

Représentation partagée et travail collaboratif en contexte C2 : monitoring d'opérateurs en situation simulée de command and control.

Baptiste Prebot

▶ **To cite this version:**

Baptiste Prebot. Représentation partagée et travail collaboratif en contexte C2 : monitoring d'opérateurs en situation simulée de command and control.. Automatique. Université de Bordeaux, 2020. Français. NNT : 2020BORD0227 . tel-03097657

HAL Id: tel-03097657
https://theses.hal.science/tel-03097657

Submitted on 5 Jan 2021

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.

université de BORDEAUX

THÈSE PRÉSENTÉE
POUR OBTENIR LE GRADE DE

DOCTEUR
DE L'UNIVERSITÉ DE BORDEAUX

École Doctorale Sciences Physiques et de l'Ingénieur
Automatique, Productique, Signal et Image, Ingénierie Cognitique
Spécialité Ingénierie Cognitique

Par Baptiste Prébot

Représentation partagée et travail collaboratif en contexte C2 : monitoring d'opérateurs en situation simulée de command and control.

Equipe de recherche Cognitique et Ingénierie Humaine (CIH) – IMS

Sous la direction des Professeurs Bernard Claverie et Jean-Marc Salotti

Soutenue le 04 Décembre 2020

Membres du jury :

Pr. CAUSSE, Mickaël	ISAE-SUPAERO	Rapporteur
Dr. FERRARI, Vincent	Centre de Recherche de l' École de l'Air	Rapporteur
Dr. CHAUDRON Laurent	Centre de Recherche de l' École de l'Air	Examinateur
Pr. DESCHODT-ARSAC, Véronique	IMS, Université de Bordeaux	Présidente
Pr. CLAVERIE, Bernard	IMS, ENSC Institut Polytechnique de Bordeaux	Directeur de thèse
Pr. SALOTTI, Jean-Marc	IMS, ENSC Institut Polytechnique de Bordeaux	Directeur de thèse
M. MASSON, Jean-Daniel	Agence Innovation Défense - DGA	Membre invité
M. SABENE, Julien	Armée de l'air et de l'espace	Membre invité

Lorem Ipsum, Dolor Sit Amet

Résumé

L'évolution des technologies de l'information et de la communication (TIC) a permis le développement du travail collaboratif dans quasiment tous les secteurs de l'activité humaine. Pour assurer la performance du collectif et minimiser le risque d'erreurs, il est crucial que les individus qui collaborent partagent une même représentation de la situation dans laquelle ils sont engagés. Ceci est d'autant plus vrai en situation de crise militaire, comme il en existe au sein des structures de commandement et contrôle d'opérations (C2).

Dans un environnement marqué par la complexité grandissante des conflits, l'enjeu est aujourd'hui à la fois scientifique et hautement applicatif. L'avancée de l'étude de la cognition collective, cœur du travail collaboratif, est porteuse d'un potentiel certain qui doit se traduire par des applications concrètes au service de l'optimisation de la gestion et de la réalisation des tâches collectives. L'évaluation en temps réel de la cognition des individus et des équipes permet d'envisager des outils et des systèmes adaptatifs pour gagner en efficacité, en performance et en agilité. Face à ces enjeux, notre objectif, sur commande de la DGA, est de trouver des mesures appropriées qui permettraient une évaluation de la dynamique du partage des consciences de situation, dans le contexte très contraignant des salles de commandement et de contrôle, qui nécessite la plus faible instrumentation possible des opérateurs.

Notre contribution au domaine est double. D'une part nous proposons le concept de synchronie des consciences de situation, pour soutenir le développement théorique de l'étude de la dynamique de partage de conscience de situation. D'autre part nous mettons en évidence l'importance d'adopter une méthodologie d'ingénierie cognitique, dans une perspective de transposition des connaissances de laboratoire à une situation d'application complexe, s'apparentant autant que possible à l'environnement réel de prise de décision en C2.

Notre travail a alors consisté à explorer des mesures quantitatives du partage de conscience de situation, adaptées à l'exploitation automatisée et en temps réel par un système de diagnostic de la cognition collective. Nous avons mis en pratique l'utilisation du monitoring psychophysiologique et comportemental d'opérateurs engagés dans une tache (individuelle, puis collective) de C2, pour évaluer leur partage de conscience de situation, par l'utilisation de la pupillométrie par oculomètre (*eye tracker*).

Ces études, nous ont amenés à mesurer la sensibilité du monitoring à la dynamique des consciences de situation des opérateurs et à leur partage en environnement écologique.

Ce travail de doctorat se présente comme une mise en avant de l'intérêt et de l'applicabilité de systèmes d'évaluation du partage de cognition en environnement de travail collaboratif réaliste, et s'accompagne de propositions concernant le futur de la recherche sur le C2.

Abstract

Operators monitoring for real-time team cognition assessment in C2 crisis situation

Advances in information and communication technologies has enabled the development of collaborative work in almost all sectors of human activity. To ensure the performance of the group and minimize the risk of errors, it is crucial that the team members share a common understanding of the situation in which they are involved. This is particularly true in military crisis situations, such as those that exist within command and control (C2) structures.

Within an environment characterized by the growing complexity of conflicts, the challenge today is both scientific and highly applicative. Progress in the study of collective cognition, the heart of collaborative work, has a clear potential that must be translated into tangible applications to optimize the management and execution of collective tasks. Real-time evaluation of the cognition of individuals and teams allows to envisage adaptive tools and systems to improve efficiency, performance and agility.

In light of these challenges, our objective, commissioned by the DGA, is to find appropriate measures that would enable an assessment of the dynamics of the sharing of situational awareness, in the very constraining context of command and control room operations, which require the lowest possible level of instrumentation of operators.

Our contribution to the field has been dual. We have proposed the concept of situation awareness synchrony to support the theoretical development of the study of the dynamics of situation awareness sharing. In addition, we have highlighted the importance of adopting a cognitive engineering methodology, in the perspective of transposing laboratory knowledge to a more complex application environments.

Thus, our work consisted in exploring quantitative measures of shared situational awareness, suitable for automated and real-time exploitation by a collective cognition diagnostic system. We applied psychophysiological and behavioural monitoring of operators engaged in a C2 task (individual, then collective), to evaluate their shared situation awareness, using eye tracker pupillometry.

These studies have led us to analyze the sensitivity of this monitoring to the dynamics of the operators' situational awareness and its sharing in an ecological environment.

This doctoral work is presented as a demonstration of the interest and applicability of shared cognition evaluation systems in realistic collaborative work environments, and is supported by proposals concerning the future of research on C2.

Remerciements

Je souhaite commencer par remercier mon directeur de thèse, Bernard Claverie, pour son soutien, à sa manière, et son exigence pyrénéenne qui m'ont poussé à sortir de ma zone de confort et produire un travail dont je peux aujourd'hui dire que je suis fier. Merci également pour les opportunités que vous m'avez offertes.
Je souhaite également remercier mon co-directeur de thèse Jean-Marc Salotti pour son encadrement, ses conseils avisés et pour m'avoir ouvert les portes de l'enseignement.

Véronique Lespinet-Najib et Edwige Clermont, merci pour votre bonne humeur quotidienne mais surtout pour votre écoute et votre confiance.

A Christophe Jauze, merci pour le café.

A Isabelle Sésé, Nadège Rodriguez, Chrystelle Chardron et Denis Plotton, merci pour votre disponibilité et votre bienveillance.

A Benoit Le Blanc, Jérôme Saracco et Jean-Marc André merci pour votre aide le moment venu.

Hervé Le Guyader, merci de m'avoir embarqué dans l'aventure IST, c'est une expérience des plus enrichissantes et intéressantes.

A Gilles Desclaux, merci pour tes enseignements, tes anecdotes d'un domaine qui m'était jusqu'alors inconnu, et surtout tes bons vins.

A Pierrick Legrand et Olivier Chator, je souhaite adresser des remerciements particuliers pour leur soutien technique et leur disponibilité. Je vous dois des verres !

Je remercie respectueusement l'Agence Innovation Défense, la Direction Générale de l'Armement et particulièrement Emmanuel Gardinetti, d'avoir rendu ce projet possible en le finançant.

Dominique Palabost, merci pour ta disponibilité et ton accueil.

A Pascale, merci pour ta disponibilité pour relire ce manuscrit ainsi que tes (très) nombreux et (très) judicieux retours.

Aux étudiants de l'ENSC Caroline Cavel, Laetitia Calice, Adrien Leduque et Mateo Mahaut, merci pour le sérieux de votre investissement dans le projet transdisciplinaire. Vous avez fait du beau travail.

Je souhaite maintenant adresser des remerciements plus personnels à mes amis et mes collègues devenus amis, qui m'ont accompagné au jour le jour.

Tout d'abord, à tous ceux passés par le bureau des doctorants, mais principalement Jessica, Marie, Hélène, Floriane, Théodore, Bryan et Delphine, merci pour toutes les parties de Fronton et de galerapagos au nom de la science.

A Fannaï "Popi" Peslaï, petit ange parti trop tôt. Merci, je ne sais pas pourquoi, mais merci.

A tous les membres de l'IFL, beaucoup de personnes vont penser que le vrai résultat de ces 4 ans réside dans ce manuscrit, peu de gens savent qu'en réalité le Fronton marquera beaucoup plus les esprits. Merci de m'avoir fait faire un peu de sport.

A Pepit', mon pilier dans l'aventure. Merci pour ton dynamisme.

To Jessica, Sven, Jonas, Claudia, and all my teammates and friends, at FKIE, thank you all for the true kindness with which you welcomed me in Wachtberg. You have contributed to this work way more than you imagine.

A Damien et Quentin, plus que des collègues de bureau, plus que des adversaires de Fronton, ce sont de véritables amis que j'ai rencontrés. (RPTV) Merci pour la bonne humeur et la lourdeur qui ont accompagné ces années.

A Florent "El Vesuvio" Bollon, merci pour ton enthousiasme à chaque fois que tu es venu, mais surtout, merci pour tes propos toujours très mesurés. Vamos Los Gringos IHMos !

Virgiles Landeros De Los Reies et Théo Moulièrés-Sebano, mes deux comparses, ça y est, je rejoins votre club. Qui aurait dit ça il y a 10 ans ? Merci pour tout.

A Jeannot le Parigot, merci pour ta joie de vivre et toutes les opportunités de bons repas, de bons vins, de surf et de voyages !

Ben, Quentin, Yayanne, Mel, Fan et Caro, vous avoir retrouvé fait partie de mes plus grandes satisfactions de ces 4 années. C'est pompé sur une autre, mais c'est vrai, vous êtes les meilleurs. J'espère avoir encore beaucoup d'occasions de vous remercier.

Marianne et Flo, mes relecteurs assidus et soutiens sans faille depuis le premier jour, votre écoute à des moments difficiles a été d'une aide incroyable. Merci pour tout. A votre tour maintenant !

La Moul, Mufmuf et Lolit', mes colocs, mes amis, je ne saurais exprimer à quel point je suis reconnaissant pour le soutien que vous m'avez apporté au quotidien. Vous êtes de belles personnes et une encore plus belle rencontre. Comme dirait l'autre : "Tié la famille".

Enfin à mes parents sans qui rien de tout cela n'aurait été possible. Littéralement.
Promis j'arrête les études. Mais grâce à la curiosité et l'esprit d'ouverture sur le monde que vous m'avez transmis, je n'arrêterai jamais d'apprendre.
Merci.

Pour finir, à toi, cher lecteur, merci de prendre le temps d'ouvrir ce document, même si ce n'est que pour y récupérer une source...

Table des matières

Résumé 3
Abstract 5
Remerciements 6
Table des matières 9
Introduction générale 13
Organisation du manuscrit 15
Chapitre 1 Concepts et Objectifs 19
1.1 Le travail collaboratif 19
1.2 C2 et challenges de la crise militaire moderne et future 21
1.2.1 Définitions 21
1.2.2 Mesurer la performance du C2 25
1.2.3 C2 et cognitique 26
1.3 Concepts clés de la cognition d'équipe 29
1.3.1 Vision collective, vision holistique et apport de la macrocognition 29
1.3.2 La notion de charge cognitive 31
1.3.3 La notion de conscience de situation (SA) 33
1.3.4 Lien entre charge cognitive et conscience de situation 35
1.3.5 Team workload 36
1.3.6 Team SA et Shared SA 37
1.3.7 La communication 39
1.4 Méthodes de mesure de la cognition d'équipe 40
1.4.1 Mesurer la charge cognitive individuelle et d'équipe 41
1.4.2 Mesurer la conscience de situation individuelle et d'équipe 45
1.5 Synthèse de la problématique et question de recherche 49
Chapitre 2 Approche temporelle de la différence de partage de conscience de situation 52
2.1 Besoins et limites de l'évaluation de la conscience de situation 52

2.2 SA Synchronie : Vers une évaluation de la dynamique du partage de conscience de situation 53
2.2.1 Modèle de la dynamique de partage de conscience de situation 54
2.2.2 Méthode indépendante et méthode collaborative de synchronisation des consciences de situation 56
2.3 Perspectives pour la mesure de la synchronie des SA 57
2.3.1 Quantifier la synchronie des consciences de situation 58
2.3.2 Qualifier la synchronie des consciences de situation 59
2.3.3 Utiliser la synchronie des consciences de situation 59
2.4 Précision de l'orientation de la recherche 60
Chapitre 3 Explore Mars 62
3.1 Contexte et objectifs de l'expérience Explore Mars 62
3.1.1 Contexte 62
3.1.2 Hypothèses de recherche 64
3.2 Plan d'expérimentation 65
3.2.1 Idée expérimentale 65
3.2.2 L'outil Explore Mars 65
3.2.3 Matériel et participants 66
3.2.4 Déroulement des passations et instructions aux participants 67
3.2.5 Recueil des données 68
3.3 Exploitation des données 71
3.3.1 Exactitude des consciences de situation 71
3.3.2 Similarité des consciences de situation 71
3.3.3 Questionnaire 71
3.4 Résultats 72
3.4.1 Partage de conscience de situation et réussite 72
3.4.2 Partage de conscience de situation et performance temporelle 74
3.4.3 Résultats complémentaires concernant le Travail Collaboratif 77
3.5 Discussion des résultats 78
3.5.1 Conclusion 78
3.5.2 Discussion et pistes de recherche 79
Chapitre 4 Monitoring Individuel - Expérience AMIGOS 82
4.1 Contexte et objectifs de l'expérience AMIGOS 82
4.1.1 Contexte de collaboration 82
4.1.2 Expérience AMIGOS 83
4.1.3 Problématique et hypothèses de recherche 85

4.2 Plan d'expérimentation 87
4.2.1 La tâche de surveillance d’espace aérien 87
4.2.2 Les Variables 89
4.2.3 Participants 90
4.2.4 Matériel 90
4.2.5 Déroulement des passations et instructions aux participants 91
4.3 Exploitation des données 92
4.3.1 Sélection des tâches 92
4.3.2 Méthode de détection de la focalisation visuelle 92
4.3.3 Extraction des données pupillaires 94
4.3.4 Processus de préparation des données pupillaires : sélection des cas exploitables et nettoyage 94
4.3.5 Dilatation Pupillaire 100
4.3.6 Résumé du processus de prétraitement des données pupillaires 100
4.4 Résultats 102
4.4.1 Présentation des données 102
4.4.2 Approche quantitative 104
4.4.3 Approche qualitative 105
4.5 Discussion des résultats 114
4.5.1 Conclusion 114
4.5.2 Discussion : monitoring, variabilité et interprétabilité. 115
Chapitre 5 Monitoring collectif - Expérience MOCOM 123
5.1 Contexte et objectifs de l'expérience 123
5.1.1 Contexte 123
5.1.2 Hypothèses de recherche 124
5.2 Plan d'expérimentation 126
5.2.1 La tâche collaborative : Wargame Red Dragon 126
5.2.2 Participants 129
5.2.3 Matériel 129
5.2.4 Déroulement des passations et instruction aux participants 130
5.2.5 Les variables 132
5.3 Exploitation des données 132
5.3.1 Synchronisation des données pupillométriques 132
5.3.2 Prétraitement des données pupillaires 133
5.3.3 Méthodologie d’observation comportementales : accord inter-juges 133
5.4 Résultats 134

5.4.1 Latence de synchronie des SA. 135
5.4.2 Similarité des dynamiques de dilatation pupillaire et qualification de la collaboration 135
5.4.3 Ébauche d'approche quantitative de la similarité. 144
5.5 Discussion des résultats 145
Chapitre 6 Contextualisation et perspectives 149
6.1 L'opérateur en centre de commandement et contrôle 149
6.1.1 Le poste de l'opérateur en JFAC 150
6.1.2 Le poste de l'opérateur en CDC 154
6.1.3 Applicabilité du monitoring aux opérations et au contrôle aérien : de l'observation à la recommandation 157
6.2 Expérimenter pour l'évaluation de la cognition collective en C2 164
6.2.1 Monitorer la cognition collective : quelles améliorations pour l'experimentation ? 164
6.2.2 Collaborative Warship Commander Task : proposition d'un outil collaboratif pour l'expérimentation sur le partage de conscience de situation 167
6.2.3 Proposition de recherche 174
6.3 Enjeux de la recherche pour le futur du C2 176
6.3.1 De l'intelligence naturelle à la cognition augmentée en C2 176
6.3.2 La complexité compréhensible, flagship de l'Europe de la défense 181
6.3.3 Cyberpsychologie et conflits du futur 185
Conclusion 187
Glossaire 193
Productions scientifiques et techniques 195
Bibliographie 198
Annexes 214

Introduction générale

Les progrès technologiques et la complexité croissante des environnements de travail ont fait évoluer le domaine de l'Interaction Homme-Système vers celui de l'Intégration Homme-Système. La notion d'intégration traduit la volonté de progression vers une organisation Homme-Machine symbiotique, où la compréhension mutuelle est le socle de l'hybridation réciproque et complémentaire : de l'humain par la machine et de la machine par l'humain. Dans cette optique, le cœur de la réflexion est orienté vers le développement d'outils et de méthodes permettant au système et à l'humain une « compréhension » mutuelle.

L'axe de recherche le plus problématique est celui de la « compréhension » de l'humain par la machine. Ce domaine a été ouvert récemment. Il est tourné notamment vers la mesure et l'analyse des états physiques, physiologiques et cognitifs des individus par le système afin que ce dernier puisse évoluer en fonction de l'état de l'humain et s'y adapter (systèmes adaptatifs). Tout comme le ferait un coéquipier, ces systèmes ont de ce fait pour mission d'assister les grandes fonctions de perception, de représentation et de prise de décision.

Dans ce contexte, nous présentons une thèse qui s'intéresse aux enjeux scientifiques des connaissances sur la cognition, et notamment la cognition collaborative dans la prise de décision. Elle répond à une demande socio-économique portée par la DGA, sur des apports scientifiques pour l'évolution du Command and Control (C2), dont l'avenir est justement dans cette collaboration entre humains et technologies.

Le C2, système de gestion de crise dont dépend entre autres la sécurité quotidienne de la France, est le lieu d'une évolution majeure motivée par les perspectives technologiques qu'offrent les outils d'IA, de *Big Data* et de *cognitive computing*.
Pour répondre à la complexification des situations de crise, cette structure névralgique centrale des opérations militaires cherche à se doter de systèmes d'aide à la prise de décision et d'outils technologiques permettant le développement d'une compréhension de la situation plus précise et partagée. Ces outils cognitifs (*cognitive tools*) doivent augmenter le C2 pour lui fournir un avantage informationnel et décisionnel au service de la réactivité et de l'agilité (Alberts & Hayes, 2003) nécessaires dans des conflits de plus en plus complexes, tentaculaires, technologiques et imprévisibles.

Le contexte applicatif du C2, domaine de travail collaboratif des plus exigeants, à particulièrement été traité en sciences politiques et en sciences organisationnelles mais peu de travaux s'y intéressent en sciences cognitives.

Bien que les structures de commandement et contrôle d'opérations soient naturellement empreintes de la marque de la collaboration, les rares travaux de ce domaine concernent principalement l'opérateur seul. L'étude de la cognition d'équipe, historiquement partagée entre la vision individualiste (l'équipe est la somme des individualités) et la vision holistique (la cognition de l'équipe détermine les cognitions individuelles), se retrouve depuis peu « unifiée » par le champ de la macrocognition[1] d'équipe (Letsky & Warner, 2017).

Ce dernier prône une approche pragmatique et écologique pour l'évaluation concomitante des deux points de vue grâce à des moyens adaptés. Il est souvent associé au courant du *Naturalistic Decision Making* (NDM) (Klein, 2008), qui distingue l'approche de la psychologie de laboratoire, en situation contrôlée, de celle de l'étude de l'individu ou du groupe en situation écologique ou du moins réaliste. Une grande partie de notre approche vise donc à réfléchir à la transposition des connaissances de laboratoire à une application écologique au collectif en C2.
C'est en cela que cette thèse revendique à la fois sa nécessité et son originalité.

En 2016, l'ENSC[2] et l'équipe CIH[3] de l'IMS[4] ont amorcé un mouvement de recherche autour de la cognition collective en C2. Au-delà des données permettant une connaissance fondamentale des processus de partage et de collaboration dans des situations critiques de type C2, les travaux présentés dans ce document, en tant qu'initiateurs de cette dynamique de recherche, ont permis d'ouvrir et de défricher l'expérimentation pour ce domaine.

Dans cette tâche nous avons rencontré deux difficultés majeures :

- La demande est extrêmement « sensible », frappée par le confidentiel ou le secret défense. Elle nécessite une expérience et une culture du domaine de la Défense que j'ai dû acquérir et assimiler dans le cadre de formations C2 de l'Armée de l'air et de l'espace. Si cet apprentissage est intéressant, il n'est pas évident pour un civil et il a fallu acquérir la confiance des spécialistes et des militaires.

