynderickx. 'Effects of stereoscopic viewing on presence, anxiety, and cybersickness in a virtual reality environment for public speaking'. In: *Presence: Teleoperators and Virtual Environments* 21.3 (2012), pp. 254–267.
- [167] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. Van Der Laak, B. Van Ginneken and C. I. Sánchez. 'A survey on deep learning in medical image analysis'. In: *Medical image analysis* 42 (2017), pp. 60–88.
- [168] C. Liu, P. Agrawal, N. Sarkar and S. Chen. 'Dynamic difficulty adjustment in computer games through real-time anxiety-based affective feedback'. In: *International Journal of Human-Computer Interaction* 25.6 (2009), pp. 506–529.
- [169] C. Liu, K. Conn, N. Sarkar and W. Stone. 'Online affect detection and robot behavior adaptation for intervention of children with autism'. In: *IEEE transactions on robotics* 24.4 (2008), pp. 883–896.
- [170] LooxidVR. <https://looxidlabs.com/looxidvr/>. Accessed: 2020-03-02.
- [171] D. Lopez-Martinez and R. Picard. 'Multi-task neural networks for personalized pain recognition from physiological signals'. In: *2017 Seventh International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)*. IEEE, Oct. 2017, pp. 181–184. ISBN: 978-1-5386-0680-3. DOI: [10.1109/ACIIW.2017.8272611](https://doi.org/10.1109/ACIIW.2017.8272611). URL: <http://ieeexplore.ieee.org/document/8272611/>.
- [172] T. Luong, F. Argelaguet, N. Martin and A. Lecuyer. 'Introducing Mental Workload Assessment for the Design of Virtual Reality Training Scenarios'. In: *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 2020, pp. 662–671. DOI: [10.1109/VR46266.2020.00089](https://doi.org/10.1109/VR46266.2020.00089).
- [173] T. Luong, N. Martin, F. Argelaguet and A. Lécuyer. 'Studying the Mental Effort in Virtual Versus Real Environments'. In: *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 2019, pp. 809–816. DOI: [10.1109/VR.2019.8798029](https://doi.org/10.1109/VR.2019.8798029).

- [174] T. Luong, F. Argelaguet, N. Martin and A. Lécuyer. ‘French Patent no FR2000704. Issued on 2020-01-24.’ Pat.
- [175] T. Luong, A. Lécuyer, N. Martin and F. Argelaguet. ‘A Survey on Affective and Cognitive VR’. In: *Submitted to IEEE Transactions on Visualization and Computer Graphics*, pp. 1–22.
- [176] T. Luong, N. Martin, A. Raison, F. Argelaguet, J.-M. Diverrez and A. Lécuyer. ‘Towards Real-Time Recognition of Users’ Mental Workload Using Integrated Physiological Sensors Into a VR HMD’. In: *IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. Virtual, Brazil, Nov. 2020, pp. 1–13. URL: <https://hal.archives-ouvertes.fr/hal-03008918>.

m

- [177] M. Ma, M. McNeill, D. Charles, S. McDonough, J. Crosbie, L. Oliver and C. McGoldrick. ‘Adaptive virtual reality games for rehabilitation of motor disorders’. In: *international conference on universal access in human-computer interaction*. Springer. 2007, pp. 681–690.
- [178] A. MacQuarrie and A. Steed. ‘Cinematic virtual reality: Evaluating the effect of display type on the viewing experience for panoramic video’. In: *2017 IEEE Virtual Reality (VR)*. IEEE. 2017, pp. 45–54.
- [179] M. F. Macedonio, T. D. Parsons, R. A. Digiuseppe, B. A. Weiderhold and A. A. Rizzo. ‘Immersiveness and physiological arousal within panoramic video-based virtual reality’. In: *Cyberpsychology & Behavior* 10.4 (2007), pp. 508–515.
- [180] E. Malbos, R. M. Rapee and M. Kavakli. ‘Behavioral presence test in threatening virtual environments’. In: *Presence: Teleoperators and Virtual Environments* 21.3 (2012), pp. 268–280.
- [181] K. Mania, D. Wooldridge, M. Coxon and A. Robinson. ‘The effect of visual and interaction fidelity on spatial cognition in immersive virtual environments’. In: *IEEE transactions on visualization and computer graphics* 12.3 (2006), pp. 396–404.
- [182] J. Marín-Morales, J. L. Higuera-Trujillo, A. Greco, J. Guixeres, C. Llinares, C. Gentili, E. P. Scilingo, M. Alcañiz and G. Valenza. ‘Real vs. immersive-virtual emotional experience: Analysis of psychophysiological patterns in a free exploration of an art museum’. In: *PloS one* 14.10 (2019).
- [183] J. Marín-Morales, J. L. Higuera-Trujillo, A. Greco, J. Guixeres, C. Llinares, E. P. Scilingo, M. Alcañiz and G. Valenza. ‘Affective computing in virtual reality: emotion recognition from brain and heart-