- A un niveau purement scientifique, le concept de cognition collective n'est que peu étudié. Comparativement aux connaissances acquises sur l'opérateur Homme-Machine, la dimension du partage n'est que peu traitée. La difficulté de son étude est critique. C'est à la fois l'opérateur, en lien direct avec la machine en tant que support, outil et coéquipier, mais également l'opérateur dans son interaction au et avec le groupe, qu'il s'agit d'étudier. Cette démarche demande la mise en place de structures de recherche complexes, lourdes expérimentalement à acquérir et à maîtriser.

[1] Klein et al. (2000) ont défini la macrocognition comme un domaine distinct d'étude des processus cognitifs des personnes qui font face à des prises de décisions difficiles dans des contextes complexes (naturels), sous la pression du temps et dans l'incertitude (ex, pompiers, pilotes, infirmières) (pour un historique complet, voir Hoffman & McNeese, 2009).
[2] Ecole Nationale Supérieure de Cognitique
[3] Cognitique et Ingénierie Humaine
[4] Laboratoire d'Intégration du Matériau au Système

Notre travail de recherche ambitionne tirer parti des connaissances concernant l'évaluation de la cognition collective, et les enrichir au profit de la définition et de la conception de systèmes visant à améliorer la performance par une assistance outillée des processus cognitifs.

En 2014, l'évaluation de la cognition d'équipe en situation écologique est officiellement identifiée par un groupe de travail de l'OTAN[5] comme « *a necessary next step* » pour le C2 (Berggren, Kermarrec, Banko, Wikberg, & Oleksandur, 2014). Dans toute situation de travail collaboratif, le partage d'une représentation commune de la situation courante est un facteur de performance de la prise de décision d'équipe. La notion de représentation soulève nécessairement la question de la représentation mentale et, de manière plus large, celle de la conscience.

En particulier en ce qui concerne son partage entre des coéquipiers qui collaborent, la représentation est à la fois l'objet et le vecteur du partage de conscience. En nous intéressant au partage de représentation il est donc nécessaire de commencer par se poser la question de la conscience et de la nature de la cognition.

Parmi d'autres travaux (le lecteur pourra trouver un excellent recueil critique, à propos des neurosciences et de la neuropsychologie, dans Missa (1993)), les écrits de Bergson (1939) disposent que la première fonction de la conscience est de retenir ce qui n'est déjà plus et d'anticiper ce qui n'est pas encore, dans le but d'effectuer un choix. En termes modernes, on parlera d'expérience et d'anticipation pour la prise de décision. Ils ancrent ainsi la conscience dans une dualité que rejoint Nagel (1989), pour qui elle est un phénomène incarné, une expérience sensible personnelle qui est à l'origine du sentiment d'un « effet que cela fait » d'être. Cette dualité réside dans le caractère subjectif de l'expérience phénoménale et oppose deux points de vue de la même chose, l'un extérieur et notamment celui de l'étude des activités biologiques, neurobiologiques et physiologiques, l'autre correspondant à l'expérience vécue, subjective et psychologique.

L'utilisation, certes limitée, du monitoring de l'opérateur comme moyen d'étude de sa cognition, répond pourtant à cette double approche, à la fois physiologique et psychologique.

Cette thèse propose donc d'étudier pourquoi et comment le monitoring peut être utilisé en tant que moyen d'évaluation du partage de cognition d'opérateurs de C2, dans une perspective d'utilisation des données en temps réel par des systèmes d'entraînement ou des systèmes adaptatifs.

Ce travail de recherche aboutit d'une part à des résultats expérimentaux et d'autre part à des recommandations théoriques, dont ce mémoire écrit se veut le témoin. Chaque chapitre présentera donc des intérêts scientifiques et des intérêts pour le C2.

Organisation du manuscrit

Ce mémoire de thèse comporte six chapitres s'articulant autour des hypothèses, méthodes et résultats de trois expérimentations.

Le **Chapitre 1** présente le C2, imprégné par le travail collaboratif, et les enjeux liés à la cognitique[6] pour ce domaine. Il identifie et détaille ensuite les composantes clés ainsi que des méthodologies d'étude

[5] Panel Human-Factors and Medicine (HFM) de la Science & Technology Organization (STO)
[6] La cognitique est une science interdisciplinaire qui s'intéresse au traitement de la connaissance, qu'elle soit humaine ou informatisée, et à leurs interactions (Claverie, 2005)

de la cognition collective. Cette revue de littérature accompagnera le lecteur dans la compréhension de la problématique amenant à la formulation de notre question de recherche

Le **Chapitre 2** sera plus particulièrement dédié à une réflexion autour de la dynamique du partage de conscience de situation. Après avoir identifié la criticité de ce concept au regard de l'évaluation en temps réel de la cognition collective, nous chercherons à comprendre en quoi la nature même de la conscience de situation limite les méthodes de mesure actuelles. Cette réflexion nous amènera à la présentation d'un nouveau concept de « synchronie des consciences de situation », servant son étude. Nous en préciserons donc les hypothèses de recherche.

Les trois chapitres suivant développent le travail expérimental réalisé durant notre recherche.

Ce travail repose sur les trois concepts majeurs que sont :

- la situation de travail collaboratif pour l'étude de la cognition collective ;
- la conscience de situation et son partage ;
- le monitoring d'opérateur ;

l'ensemble appliqué à un contexte particulier, celui du commandement et contrôle d'opérations militaires (C2).

Les trois expériences s'articulent autour de ces trois concepts, la Figure 1 montre leur complémentarité.

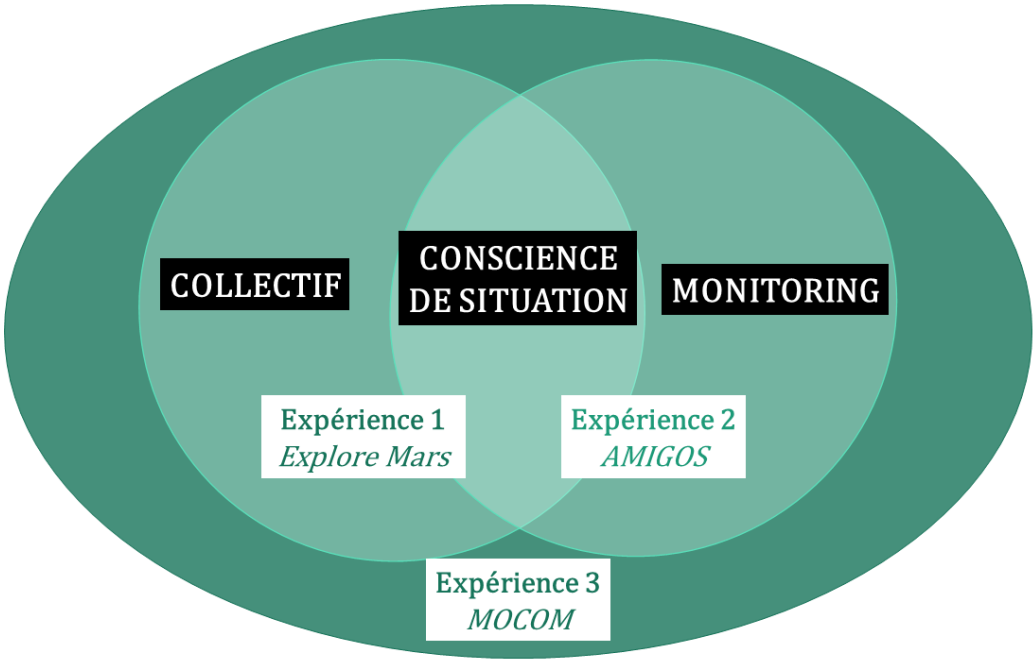


Figure 1. Organisation des expériences autour des concepts clés de la thèse.

Le premier chapitre expérimental (**Chapitre 3**) se concentre sur l'évaluation du partage de conscience de situation de deux individus collaborant dans une tâche de commandement, sans monitoring. Cette expérience a pour but d'étudier l'utilisation d'une métrique objective issue de méthode classique d'évaluation de la conscience de situation, dont le traitement soit automatisable.

Le second chapitre expérimental (**Chapitre 4**) prend la mesure de l'utilisation du monitoring pour l'évaluation de la conscience de situation d'un individu seul impliqué dans une tâche de C2. Pour cela,

nous présentons l'expérience réalisée en collaboration avec le laboratoire allemand FKIE[7].

Le **Chapitre 5**, troisième chapitre expérimental, tentera de conjuguer les trois contraintes en étudiant l'utilisation du monitoring pour évaluer le partage de conscience de situation dans une tâche de travail collaboratif de type C2.

Le **Chapitre 6**, à la suite, ouvre trois points de réflexion. D'une part, il traite de l'applicabilité de nos recherches au travail en C2. D'autre part, il propose des recommandations sur la structure de recherche nécessaire à l'expérimentation sur l'évaluation de la cognition collective par monitoring. Enfin il recontextualise les travaux au regard des enjeux technologiques et humains pour le C2, et émet des projections sur la place de la cognition dans les futurs conflits.

En résumé, cette thèse s'attèle à l'étude du partage de représentation pour la collaboration, possédant des implications concrètes pour la gestion des équipes en situation de crise. Plus particulièrement, l'intérêt de la DGA se porte sur un C2 du futur où il serait possible de détecter, en temps réel, des différences de représentations entre opérateurs. Ce besoin nous amène à nous concentrer sur la partage de représentation de la situation en cours des individus qui collaborent, c'est à dire leur conscience de situation, concept central du domaine de la cognition collective et fil conducteur de nos travaux.

Nous soutenons l'idée qu'il est théoriquement possible d'estimer le niveau de partage de représentation de la situation des coéquipiers, et d'identifier des divergences, par la mesure de la synchronisation de leurs activités (physiologiques et comportementales). Afin d'ancrer nos travaux dans un socle scientifique cohérent, nous avons introduit le concept de « synchronie des consciences de situation ». Cela nous a permis d'éprouver nos hypothèses via trois situations expérimentales, une individuelle et deux collectives, composées de tâches de C2 réalistes, en environnement de plus en plus complexe.
Cette mise à l'épreuve a révélé la sensibilité du monitoring aux modifications de conscience de situation, mais aussi la complexité de sa mise en œuvre en tant que méthode de mesure pourtant la plus adaptée à l'environnement de C2. Cela est venu enrichir nos réflexions sur le futur du C2 et de la recherche le concernant.

Bien conscients que la tâche restante est immense, ce document témoigne de quatre années[8] de recherche intense, dans un milieu difficile, sur une problématique peu étudiée, et pour des enjeux de sécurité globale. Nous sommes heureux qu'il contribue à la connaissance d'une problématique cognitique originale, et qu'il ouvre la voie à une recherche duale et à des applications concrètes.

Le schéma présenté en Figure 2 reprend l'organisation du document.

[7] Institut allemand Fraunhofer for Communication, Information processing and Ergonomics
[8] Trois années financées par la DGA, et une par Bordeaux INP dans le cadre de la convention ENSC/Armée de l'air et de l'espace/OTAN.

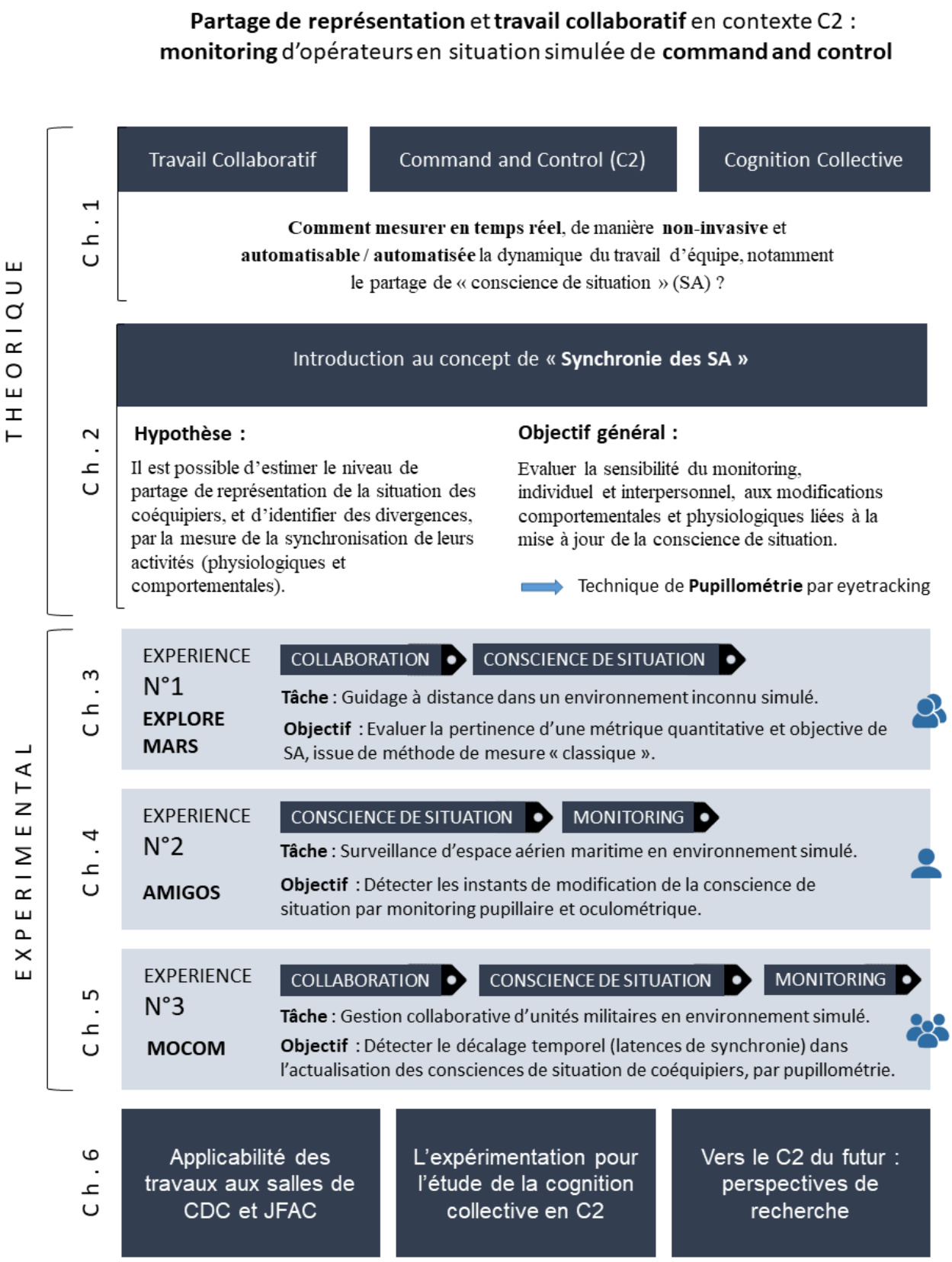


Figure 2. Schéma organisationnel de la thèse.

Chapitre 1 Concepts et Objectifs

Nous ne saurions débuter ce travail sans apporter en préambule des éléments de compréhension de la problématique illustrant notre démarche. Nous développerons donc les concepts de travail collaboratif, de commandement et contrôle (C2), de cognition collective ainsi que les notions qui leur sont associées. Après avoir présenté des méthodes actuelles de mesure de la cognition collective dans leur diversité et leurs contraintes, nous aboutirons à la formulation des premières hypothèses de travail.

1.1 Le travail collaboratif

Il n'existe aujourd'hui quasiment plus de poste de travail isolé. Tant dans la conception, que dans le contrôle, l'administration ou la gestion de dispositifs industriels, le travail collaboratif, forme de travail qui associe ou implique un nombre variable d'acteurs individuels ou de groupes d'individus dans une tâche concourante, est de plus en plus répandu. Ainsi la co-construction, la co-conception ou encore le co-management ont conquis la majorité des secteurs socio-économiques modernes.

Le travail collaboratif peut être formellement défini comme un ensemble de processus mis en jeu avec méthode, par des hommes d'expertises diverses qui exploitent des moyens dans le but commun de concevoir, de réaliser, de mettre en œuvre ou d'effectuer un ensemble de produits de transformation. Ces produits peuvent être d'ordre matériel ou immatériel, c'est-à-dire de nature physique ou intellectuelle. Ils peuvent être conçus par certains comme une somme de travaux individuels identifiés en relation ou organisés entre eux, ou pour d'autres comme un ensemble unitaire dans lequel la partie relative à chaque travailleur est plus abstraite.

Néanmoins, ce domaine est très faiblement instrumenté et rarement étudié de manière objective au sens de l'ingénierie cognitique.

La collaboration peut être vue comme la forme intégrative de la coopération. Schmidt (1991) propose en effet trois formes de coopération :

- la forme **augmentative**, où les agents ont le même savoir-faire mais doivent être plusieurs à effectuer la même tâche pour absorber la charge de travail (ex : deux banquiers au guichet.) ;
- la forme **débative**, dans laquelle les agents ont le même savoir-faire et réalisent une même tâche indivisible (ex : deux médecins discutant d'un patient) ;
- la forme **intégrative** où les agents ont des savoir-faire différents et complémentaires, chacun réalise une sous-tâche de la même tâche principale (ex : un anesthésiste et un chirurgien lors d'une même opération).

La définition de cette dernière forme de coopération se rapproche de celle de la notion d'équipe donnée par Salas et al. (1992), nous permettant d'assimiler ici le travail collaboratif au travail d'équipe. Ces derniers définissent en effet une équipe comme :

"a distinguishable set of two or more people who interact dynamically, interdependently, and adaptively toward a common and valued goal/object/mission, who have each been assigned specific roles or functions to perform, and who have a limited life span of membership" (pp. 126-127)[9].

Une équipe est donc une entité psychosociale composée de plusieurs membres possédant (1) des tâches interdépendantes, (2) des interactions sociales dynamiques, (3) un but commun, (4) des rôles et expertises spécifiques et distribués, (5) une durée d'appartenance limitée. L'équipe n'existe qu'au regard de la tâche qui lui est confiée et dans la temporalité de celle-ci. Elle est formée pour accomplir un but précis et peut au besoin être restructurée afin que l'expertise de ses membres serve la stratégie mise en place.
Pour membre de l'équipe on peut alors distinguer deux types activités (Bowers & Braun, 1997; Gllckman et al., 1987), appelées en anglais *taskwork* et *teamwork* [10]. Le *taskwork* est la partie individuelle de la réalisation de la tâche, ne nécessitant pas d'interaction avec d'autres membres de l'équipe. Elle fait appel au rôle et à l'expertise propre de l'individu. Le *teamwork* quant à lui fait référence à la partie de la tâche demandant une interaction avec l'équipe afin de coordonner la performance de celle-ci. C'est une partie relationnelle de partage, de communication et de leadership.

Ainsi la performance d'équipe peut être définie comme l'ensemble des processus de management de ces deux activités durant la tâche. L'évaluation des résultats traduit l'efficacité de l'équipe tandis que l'évaluation des ressources engagées traduit son efficience.

La performance imputable au travail d'équipe est une question centrale pour les organisations modernes chargées de tâches de sécurité, civiles comme militaires, et revêt une importance toute particulière pour le domaine du Command and Control (C2).

La notion de collaboration est d'ailleurs au cœur du devenir du C2. Les organisations militaires modernes se sont adaptées et ont évolué au cours des siècles. Pour répondre à la complexité grandissante des missions du XXI^{e} siècle, le C2 s'éloigne de la représentation historique de la chaîne de commandement unique et du flux d'information très contraint qui lui était associé. Il adopte progressivement une structure en réseau (Figure 3), centrée sur le partage de l'information et de la connaissance.

[9] « un ensemble de deux personnes ou plus qui interagissent dynamiquement, de manière interdépendante et adaptative envers un objectif commun et valorisé, qui ont chacun reçu des rôles ou des fonctions spécifiques à remplir et dont la période de participation est limitée dans le temps »
[10] Bien qu'une traduction littérale ne soit pas disponible, *taskwork* pourrait être traduit par « travail de l'équipe » et *teamwork* par « travail en équipe ». Cette traduction permet de conserver la distinction entre le travail relatif aux tâches propres au but de l'équipe (*taskwork*), et le travail relatif aux tâches de collaboration nécessaires pour l'accomplir (*teamwork*).

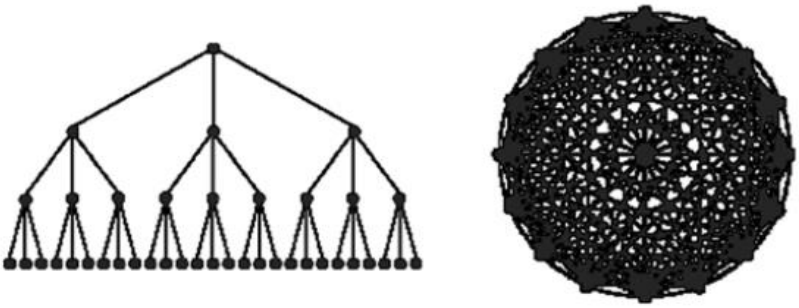


Figure 3. La structure hiérarchique héritée de l'ère industrielle (gauche) présente très peu de connexions comparée à une structure inspirée de l'ère de l'information (droite), pleinement connectée possédant pourtant le même nombre de nœuds (Alberts & Hayes, 2003).

Dans *Power to the Edge*, Alberts et Hayes (Alberts & Hayes, 2003), théoriciens reconnus du C2, affirment que l'agilité requise pour gérer les conflits modernes repose en grande partie sur la collaboration et les représentations partagées de la situation entre les différents nœuds du réseau. Ils promeuvent ainsi le développement d'une approche organisationnelle collaborative du commandement et contrôle.

1.2 C2 et challenges de la crise militaire moderne et future

1.2.1 Définitions

Le C2 (Commandement et Contrôle ou *Command and Control*) est un ensemble de procédures, d'outils et de moyens de haute technologie conçus pour la gestion d'une situation complexe (gestion de crise, conduite d'opérations, etc.). Il permet de planifier, de coordonner, de diriger et de contrôler l'ensemble des ressources, des informations, des individus et des organisations à disposition, vers l'accomplissement d'une tâche, d'un objectif ou d'un but. Une telle structure peut être permanente, (pour la conquête spatiale ou la surveillance de l'espace aérien ou naval national), ou mise en place pour gérer une situation de crise ponctuelle (lors de grandes catastrophes naturelles ou de conflits militaires). Il reste cependant un concept privilégié du domaine de la défense.

Claverie et Desclaux (2015), dans la lignée de Alberts et Hayes (Alberts & Hayes, 2006), différencient clairement le C2, opération théorique, du C4ISR-TAR, conditions et moyens concrets, techniques et humains mis en œuvre. Le C2 est constitué de grands ensembles informatiques mobilisés dans la réception des informations, dans leur analyse, dans l'aide à la décision, dans la transmission des informations et des directives, dans la commande de l'action et son contrôle pour ajuster cette action en permanence aux objectifs recherchés. Il est également défini comme une organisation de gestion et un ensemble de procédures. Cette seconde approche permet de déterminer un ensemble hiérarchique et de délégation en commandeurs et opérateurs pour organiser l'action des forces. Ces deux dimensions complémentaires, l'une structurelle et l'autre fonctionnelle, constituent ce que l'on désigne aujourd'hui sous l'acronyme C4ISR-TAR. Cette structure associe aux caractéristiques du Commandement et de la conduite des opérations les dimensions et les outils de *Control*, *Communication*, *Computers*, *Intelligence*, *Surveillance*, *Reconnaissance*, *Target Acquisition* et *Reconnaissance* (acquisition de cibles).

Claverie et Desclaux conçoivent alors le C2 comme une machine théorique visant à produire les effets

escomptés d'une stratégie. Cette machine repose sur trois piliers, présentés comme les trois côtés du triangle de la Figure 4, permettant l'entrée, la transformation puis la sortie d'information traitée pour commander l'action de modification de l'environnement. La dominance informationnelle (*information dominance*), la confiance accordée aux systèmes numériques (*cyber confidence*), et la supériorité décisionnelle (*decision superiority*), doivent assurer la traduction des objectifs politiques et de la stratégie en effets désirés dans la constante de temps nécessaire et avec la précision temporelle souhaitée.

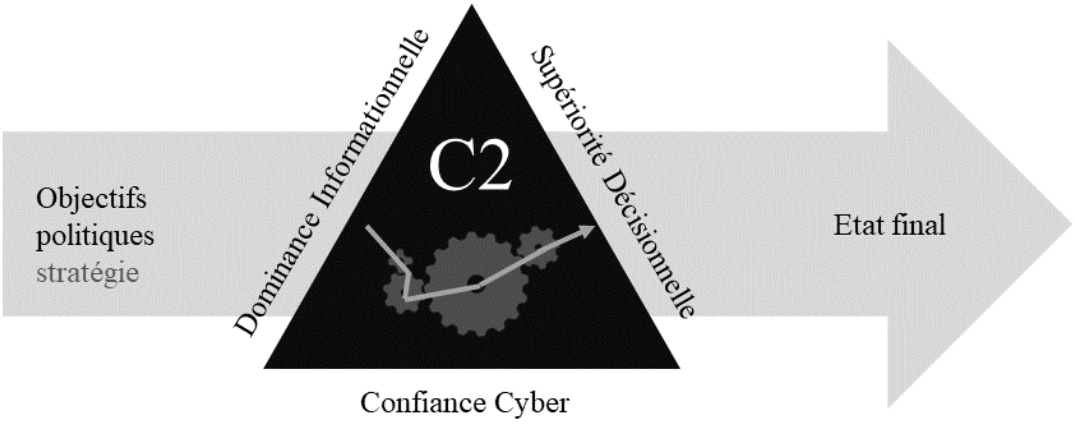


Figure 4. La machine théorique du C2 (Claverie, Desclaux 2015)

Le C2 est donc une machine décisionnelle, dans laquelle des opérateurs traitent l'information au regard de la stratégie et des objectifs finaux afin de déterminer les actions les plus pertinentes dans une temporalité donnée. Ce concept est traditionnellement illustré par la boucle OODA (*OODA Loop*, Figure 5), pour *Observe*, *Orient*, *Decide* and *Act* (« observer, orienter, décider et agir ») (Boyd, 1976, 1987). Il s'agit d'une adaptation aux situations rapides de la roue de Deming[11] (Deming, 2019), apparue dans le domaine du management de la qualité dans les années 1950. Boyd formalise ainsi initialement le cycle de décision permettant d'obtenir la supériorité face à un pilote de chasse ennemi.

Désormais valable à différents niveaux d'abstraction, cette boucle reprend les grandes étapes de toute prise de décision au sein du C2, et en intègre son aspect continu et séquentiel. Cette représentation suggère que la réalisation de cette boucle plus rapidement que l'adversaire permet de lui apparaître imprévisible et ainsi que prendre l'avantage en l'empêchant de construire une représentation claire de la dynamique des forces qu'il affronte.

[11] La roue de Deming est une transposition graphique de la méthode de gestion de la qualité dite PDCA (*plan*, *do*, *check*, *act*, « préparer, développer, contrôler, agir »), moyen mnémotechnique permettant de repérer avec simplicité les étapes à suivre pour améliorer la qualité dans une organisation.

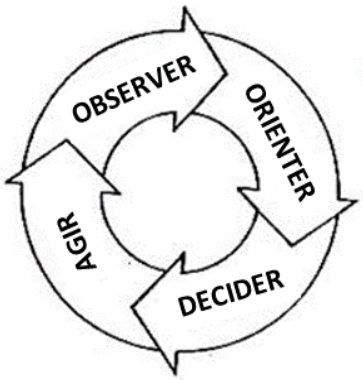


Figure 5. Boucle OODA ou cycle de Boyd

Cependant, sa représentation simplifiée fait l'objet de certaines critiques. En tant que modèle de la prise de décision, elle offre une granularité trop faible pour expliquer les différents processus cognitifs en jeu. Elle n'illustre également pas les itérations et boucles au sein et entre chaque phase, ni la dimension d'équipe, pourtant fondamentale au C2 (Breton & Rousseau, 2007).

Ainsi, certaines variantes ont été proposées. Brehmer (2005) et Baroutsi (2015) mettent en avant trois fonctions critiques du C2 et leurs relations (Figure 6).

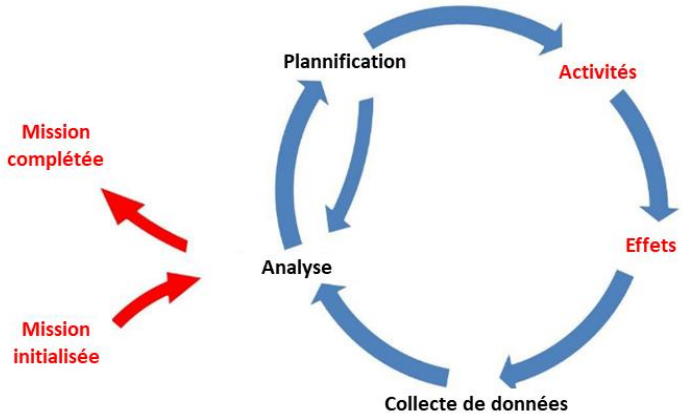


Figure 6. La OODA-loop *dynamique illustrant les trois fonctions critiques de C2 et leurs relations logiques, d'après Baroutsi 2015 adapté de Brehmer 2005.*

Desclaux (2016 - Figure 7) décrit les phases du processus par les actions à réaliser durant celles-ci.

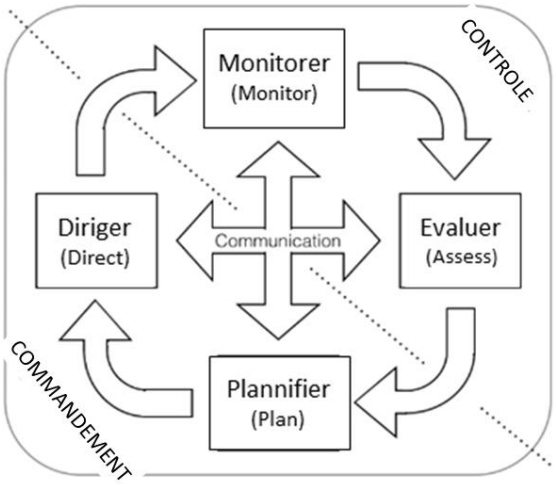


Figure 7. Cycle de la décision en C2, Desclaux 2016

Quel que soit le modèle, quatre grandes séquences restent identifiables : une phase de collecte de données (*Observe/Data collection/Monitor*), une phase d'interprétation de ces données (*Orient/Sense-making/Assess*), une phase de Planification (*Decide/Planning/Plan*) et une phase d'action (*Act/Activities & Effects/Direct*), propres à toute boucle décisionnelle.

S'il est évident qu'un modèle générique ne peut aborder toutes les dimensions du C2 théorique, les autres propositions permettent de rendre compte de la séquentialité du C2 actuel. Dans ce dernier cas le cycle décisionnel reflète la lourdeur, le manque de réactivité et de flexibilité des processus. Il apparait peu adapté aux conflits modernes, à l'agilité et la vitesse de réponse nécessaires. Même si les boucles du modèle actuel sont itératives, le temps de réponse n'est plus adapté aux exigences de rapidité. Leur séquentialité est un frein alors qu'une simultanéité et une permanence des processus est nécessaire.

La structure actuelle du C2 peut également être vue sous l'angle de la temporalité. Les doctrines militaires définissent deux boucles intriquées avec différentes échelles temporelles de réalisation.

Le premier temps est celui de la « Stratégie à la Tâche » (*Strategy to Task* - S2T) : les directives stratégiques sont traduites en objectifs et tâches concrètes à réaliser. C'est une phase d'analyse des objectifs et de planification, en temps longs.

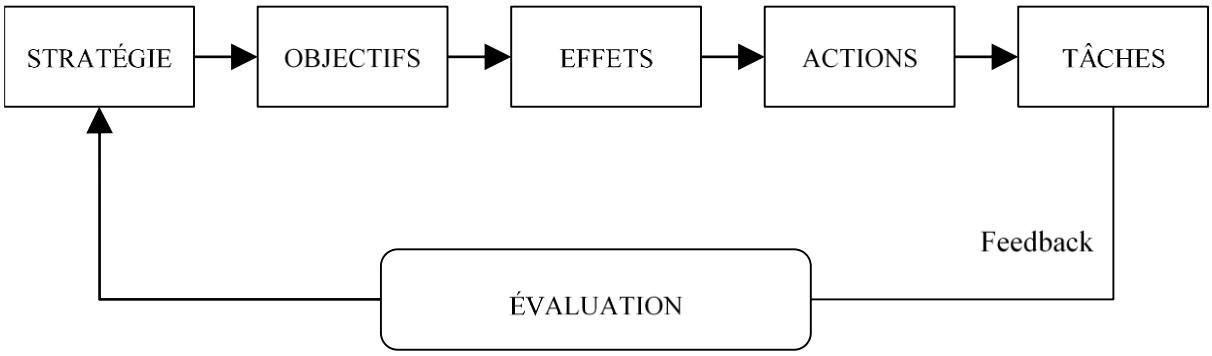


Figure 8. Boucle « Strategy to Task » (S2T), d'après Desclaux et al. (2016a)

Le second temps est celui du temps réel, des « Données à la prise de Décision » (*Data to Decision* - D2D), transformant les données brutes perçues dans l'environnement en informations sur lesquelles baser la prise de décision et leur exécution.

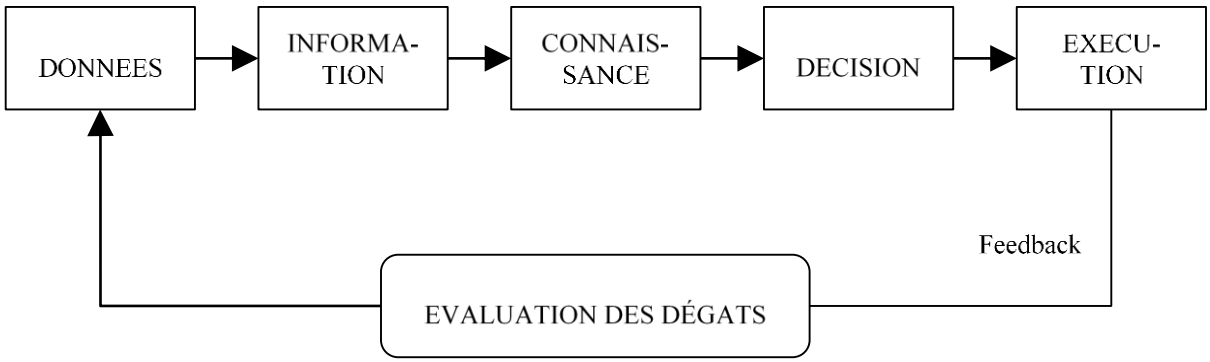


Figure 9. Boucle « Data to Decision » (D2D), d'après Desclaux et al. (2016a)

Le contexte concret dans lequel s'ancre nos travaux de recherche est celui du contrôle des opérations aériennes. Nos travaux portent donc sur le travail fourni par les opérateurs utilisant les systèmes technologiques et collaborant en « temps réel » (temps court), dans cette situation de réponse immédiate et d'action du C2.

1.2.2 **Mesurer la performance du C2**

La question de la performance du C2, et par extension de sa mesure, est un thème récurrent dans la littérature. Comme tout processus, l'amélioration de son efficacité est un but continu source de nombreux travaux, études et frameworks. Les recherches sur le C2 peuvent être réparties en deux types : recherches descriptives et recherches normatives.

Les recherches *descriptives* étudient la façon dont le C2 est effectivement réalisé. Les recherches *normatives* concernent la façon dont il devrait être conduit. Alors que, selon la définition officielle du DoD (Département de la Défense américain), la seule façon d'évaluer la qualité du C2 est de l'assimiler à l'accomplissement de la mission (Alberts & Hayes, 2006), les recherches normatives cherchent à

identifier les fonctions critiques qui devraient être réalisées pour que le C2 achève son but de manière optimale. En abordant le C2 comme un système complexe, il est alors possible d'en différencier les fonctions critiques et de les évaluer au regard de buts individuels (Brehmer, 2007).

L'évaluation du C2 souffre évidemment aussi des mêmes limitations que tout autre domaine. Ainsi la dichotomie classique entre étude en laboratoire et étude en situation écologique est d'autant plus problématique que le C2 militaire est un environnement complexe pour lequel le dynamisme est à la fois condition nécessaire au réalisme des tâches et du contexte, et source de manque de reproductibilité des conditions d'expérience militaire et donc du caractère généralisable des résultats.

Plus la situation est contrôlée, plus les mesures sont scientifiquement valides et fiables, mais en contrepartie, l'éloignement des conditions réelles restreint d'autant l'applicabilité de ces mesures.

En 2014, le groupe de travail HFM[12]-156 de la STO[13] de l'OTAN publie une étude détaillée sur les méthodes de mesure actuelles du C2 (Berggren et al., 2014). Ces méthodes se focalisent sur l'évaluation des résultats et des processus, permettant de mesurer ce qui arrive (*what*) et pourquoi (*why*). Cependant Berggren *et al*. précisent que le développement des mesures psychophysiologiques et des processus cognitifs doit permettre d'évaluer le comment (*how*) et le quand (*when*), aspects également centraux pour l'aide à la décision.

De telles méthodes, bien que difficiles à mettre en place, sont de plus en plus associées aux méthodes classiques telles que l'observation directe, les questionnaires ou l'utilisation d'indicateurs et de checklists, et apparaissent comme les futures méthodes de mesure incontournables du C2. De cette étude ressort entre autres le besoin de développer les méthodes de mesure directes et non invasives psychophysiologiques en neurosciences cognitives ou cognition augmentée et les méthodes de recherche en situation écologique. Nous verrons plus loin en quoi le sujet de cette thèse s'ancre pleinement dans les recommandations de recherche formulées par les chercheurs de l'OTAN.

1.2.3 C2 et cognitique

Les processus de collaboration et de décision mis en œuvre dans le C2 reposent sur la mobilisation de nombreux acteurs, humains ou informatiques, et l'analyse d'une grande quantité d'informations. Capteurs sophistiqués, armes intelligentes, systèmes de suivi et communications digitales renforcées sont au cœur des guerres en réseau actuelles (*network-centric warfare*). L'impressionnant volume de données produites par le C4ISR dans ce contexte riche en informations et en décisions, menace de surcharger les réseaux comme les opérateurs humains et notamment les combattants.

Cette complexité dimensionnelle, issue du volet instrumental du C2, met en avant l'importance du domaine cyber dans la fiabilité de l'information et la pertinence de son exploitation.

Cependant, il ne s'agit pas seulement de questions d'intégrité des réseaux, des calculateurs et des programmes. L'incertitude technologique est complétée par l'incertitude du facteur humain, plus difficile à évaluer, à comprendre et à maitriser. Il est cause de biais et d'erreurs qui créent de l'incertitude de manière d'autant plus dangereuse qu'elle est souvent produite de bonne foi, susceptible d'être mal

[12] HFM : Human Factors and Medicine
[13] STO : Science and Technology Organisation. Organisme civil de l'OTAN en charge de la recherche scientifique et technologique, réunissant des scientifiques des pays de l'Alliance.

maitrisée et de passer inaperçue. Les membres d'une équipe, qu'ils soient dans les centres de *command and control*, dans des véhicules ou sur le terrain, sont connectés et communiquent par exemple par liaisons électroniques, radio, messages ou position GPS. Ces membres doivent dans le même temps participer à une cognition collective et à la coordination d'une équipe, dans un environnement opérationnel complexe. De telles capacités exceptionnelles, pourtant déjà requises, sont loin d'être assurées aujourd'hui. Dans ce cadre, l'enjeu des domaines des facteurs humains et de l'ingénierie humaine est de traiter les causes de l'incertitude, en travaillant sur la robustesse des procédures, sur le partage des représentations et sur la confiance relationnelle pour la maitrise de l'erreur dans les systèmes complexes (Amalberti, 1998; Strauch, 2017).

Au regard de ces besoins le C2 du futur doit être pensé comme un cycle de décision, augmenté par la symbiose entre l'homme et les systèmes (*Human-Autonomy Teaming),* pour mieux aborder l'extrême complexité dans laquelle il s'ancre. Cela revient à faire passer le C2 de l'ère des processus mécanisés à celle de la symbiose homme-machine. Dans leur définition de l'agilité et de la performance globale du C2, Alberts et Hayes l'ont caractérisé par la distribution des connaissances, par l'attribution de droits de décision et par la capacité des unités à interagir entre elles. De notre point de vue, la performance du futur C2 sera mesurée par sa capacité à traiter des données massives et complexes, par la charge de travail cognitif qu'il peut produire et par sa capacité à générer des décisions complexes, sous de fortes contraintes de temps, tout en supposant une connaissance complète de leurs conséquences (Figure 10).

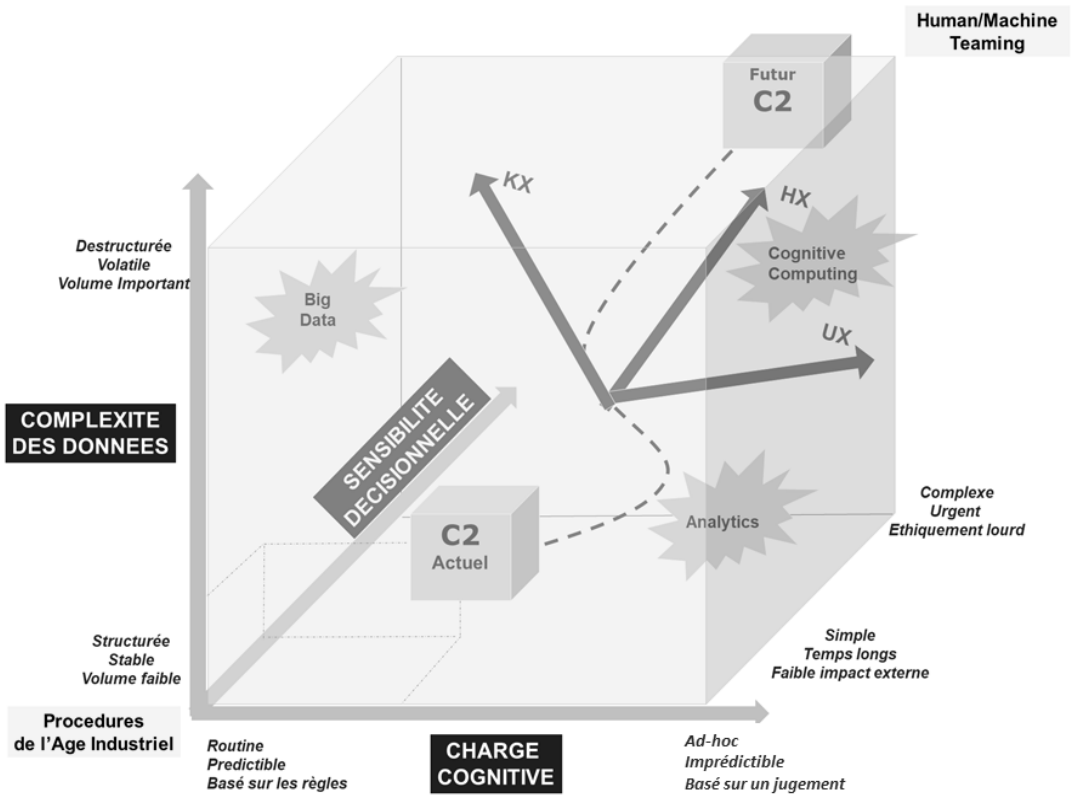


*Figure 10. Projection de l'évolution du C2, de l'ère des processus mécanisés à celle de la symbiose homme-système, augmenté par l'apport des outils cognitifs (*cognitive tools) (Desclaux et al., 2016).