- beat dynamics using wearable sensors'. In: *Scientific reports* 8.1 (2018), pp. 1–15.
- [184] I. B. Mauss and M. D. Robinson. 'Measures of emotion: A review'. In: *Cognition and emotion* 23.2 (2009), pp. 209–237.
- [185] I. Mavridou, J. T. McGhee, M. Hamed, M. Fatoorechi, A. Cleal, E. Ballaguer-Balester, E. Seiss, G. Cox and C. Nduka. 'FACETEQ interface demo for emotion expression in VR'. In: *Virtual Reality (VR), 2017 IEEE*. IEEE. 2017, pp. 441–442.
- [186] I. Mavridou, E. Seiss, T. Kostoulas, C. Nduka and E. Balaguer-Ballester. 'Towards an effective arousal detection system for virtual reality'. In: *Proceedings of the Workshop on Human-Habitat for Health (H3): Human-Habitat Multimodal Interaction for Promoting Health and Well-Being in the Internet of Things Era*. 2018, pp. 1–6.
- [187] L. K. McCorry. 'Physiology of the autonomic nervous system'. In: *American journal of pharmaceutical education* 71.4 (2007).
- [188] R. McCraty and F. Shaffer. 'Heart rate variability: new perspectives on physiological mechanisms, assessment of self-regulatory capacity, and health risk'. In: *Global Advances in Health and Medicine* 4.1 (2015), pp. 46–61.
- [189] D. McDuff, S. Gontarek and R. W. Picard. 'Remote detection of photoplethysmographic systolic and diastolic peaks using a digital camera'. In: *IEEE Transactions on Biomedical Engineering* 61.12 (2014), pp. 2948–2954.
- [190] P. E. McSharry, G. D. Clifford, L. Tarassenko and L. A. Smith. 'A dynamical model for generating synthetic electrocardiogram signals'. In: *IEEE transactions on biomedical engineering* 50.3 (2003), pp. 289–294.
- [191] M. Meehan, S. Razzaque, B. Insko, M. Whitton and F. P. Brooks. 'Review of four studies on the use of physiological reaction as a measure of presence in stressful virtual environments'. In: *Applied psychophysiology and biofeedback* 30.3 (2005), pp. 239–258.
- [192] A. Mehrabian and J. A. Russell. *An approach to environmental psychology*. the MIT Press, 1974.
- [193] M. Mendl, O. H. Burman and E. S. Paul. 'An integrative and functional framework for the study of animal emotion and mood'. In: *Proceedings of the Royal Society of London B: Biological Sciences* 277.1696 (2010), pp. 2895–2904.
- [194] Z. Merchant, E. T. Goetz, L. Cifuentes, W. Keeney-Kennicutt and T. J. Davis. 'Effectiveness of virtual reality-based instruction on students' learning outcomes in K-12 and higher education: A meta-analysis'. In: *Computers & Education* 70 (2014), pp. 29–40.
- [195] N. Moray. *Mental workload: Its theory and measurement*. Vol. 8. Springer Science & Business Media, 2013.

- [196] C Mühl, C. Jeunet and F. Lotte. ‘Design and Validation of a Mental and Social Stress Induction Protocol-Towards Load-invariant Physiology-based Stress Detection’. In: (2014).
- [197] G Mulder. ‘The concept and measurement of mental effort’. In: *Energetics and human information processing*. Springer, 1986, pp. 175–198.
- [198] J. E. Muñoz, T. Paulino, H. Vasanth and K. Baras. ‘PhysioVR: A novel mobile virtual reality framework for physiological computing’. In: *e-Health Networking, Applications and Services (Healthcom), 2016 IEEE 18th International Conference on*. IEEE. 2016, pp. 1–6.

n

- [199] L. E. Nacke, M. Kalyn, C. Lough and R. L. Mandryk. ‘Biofeedback game design: using direct and indirect physiological control to enhance game interaction’. In: *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM. 2011, pp. 103–112.
- [200] M. Narayan, L. Waugh, X. Zhang, P. Bafna and D. Bowman. ‘Quantifying the benefits of immersion for collaboration in virtual environments’. In: *Proceedings of the ACM symposium on Virtual reality software and technology*. ACM. 2005, pp. 78–81.
- [201] E. Niedermeyer and F. L. da Silva. *Electroencephalography: basic principles, clinical applications, and related fields*. Lippincott Williams & Wilkins, 2005.
- [202] F. Noroozi, C. A. Corneanu, D. Kamińska, T. Sapiński, S. Escalera and G. Anbarjafari. ‘Survey on emotional body gesture recognition’. In: *arXiv preprint arXiv:1801.07481* (2018).
- [203] D. Novak, M. Mihelj and M. Munih. ‘A survey of methods for data fusion and system adaptation using autonomic nervous system responses in physiological computing’. In: *Interacting with Computers* 24.3 (2012), pp. 154–172. ISSN: 0953-5438. DOI: [10.1016/j.intcom.2012.04.003](https://doi.org/10.1016/j.intcom.2012.04.003).
- [204] D. Novak, M. Mihelj and M. Munih. ‘A survey of methods for data fusion and system adaptation using autonomic nervous system responses in physiological computing’. In: *Interacting with computers* 24.3 (2012), pp. 154–172.