L'axe de recherche dans lequel s'inscrivent nos travaux repose ainsi sur l'idée que l'intégration au C2 d'outils d'intelligence artificielle, orientés vers une collaboration symbiotique par le partage des

connaissances entre l'humain et la machine, est la seule façon de gérer une telle complexité (Desclaux et al., 2016). *Big Data, data analytics, cognitive computing* et autres *cognitives tools*, comme nous les appellerons, doivent permettre de fusionner et d'interpréter en quasi temps réel, les sources de renseignements habituelles, les données de terrain et les informations recueillies via l'analyse des réseaux sociaux, de « l'internet des objets » et des *Open Data*. Les outils de *cognitive computing*, tel que le système IBM Watson, permettent d'augmenter la compréhension de la dynamique du conflit et de la stratégie ennemie. Les *Big Data* et les *Analytics* aident l'humain à surveiller les activités, à automatiser des tâches, à surveiller et à prédire les CCIRs[14], construisant une représentation de la situation fine et prédictive.

L'ensemble de ces outils, réunis au sein du concept de *cognitive engine* (moteur cognitif) (Figure 11), doit affecter profondément les processus de prise de décision en fournissant au C2 l'agilité, la réactivité et la capacité de prédiction nécessaires pour mener les opérations d'aujourd'hui et de demain, reposant sur une représentation partagée de la dynamique de crise.

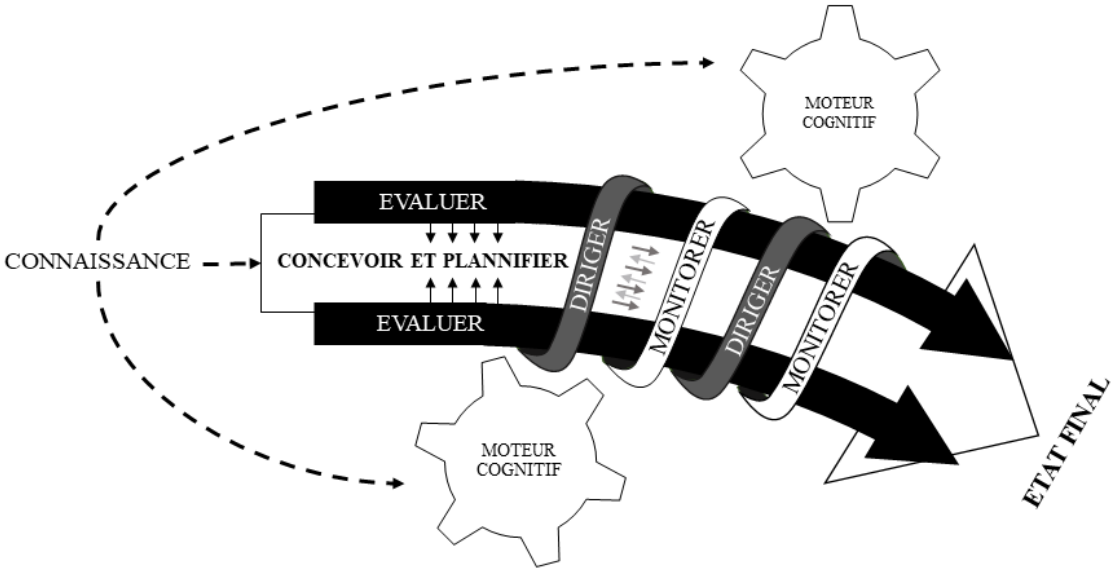


Figure 11. Cycle augmenté de la décision (Desclaux & Prébot, 2018)

La réponse tardive à l'ouragan Katrina, l'accident de Three Mile Island, l'incident de la navette Challenger, ou encore l'abattage d'un Airbus Iranien en 1988 par l'USS Vincennes sont autant d'exemples connus de catastrophes dont les causes ont pu être reliées à un dysfonctionnement dans la cognition collaborative d'équipe.

Cette cognition d'équipe (*Team Cognition*) joue un rôle majeur dans la performance et l'efficacité d'équipe et dans celle de la plupart des systèmes sociotechniques complexes (J. C. Gorman et al., 2007). Il est donc devenu nécessaire de l'évaluer, de la comprendre, d'en simplifier certains aspects et de la maitriser pour une meilleure performance et fiabilité.

Dans ce contexte, l'évaluation en temps réel des processus cognitifs de l'opérateur et de l'équipe devient progressivement un sujet de recherche majeur. De nombreux travaux indiquent un intérêt grandissant pour des systèmes adaptatifs en temps réel (Schwarz, Fuchs, & Flemisch, 2014; Vidulich & Tsang,

[14] Commander's Critical Information Requirements : ensemble d'informations clés identifiées par le commandant comme étant essentielles pour faciliter la prise de décision en temps utile.

2015), où le monitoring et l'évaluation de la cognition de l'opérateur sont au cœur du processus de collaboration homme-machine. Dans leur framework « *sens, asses, and augment* », Parasuraman et Galster (2013) proposent que l'état cognitif de l'opérateur soit dans un premier temps **perçu** par des capteurs et mesuré afin d'être ensuite **évalué** en fonction de la mission, du système ou des performances requises. La troisième étape étant l'**adaptation** en continu des systèmes selon la précédente évaluation. C'est donc par l'évaluation de la cognition individuelle et d'équipe que peut être créée une véritable équipe homme-machine augmentée.

1.3 Concepts clés de la cognition d'équipe

1.3.1 Vision collective, vision holistique et apport de la macrocognition

« *Une équipe d'experts ne suffit pas à former une équipe experte*. »

C'est souvent par cette phrase qu'est illustrée la problématique majeure concernant l'étude du travail d'équipe. Autrement dit, il s'agit de caractériser et de modéliser les subtiles différences inhérentes à la collaboration qui font l'atout de ce type de travail.

Historiquement, la recherche sur la cognition d'équipe se décompose en deux approches théoriques distinctes.

La première, la plus représentée, prend racine dans le domaine de la psychologie Industrielle/Organisationnelle (I/O)[15]. Dans cette approche, l'équipe est vue comme une unité de traitement d'information et la cognition d'équipe est étudiée au regard du modèle standard *Input-Process-Output* (I-P-O) (Hackman, 1987). Chaque membre est alors considéré comme une sous-unité de traitement possédant un certain degré de connaissances et de partage avec ses coéquipiers. Cette approche collective, centrée sur la connaissance et la représentation, suggère que la connaissance de l'équipe est la somme des connaissances individuelles. Les méthodes utilisées tentent donc de comprendre et de prédire la performance de l'équipe par des mesures au niveau individuel, en évaluant par exemple le contenu des représentations et connaissances internes des coéquipiers ainsi que les modifications apportées à leurs processus cognitifs dans ce contexte. Cette approche est aussi appelée approche *in the head*[16] (ITH) (Cooke, Gorman, & Kiekel, 2017) en référence à sa focalisation sur les processus internes aux membres de l'équipe.

La seconde approche théorique est issue du domaine de la psychologie écologique[17]. Elle situe la cognition d'équipe au niveau des processus d'interaction impliqués dans l'activité de l'équipe. Cette

[15] La psychologie industrielle/organisationnelle se caractérise par l'étude des comportements humains au sein des organisations et des espaces de travail. Cette spécialité se concentre sur la définition des principes de comportement individuel, de groupe et d'organisation et sur l'application de ces connaissances à la résolution de problèmes au travail.

[16] « Dans la tête »

[17] La psychologie écologique est une approche incarnée, située et non représentative dont les pionniers sont J. J. Gibson et E. J. Gibson. Cette théorie vise à offrir une troisième voie au-delà du cognitivisme et du behaviorisme pour comprendre la cognition. Les grands principes de la psychologie écologique sont la continuité de la perception et de l'action, le système organisme-environnement comme unité d'analyse, l'étude des affordances comme objets de perception, combinée avec un focus sur l'apprentissage et le développement perceptuels

méthode holistique suggère que la connaissance globale de l'équipe résulte d'un traitement cognitif sous forme de communications, d'évaluations de la situation et de coordination (Cooke, Salas, Cannon-Bowers, & Stout, 2000). Ce sont donc les processus externes de l'équipe qui sont mesurés, afin d'en prédire la performance. En réponse à la première approche, celle-ci est parfois nommée *between the heads*[18] (BTH).

Cette dichotomie s'ancre pleinement dans le dualisme symboliste caractéristique en sciences cognitives, en transposant à l'échelle de l'équipe les concepts d'émergence et de réductionnisme propres aux courants du connexionnisme et du cognitivisme (Sallaberry, 2020; Sallaberry & Claverie, 2018). En effet, pour Varela, l'émergence, qui voit la cognition comme résultante de l'interaction des éléments simples, s'oppose au réductionnisme qui conçoit le tout décomposable en une somme d'éléments singuliers hébergeant la cognition (Varela & Lavoie, 1996).

Comme le cognitivisme, l'approche ITH repose sur la notion de représentation et réduit le système complexe (ici l'équipe) à une somme d'individualités. En revanche, selon l'approche BTH et le connexionnisme, la connaissance émerge de l'interaction entre les « unités d'information » (neurones pour le connexionnisme, individus pour BTH).

Néanmoins, au-delà du débat épistémologique, le domaine du *Naturalistic Decision Making* propose le concept unificateur de macrocognition (Klein et al., 2003; Zsambok & Klein, 2014) dans une perspective de recherches sur la cognition d'équipe en situation écologique. L'intérêt porté par ce concept est de recentrer la problématique autour du besoin d'évaluation en situation écologique complexe, nécessitant la prise en compte à la fois des interactions de l'individu avec son environnement (coéquipiers, pression temporelle, objectifs dynamiques…) et des processus individuels. C'est dans cette optique que s'inscrivent la plupart des travaux que nous citons par la suite.

La macrocognition est définie par Warner et Letsky (2017) comme *« the internalized and externalized high-level mental processes employed by teams to create new knowledge during complex, one-of-a-kind collaborative problem-solving. »*[19]. Dans cette définition, les auteurs font référence à des processus cognitifs de haut niveau (c'est-à-dire les processus de visualisation et d'agrégation et de combinaison des informations visant à créer de la connaissance), internes et externes.

Les processus internes sont les processus cognitifs se produisant au niveau de l'individu ou de l'équipe, et qui ne sont pas exprimés à l'extérieur. Ils ne peuvent être mesurés que de manière indirecte par des mesures qualitatives telles que des protocoles de « pensée à voix haute » (*think aloud protocols*), ou des mesures quantitatives de substitution (par exemple le diamètre pupillaire, la réponse électrodermale…). Ces processus peuvent être externalisés de façon à être interprétables par les coéquipiers.

Les processus externes sont donc les processus qui se produisent au niveau individuel ou collectif, associés uniquement à des actions observables et mesurables de manière cohérente.

Du point de vue de la macrocognition, se concentrer sur les états des représentations internes plutôt que sur les relations entre les membres de l'équipe (ou inversement) conduit à différentes hypothèses sur la manière dont les processus cognitifs peuvent être observés et mesurés le plus directement. ITH et BTH sont donc deux paradigmes de recherche, deux façons de voir le monde qui ne sont pas nécessairement en compétition ou en opposition mais doivent plutôt être utilisés de manière complémentaires (Cooke

[18] « Entre les têtes »

[19] « La macrocognition est définie comme les processus mentaux de haut niveau, intériorisés et externalisés, employés par les équipes pour créer de nouvelles connaissances lors de la résolution de problèmes complexes et uniques en leur genre »

et al., 2017). C'est également ce qui est avancé par Sallaberry (2020), pour qui cognition individuelle et cognition collective doivent être conçus ensembles, pensées comme articulées autour de la notion de représentation.

C'est autour de trois concepts majeurs à la fois micro et macro, internes et externes, que se construit l'évaluation de la cognition d'équipe.

Les deux premiers concepts reconnus sont la **charge cognitive** et la **conscience de situation**, ainsi que leurs équivalents collectifs (Berggren, Prytz, Johansson, Nahlinder, & Nählinder, 2011; Tsang & Vidulich, 2006). À ces composantes, Berggren *et al*. (2011) ajoutent également l'étude d'une composante purement collective : les **interactions** interpersonnelles, notamment les communications. Ainsi, l'étude conjointe de ces trois mesurables doit permettre une meilleure compréhension et prédiction des performances homme-système dans des systèmes complexes (Parasuraman, Sheridan, & Wickens, 2008; Vidulich & Tsang, 2015).

1.3.2 La notion de charge cognitive

La notion de charge couvre un large éventail d'activités humaines dans ce contexte. La charge cognitive concerne les activités qui sont principalement de nature mentale pouvant impliquer une coordination physique. Elle exclut les activités entraînant uniquement une fatigue musculaire. Plus formellement Chevalier et Cegarra (2008) la définissent ainsi :

« La charge cognitive présuppose que les processus cognitifs ont un coût tiré d'une réserve limitée de ressources cognitives. En conséquence, la charge cognitive est souvent définie comme le rapport entre la demande de la tâche et les ressources humaines disponibles »

Cette définition a l'avantage de confronter le coût de la tâche et les ressources disponibles dans un stock limité. La charge cognitive n'est pas seulement inhérente à la tâche ou à l'humain mais bien un rapport entre les deux et est donc influencée par la perception que l'individu a de la tâche.

Bien qu'elle ne soit pas spécifiée dans la définition de Cegarra et Chevalier, la notion de performance est intimement liée à celle de charge cognitive. En effet, le coût de la tâche est directement influencé par le niveau de performance attendue dans sa réalisation. Ainsi des contraintes extérieures à la tâche elle-même (temps imparti, finances, qualité...) viennent pondérer son coût. De plus, la gestion de ces ressources cognitives influera sur la performance effective de l'individu réalisant la tâche. Yerkes et Dodson (1908) mettent en lumière le fait que le lien entre le niveau de charge cognitive et la performance n'est pas linéaire et qu'une performance faible peut aussi bien être due à une surcharge qu'à une sous-charge (ennui). Il existe donc un certain niveau de charge pour lequel la performance de l'opérateur sera maximale, dépendant à la fois de la tâche, de l'individu et de son environnement. Pouvoir l'identifier présente un intérêt pour tendre vers une performance optimale (Figure 12).

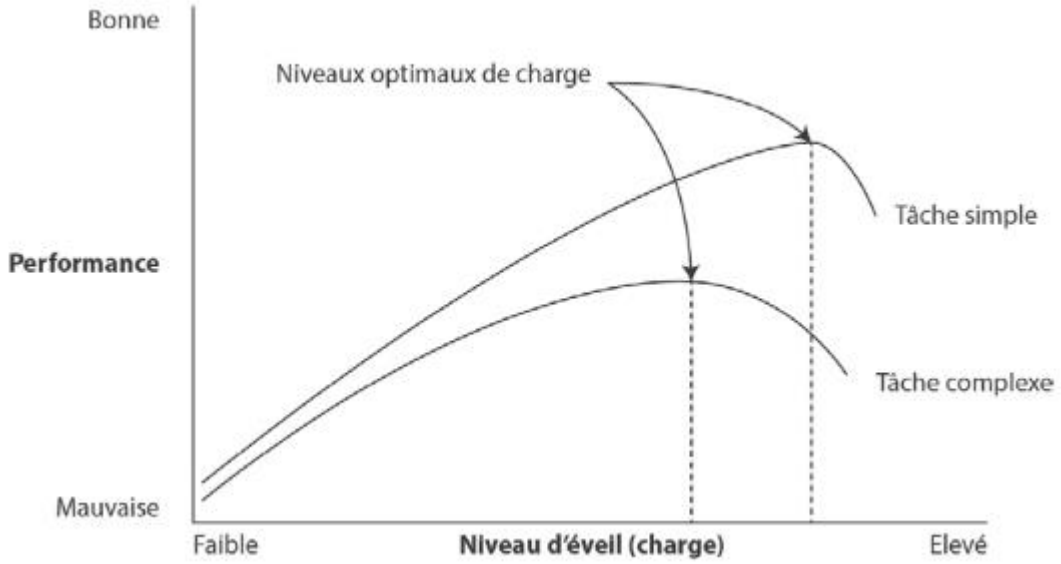


Figure 12. Loi de Yerkes et Dodson (1908), d'après Lini (2013).

Pour Megaw et Sharples (2015), l'individu surveille sa propre performance en continu, à la fois consciemment et inconsciemment. Combiné aux feedbacks qu'il peut recevoir de différentes sources, ce monitoring peut influer sur sa perception de la tâche et lui permettre d'ajuster en permanence sa stratégie. Ainsi, la charge cognitive est un concept dynamique, son niveau fluctue en continue pour adapter le système aux contraintes et besoins perçus ou ressentis.

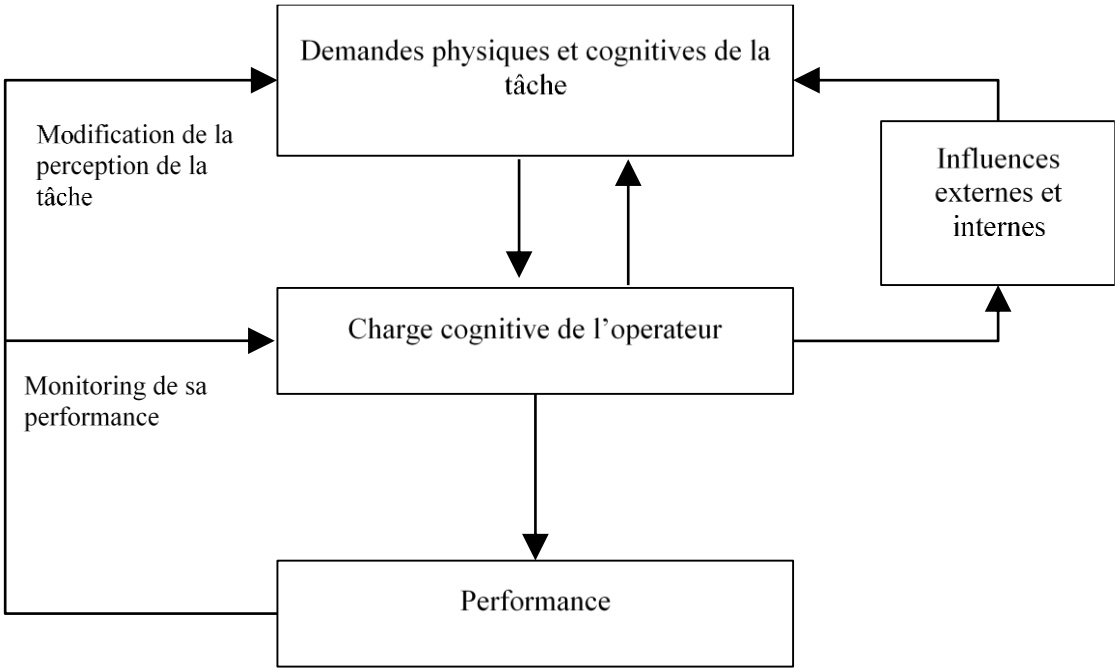


Figure 13. Framework pour la définition et de l'évaluation de la charge cognitive - Megaw & Sharples, 2015.

Ces fluctuations de charge sont les témoins de l'influence de différents facteurs et modifications de l'état cognitif. Leur mesure présente donc un intérêt particulier pour l'évaluation de la cognition.

En résumé, le niveau de charge cognitive peut être conçu comme le ratio entre les demandes et les capacités. Les demandes proviennent de l'environnement et des objectifs et sont notamment déterminées par le nombre de tâches à effectuer, le degré d'attention nécessaire et le temps disponible. Les capacités relèvent des compétences et de la formation de l'opérateur, mais peuvent également être influencées par des facteurs internes ou externes tels que la fatigue, le stress ou encore le bruit. (Hancock & Meshkati, 1988). La représentation que l'individu se fait de la demande, c'est-à-dire de la tâche et des contraintes (*input*), et celle de sa propre performance (*output*) sont centrales dans le processus de régulation de la charge.

1.3.3 La notion de conscience de situation (SA)

Le deuxième concept central dans l'étude de la cognition d'équipe est celui de conscience de situation. Usuellement référée par son terme anglais de *situational awareness* (SA), la conscience de situation fait l'objet d'un intérêt approfondi dans les domaines du facteur humain et de la psychologie depuis les années 80. Quatre grandes familles de modèles sont identifiées dans la littérature (Chalandon, 2007), plus ou moins complexes, basées sur des visions différentes selon leur objet d'étude. Nous appuierons notre travail sur le modèle des trois niveaux de SA (Endsley, 1995), régulièrement pris comme référence internationale et sur lequel repose la majorité des méthodes de mesure. La conscience de situation y est définie comme étant

"the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning and the projection of their status in the near future"[20].

La conscience de situation peut donc être comprise comme la connaissance active de la situation dans laquelle une personne est impliquée et lui permettant d'anticiper son évolution. Pour Endsley, la conscience de situation de l'opérateur est le produit interne du processus d'évaluation de la situation (*situation assessement process*). Son modèle (Figure 14) distingue trois niveaux hiérarchisés de ce processus : perception, compréhension et projection.

[20] « la perception des éléments de l'environnement dans un volume de temps et d'espace, la compréhension de leur sens et la projection de leur état dans le futur proche ».

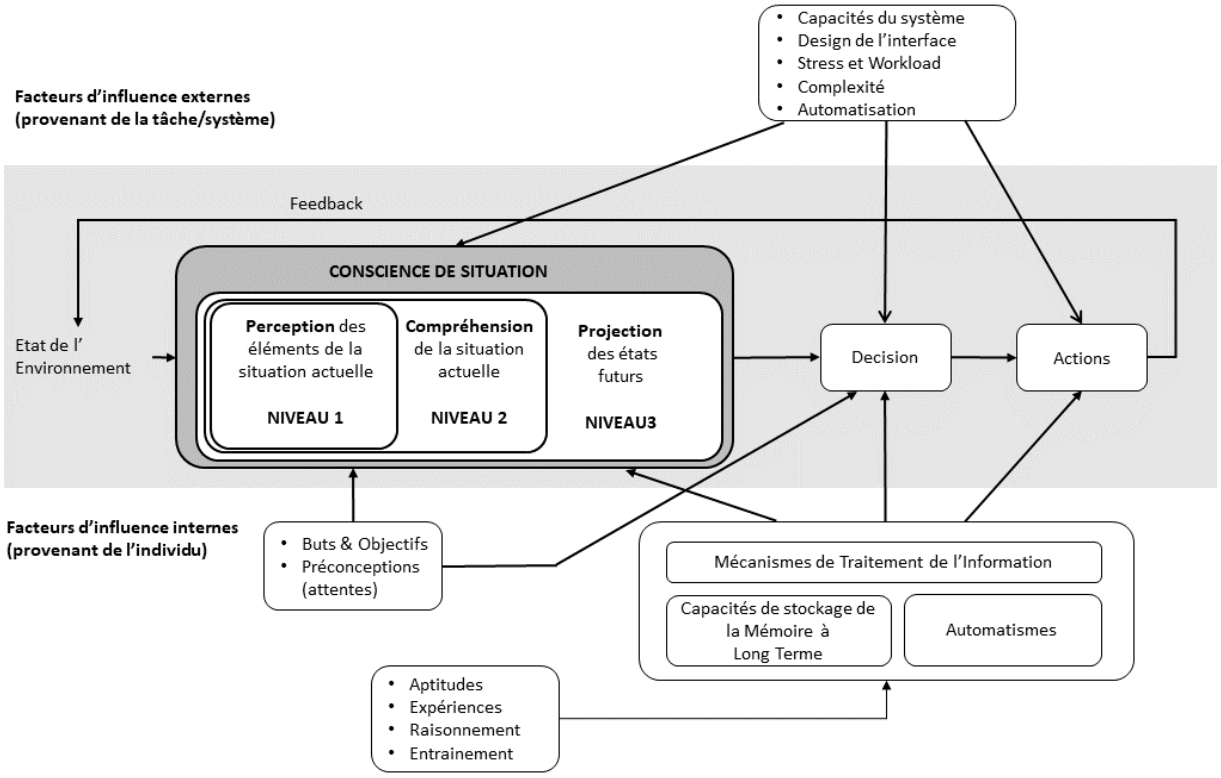


Figure 14. Modèle des 3 niveaux de conscience de situation (SA) de Endsley (1995)

La première étape est celle de la perception (Level 1 SA) des éléments de la situation présente. Il s'agit d'une phase sélective au cours de laquelle l'individu extrait de l'environnement des informations sur l'état et la dynamique des éléments qui l'intéressent. La deuxième étape consiste en la compréhension (Level 2 SA) de la situation. Le sujet construit une représentation de la situation basée sur l'interprétation et l'agrégation des éléments récoltés lors de la première étape. Une signification est ainsi donnée aux objets et aux événements, construisant une représentation holistique de la situation. Enfin, le troisième niveau de conscience de situation est celui de la projection de l'évolution des éléments de la situation dans un futur proche (Level 3 SA). L'individu anticipe les futurs états possibles des éléments de l'environnement en se basant sur l'instantané de la situation produit lors la deuxième étape, ainsi que sur les précédentes dynamiques d'évolution des éléments. C'est ensuite au regard de cette projection à court terme de l'évolution de l'environnement que l'individu décide des actions à réaliser ainsi que de leur temporalité.

Dans ce modèle, la conscience de situation apparaît donc comme un prérequis à la prise de décision humaine. L'acquisition et le maintien de la conscience de situation sont influencés à la fois par des facteurs individuels tels que l'expérience, la formation ou la charge de travail, et par des facteurs liés aux tâches et à l'environnement, telles que la complexité de l'interface.

Le modèle de Endsley rend compte non seulement de la théorisation possible de la représentation d'une situation externe mais considère également les buts et objectifs des individus impliqués. Il traduit la construction d'une représentation juste à l'instant donné en fonction du but poursuivi. Pour Endsley, la conscience de situation est donc le produit de l'ensemble des processus cognitifs en jeu dans la représentation de la situation.

Il est intéressant de noter ici le parallèle qui peut être fait avec le cycle D2D du C2 (Figure 9). On y retrouve en effet les mêmes étapes de :

- perception de l'information ;

- traitement, création de sens ;

- raisonnement et projection ;

- prise de décision ;

- action et mesure de leur performance, ainsi qu'une boucle de feedback.

Le cycle D2D n'est autre qu'une boucle décisionnelle basée sur la compréhension de la situation actuelle de l'opération et de sa dynamique par l'équipe. Il n'est donc pas étonnant que la représentation de la situation soit un élément majeur du C2 et un facteur de performance des opérations en temps court.

Dans le domaine militaire, la conscience de situation est souvent assimilée à la *common operational picture* (COP), un instantané clair, précis et commun de l'espace de combat, permettant à tous de répondre aux questions « Où suis-je ? », « Où sont mes amis ? », « Où sont mes ennemis ? » (Nofi, 2000). Cette vision concise et statique cherche à définir le niveau de SA minimal à posséder pour conserver une dominance informationnelle et ainsi éviter d'être pris de court par la situation. Elle ne tente en aucun cas de traduire les processus cognitifs desquels découlent la construction de la représentation de la situation.

1.3.4 Lien entre charge cognitive et conscience de situation

Charge cognitive et conscience de situation sont deux concepts distincts mais profondément intriqués (Tsang & Vidulich, 2006; Wickens, 2001). Endsley (Endsley, 1993) les présente d'ailleurs comme les deux faces d'une même pièce, chacun affectant et étant affecté par l'autre. Tsang et Vidulich (Tsang & Vidulich, 2006) ont développé un framework théorique présentant les relations entre ces deux concepts (Figure 15).

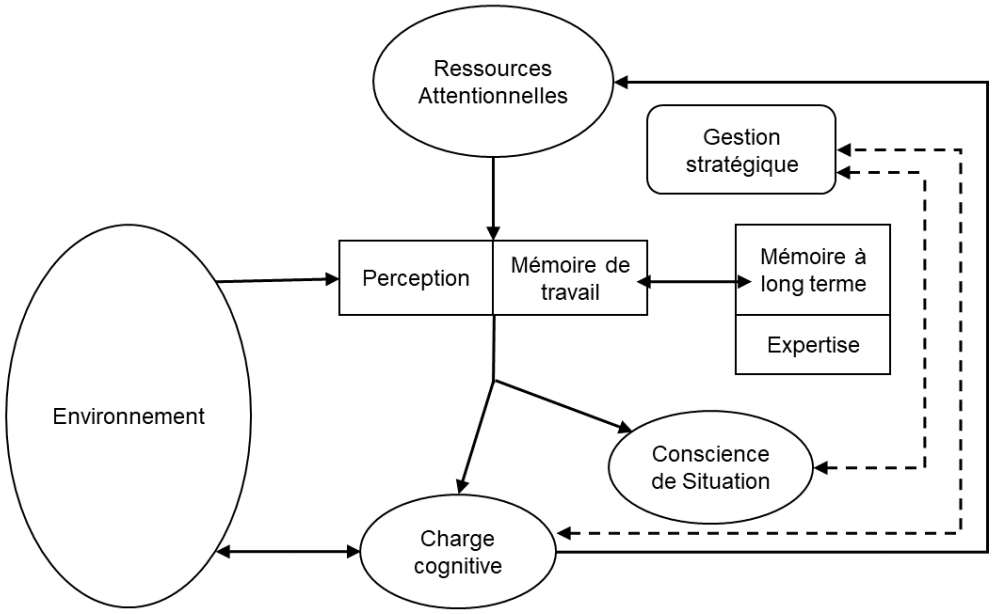


Figure 15. Framework conceptuel illustrant le lien entre charge cognitive et conscience de situation (Tsang & Vidulich, 2012)

Il permet de faire apparaitre que la qualité de la conscience de situation et le niveau de charge mentale sont influencés par des facteurs communs : d'une part les variables humaines, endogènes, telles que l'expertise, la fatigue, la limite des ressources cognitives ou de la mémoire de travail ; d'autre part les variables du système, exogènes, telles que la complexité de la situation, les contraintes du système ou la demande imposée par la tâche ou l'objectif de performance. Ainsi, dans une situation d'une même complexité, une même tâche imposerait probablement à un opérateur peu qualifié une charge plus élevée qu'a un opérateur expert et provoquerait également une plus mauvaise conscience de situation (Wickens, 2001).

Bien que la majorité des travaux menés par Endsley mette l'accent sur la difficulté générale de prédire l'évolution de l'un de ces concepts par l'autre (Endsley, 1993), il a été observé qu'une augmentation de la charge cognitive a une influence négative sur la qualité de la conscience de situation (Berggren et al., 2011; Endsley, 1996).

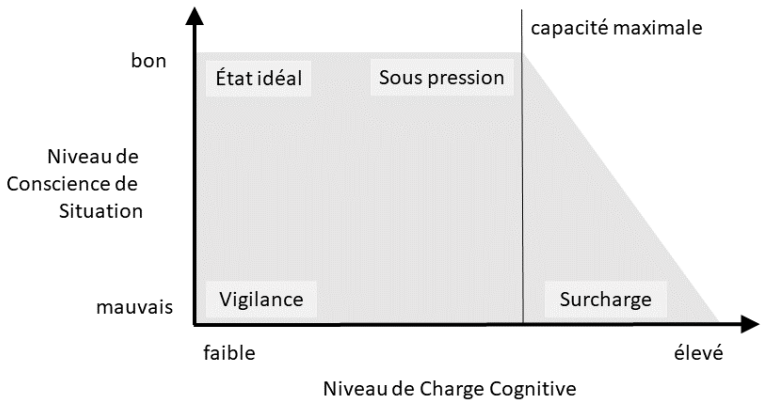


Figure 16. Relation entre conscience de situation (SA) et niveau de charge (workload)(Endsley, 1996)

En se basant sur la « théorie de stock limité de ressources attentionnelles » (Egeth & Kahneman, 1975), il est intuitivement compréhensible que plus le niveau de charge est élevé plus la quantité de ressources nécessaire pour conserver le niveau de performance souhaité dans la réalisation de la tâche est important. Or, d'après cette théorie, il n'existe qu'une quantité limitée de ressources attentionnelles. De fait, toute ressource mobilisée en plus pour répondre à un accroissement de charge est une ressource attentionnelle de moins disponible pour l'évaluation de la situation. La conscience de situation peut donc entrer en compétition avec la charge cognitive pour l'attribution des ressources attentionnelles. Une charge cognitive élevée peut ainsi provoquer une perte de conscience de situation globale, l'individu concentrant son attention sur la réalisation de la tâche qui le charge au détriment de la compréhension de son environnement.

L'étude conjointe de la charge cognitive et de la conscience de situation est particulièrement privilégiée pour aider à mieux identifier les contraintes et à expliquer la variabilité de la performance homme-système, en particulier dans le cas de systèmes fortement automatisés. Nos travaux s'ancrant dans un contexte d'étude similaire à l'échelle de l'équipe, il nous faut élargir ces concepts au collectif.

1.3.5 Team workload

La dissociation entre point de vue collectif ou additif et point de vue holistique est valable à tout niveau

et pour tout concept touchant aux processus cognitifs de l'équipe. Le débat autour des différents modèles de la charge mentale d'équipe ne fait donc pas exception.

Le travail d'équipe, examiné du point de vue individuel, est comparable à une situation de double-tâche pour chaque membre de l'équipe. Les membres doivent partager leur temps entre les besoins en interactions interpersonnelles, communications et échanges d'informations (*teamwork*) et les demandes imposées par la réalisation de la tâche individuelle qui leur est assignée (*taskwork*)(Funke, Knott, Salas, Pavlas, & Strang, 2012).

La charge cognitive d'équipe (*team workload*) peut donc être exprimée comme la charge demandée par le *teamwork* et le *taskwork*, ajoutée à celle demandée par la gestion de la répartition du temps entre ces deux activités (Bowers & Braun, 1997). Ainsi, même si chaque membre expérimente un certain niveau de charge de travail individuel imposé par la tâche, l'équipe subit une demande globale supérieure à la somme des demandes individuelles.

1.3.6 Team SA et Shared SA

La conscience de situation d'un opérateur individuel semble aujourd'hui bien définie ; il n'en va cependant pas de même avec le penchant collaboratif de la conscience de situation d'un groupe d'opérateurs. Ces dernières années sont apparus de nombreux concepts dont des définitions reconnues et acceptées restent encore à formuler (Nofi, 2000; Saner, Bolstad, Gonzalez, & Cuevas, 2009; Stanton, Salmon, Walker, Salas, & Hancock, 2017). Parmi eux, les concepts de conscience de situation d'équipe (*Team SA* - TSA) et de conscience de situation partagée (*Shared SA* - SSA) sont certainement les plus présents. Cependant, comme le soulignent Salas et al. (1995), ces concepts semblent aller bien au-delà de la simple agrégation des consciences de situation individuelles des individus de l'équipe.

Chaque membre possède un rôle et des sous-objectifs individuels constituants de l'objectif général de l'équipe. À ce sous-objectif est associé un ensemble d'éléments de situation, d'informations requises pour conduire à bien cet objectif. Endsley et Jones (1997) les nomment *SA requirements* ; Cain et Schuster (2016) s'y réfèrent en tant que *necessary knowledge elements* (NKE).

Dans la mesure où les membres d'une équipe sont interdépendants pour atteindre l'objectif global de celle-ci, il existe nécessairement un chevauchement de ces besoins en informations entre plusieurs membres. Une partie de la conscience de situation peut donc être commune ou partagée avec d'autres membres de l'équipe. Elle repose alors sur la mise en commun des informations individuelles nécessaires également aux autres membres pour la réalisation de leur tâche et par extension sur la construction de leur conscience de situation. Ainsi Endsley et Jones (2001) définissent la *Shared SA* (SSA) comme le degré de similarité des consciences de situation sur ces éléments partagés. Ces informations et ces connaissances sont distribuées par le biais de processus d'équipe tels que la communication ou la coordination.

Afin d'initier ce partage, chaque membre doit également posséder une conscience de situation liée aux autres, y compris une connaissance de leurs tâches, de leurs rôles et de leurs responsabilités, ainsi qu'une représentation des objectifs et des performances de l'équipe dans son ensemble.

Cette représentation globale intègre à la fois la conscience de situation individuelle et les connaissances concernant l'état de l'équipe à cet instant. La *Team SA* (TSA) est ainsi définie comme le degré auquel chaque membre de l'équipe possède la conscience de situation requise pour ses propres responsabilités (Endsley, 1995) ainsi que pour la coordination de la perception et des actions au sein de l'équipe. (J. Gorman, Cooke, & Winner, 2006, p. 1314). La conscience de situation d'équipe est donc l'ensemble des

informations, individuelles et partagées, sur l'ensemble des connaissances de l'état de l'équipe et de la situation (Salmon, Stanton, Walker, & Jenkins, 2008).

Cette définition prend donc en compte l'aspect métacognitif de la conscience de situation d'équipe en incluant la connaissance de « Qui sait quoi » (mémoire transactive) et de « Qui a besoin de savoir quoi », nécessaire à l'initiation du partage d'information. La connaissance nécessaire du rôle et de la tâche des autres membres fait l'objet de nombreuses recherches sur les méthodes de formations optimales pour l'amélioration de la conscience de situation et de la performance d'équipe particulièrement au sein du C2 (Strang, Funke, Knott, Galster, & Russell, 2012). Pour résumer, toute l'équipe n'a pas besoin de tout savoir, mais chaque membre doit, en plus de posséder toutes les informations dont il a besoin pour sa conscience de situation individuelle, être conscient à tout moment de ce que chacun sait, ne sait pas et a besoin de savoir parmi les informations disponibles (Figure 17).

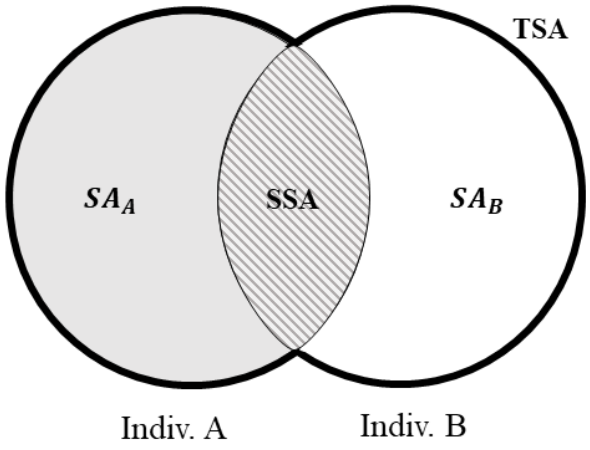


Figure 17. SA d'un collectif - d'après Bolstad 2009
SA : Situation Awareness (individuelle), SSA : Shared Situation Awareness (commune), TSA : Team Situation Awareness (équipe)

Salmon et al. (Figure 18) illustrent la construction et l'organisation des consciences de situation individuelles des membres et de la conscience de situation d'équipe, au regard de la décomposition en *taskwork* et *teamwork*, partie individuelle et partie collective de la tâche comme vu au 1.3.5, permettant de lier à l'échelle collective les concepts de charge mentale et de conscience de situation.

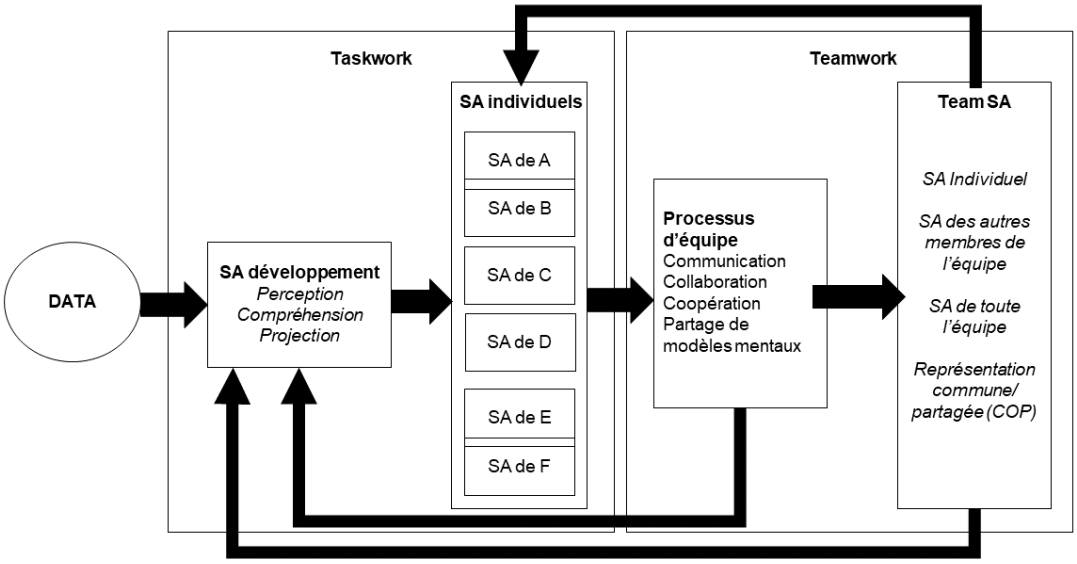


Figure 18. Organisation des consciences de situation individuelles (SA) et de l'équipe (Team SA) au regard du travail d'équipe – d'après (Salmon et al., 2007).

Les différents modèles de conscience de situation décrits et discutés jusqu'ici présentent l'acquisition d'une représentation à un instant précis ou concernant un élément situationnel particulier.

Pourtant, les principaux chercheurs du domaine s'accordent à dire que la conscience de situation, qu'il s'agisse d'un individu ou d'un groupe, n'est pas un phénomène stable, mais plutôt un état dynamique (Nofi, 2000). La situation change au fur et à mesure qu'elle est vécue. Ainsi, bien que l'on évoque « l'état » de la conscience de situation d'un individu, il s'agit en réalité d'un phénomène temporaire. La conscience de situation est en constante évolution. Un individu peut posséder une excellente conscience de la situation à un moment donné et une très mauvaise le moment suivant, selon la façon dont la situation évolue et l'efficacité avec laquelle les informations disponibles sont intégrées dans le processus continu de perception, de compréhension, de projection et de prédiction.

À notre connaissance, la dynamique de la conscience de situation est très peu étudiée. Pourtant son évaluation présente un intérêt central dans la perspective de l'évaluation en temps réel de l'état de l'opérateur, que ce soit dans un contexte de travail individuel ou collaboratif. C'est autour de ce constat que nous axons une partie de notre travail, par le biais d'une analyse et d'un modèle que nous présentons au chapitre 2.

1.3.7 La communication

Nous avons vu que l'échange d'informations et les interactions interpersonnelles, propriétés intrinsèques à la collaboration, forment le socle du travail d'équipe et sont donc des domaines d'étude privilégiés pour l'évaluation de la cognition collective.

Particulièrement, la communication, au sens de l'échange d'information entre plusieurs individus quel que soit le canal employé (vocal, numérique, comportemental) peut fournir une représentation dynamique et détaillée de la cognition collective. Quand les membres d'une équipe interagissent, le flux

d'échanges contient des informations sur le contenu de leurs connaissances mais également sur leur conscience de la situation, leurs doutes, le stress et autres états cognitif ou psychologique. L'étude de la communication se révèle également plus accessible que l'étude des structures cognitives individuelles telles que la charge mentale ou la conscience de situation puisque reposant sur l'analyse des échanges entre individus (Kiekel, Cooke, Foltz, & Shope, 2001).