O

- [205] R. D. O'Donnell. 'Workload assessment methodology'. In: *Cognitive processes and performance* (1986).
- [206] R. D. O'Donnell and F. T. Eggemeier. 'Workload assessment methodology'. In: *Handbook of perception and human performance, Vol. 2: Cognitive processes and performance*. John Wiley & Sons, 1986, pp. 1–49. ISBN: 978-0-471-82956-0.
- [207] G. D. Ogden, J. M. Levine and E. J. Eisner. 'Measurement of workload by secondary tasks'. In: *Human Factors* 21.5 (1979), pp. 529–548.
- [208] T. E. Oliphant. *A guide to NumPy*. Vol. 1. Trelgol Publishing USA, 2006.
- [209] T. Oliveira, P. Noriega, F. Rebelo and R. Heidrich. 'Evaluation of the Relationship Between Virtual Environments and Emotions'. In: *International Conference on Applied Human Factors and Ergonomics*. Springer. 2017, pp. 71–82.
- [210] K. Olszewski, J. J. Lim, S. Saito and H. Li. 'High-fidelity facial and speech animation for VR HMDs'. In: *ACM Transactions on Graphics (TOG)* 35.6 (2016), pp. 1–14.
- [211] A. Ortony and T. J. Turner. 'What's basic about basic emotions?' In: *Psychological review* 97.3 (1990), p. 315.
- [212] S. A. Osimo, R. Pizarro, B. Spanlang and M. Slater. 'Conversations between self and self as Sigmund Freud—A virtual body ownership paradigm for self counselling'. In: *Scientific reports* 5 (2015), p. 13899.
- [213] H. Osking and J. A. Doucette. 'Enhancing Emotional Effectiveness of Virtual-Reality Experiences with Voice Control Interfaces'. In: *International Conference on Immersive Learning*. Springer. 2019, pp. 199–209.

p

- [214] F. G. Paas and J. J. Van Merriënboer. 'Instructional control of cognitive load in the training of complex cognitive tasks'. In: *Educational psychology review* 6.4 (1994), pp. 351–371.
- [215] F. Paas, J. E. Tuovinen, H. Tabbers and P. W. Van Gerven. 'Cognitive load measurement as a means to advance cognitive load theory'. In: *Educational psychologist* 38.1 (2003), pp. 63–71.

- [216] F. Pallavicini, A. Ferrari, A. Pepe, G. Garcea, A. Zancchi and F. Mantovani. 'Effectiveness of virtual reality survival horror games for the emotional elicitation: Preliminary insights using Resident Evil 7: Biohazard'. In: (2018), pp. 87–101.
- [217] Y. Pan and A. Steed. 'Avatar Type Affects Performance of Cognitive Tasks in Virtual Reality'. In: *25th ACM Symposium on Virtual Reality Software and Technology*. 2019, pp. 1–4.
- [218] K. S. Park. *Human reliability: Analysis, prediction, and prevention of human errors*. Elsevier, 2014.
- [219] T. D. Parsons and C. G. Courtney. 'Interactions between threat and executive control in a virtual reality stroop task'. In: *IEEE Transactions on Affective Computing* 9.1 (2016), pp. 66–75.
- [220] T. D. Parsons and J. L. Reinebold. 'Adaptive virtual environments for neuropsychological assessment in serious games'. In: *IEEE Transactions on Consumer Electronics* 58.2 (2012).
- [221] T. D. Parsons and A. A. Rizzo. 'Affective outcomes of virtual reality exposure therapy for anxiety and specific phobias: A meta-analysis'. In: *Journal of behavior therapy and experimental psychiatry* 39.3 (2008), pp. 250–261.
- [222] M. Paul, A. F. Mota, C. H. Antink, V. Blazek and S. Leonhardt. 'Modeling photoplethysmographic signals in camera-based perfusion measurements: optoelectronic skin phantom'. In: *Biomedical Optics Express* 10.9 (2019), pp. 4353–4368.
- [223] F. Pedregosa et al. 'Scikit-learn: Machine Learning in Python'. In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830.
- [224] J. S. Peper, R. M. Brouwer, D. I. Boomsma, R. S. Kahn, H. Pol and E. Hilleke. 'Genetic influences on human brain structure: a review of brain imaging studies in twins'. In: *Human brain mapping* 28.6 (2007), pp. 464–473.
- [225] D.-P. Pertaub, M. Slater and C. Barker. 'An experiment on public speaking anxiety in response to three different types of virtual audience'. In: *Presence: Teleoperators & Virtual Environments* 11.1 (2002), pp. 68–78.
- [226] R. W. Picard, E. Vyzas and J. Healey. 'Toward machine emotional intelligence: analysis of affective physiological state'. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 23.10 (Oct. 2001), pp. 1175–1191. ISSN: 0162-8828. DOI: [10.1109/34.954607](https://doi.org/10.1109/34.954607).
- [227] R. W. Picard, E. Vyzas and J. Healey. 'Toward machine emotional intelligence: Analysis of affective physiological state'. In: *IEEE transactions on pattern analysis and machine intelligence* 23.10 (2001), pp. 1175–1191.
- [228] R. W. Picard et al. 'Affective computing'. In: (1995).