En tant que mécanisme central du partage et du traitement de l'information au sein d'une équipe, le comportement de communication au sein de celle-ci est naturellement influencée par de nombreux facteurs. Des liens avec la charge mentale, la difficulté et la nature de la tâche, mais également la structure de l'équipe, sa cohésion ou encore la formation des coéquipiers, ont tous été démontrés (Kiekel et al., 2001). Plusieurs études indiquent notamment une relation positive entre la fréquence de communication et la conscience de situation (Costello, Strater, Bolstad, Cuevas, & Endsley, 2006; Saner, Bolstad, Gonzalez, & Cuevas, 2010) ainsi qu'entre la fréquence de communication et la performance de l'équipe (Orasanu & Fischer, 1991; Sexton & Helmreich, 2000). Ainsi, sous forte charge cognitive, une augmentation de la fréquence de communication au sein de l'équipe a pu être mise en évidence. Bailey et al. (2002) suggèrent que cette augmentation est nécessaire afin de conserver une bonne conscience de situation. Inversement, Schreiber et al. (1998) ont pu établir que plus la conscience de situation d'un pilote d'avion de chasse est élevée, moins il a tendance à communiquer avec ses coéquipiers (autres pilotes, contrôleurs…).

Il a également été montré que l'équipe a tendance à augmenter sa communication pour répondre aux contraintes liées à une nécessité de correction d'erreur. Ainsi, une augmentation de la fréquence de communication peut également être la conséquence d'une baisse passagère de performance et traduire la stratégie de résilience mise en place par le groupe (Thorntron (1992) cité par Wickens et al. 2009).

L'interprétation des observations dépend donc du contexte et de la temporalité du changement de la dynamique de communication de l'équipe (J. Gorman et al., 2020).

Deux types de données de communication peuvent être utilisés pour étudier la cognition d'équipe : d'une part, les mesures « physiques », quantitatives, comme la fréquence des communications ou la durée de parole ; d'autre part, les mesures de contenu, qualitatives, comme le type d'informations transmises ou le ton employé. Alors que l'analyse des premières peut aisément être automatisée et mise en place en temps réel, les dernières demandent une analyse *a posteriori* coûteuse en temps. Mais l'évolution des outils d'intelligence artificielle n'exclut pas l'utilisation prochaine en temps réel de données qualitatives issues des communications, dans lesquelles résident la plupart des données pertinentes sur les échanges (Foltz, 2018; Kiekel et al., 2001; Spain, Geden, Min, Mott, & Lester, 2019; Vukovic et al., 2019).

1.4 Méthodes de mesure de la cognition d'équipe

Les nombreuses méthodes utilisées pour étudier l'humain dans son environnement de travail découlent principalement des approches "classiques" issues de l'ergonomie et de l'étude du facteur humain. Analyses de tâche, entretiens ou encore observations sont généralement les méthodes utilisées pour évaluer le travail d'un opérateur dans son environnement.

Ces méthodes ont l'avantage de minimiser l'impact de la mesure sur le mesuré en conservant une situation de travail et un contexte aussi naturel et authentique que possible. Plus l'individu peut évoluer dans ses conditions habituelles durant l'évaluation, plus la mesure en sera pertinente.

Si de telles méthodes se sont montrées appropriées pour des petits groupes de travail co-localisés, leur

mise en place et leur exploitation deviennent incompatibles avec des collectifs de travail de grande, voire de très grande taille. Il ne paraît en effet pas possible, par exemple, d'assigner un, voire plusieurs, observateurs pour chaque opérateur. De plus, le traitement des informations recueillies ainsi ne pouvant être réalisé qu'a *posteriori*, ces méthodes perdent tout intérêt pour une évaluation en temps réel. C'est un écueil majeur pour l'étude de l'état de la collaboration dans le cadre d'une gestion continue des risques.

Une rupture de méthode dans l'acquisition et l'interprétation des données s'impose donc.

Que ce soit pour l'évaluation de la performance du C2 (Alberts & Hayes, 2003; Berggren et al., 2014) ou plus spécifiquement celles de la conscience de situation et de la charge mentale (Parasuraman et al., 2008), les recommandations soulignent le besoin de diversification des méthodes de mesure et d'évaluation. L'approche proposée repose sur l'agrégation de mesures issues des domaines physique (exemple : le produit, la performance), informationnel (la communication) et cognitif.

1.4.1 Mesurer la charge cognitive individuelle et d'équipe

La charge cognitive est certainement l'un des phénomènes cognitifs les plus étudiés et dont les méthodes de mesure sont très variées. Ces méthodes sont habituellement classées en trois catégories (Guhe et al., 2005) : les mesures **subjectives**, la mesure de **performance** et les mesures **psychophysiologiques**.

Les mesures subjectives, telles que les questionnaires, demandent à l'individu d'estimer lui-même son niveau de charge.

Le NASA-Task Load indeX (NASA-TLX), développé dans les années 80, est le plus connu et plus utilisé. Il reste la méthode de référence pour la mesure du niveau de charge individuel.

Pour la mesure de la charge cognitive d'équipe de nombreuses variations ont été proposées (Funke et al., 2012). Par exemple, une version modifiée du NASA-TLX ajoute quatre sous-échelles aux six échelles habituelles, afin d'évaluer les demandes attribuées au travail en équipe en ce qui concerne : la communication, le monitoring, le contrôle et la coordination (Sellers, Helton, Näswall, Funke, & Knott, 2014). Ces questionnaires sont généralement soumis aux sujets durant et/ou après la tâche. L'estimation de la charge d'équipe se fait alors par moyenne des charges individuelles. Certaines versions du TLX ou certains questionnaires alternatifs (e.x., *Team Workload Awareness Questionnaire* - TWOQ, Entin, Serfaty, and Kerrigan 1998) demandent également à l'individu d'estimer la charge cognitive de ses coéquipiers et de l'équipe en plus du sien. Cela fait sens si l'on considère l'aspect métacognitif de la collaboration comme nous l'avons vu plus tôt.

Ces méthodes, aussi reconnues et fiables soient-elles, ne sont cependant pas applicables pour une évaluation en temps réel.

Les mesures de performance visent à surveiller les résultats de performance lors de la manipulation des demandes de la tâche. Se basant sur la *Multiple Resources Theory* (Wickens, Hollands, Banbury, & Parasuraman, 1992) le principe de ces méthodes est d'ajouter une tâche secondaire, censée élever le niveau de charge de l'individu. La performance de l'opérateur est alors mesurée suite à cette perturbation et il est supposé qu'une modification de la performance de la tâche principale ou secondaire est un indicateur indirect du niveau de charge. Cependant, il est difficile d'établir des relations de cause à effet entre ces changements de performances

avec de véritables changements dans l'investissement des ressources de l'équipe ainsi que la source de la charge. De par leur intrusivité dans le scénario ces méthodes sont mieux adaptées aux simulations et études en laboratoire qu'aux situations réelles.

Les mesures psychophysiologiques s'intéressent aux métriques physiologiques et comportementales susceptibles d'être corrélées avec l'activité cognitive. Ce principe se base sur la réflexion de l'augmentation du *workload* sur l'activité du système nerveux autonome.

Classifiables en 6 catégories en fonction de l'activité qu'elles mesurent (Megaw & Sharples, 2015; Rossi, 1999), ces méthodes ont l'avantage de fournir une évaluation objective de la charge cognitive.

- **Activité cardio Respiratoire** : le rythme cardiaque (ECG), la respiration ou encore la pression sanguine sont les méthodes de mesure les plus utilisées car relativement abordables, simples et assez sensibles aux changements de charge mentale.
- **Activité cérébrale :** l'imagerie cérébrale est à la base des méthodes de neuro-ergonomie les plus utilisées. Les trois principales se basent soit sur l'activité électromagnétique du cerveau (Electroencéphalographie - EEG), soit sur la variation de l'oxygénation des flux sanguins (fNIRs et IRMf). Elles ont l'avantage de posséder la résolution temporelle la plus fine, cependant elle sont extrêmement coûteuses (Mehta & Parasuraman, 2013) et intrusives.
- **Fluides corporels :** peu fréquemment utilisées du fait de la complexité du recueil d'échantillon et de la latence dans leur analyse, ces méthodes étudient notamment les changements hormonaux et les niveaux de glucose dans le sang comme indicateurs du niveau de charge mentale.
- **Activité électrodermale :** la méthode principale vise à mesurer l'activité électrodermale (Réponse Electrodermale - RED), activité électrique biologique à la surface de la peau, témoin de l'activité des glandes de la sudation. Il a été démontré que, comme le rythme cardiaque, la conductance de la peau augmente avec la charge mentale (Reimher et Mehler 2001, cités par Sharples 2014).
- **Activité musculaire et mouvements :** basées sur le lien entre effort physique et mental, ces mesures étudient la posture, l'activité musculaire ou encore la démarche de l'individu. Il a par exemple été démontré qu'une différence de position des épaules et de la tête sur des opérateurs assis devant un ordinateur indiquerait une perte de vigilance (Witch et al., 2013, cités par Sharples 2014). Les avancées en intelligence artificielle et vision par ordinateur vont permettre d'explorer ces méthodes plus en profondeur.
- **Activité oculaire :** la mesure des mouvements et réponses volontaires et involontaires des yeux est une méthode de plus en plus utilisée pour évaluer et prédire la charge cognitive, principalement car elle est peu invasive. Elle présente l'inconvénient majeur d'être très sensible aux variations de lumière ambiante (Hogervorst, Brouwer, & van Erp, 2014; Wickens et al., 1992; Zagermann, Pfeil, & Reiterer, 2016). Cependant, il a été prouvé de nombreux liens entre élévation de charge et activités oculaires. Les principaux sont résumés dans le Tableau 1.

Indicateurs d'augmentation de la charge cognitive	
Diamètre pupillaire	augmente
Dilatation pupillaire	augmente
Fréquence de fixation	augmente
Durée de fixation	augmente
Durée de clignement	augmente
Intervalle de clignement	augmente
Fréquence de clignement	augmente
Vitesse moyen de clignement	diminue
Vitesse moyenne de saccade oculaire	augmente

Tableau 1. Résumé de l'effet de l'augmentation de la charge sur les principales mesures oculaires

Bien que EEG et fNIRs soient reconnus comme de très bons indicateurs, ces méthodes sont difficilement applicables et utilisables dans la situation écologique que nous étudions. On comprend aisément qu'il ne soit pas concevable, en réelle situation de crise, de couvrir les opérateurs de capteurs, électrodes et autres casques.

Bien que les évolutions technologiques permettent de limiter l'intrusivité et d'augmenter la portabilité des capteurs (*wearables devices*), les signaux recueillis restent extrêmement sensibles aux bruits, aisément maîtrisables en situation contrôlée mais difficilement applicable à un réel environnement collaboratif large.

L'oculométrie et la pupillométrie sont alors des solutions alternatives relativement fiables (St. John 2003), non invasive, moins coûteuses et dont l'analyse peut être automatisée et généralisée à un grand nombre de postes de travail sans souffrir d'un excès de contraintes. Pour ces raisons nous privilégierons cet outil de recueil de données psychophysiologiques, nous exploiterons plus particulièrement le diamètre pupillaire, indicateur dont la mesure brute est directement accessible contrairement aux clignements, fixations et saccades, qui demandent l'application d'algorithmes supplémentaires aux données brutes de position du regard.

À ces méthodes peut être ajoutée l'analyse de l'interaction homme-système. Souvent qualifiée de physio-behavioriste, elle cherche à mesurer le comportement de l'individu en lien avec son poste de travail. Dans le cas du travail sur ordinateur, on s'intéresse en autres à la vitesse de

frappe sur le clavier, la fréquence de clics de souris ou la pression exercée sur celle-ci, comme traductions de l'état cognitif de l'opérateur. St John et al. (2004) ont par exemple observé une corrélation positive entre la pression exercée sur la souris et le niveau de charge mentale de l'individu.

En outre, un champ émergeant de recherche s'intéresse à la synchronisation neurophysiologique et comportementale entre les sujets travaillant en collaboration (Stevens et al., 2011; Tognoli, Lagarde, DeGuzman, & Kelso, 2007), également appelée synchronisation socio-psychophysiologique (Henning, Armstead, & Ferris, 2009) ou synchronicité physio-comportementale (Strang, Funke, Russell, & Thomas, 2013). Il s'agit de l'observation des changements synchrones dans les états physio-comportementaux des membres d'une équipe pendant l'exécution d'une tâche collaborative.

Dans cette optique, l'évaluation de la cognition d'équipe se fait par comparaison des activités mesurées des différents opérateurs, issues de différents capteurs physiologiques, neurologiques ou comportementaux.

La synchronicité des activités semble indiquer le degré de coordination interpersonnelle entre les coéquipiers et donc être un témoin de partage de cognition (Tognoli, 2008). Une plus grande synchronicité améliorerait également la performance de l'équipe (Elkins et al., 2009; Henning et al., 2009). Les fluctuations de la synchronicité peuvent alors servir d'indicateurs utiles de l'évolution de la charge de travail des équipes.

Des techniques qualitatives, comme l'analyse de la tâche par un expert ou l'observation, peuvent également être utilisées. Cependant, ces mesures principalement subjectives sont basées sur une vision extérieure de la charge cognitive. Concentrées sur la charge induite par la tâche elle-même et l'environnement elles oblitèrent la perception de celle-ci par l'individu. De telles méthodes ne sont donc pas directement adaptées à notre problématique. Elles s'avèrent néanmoins nécessaires dans l'évaluation et la conception de systèmes complexes.

Le Tableau 2 compare ces différentes techniques au regard de leurs caractéristiques.

Type de technique	Moment d'administration	Caractéristiques
Subjectives – Questionnaires d'auto-évaluation (e.g., NASA-TLX, TWOQ)	Durant la tâche (interruption ou non)/Après la tâche	Intrusives pour la tâche si pendant. Ponctuelles (1 à n fois) Subjectives (sujet)
Performance	Après la tâche	Non intrusives Ponctuelles (1à n fois) Objectives
Physio-comportementales (EEG, ECG, fNIR, EOG, GSR…)	Durant la tâche	Plus ou moins intrusives sur le sujet selon la méthode. Intrusives pour la tâche selon le type d'appareillage. Continues Objectives
Observations	Durant la tâche/Après la tâche	Non intrusives Continues Subjectives (expert)

Tableau 2. Principales catégories de méthodes de mesure de la charge cognitive, d'après (Funke et al., 2012)

1.4.2 Mesurer la conscience de situation individuelle et d'équipe

De par sa nature qualitative et subjective, la conscience de situation est principalement mesurée à l'aide de questionnaires visant à faire expliciter par l'opérateur le contenu de sa vision de la situation actuelle afin de la comparer avec l'état réel. Cela requiert une connaissance omnisciente de la situation, donc une pré-analyse visant à identifier les éléments de SA nécessaires. La méthode d'approche la plus souvent choisie est la *Goal Directed Task Analysis* (GDTA) (Endsley, 2000). Cette pré-analyse se fait en interviewant des opérateurs ou experts de la tâche afin d'identifier leurs besoins d'informations pour chaque but et sous-but.

Si de nombreuses méthodes d'évaluation ont fait leurs preuves selon les contextes, la *Situation Awareness Global Assessment Technique* (SAGAT) (Endsley, 1988), méthode objective et directe, est aujourd'hui la plus utilisée, notamment dans les études sur l'environnement de C2 et C4i (Salmon, Stanton, Walker, & Green, 2006; Stanton et al., 2017). En 2018, Meireles *et al*. (Meireles, Alves, & Cruz, 2018) ont référencé cinquante-quatre techniques de mesure de la conscience de situation dans huit domaines d'application (conduite automobile, aviation, militaire, médical, sportif...). Cette diversité de techniques d'évaluation illustre la nécessité de mettre en place des méthodes adaptées au contexte. Cependant, la grande majorité des techniques repose sur des évaluations discrètes, soit pendant, soit après la tâche. Techniques d'auto-évaluation, mesures de performance, techniques de sondage, évaluation par l'observateur - chacune de ces techniques présente des avantages et des inconvénients particuliers. Nous en présentons ci-dessous les grands types, au regard de la temporalité dans laquelle

elles sont appliquées.

- Les techniques post-expérimentales regroupent un ensemble de méthodes d'auto-évaluation de leur conscience de situation, pratiquée immédiatement après la tâche par les participants eux-mêmes, sur la base d'une échelle de notation. La *Situation Awareness Rating Technique* (SART) (Taylor, 1990), la *Situation Awareness Rating Scale* (SARS)(Waag & Houck, 1994), la *Situation Awareness Subjective Workload Dominance* (SA-SWORD)(Vidulich & Hughes, 1991) et la *Mission Awareness Rating Scale* (MARS)(Matthews & Beal, 2002) sont quelques-unes des techniques d'auto-évaluation les plus utilisées. Les mesures de performance ont également été étudiées comme méthode d'évaluation a posteriori. La conscience de situation est alors déduite de la performance du sujet sur un ensemble d'objectifs de la tâche (Salmon et al., 2006).

  N'étant pas intrusives, les méthodes d'évaluation postérieures semblent bien adaptées à une utilisation « sur le terrain » et sont facilement applicables dans un contexte d'équipe. Cependant, elles présentent des inconvénients majeurs : les techniques d'auto-évaluation sont fortement influencées par la performance des sujets dans la tâche et leur capacité limitée à évaluer leur propre conscience de situation (Endsley, Selcon, Hardiman, & Croft, 1998). Les mesures de performance sont basées sur l'hypothèse controversée selon laquelle la performance et l'évaluation de la performance sont réciproquement liées.

- Une deuxième catégorie de techniques de mesure consiste à demander aux sujets de répondre à des questions sur la situation pendant la tâche. Appelées techniques d'interrogation ou de sondage (*query* ou *probes techniques*), ces méthodes peuvent être séparées en deux catégories. Les ***Freeze probe techniques***, dont fait notamment partie la *Situation Awareness Gobal Assessment Technique* (SAGAT)(Endsley, 1988), interrompent la tâche de manière aléatoire. Durant cette interruption, les écrans des opérateurs sont obturés, « gelés », et une série de questions relatives à la situation actuelle sont alors posées. L'opération peut être répétée plusieurs fois. Cette famille de techniques comprend également la SALSA, développée pour des applications de contrôle du trafic aérien (Hauss, Gauss, & Eyferth, 2001) et la *Situation Awareness Control Room Inventory* (SACRI) (Hogg, Follesø, Volden, & Torralba, 1994).

  Alternativement, les ***Real-time probe techniques*** visent également à administrer des questions sur la situation durant la tâche, cette fois-ci sans interruption de cette dernière. Les questions sont posées au moment jugé opportun et au fur et à mesure, par le biais de l'interface. L'évaluation se fait sur le contenu et le temps de réponse. La *Situation-Present Assessment Method* (SPAM)(Durso, Hackworth, & Truitt, 1999) utilise en particulier la latence des réponses comme mesure d'évaluation principale. Cependant, les questions peuvent orienter l'attention de l'individu et biaiser les résultats. De plus, l'évaluation se faisant en parallèle de la tâche, son impact sur la charge mentale de l'opérateur est non négligeable.

  Bien que ces techniques permettent d'évaluer objectivement ce que les sujets savent de la situation à des moments critiques, elles sont confrontées à des problèmes d'interruption ou de commutation de tâches, et sont donc profondément intrusives et moins adaptées à une utilisation *in situ*.

- Les méthodes permettant une évaluation potentiellement continue de la conscience de situation sont les plus difficiles. Une technique d'évaluation par les observateurs, telle que la *Situation Awareness Behaviorally Rating Scale* (SABARS) (Matthews & Beal, 2002), est conçue pour

déduire la qualité de la conscience de situation de l'individu à partir d'un ensemble d'indicateurs comportementaux. Ces méthodes directes nécessitent la participation d'experts et aboutissent souvent à une évaluation complexe et très subjective.

Les mesures physiologiques et comportementales peuvent également servir de techniques indirectes mais potentiellement continues de mesure de la conscience de situation. Cependant, contrairement à la charge de travail pour laquelle les mesures psychophysiologiques et neuropsychologiques dont on a pu prouver le fiabilité (Roscoe, 1993; Veltman & Gaillard, 1998), une mesure continue objective et valide de la conscience de situation n'est toujours pas disponible (Nofi, 2000; Stanton et al., 2017).

En résumé, les méthodes telles que les questionnaires post-situationnels ou l'observation sont généralement non intrusives par rapport à la tache mais subjectives, tandis que les méthodes objectives nécessitent une intrusion dans la tâche pour poser les questions sur le contenu de la représentation (Tableau 3).

Type de technique	Moment d'administration	Caractéristiques
Self-Rating Techniques (e.g., SART, SARS, SA-SWORD, MARS)	Après la tâche	Non intrusives par rapport à la tâche Ponctuelles (1 fois) Subjectives
Performance	Après la tâche	Non intrusives Ponctuelles (1-fois) Objectives
Probe Techniques -Freeze Probes (SAGAT, SALSA, SACRI) - Real-time Probes (SPAM)	Durant la tâche	Intrusives Ponctuelles (n fois) Objectives
Observer-Rating Techniques (e.g. SABARS)	Durant la tâche/Après la tâche	Non-intrusives Continues Subjectives

Tableau 3. Principales catégories de méthodes de mesure de la conscience de situation

Méthodes de mesure de la conscience de situation d'une équipe

Une partie des méthodes que nous venons de présenter a, par la suite, été adaptée pour inclure également l'évaluation de la conscience de situation des autres membres de l'équipe. Ainsi à ce jour, la majorité des mesures existantes de la *Team Situation Awareness* (TSA) ou de la *Shared Situation Awareness* (SSA) sont des variantes de méthodes de mesure de la conscience de situation individuelle et aucune mesure n'a été formellement validée (Cooke, Stout, & Salas, 2018; Endsley & Garland, 2000; Höglund,

Berggren, & Nählinder, 2010; Salmon, Stanton, Walker, Jenkins, et al., 2008). Le partage de conscience de situation peut être considéré comme dépendant à la fois de connaissances individuelles et de coordination, différenciant ainsi deux niveaux de mesure (Bolstad et al., 2007; Saner et al., 2009) : (1) le degré de précision ou d'exactitude de la conscience de situation individuelle et (2) la similarité des consciences de situation des coéquipiers. Bien entendu cette évaluation est effectuée sur les éléments de connaissance nécessairement communs. C'est ainsi que, par exemple, le *Shared Awareness Questionnaire* (Prytz et al., 2015) évalue la SSA des coéquipiers sur l'accord et l'exactitude de leurs réponses aux questions objectives concernant la tâche.

L'évaluation de l'exactitude de la conscience de situation est essentiellement ce sur quoi portent la plupart des techniques individuelles. La compréhension de la situation est comparée à l'état réel de l'environnement au moment de l'évaluation, ce qui conduit à évaluer la conscience de situation comme un degré de congruence avec la réalité.

L'évaluation de la similarité des consciences de situation est généralement basée sur la comparaison directe de la compréhension que les coéquipiers ont des éléments de la situation qui les concerne. La connaissance relative à un élément est considérée comme partagée s'ils en ont une compréhension similaire.

Dans le contexte de travail en équipe, afin de prévenir des erreurs, détecter le non-partage de la conscience de situation est aussi important que de détecter son inexactitude (Saner et al., 2009). Cependant, par nature, ces méthodes souffrent des mêmes limites que les techniques individuelles et aucune méthode validée et fiable n'est encore disponible (Salas, Cooke, & Rosen, 2008).

Par ailleurs, cette distinction entre exactitude et similarité des consciences de situation des coéquipiers créé mathématiquement plusieurs états de partage de conscience de situation possibles. Par exemple, si deux personnes collaborent, la comparaison de leurs SA suite à leur mesure entraine les quatre cas suivant (Saner et al., 2009) :

- les deux sujets répondent correctement – les consciences de situation sont justes et partagées, c'est le cas idéal recherché ;
- l'un des deux individus est exact et le second se trompe – SA non partagée ;
- aucun des deux ne répond correctement mais leurs réponses sont identiques - SA inexact mais partagée ;
- aucun des deux ne répond correctement et leurs réponses sont différentes - leurs SA sont inexacts et non partagés.

En résumé sur les méthodes

Comme cela a été évoqué, l'étude de la qualité ou de l'exactitude de la conscience de situation requiert une connaissance omnisciente de la situation. Or, cette connaissance aisément disponible en situation contrôlée de laboratoire est plus difficilement accessible en situation écologique ou proche de la réalité. Ainsi, afin de s'approcher de cette contrainte dans le cadre de nos recherches sur le travail collaboratif, nous ne nous focaliserons pas sur l'évaluation de l'exactitude de la SA. Notre apport reposera sur la détection du partage ou des divergences de la représentation de la situation, si possible en incorporant une dimension continue à celle-ci de manière à étudier la dynamique de partage de conscience de situation au sein du groupe.

Nous nous plaçons dans une perspective d'automatisation nécessitant une utilisation de données quantitatives et objectives. Or, la conscience de situation est par nature une donnée qualitative à verbaliser. Il est cependant possible d'en assimiler certains éléments à des données quantitatives effectivement mesurables. Par exemple, des études menées à l'Army Research Laboratory[21] utilisent entre autres la distance séparant la position réelle de l'ennemi et celle estimée par l'opérateur (Ownby & Kott, 2006). D'autres études utilisent des approches similaires dans le cadre de l'évaluation de la conscience de situation de conducteurs automobiles (Johannsdottir & Herdman, 2010) ou de sapeurs-pompiers (Artman, 1999). Nous appliquerons cette approche à la mesure de partage de conscience de situation dans le chapitre 3.

1.5 Synthèse de la problématique et question de recherche

Le secteur militaire, l'industrie aéronautique et aérospatiale, les structures de santé, les entreprises, les banques et les établissements d'enseignement sont aujourd'hui plus que jamais tributaires des équipes et du travail collaboratif. Un travail d'équipe efficace crée des connaissances, minimise les erreurs, encourage l'innovation, sauve des vies, améliore la productivité, augmente la satisfaction au travail et assure le succès et la compétitivité. Les équipes déployées, formées et dirigées correctement, constituent un puissant atout. Étudier la cognition d'équipe permet de révéler les facteurs sous-jacents à l'acquisition d'une compétence complexe par le collectif, et à sa réalisation. Cela constitue un élément clé pour diagnostiquer et évaluer les sources des succès et échecs d'une équipe. La compréhension de la cognition d'équipe permet ainsi de développer des entraînements et de concevoir des systèmes optimisés pour la performance du travail collectif. Les besoins actuels en terme de mesure du travail collaboratif sont résumés de la façon suivante par Salas et al. (Salas et al., 2008) :

"The pursuit of dynamic and adaptive systems that are sensitive to team-level performance requires unobtrusive and real-time measures of team performance that can be practically implemented, especially in the field." [22]

En particulier l'évaluation des performances d'équipe dans le contexte du C2 créent trois principales contraintes auxquelles les systèmes de mesure sélectionnés devront répondre. Les observations et recommandations formulées à la fois par Stanton *et al*. dans leur récente revue de la littérature des 20 dernières années (Stanton et al., 2017), par le groupe de travail HFM-156 de l'OTAN en 2014 (Berggren et al., 2014) et par la plupart des références des domaines du C2 et de la cognition d'équipe ces dernières années, se rejoignent pour affirmer le besoin de :

mesurer en temps réel, de manière **non-invasive** et **automatisable puis automatisée** la dynamique du travail d'équipe.

Au regard de ces contraintes exprimées ici et des différentes méthodes de mesure présentées dans ce chapitre, le monitoring de l'opérateur apporte une réponse au besoin d'évaluation en temps réel et automatisée requis par le contexte dynamique du C2. Le monitoring ou monitorage peut en effet être défini comme **la surveillance en continu de l'état d'un système ou d'un individu via un réseau de capteurs fonctionnant de manière automatique.**

[21] Laboratoire de recherche fondamentale et technologique de l'armée américaine - ARL

[22]« La recherche de systèmes dynamiques et adaptatifs sensibles à la performance des équipes exige des mesures non intrusives et en temps réel de la performance des équipes qui peuvent être mises en œuvre de manière pratique, en particulier sur le terrain. »

On retrouve deux niveaux possibles de monitoring, déjà présents dans les méthodes de mesure de la charge cognitive (1.4.1).

Il peut s'agir d'un monitoring psycho-physiologique, reposant sur la mesure de signaux produits par l'organisme ou le cerveau au cours de son fonctionnement normal. Citons par exemple la fréquence cardiaque, la température corporelle, la dilatation pupillaire ou le rythme respiratoire.

Il peut également s'agir d'un monitoring comportemental, via l'utilisation de caméras, d'enregistrements vocaux, de dispositifs de suivi oculaires (oculomètres ou *eye trackers*) ou de mesure d'interactions avec le système (par exemple l'activité de la souris).

L'objectif sous-jacent étant de prédire la performance de l'équipe par l'évaluation de la cognition collective, cela nous amène à formuler la question de recherche **Q_0** :

Comment analyser la **performance** d'un groupe (équipe ou ensemble social) par **l'évaluation en temps réel** de la **cognition collective** issue du **monitoring non-invasif** de ses individus collaborant (ou de certains de ses individus) ?

Nous prenons le parti de désigner comme « cognition collective » d'un groupe, équipe ou ensemble social, les dimensions collectives de la cognition individuelle des individus collaborant au sein du même groupe.

En nous appuyant sur la littérature et plus particulièrement les travaux de Berggren et al. (2011), nous pouvons préciser cette question par l'hypothèse suivante **H_0** :

La **performance** d'un groupe peut être analysée par l'**évaluation en temps réel** des indicateurs de **charge cognitive**, de **conscience de situation partagée** et des **communications** des individus collaborant, via des **mesures de monitoring non-invasives**.

Dans ces mêmes travaux sur l'évaluation de la cognition collective, les auteurs proposent une modélisation (Figure 19) des interactions et influences mesurées entre ces concepts.

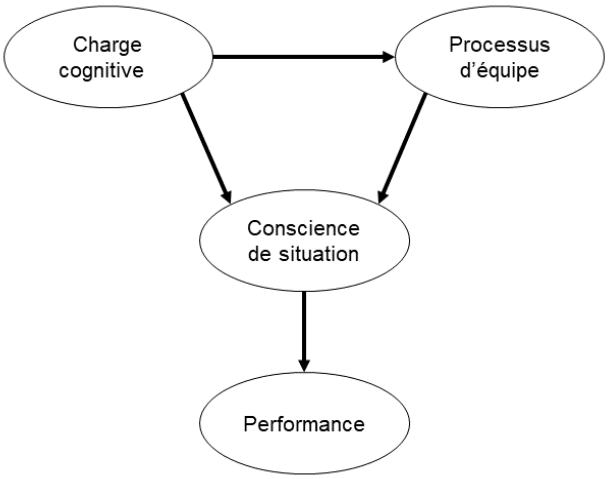


Figure 19. Modèle relationnel de la charge cognitive, la conscience de situation, les processus d'équipe (Teamwork) et la performance dans l'équipe, d'après Berggren et al. (2011)

Cette représentation est hautement critiquable, notamment par l'absence remarquable des influences réciproques que nous avons évoquées précédemment. Néanmoins elle illustre le rôle charnière de la

conscience de situation au sein de la cognition d'équipe. D'un point de vue individuel, la conscience de situation est un prérequis à la prise de décision ; d'un point de vue collectif le partage de représentations correctes et similaires de la situation entre coéquipiers et un facteur majeur de cohérence des décisions individuelles et donc de performance du groupe.

Le partage de conscience de situation nous semble donc être un témoin intéressant de partage de cognition. Ainsi, nous proposons d'aborder notre travail par l'angle de l'évaluation du partage de conscience de situation, en cherchant à intégrer les contraintes d'applicabilité au temps réel et de traitement automatisable exprimées précédemment.

Nous avons mentionné plus haut (1.3.3) le caractère intrinsèquement dynamique de la conscience de situation. Le prochain chapitre sera consacré à une analyse approfondie des besoins en terme d'évaluation de la dynamique du partage de conscience de situation, pour laquelle nous proposons le concept de synchronie des SA. Cela nous amènera à préciser certaines hypothèses.

Chapitre 2 Approche temporelle de la différence de partage de conscience de situation

La principale limite à laquelle se heurte l'évaluation de la conscience de situation est la mesure de sa dynamique. Nous verrons pourquoi cette mesure serait particulièrement pertinente dans le cadre du travail collaboratif. Nous introduirons le concept de synchronie des consciences de situation (SA[23]) et proposerons des pistes pour le mesurer. Ce cheminement nous amènera à précisera nos hypothèses de recherche.

2.1 Besoins et limites de l'évaluation de la conscience de situation

Au fil des ans, les besoins de mesure de la conscience de situation ont évolué. De l'accidentologie à la fin des années 80 (Foushee & Helmreich, 1988), la conscience de situation est devenue un sujet d'évaluation "in-situ" dans des contextes de formation ou de conception (Chen, Qian, & Lei, 2016; Endsley, 2016; Endsley, Bolstad, Jones, & Riley, 2003; Salas, Cannon-Bowers, & Johnston, 1997). Sa place centrale dans le processus de prise de décision, individuelle et d'équipe, fait de son évaluation un élément clé dans la prédiction et l'analyse des performances.

Alors que la technologie devient de plus en plus personnalisable, les objectifs des chercheurs et des ingénieurs ont progressé vers une évaluation en temps réel de l'état de l'utilisateur et de l'équipe. Le but est notamment d'offrir aux systèmes techniques des leviers de réaction automatiques pour palier à des états de l'utilisateur présentant des risques pour la performance. Par exemple dans les systèmes d'enseignement adaptatifs, des mesures en temps réel peuvent être utilisées pour surveiller et garantir un état optimal pour l'apprentissage, notamment en mesurant et en ajustant en continu le niveau de sollicitation attentionnelle (Carneiro, Pimenta, Gonçalves, Neves, & Novais, 2016; Szafir & Mutlu, 2012). Dans les contextes opérationnels, les mesures en continu offrent la possibilité de soulager

[23] SA - *Situation Awareness* : acronyme anglais pour conscience de situation.
Pour alléger le texte, les termes « SA » et « conscience de situation » sont tous deux employés dans ce chapitre.

l'utilisateur en adaptant dynamiquement les niveaux d'automatisation, les modes d'interaction ou de communication (Scerbo, 1996).

L'examen des mesures traditionnelles de la conscience de situation effectué au Chapitre 1, révèle une limite commune à toutes ces méthodes. En étant conçues pour détecter ce qui est erroné dans la représentation que l'individu se fait d'une situation réelle donnée, les méthodes se sont concentrées sur l'exactitude de la conscience de situation bien plus que sur le moment ou la rapidité avec laquelle elle est acquise, ce qui a souvent donné lieu à des évaluations principalement qualitatives. La conscience de situation étant une construction cognitive, cette évaluation de cette représentation de la situation requiert l'accès à son contenu (Endsley et al., 2003). Par conséquent, toute mesure directe nécessite la verbalisation de la représentation interne, la rendant inévitablement intrusive et nécessairement ponctuelle.

Cependant, à mesure que l'environnement évolue, la conscience de situation doit être reconstruite et mise à jour en permanence pour intégrer les nouveaux événements, informations et objectifs pertinents. Ainsi, la conscience de situation est intrinsèquement dynamique (Hjelmfelt & Pokrant, 1998; Nofi, 2000) et constitue une structure en constante évolution. L'incapacité à prendre en compte la nature dynamique de la conscience de situation est la principale critique exprimée à l'égard des techniques d'évaluation actuelles (Stanton et al., 2017). Toutefois, les évaluations quantitatives sont importantes pour savoir quand s'adapter. En particulier dans le cadre d'une équipe, une évaluation continue de la conscience de situation est nécessaire pour déterminer quand la performance de l'équipe risque de souffrir de consciences de situation asynchrones. De nombreux chercheurs s'accordent à dire que l'élaboration de mesures objectives de la conscience de situation, non intrusives et en temps réel est une prochaine étape logique et nécessaire (de Winter, Eisma, Cabrall, Hancock, & Stanton, 2019; Nofi, 2000). Malgré ce consensus, nous ne connaissons aucune étude qui ait examiné de plus près ce que signifierait la mesure de la dynamique de la conscience de situation.

Pour répondre à cette demande, nous proposons le concept de **synchronie des consciences de situation (*SA Synchrony*)** comme mesure de la dynamique temporelle de partage de conscience de situation. Ce concept a fait l'objet d'une présentation à une conférence internationale avec publication dans les actes (Prébot, Schwarz, Fuchs, & Claverie, 2020).

Les concepts de similarité et de synchronie des consciences de situation concernent nécessairement les éléments de connaissances devant être communs aux coéquipiers. On ne s'intéresse donc pas à l'ensemble de la conscience de situation des individus (SA), mais à la partie devant être partagée (SSA). Pour des raisons de simplification linguistique, nous désignerons ici la « conscience de situation partagée » (SSA) par l'expression raccourcie « conscience de situation ». Cela nous permettra de distinguer plus clairement les situations de partage et non-partage de « conscience de situation partagée » (SSA).

2.2 SA Synchronie : Vers une évaluation de la dynamique du partage de conscience de situation

La construction de la représentation de la situation est un processus profondément dynamique puisqu'il repose sur la perception et l'interprétation continue des éléments de l'environnement (Salas et al., 1995). Dans un contexte où *Human-Autonomy Teaming* (HAT) et systèmes adaptatifs, utilisés en formation et en situation opérationnelle, ont un besoin urgent de connaissance en temps réel de l'état de l'opérateur, la question de l'évaluation des modifications temporelles de la conscience de situation (dynamique de la SA) est devenue centrale (Adams, Tenney, & Pew, 1995; Ziemke, Schaefer, & Endsley, 2017). Dans le

cas de la collaboration, la fluctuation du degré de partage de représentation de la situation est un élément clé sur lequel se pencher. Savoir quand et pendant combien de temps les SA sont (ou ne sont pas) synchronisés peut-être un complément utile pour évaluer le partage de conscience de situation et prévenir l'erreur humaine dans un contexte d'équipe.

Ainsi, en complément de l'**exactitude** et de la **similarité** des SA habituellement mesurés, nous proposons de considérer le concept de **synchronie** des SA, définissant un indicateur de la dynamique du partage de conscience de situation dans l'équipe.

2.2.1 Modèle de la dynamique de partage de conscience de situation

Selon le modèle de conscience de la situation d'Endsley (1995), nous comprenons que la construction d'une conscience de situation partagée repose sur la perception et l'intégration similaire des bons éléments situationnels par les coéquipiers. Cain et Schuster désignent ces éléments d'information sous le nom d' « éléments de connaissance nécessaires » (*Necessary Knowledge Elements* - NKE) (Cain et al., 2016). Nous proposons d'étendre le concept à l'« élément de connaissance partagée nécessaire » (*Necessary Shared Knowledge Elements* - NSKE, Figure 20), définissant un élément d'information dont la connaissance est nécessaire à plusieurs membres d'une équipe pour accomplir une partie collaborative de leurs tâches, en d'autres termes, les "*shared SA requirements*" selon la définition de Endsley & Jones (1.3.6).

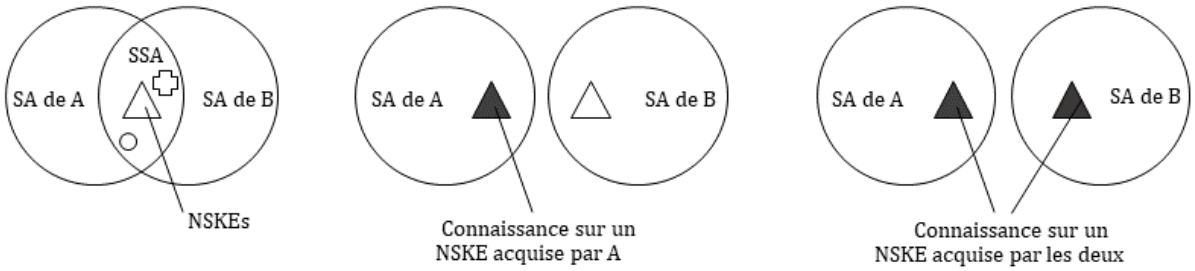


Figure 20. Illustration des trois états de connaissance possible sur un élément de connaissance partagée nécessaire (Necessary Shared Knowledge Elements - NSKE). Soit aucun des deux coéquipiers ne possède une connaissance à jour sur l'élément (schéma de gauche), soit l'un la possède et pas l'autre (schéma central), soit les deux sont à jour et partagent la connaissance de cet NSKE (schéma de droite).

Ces éléments de connaissance nécessaires dépendent de l'interconnexion des tâches des individus, ils sont donc préalablement clairement identifiés. La formation et l'apprentissage en équipe permettent notamment de construire les modèles mentaux partagés requis, comprenant entre autres cette connaissance des besoins mutuels en terme d'information. Cependant, les NSKE sont rarement perçus simultanément par chacun (Cain et al., 2016; Endsley & Jones, 2001). La perception étant la première étape du processus d'évaluation de la situation, il existe donc de fait toujours une latence entre la mise à jour de la conscience de situation de chacun des coéquipiers.

Prenons un exemple où toutes les informations nécessaires sont à la disposition de deux coéquipiers (A et B) et où ils ont tous deux réussi à en former la même représentation, partageant ainsi un même niveau de conscience de situation initial. Chaque fois qu'un nouvel NSKE apparaît, il invalide l'état de la conscience de situation actuelle jusqu'à ce qu'il soit intégré à la conscience de situation des coéquipiers,

pour obtenir une conscience de situation partagée à jour (Figure 21). Dans ce modèle, nous pouvons distinguer 4 phases, créant trois latences intéressantes pour l'évaluation de la conscience de situation partagée (SSA).

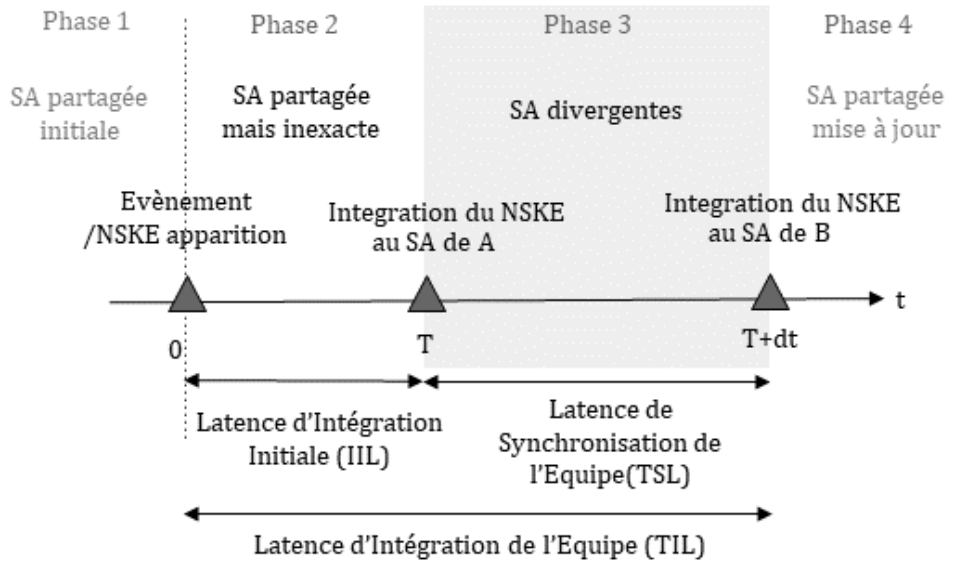


Figure 21. Illustration de l'évolution temporelle des SSA et les latences associées : Latence d'Intégration Initiale (Initial Integration Latency -IIL), Latence de Synchronisation de l'Équipe (Team Synchronization Latency - TSL) et Latence d'Intégration de l'Équipe (Team Integration Latency - TIL).

La latence d'intégration initiale (*Intial Integration Latency* - IIL) est le temps initial requis par le premier coéquipier pour percevoir et intégrer le nouveau NSKE à sa conscience de situation.