- [229] P. Pinti, I. Tachtsidis, A. Hamilton, J. Hirsch, C. Aichelburg, S. Gilbert and P. W. Burgess. 'The present and future use of functional near-infrared spectroscopy (fNIRS) for cognitive neuroscience'. In: *Annals of the New York Academy of Sciences* 1464.1 (2020), p. 5.
- [230] R. Plutchik. *The emotions*. University Press of America, 1991.
- [231] S. Poria, E. Cambria, R. Bajpai and A. Hussain. 'A review of affective computing: From unimodal analysis to multimodal fusion'. In: *Information Fusion* 37 (2017), pp. 98–125.
- [232] H. F. Posada-Quintero and K. H. Chon. 'Innovations in electrodermal activity data collection and signal processing: A systematic review'. In: *Sensors* 20.2 (2020), p. 479.
- [233] J. Psotka. 'Immersive training systems: Virtual reality and education and training'. In: *Instructional science* 23.5-6 (1995), pp. 405–431.
- [234] *Pupil Constriction and Dilation*. <http://www.math.ubc.ca/~cass/courses/m309-03a/m309-projects/bajwa/Bajwa.htm>. Accessed: 2020-11-27.
- [235] F. Putze, C. Herff, C. Tremmel, T. Schultz and D. J. Krusienski. 'Decoding mental workload in virtual environments: a fNIRS study using an immersive n-back task'. In: *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE. 2019, pp. 3103–3106.
- [236] S. Putze, D. Alexandrovsky, F. Putze, S. Höffner, J. D. Smeddinck and R. Malaka. 'Breaking The Experience: Effects of Questionnaires in VR User Studies'. In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 2020, pp. 1–15.

q

- [237] Y. Qi. 'Random Forest for Bioinformatics'. en. In: *Ensemble Machine Learning*. Ed. by C. Zhang and Y. Ma. Boston, MA: Springer US, 2012, pp. 307–323. ISBN: 978-1-4419-9325-0 978-1-4419-9326-7. DOI: [10.1007/978-1-4419-9326-7_11](https://doi.org/10.1007/978-1-4419-9326-7_11). URL: http://link.springer.com/10.1007/978-1-4419-9326-7_11 (visited on 15/05/2020).

r

- [238] R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing, 2018. URL: <https://www.R-project.org>.
- [239] S. Raghavachari, M. J. Kahana, D. S. Rizzuto, J. B. Caplan, M. P. Kirschen, B. Bourgeois, J. R. Madsen and J. E. Lisman. ‘Gating of human theta oscillations by a working memory task’. In: *Journal of Neuroscience* 21.9 (2001), pp. 3175–3183.
- [240] G. B. Reid and T. E. Nygren. ‘The subjective workload assessment technique: A scaling procedure for measuring mental workload’. In: *Advances in psychology*. Vol. 52. Elsevier, 1988, pp. 185–218.
- [241] C. Reynolds and R. Picard. ‘Affective sensors, privacy, and ethical contracts’. In: *CHI’04 extended abstracts on Human factors in computing systems*. 2004, pp. 1103–1106.
- [242] D. J. Rezende, S. Mohamed and D. Wierstra. ‘Stochastic backpropagation and approximate inference in deep generative models’. In: *arXiv preprint arXiv:1401.4082* (2014).
- [243] G. Riva. ‘Virtual reality in psychotherapy’. In: *Cyberpsychology & behavior* 8.3 (2005), pp. 220–230.
- [244] G. Riva, F. Mantovani, C. S. Capideville, A. Preziosa, F. Morganti, D. Villani, A. Gaggioli, C. Botella and M. Alcañiz. ‘Affective interactions using virtual reality: the link between presence and emotions’. In: *CyberPsychology & Behavior* 10.1 (2007), pp. 45–56.
- [245] P. Robitaille and M. J. McGuffin. ‘Increased affect-arousal in VR can be detected from faster body motion with increased heart rate’. In: *Proceedings of the ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games*. 2019, pp. 1–6.
- [246] F. Rose, E. A. Attree, B. Brooks, D. Parslow and P. Penn. ‘Training in virtual environments: transfer to real world tasks and equivalence to real task training’. In: *Ergonomics* 43.4 (2000), pp. 494–511.
- [247] O. Russakovsky et al. ‘ImageNet Large Scale Visual Recognition Challenge’. In: *International Journal of Computer Vision (IJCV)* 115.3 (2015), pp. 211–252. DOI: [10.1007/s11263-015-0816-y](https://doi.org/10.1007/s11263-015-0816-y).
- [248] J. A. Russell. ‘A circumplex model of affect.’ In: *Journal of personality and social psychology* 39.6 (1980), p. 1161.
- [249] H. Ruthrof. *The body in language*. Bloomsbury Publishing, 2015.

S

- [250] K. Saga. 'Structure and function of human sweat glands studied with histochemistry and cytochemistry'. In: *Progress in histochemistry and cytochemistry* 37.4 (2002), pp. 323–386.
- [251] A. C. Samson, S. D. Kreibig, B. Soderstrom, A. A. Wade and J. J. Gross. 'Eliciting positive, negative and mixed emotional states: A film library for affective scientists'. In: *Cognition and Emotion* 30.5 (2016), pp. 827–856.
- [252] A. Sano, W. Chen, D. L. Martinez, S. Taylor and R. W. Picard. 'Multimodal Ambulatory Sleep Detection Using LSTM Recurrent Neural Networks'. In: *IEEE Journal of Biomedical and Health Informatics* (2018), pp. 1–1. ISSN: 2168-2194, 2168-2208. DOI: [10.1109/JBHI.2018.2867619](https://doi.org/10.1109/JBHI.2018.2867619).
- [253] Y. Santiago-Espada, R. R. Myer, K. A. Latorella and J. R. Comstock Jr. 'The multi-attribute task battery II (MATB-II) software for human performance and workload research: A user's guide'. In: (2011).
- [254] O. C. Santos. 'Emotions and personality in adaptive e-learning systems: an affective computing perspective'. In: *Emotions and personality in personalized services*. Springer, 2016, pp. 263–285.
- [255] K. Schaaff, R. Degen, N. Adler and M. T. Adam. 'Measuring affect using a standard mouse device'. In: *Biomedical Engineering/Biomedizinische Technik* 57.SI-1 Track-N (2012), pp. 761–764.
- [256] K. R. Scherer. 'On the nature and function of emotion: A component process approach'. In: *Approaches to emotion* 2293 (1984), p. 317.
- [257] K. R. Scherer. 'Appraisal considered as a process of multilevel sequential checking'. In: *Appraisal processes in emotion: Theory, methods, research* 92.120 (2001), p. 57.
- [258] K. R. Scherer. 'What are emotions? And how can they be measured?' In: *Social science information* 44.4 (2005), pp. 695–729.
- [259] L. A. Schmidt and L. J. Trainor. 'Frontal brain electrical activity (EEG) distinguishes valence and intensity of musical emotions'. In: *Cognition & Emotion* 15.4 (2001), pp. 487–500.
- [260] F. Scholkmann, S. Kleiser, A. J. Metz, R. Zimmermann, J. M. Pavia, U. Wolf and M. Wolf. 'A review on continuous wave functional near-infrared spectroscopy and imaging instrumentation and methodology'. In: *Neuroimage* 85 (2014), pp. 6–27.
- [261] A. Schwerdtfeger. 'Predicting autonomic reactivity to public speaking: don't get fixed on self-report data!' In: *International Journal of Psychophysiology* 52.3 (2004), pp. 217–224.

- [262] S. Seinfeld, I. Bergstrom, A. Pomes, J. Arroyo-Palacios, F. Vico, M. Slater and M. V. Sanchez-Vives. 'Influence of music on anxiety induced by fear of heights in virtual reality'. In: *Frontiers in psychology* 6 (2016), p. 1969.
- [263] H. Selye et al. 'A syndrome produced by diverse nocuous agents'. In: *Nature* 138.3479 (1936), p. 32.
- [264] H. Sequeira, P. Hot, L. Silvert and S. Delplanque. 'Electrical autonomic correlates of emotion'. In: *International journal of psychophysiology* 71.1 (2009), pp. 50–56.
- [265] N. E. Seymour, A. G. Gallagher, S. A. Roman, M. K. O'Brien, V. K. Bansal, D. K. Andersen and R. M. Satava. 'Virtual reality training improves operating room performance: results of a randomized, double-blinded study'. In: *Annals of surgery* 236.4 (2002), p. 458.
- [266] Y. Shimomura, T. Yoda, K. Sugiura, A. Horiguchi, K. Iwanaga and T. Katsuura. 'Use of frequency domain analysis of skin conductance for evaluation of mental workload'. eng. In: *Journal of Physiological Anthropology* 27.4 (June 2008), pp. 173–177. ISSN: 1880-6805. DOI: [10.2114/jpa2.27.173](https://doi.org/10.2114/jpa2.27.173).
- [267] M. Slater. 'Measuring presence: A response to the Witmer and Singer presence questionnaire'. In: *Presence* 8.5 (1999), pp. 560–565.
- [268] M. Slater, A. Brogni and A. Steed. 'Physiological responses to breaks in presence: A pilot study'. In: *Presence 2003: The 6th annual international workshop on presence*. Vol. 157. Citeseer. 2003.
- [269] M. Slater, D.-P. Pertaub, C. Barker and D. M. Clark. 'An experimental study on fear of public speaking using a virtual environment'. In: *CyberPsychology & Behavior* 9.5 (2006), pp. 627–633.
- [270] M. Slater, M. Usoh and A. Steed. 'Depth of presence in virtual environments'. In: *Presence: Teleoperators & Virtual Environments* 3.2 (1994), pp. 130–144.
- [271] Society for Psychophysiological Research Ad Hoc Committee on Electrodermal Measures. 'Publication recommendations for electrodermal measurements: Publication standards for EDA'. en. In: *Psychophysiology* 49.8 (Aug. 2012), pp. 1017–1034. ISSN: 00485772. DOI: [10.1111/j.1469-8986.2012.01384.x](https://doi.org/10.1111/j.1469-8986.2012.01384.x). URL: <http://doi.wiley.com/10.1111/j.1469-8986.2012.01384.x> (visited on 23/11/2018).
- [272] E. T. Solovey, M. Zec, E. A. Garcia Perez, B. Reimer and B. Mehler. 'Classifying driver workload using physiological and driving performance data: two field studies'. en. In: *Proceedings of the 32nd annual ACM conference on Human factors in computing systems - CHI '14*. Toronto, Ontario, Canada: ACM Press, 2014, pp. 4057–4066. ISBN: 978-1-4503-2473-1. DOI: [10.1145/2556288.2557068](https://doi.org/10.1145/2556288.2557068).

- URL: <http://dl.acm.org/citation.cfm?doid=2556288.2557068> (visited on 23/12/2019).
- [273] C. D. Spielberger. 'State-Trait anxiety inventory'. In: *The Corsini encyclopedia of psychology* (2010), pp. 1–1.
- [274] C. D. Spielberger and E. C. Reheiser. 'Assessment of emotions: Anxiety, anger, depression, and curiosity'. In: *Applied Psychology: Health and Well-Being* 1.3 (2009), pp. 271–302.
- [275] C. D. Spielberger, S. J. Sydeman, A. E. Owen and B. J. Marsh. *Measuring anxiety and anger with the State-Trait Anxiety Inventory (STAI) and the State-Trait Anger Expression Inventory (STAXI)*. Lawrence Erlbaum Associates Publishers, 1999.
- [276] C. D. Spielberger, R. L. Gorsuch and R. E. Lushene. 'Manual for the state-trait anxiety inventory'. In: (1970).
- [277] A. Steed, Y. Pan, F. Zisch and W. Steptoe. 'The impact of a self-avatar on cognitive load in immersive virtual reality'. In: *2016 IEEE Virtual Reality (VR)*. IEEE. 2016, pp. 67–76.
- [278] C. L. Stephens, I. C. Christie and B. H. Friedman. 'Autonomic specificity of basic emotions: Evidence from pattern classification and cluster analysis'. In: *Biological Psychology*. The biopsychology of emotion: Current theoretical and empirical perspectives 84.3 (July 2010), pp. 463–473. ISSN: 0301-0511. DOI: [10.1016/j.biopsycho.2010.03.014](https://doi.org/10.1016/j.biopsycho.2010.03.014).
- [279] J. Steuer. 'Defining virtual reality: Dimensions determining telepresence'. In: *Journal of communication* 42.4 (1992), pp. 73–93.
- [280] C. Stinson and D. A. Bowman. 'Feasibility of training athletes for high-pressure situations using virtual reality'. In: *IEEE transactions on visualization and computer graphics* 20.4 (2014), pp. 606–615.
- [281] F.-T. Sun, C. Kuo, H.-T. Cheng, S. Buthpitiya, P. Collins and M. Griss. 'Activity-aware mental stress detection using physiological sensors'. In: *International Conference on Mobile Computing, Applications, and Services*. Springer. 2010, pp. 282–301.
- [282] S. S. Sundar, J. Kang and D. Oprean. 'Being There in the Midst of the Story: How Immersive Journalism Affects Our Perceptions and Cognitions'. In: *Cyberpsychology, Behavior, and Social Networking* 20.11 (2017), pp. 672–682.
- [283] K. Suzuki, F. Nakamura, J. Otsuka, K. Masai, Y. Itoh, Y. Sugiura and M. Sugimoto. 'Recognition and mapping of facial expressions to avatar by embedded photo reflective sensors in head mounted display'. In: *2017 IEEE Virtual Reality (VR)*. IEEE. 2017, pp. 177–185.

- [284] K. T. Sweeney, T. E. Ward and S. F. McLoone. ‘Artifact removal in physiological signals—Practices and possibilities’. In: *IEEE transactions on information technology in biomedicine* 16.3 (2012), pp. 488–500.
- [285] N. Szilas. ‘IDtension: a narrative engine for Interactive Drama’. In: *Proceedings of the technologies for interactive digital storytelling and entertainment (TIDSE) conference*. Vol. 3. 2. 2003, pp. 1–11.

t

- [286] T. Tamura, Y. Maeda, M. Sekine and M. Yoshida. ‘Wearable photoplethysmographic sensors—past and present’. In: *Electronics* 3.2 (2014), pp. 282–302.
- [287] A. J. Tattersall and P. S. Foord. ‘An experimental evaluation of instantaneous self-assessment as a measure of workload’. In: *Ergonomics* 39.5 (1996), pp. 740–748.
- [288] G. Teo, G. Matthews, L. Reinerman-Jones and D. Barber. ‘Adaptive aiding with an individualized workload model based on psychophysiological measures’. In: *Human-Intelligent Systems Integration* (2019), pp. 1–15.
- [289] G. Teo, G. Matthews, L. Reinerman-Jones and D. Barber. ‘Adaptive aiding with an individualized workload model based on psychophysiological measures’. en. In: *Human-Intelligent Systems Integration* (Nov. 2019). ISSN: 2524-4876, 2524-4884. DOI: [10.1007/s42454-019-00005-8](https://doi.org/10.1007/s42454-019-00005-8). (Visited on 15/04/2020).
- [290] E. R. Thompson. ‘Development and validation of an internationally reliable short-form of the positive and negative affect schedule (PANAS)’. In: *Journal of cross-cultural psychology* 38.2 (2007), pp. 227–242.
- [291] B. S. Todd and D. C. Andrews. ‘The Identification of Peaks in Physiological Signals’. In: *Computers and Biomedical Research* 32.4 (Aug. 1999), pp. 322–335. ISSN: 0010-4809. DOI: [10.1006/cbmr.1999.1518](https://doi.org/10.1006/cbmr.1999.1518).
- [292] D. Traum, J. Rickel, J. Gratch and S. Marsella. ‘Negotiation over tasks in hybrid human-agent teams for simulation-based training’. In: *Proceedings of the second international joint conference on Autonomous agents and multiagent systems*. ACM. 2003, pp. 441–448.
- [293] C. Tremmel, C. Herff, T. Sato, K. Rechowicz, Y. Yamani and D. J. Krusienski. ‘Estimating Cognitive Workload in an Interactive Virtual Reality Environment Using EEG’. In: *Frontiers in Human Neuroscience* 13 (2019).

- [294] P. S. Tsang and V. L. Velazquez. 'Diagnosticity and multidimensional subjective workload ratings'. In: *Ergonomics* 39.3 (1996), pp. 358–381.
- [295] P. S. Tsang and M. A. Vidulich. 'Mental workload and situation awareness.' In: (2006).
- [296] C. J. Turner, W. Hutabarat, J. Oyekan and A. Tiwari. 'Discrete event simulation and virtual reality use in industry: New opportunities and future trends'. In: *IEEE Transactions on Human-Machine Systems* 46.6 (2016), pp. 882–894.
- [297] M. L. Turner and R. W. Engle. 'Is working memory capacity task dependent?' In: *Journal of memory and language* 28.2 (1989), pp. 127–154.
- [319] T. pandas development team. *pandas-dev/pandas: Pandas*. Version latest. Feb. 2020. DOI: [10.5281/zenodo.3509134](https://doi.org/10.5281/zenodo.3509134). URL: <https://doi.org/10.5281/zenodo.3509134>.

V

- [298] W. B. Verwey and H. A. Veltman. 'Detecting short periods of elevated workload: A comparison of nine workload assessment techniques.' In: *Journal of experimental psychology: Applied* 2.3 (1996), p. 270.
- [299] P. Virtanen et al. 'SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python'. In: *Nature Methods* 17 (2020), pp. 261–272. DOI: <https://doi.org/10.1038/s41592-019-0686-2>.

W

- [300] D. Waller, E. Hunt and D. Knapp. 'The transfer of spatial knowledge in virtual environment training'. In: *Presence* 7.2 (1998), pp. 129–143.
- [301] J. Wang, Y. Chen, S. Hao, X. Peng and L. Hu. 'Deep learning for sensor-based activity recognition: A survey'. In: *Pattern Recognition Letters* 119 (2019), pp. 3–11.
- [302] D. Watson, L. A. Clark and A. Tellegen. 'Development and validation of brief measures of positive and negative affect: the PANAS scales.' In: *Journal of personality and social psychology* 54.6 (1988), p. 1063.
- [303] C. D. Wickens. 'Multiple resources and performance prediction'. In: *Theoretical issues in ergonomics science* 3.2 (2002), pp. 159–177.

- [304] C. D. Wickens. 'Multiple resources and mental workload'. In: *Human factors* 50.3 (2008), pp. 449–455.
- [305] D. A. Wiegmann and S. A. Shappell. *A human error approach to aviation accident analysis: The human factors analysis and classification system*. Routledge, 2017.
- [306] W. W. Wierwille and F. T. Eggemeier. 'Recommendations for mental workload measurement in a test and evaluation environment'. In: *Human factors* 35.2 (1993), pp. 263–281.
- [307] A. F. Winfield, K. Michael, J. Pitt and V. Evers. 'Machine ethics: the design and governance of ethical AI and autonomous systems'. In: *Proceedings of the IEEE* 107.3 (2019).
- [308] C.-H. Wu, Y.-M. Huang and J.-P. Hwang. 'Review of affective computing in education/learning: Trends and challenges'. In: *British Journal of Educational Technology* 47.6 (2016), pp. 1304–1323.
- [309] D. Wu, C. G. Courtney, B. J. Lance, S. S. Narayanan, M. E. Dawson, K. S. Oie and T. D. Parsons. 'Optimal arousal identification and classification for affective computing using physiological signals: Virtual reality stroop task'. In: *IEEE Transactions on Affective Computing* 1.2 (2010), pp. 109–118.

X

- [310] B. Xie and G. Salvendy. 'Review and reappraisal of modelling and predicting mental workload in single-and multi-task environments'. In: *Work & stress* 14.1 (2000), pp. 74–99.

y

- [311] R. M. Yerkes and J. D. Dodson. 'The relation of strength of stimulus to rapidity of habit-formation'. In: *Journal of comparative neurology and psychology* 18.5 (1908), pp. 459–482.
- [312] M. S. Young, K. A. Brookhuis, C. D. Wickens and P. A. Hancock. 'State of science: mental workload in ergonomics'. In: *Ergonomics* 58.1 (2015), pp. 1–17.

Z

- [313] M. Zahabi and A. M. A. Razak. 'Adaptive virtual reality-based training: a systematic literature review and framework'. In: *Virtual Reality* (2020), pp. 1–28.
- [314] R. Zangróniz, A. Martínez-Rodrigo, J. M. Pastor, M. T. López and A. Fernández-Caballero. 'Electrodermal Activity Sensor for Classification of Calm/Distress Condition'. In: *Sensors* 17.10 (2017), p. 2324.
- [315] C. Zeagler. 'Where to wear it: functional, technical, and social considerations in on-body location for wearable technology 20 years of designing for wearability'. In: *Proceedings of the 2017 ACM International Symposium on Wearable Computers*. ACM. 2017, pp. 150–157.
- [316] Z. Zeng, M. Pantic, G. I. Roisman and T. S. Huang. 'A survey of affect recognition methods: Audio, visual, and spontaneous expressions'. In: *IEEE transactions on pattern analysis and machine intelligence* 31.1 (2008), pp. 39–58.
- [317] J. Zhou, S. Z. Arshad, S. Luo, K. Yu, S. Berkovsky and F. Chen. 'Indexing Cognitive Load using Blood Volume Pulse Features'. In: *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. ACM. 2017, pp. 2269–2275.
- [318] F. R. H. Zijlstra. 'Efficiency in work behaviour: A design approach for modern tools'. PhD thesis. 1993.

Titre : Vers l'Exploitation de la Charge Mentale de Travail dans les Systèmes de Réalité Virtuelle

Mots-clés : Réalité Virtuelle, Charge Mentale de Travail, Informatique Physiologique, Formation

Résumé : Malgré l'émergence rapide des systèmes automatisés et le développement de l'informatique affective, très peu d'études considèrent la charge mentale de travail dans la conception de scénarios de formation en RV. Cette thèse a pour objectif de contribuer au développement des systèmes adaptatifs en RV, basés sur la charge mentale de travail des utilisateurs. Nous proposons 3 axes de recherches : induction, reconnaissance, et exploitation de la charge mentale de travail en RV, ainsi qu'une définition de la "Réalité Virtuelle Affective et Cognitive". Dans un premier temps, nous étudierons l'impact du port de casque de RV sur l'effort mental des utilisateurs. De plus, l'influence potentielle de la marche et de l'effet d'accommodation en RV seront analysés. Puis, nous proposerons une approche méthodologique pour introduire

l'évaluation de la charge mentale de travail dans la conception de scénarios de formation en RV. Cette méthodologie permettra notamment de moduler le niveau de charge mentale de travail des utilisateurs au cours du temps. Des études utilisateurs seront menées dans un simulateur de vol en RV afin d'évaluer cette approche. Finalement, nous proposerons une solution tout-en-un afin d'estimer la charge mentale de travail des utilisateurs en temps-réel en utilisant des capteurs intégrés aux casques de RV. Cette configuration sera comparée aux systèmes plus répandus dans le commerce vis-à-vis des performances de prédiction. Les influence du types de mesures, des capteurs, et des méthodes de normalisation des signaux seront également analysées.

Title: Towards the Exploitation of Mental Workload in Virtual Reality Systems

Keywords: Virtual Reality, Mental Workload, Physiological Computing, Training

Abstract: Despite the rapid emergence of autonomous systems and the development of affective computing, little studies considered mental workload in the design of VR training scenarios. This thesis aims to contribute to the development of adaptive VR systems based on the users' mental workload. We propose 3 research axes: induction, recognition, and exploitation of mental workload in VR, and a definition of "Affective and Cognitive Virtual Reality". First, we study the impact of wearing a VR HMD on the users' mental efforts. In addition, the potential influence of walking and accommodation to VR effect are analysed. Then, we propose a methodological approach to intro-

duce mental workload assessment in the design of VR training scenarios in complex immersive virtual environments. This methodology enables to modulate users' mental workload levels over time. User studies in a VR flight simulator are conducted to evaluate the approach. Finally, we propose an all-in-one solution to recognize users mental workloads in real-time using integrated physiological sensors into a VR HMD. This setup is compared to a commercial-grade system with regards to the mental workload recognition accuracy. The influences of the type of measures, sensors, and signal normalization methods is also investigated.