L'intervalle entre l'apparition du NSKE et son intégration dans la conscience de situation de A (phase 2 de la Figure 21) représente une période de conscience de situation partagée mais inexacte qui s'accompagne d'une probabilité accrue de prise de décision erronée. Pendant cette période, les coéquipiers possèdent toujours une représentation commune de la situation. Les décisions individuelles sont cohérentes entre-elles et les décisions collectives sont cohérentes avec la stratégie en cours. Cependant, la représentation de la situation n'est plus à jour. Cette différence avec la réalité augmente le risque de prise de décision inappropriée. La durée de cette latence est influencée par les mêmes facteurs concernant l'attention et le système perceptivo-sensoriel qui ont un impact sur la conscience de situation de niveau 1 (Level 1 SA) (Endsley & Garland, 2000) : stress, fatigue, charge de travail ou complexité de l'interface.

La seconde latence, appelée latence de synchronisation de l'équipe (*Team Synchronization Latency* - TSL), représente le temps mis par le second coéquipier pour percevoir et intégrer le nouveau NSKE à sa conscience de situation après que le premier coéquipier l'ait fait (phase 3 de la Figure 21). En tenant compte de la première latence, cela crée un intervalle de temps situé les deux mises à jours des SA, durant lequel les représentations de la situation divergent. Pendant cet intervalle de temps, en plus d'être inexacte pour au moins un des coéquipiers, la conscience de situation n'est pas non plus partagée, ce qui augmente la probabilité d'une prise de décision incohérente. Dans ce cas, deux coéquipiers, dont l'un est à jour avec la situation et pas l'autre, pourraient envoyer des instructions contradictoires à un troisième.

Enfin, la latence d'intégration de l'équipe (*Team Integration Latency* - TIL), est la somme des deux premières latences. Elle représente le temps écoulé entre l'apparition du NSKE et son intégration par le dernier membre de l'équipe concerné (Phase 2 + Phase 3 dans la Figure 21). Elle reflète la durée pendant laquelle les membres de l'équipe ne possèdent pas tous une conscience de situation exacte.

2.2.2 Méthode indépendante et méthode collaborative de synchronisation des consciences de situation

La conscience de situation est de nouveau partagée lorsque le deuxième membre de l'équipe acquiert le NSKE (c'est-à-dire après la deuxième latence). On peut distinguer deux modes d'acquisition de l'information : la méthode **indépendante** et la méthode **collaborative**.

La méthode de synchronisation **indépendante** (Figure 22) consiste en l'acquisition autonome du NSKE par les deux coéquipiers. Comme la perception du nouvel élément de situation se fait sans l'aide du coéquipier, l'IIL, la TSL et par extension la TIL sont soumises à des facteurs d'influence similaires.

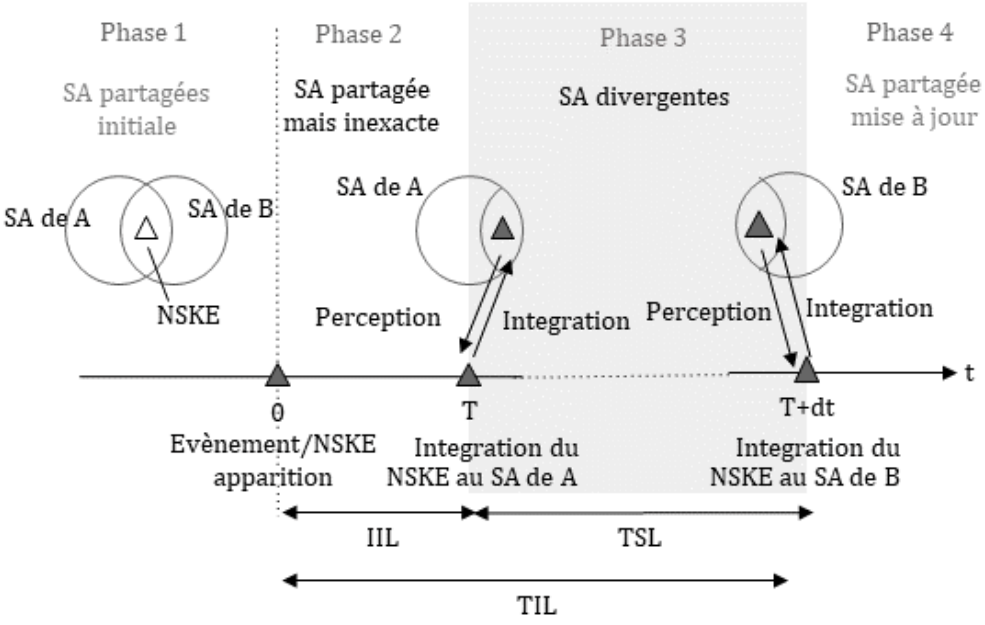


*Figure 22. Méthode **indépendante** d'acquisition d'un NSKE et retour à une conscience de situation partagée*

En revanche, l'acquisition **collaborative** du NSKE (Figure 23) est basée sur l'échange actif de l'information entre coéquipiers. Les recherches ont montré que la communication verbale ou électronique est essentielle dans le processus de construction et de maintien d'une SA partagée (Ososky et al., 2012; Perla, Markowitz, Nofi, Weuve, & Loughran, 2000). Le premier individu à percevoir un NSKE (coéquipier A) le communique à son coéquipier B quelque temps après l'avoir intégré dans sa propre SA. La communication peut comprendre l'élément de la situation lui-même (par exemple, "Il y a une nouvelle piste d'avion non identifiée" ; SA de niveau 1) ou inclure déjà des informations de plus haut niveau basées sur le traitement et l'interprétation faits par A (par exemple, "Identification nécessaire sur la nouvelle piste" ; SA de niveau 2).

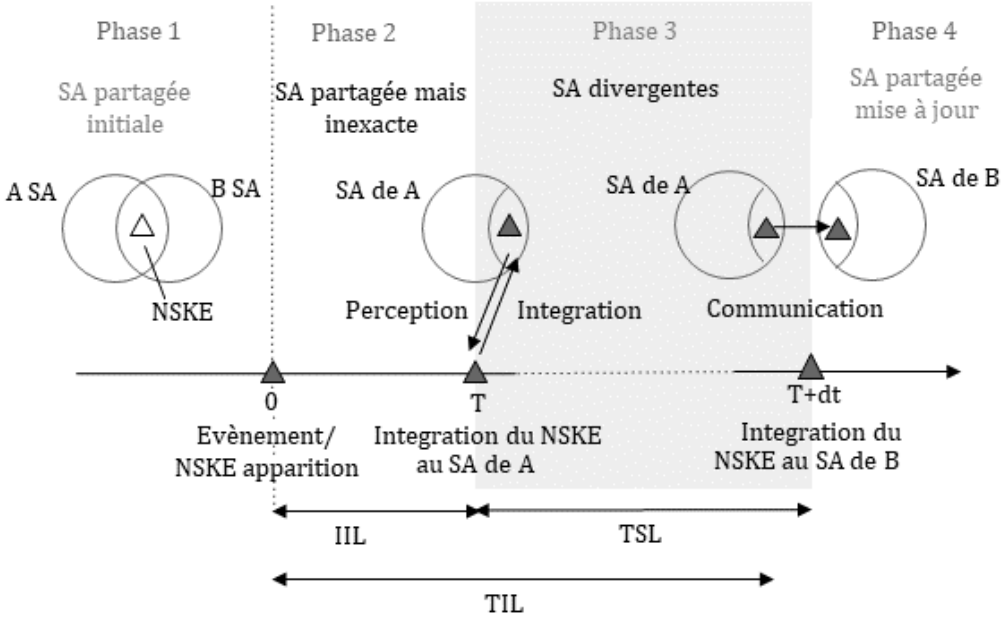


*Figure 23. Méthode **collaborative** d'acquisition d'un NSKE et retour à une conscience de situation partagée*

Dans ce cas, l'acquisition du NSKE par B dépend de son acquisition préalable par A, de sorte que la durée de la seconde latence est influencée par des facteurs supplémentaires. Si la perception et l'attention peuvent toujours être influencées par les facteurs mentionnés ci-dessus, il existe d'autres facteurs spécifiques à l'échange d'informations qui peuvent avoir un impact sur la communication et son contenu. Il peut s'agir, par exemple, de la charge de travail de A, de la priorité du NSKE par rapport aux autres tâches de A, de la qualité des modèles mentaux partagés ou de la capacité de B à recevoir et à traiter les communications.
La reconnaissance et l'échange efficace de NSKE exigent également une connaissance suffisante des tâches du coéquipier et de ses besoins d'information, ainsi qu'une estimation continue de sa conscience de situation actuelle. Cette méta-connaissance est par ailleurs un élément central de la conscience de situation d'équipe (*Team SA*)(Stanton et al., 2017). En étant inhérentes au processus de mise à jour et de partage de la conscience de situation, ces latences illustrent l'importance des propriétés dynamiques de celle-ci.

Ce modèle est applicable aussi bien à une équipe co-localisée qu'à une équipe distribuée et peut également convenir à l'étude de l'intégration Homme-Machine en considérant une des entités comme un système autonome (*Human-Autonomy Teaming* - HAT).

2.3 Perspectives pour la mesure de la synchronie des SA

Les mesures objectives directes de la conscience de situation ont été largement validées dans de nombreux domaines ; toutefois, elles reposent sur la connaissance d'un vérité objective de la situation (*Ground Truth*) à laquelle comparer les réponses. Ces mesures emploient ainsi une standardisation méthodologique pour objectiver une représentation autrement subjective de la situation. La conscience de situation étant une construction cognitive, l'évaluation du modèle de situation nécessite l'accès à son contenu (Endsley et al., 2003). Par conséquent, toute mesure directe nécessite la verbalisation de la

représentation, la rendant inévitablement intrusive et nécessairement ponctuelle. La conscience de situation jugée est donc celle d'un individu à un moment précis. Même si les différentes techniques d'évaluation définissent des méthodes pour identifier les périodes les plus pertinentes et critiques pour effectuer la mesure, elles ne traduisent pas l'évolution temporelle continue de la conscience de situation. Or l'évaluation doit être faite au regard de la demande imposée par la situation et la tâche. Un opérateur peut posséder une bonne conscience de situation à un instant t, mais ne l'avoir acquise que bien trop tard par rapport à l'apparition de l'événement dans l'environnement. Cette contextualisation, pourtant primordiale dans l'évaluation du risque encouru est actuellement absente des méthodes de mesure.

Bien que dans certaines situations il puisse être suffisant d'évaluer le degré d'exactitude des SA, détecter et mesurer les latences susmentionnées par des méthodes indirectes ouvrirait de nouvelles voies pour quantifier, qualifier et répondre en temps réel aux problèmes de partage des consciences de situation.

2.3.1 Quantifier la synchronie des consciences de situation

En 1997, Cooke et al. (1997) ont proposé une approche d'ingénierie cognitive pour la mesure de la conscience de situation des individus et des équipes comprenant, entre autres, l'analyse du comportement, des processus ou encore des démarches de raisonnement à voix haute. Certaines de ces méthodes trouvent un écho dans l'approche actuelle de monitoring humain en ingénierie cognitive. Le suivi des activités des membres de l'équipe (y compris les activités cognitives, physiologiques et comportementales) et leur comparaison, pourraient fournir un moyen indirect d'évaluer la conscience de situation aidant à surmonter les contraintes d'évaluation inhérentes à sa nature.

La synchronisation des consciences de situation peut être considérée comme la comparaison temporelle des réactions et des comportements liés au partage de conscience de situation entre deux individus dans une situation de travail collaboratif (Delaherche et al., 2012).

La perception d'un nouvel élément situationnel définit trois latences centrales dans le processus de synchronisation des SA, comme illustré au 2.2.2. Des techniques orientées vers la détection de cette perception pourraient ainsi être des solutions adaptées pour mesurer les latences. Certaines mesures comportementales et physiologiques présentent l'avantage d'être continues et sont déjà utilisées pour la quantification de l'état et de l'activité de l'utilisateur (Jorna, 1993; Schwarz & Fuchs, 2017; Tomarken, 1995). En tant que mesures continues, leur évolution temporelle peut être facilement comparée et des latences entre les réactions ou les comportements peuvent être observées.

En 2019, de Winter et al (2019) ont établi le bien-fondé de l'évaluation continue de la conscience de situation en utilisant une mesure dérivée des mouvements oculaires comme mesure prometteuse, bien que très améliorable.

Plus généralement, le suivi de la position du regard est considéré comme un outil de mesure de la conscience de situation potentiellement non intrusif. Des applications récentes dans l'aviation (Kilingaru, Tweedale, Thatcher, & Jain, 2013; Moore & Gugerty, 2010; van de Merwe, van Dijk, & Zon, 2012) et la conduite automobile (Hauland, 2019) ont permis de déduire la focalisation attentionnelle d'un individu à partir de modèles de fixation du regard qui peuvent ensuite être comparés entre coéquipiers. De même, les temps de réaction peuvent également être enregistrés par le suivi de la souris ou d'autres comportements l'interaction avec l'interface, comme exploré dans (Calcagnì, Lombardi, & Sulpizio, 2017; Freeman & Ambady, 2010; Frisch, Dshemuchadse, Görner, Goschke, & Scherbaum, 2015; Kieslich, Wulff, Henninger, Haslbeck, & Schulte-Mecklenbeck, 2018).

Compte tenu de toutes les techniques déjà explorées et de leurs résultats mitigés, une approche multi-mesures semble être nécessaire pour saisir des constructions complexes telles que la conscience de situation (Salmon et al., 2006). Une approche multidimensionnelle du monitoring de l'opérateur fournit la base d'une évaluation continue et objective en temps réel. Nous reconnaissons cependant qu'en se concentrant sur le comportement résultant (la réaction), la mesure pourrait être fortement influencée par des facteurs autres que la conscience de situation (Cooke et al., 1997). Les méthodes décrites ci-dessus présentent encore des limites d'utilisation en situation écologique et de nombreuses questions doivent être prises en compte, comme le phénomène "regarder mais ne pas voir" par lequel l'individu peut fixer un objet sans nécessairement le percevoir. Les techniques étant déjà très sensibles et complexes, la nécessité de les combiner rend leur application d'autant plus difficile. Nous avons été confrontés à ce problème dans les expériences présentées dans les chapitres suivants. Néanmoins vous verrons quelles stratégies nous avons dû mettre en place pour y remédier.

2.3.2 Qualifier la synchronie des consciences de situation

Si le concept de synchronie des SA des différents opérateurs se veut avant tout une mesure objective quantitative, son interprétation et sa qualification n'en sont pas moins intéressantes. Il est important de noter qu'étant donné que les communications ajoutent une latence inhérente, une synchronisation parfaite des consciences de situation entre les membres de l'équipe n'est ni réaliste ni nécessairement souhaitable (Salas et al., 1995; Salmon, Stanton, Walker, Jenkins, & Rafferty, 2010; Sonnenwald, Maglaughlin, & Whitton, 2004). Nous comprenons que lors d'une collaboration, la pertinence du NSKE, son interprétation et la hiérarchisation des tâches sont fonction des stratégies et des objectifs individuels. Ainsi, afin d'identifier des problèmes de conscience de situation, il peut être nécessaire d'évaluer l'écart par rapport à une latence attendue. L'interprétation ou la qualification de la mesure de synchronisation des SA requiert une compréhension approfondie des tâches individuelles et collectives, des processus et des communications. Comme l'indiquent Salas et al.(2017), la qualification des marqueurs comportementaux doit être contextualisée par rapport à l'environnement dans lequel ils sont appliqués. De même que l'exactitude optimale théorique (Hooey et al., 2011), une synchronisation optimale théorique des consciences de situation pourrait être définie sur la base de l'analyse des tâches d'équipe. Afin d'être qualifié, le lien entre la synchronisation des SA et la performance doit également être étudié.

2.3.3 Utiliser la synchronie des consciences de situation

Comme la plupart des mesures quantifiables, la synchronie des SA peut être utilisée comme descripteur de la collaboration ou de la performance. Elle convient à la fois à l'évaluation a posteriori du comportement global de l'équipe pendant les tâches et à l'évaluation en temps réel pour un système adaptatif ou fournir un feedback à l'équipe. Des états optimisés de synchronisation entre les coéquipiers ou avec la réalité peuvent être définis, ce qui permet d'identifier les périodes problématiques au cours des processus de collaboration et éventuellement définir des temps de latences optimaux. Dans un contexte de formation ou d'entraînement les scénarios et les NSKE sont souvent connus a priori, ce qui permet de définir des réponses anticipées pour mesurer l'IIL, la TSL et la TIL. Ainsi, les scénarios de formation peuvent permettre une évaluation plus facile de la synchronie des SA que les situations écologiques. Des interventions adaptatives pourraient alors être conçues pour réduire les latences problématiques.

2.4 Précision de l'orientation de la recherche

La nature dynamique de la conscience de situation est unanimement reconnue et son évolution temporelle fait l'objet de nombreuses discussions. Cependant, son évaluation a reçu peu d'attention par rapport à l'évaluation de l'exactitude. En se concentrant sur le contenu de la conscience de situation, les méthodes de mesure actuelles sont de fait limitées par la nature subjective et interne de la représentation de la situation ; rendant très compliquée une application à l'évaluation en temps réel et en situation écologique.

Ainsi, nous avons proposé de considérer le concept de synchronie des consciences de situation comme un indicateur de leur dynamique dans l'équipe. N'évaluant pas nécessairement la qualité du contenu de la représentation mentale, la mesure de la synchronie peut théoriquement dispenser le chercheur du besoin de verbalisation et donc d'interruption ou d'intrusion dans la tâche. Dans cette perspective, le monitoring d'opérateurs offre des perspectives de mesure, certes indirectes, mais applicables en temps réel.

En proposant le concept de synchronie des SA, nous souhaitons établir une base de travail et de réflexion sur l'évaluation de la dynamique du partage de conscience de situation, ainsi que sur les perspectives d'application qui en découlent en environnement simulé ou écologique.

Les travaux des trois chapitres expérimentaux qui suivent ont pour but de prendre la mesure de l'applicabilité de ce concept.

L'étude de la cognition collective et en particulier son évaluation par monitoring, est une thématique nouvelle et en développement au sein du laboratoire. Le besoin émanant aujourd'hui à la fois des milieux industriels et de la recherche, est d'identifier les outils et les métriques adaptés à leur évaluation en temps réel dans une perspective d'automatisation de celle-ci. Cette thèse se présente donc comme un « défrichage » de la thématique, que nous avons choisi d'orienter plus spécifiquement sur le partage de conscience de situation.

En vue d'explorer des pistes de réponse, notre première expérimentation (Chapitre 3) consistera à tester l'utilisation d'une métrique objective de mesure de partage de conscience de situation issue d'une méthode classique de type *freeze-probe*.

Nous consacrerons les deux expériences suivantes (Chapitre 4 et Chapitre 5) à évaluer la sensibilité du monitoring, individuel et collectif, aux modifications comportementales et physiologiques liées à la mise à jour de la conscience de situation.

Pour cela nous précisons l'hypothèse formulée au chapitre précédent de la manière suivante :

H1 : La **modification similaire d'un même indicateur** suite à l'apparition d'un nouvel élément de situation, est un indice de **partage de cognition** entre deux individus collaborant, en **particulier de partage de conscience de situation.**

Le domaine du monitoring physio-comportemental couvre un ensemble complexe d'outils, de capteurs, de métriques et de méthodes d'analyse dont la maîtrise dépasse les contraintes temporelles d'une simple thèse. Ne pouvant décemment pas explorer toutes les possibilités offertes, une orientation a donc dû être privilégiée.

Nous avons choisi l'utilisation de la pupillométrie par *eye tracking* comme principale mesure la cognition. La pupille est connue pour réagir (se contracter ou se dilater) en réponse à trois grands facteurs

: la lumière, l'émotion et la cognition (Beatty, 1982; Beatty & Lucero-Wagoner, 2000; McGuigan & Andreassi, 1980; Rossi, 1999). Sa dynamique ainsi que les latences de réaction à la présentation visuelle de stimulus ou de tâches ont fait l'objet de nombreuses études (Beatty, 1982; Klingner, Kumar, & Hanrahan, 2008), nous permettant d'ancrer nos travaux dans une continuité expérimentale par rapport à la littérature.

Ce choix s'explique également par des raisons techniques liées au matériel d'acquisition et de traitement disponible au laboratoire. Le diamètre pupillaire présente l'avantage d'être l'indicateur issu de l'*eye tracking* dont la mesure est la plus directe, à l'inverse des saccades oculaires ou fixations qui sont algorithmiquement dérivées de la position du regard.

Nous la complèterons par d'autres mesures selon les expériences.

Le chapitre 4 met en pratique la mesure de la pupillométrie individuelle associée à l'oculométrie (position du regard), pour identifier la réaction à la détection visuelle d'un nouveau NSKE.

Dans le chapitre 5, une analyse des similarités d'évolution des dilatations pupillaires de coéquipiers est entreprise.

Chapitre 3 Explore Mars

L'une des principales difficultés rencontrées durant ces travaux fut la disponibilité d'outils de simulations de tâches collaboratives pertinentes. Le protocole et l'expérience présentés dans ce chapitre tirent parti d'un outil de mise en situation présent au sein du laboratoire en l'utilisant comme support d'un travail collaboratif de type C2. Avant de nous orienter dans les chapitres suivants sur l'utilisation du monitoring, nous éprouvons les limites des méthodes des mesures de conscience de situation actuelles discutées dans le chapitre 2.

3.1 Contexte et objectifs de l'expérience Explore Mars

3.1.1 Contexte

L'exploration martienne n'est à première vue pas notre sujet. Cependant, elle constitue une thématique de recherche appliquée du laboratoire et nous avons eu l'opportunité, à ce propos, d'éprouver certaines méthodes de mesure de la conscience de situation dans ce contexte. Par ailleurs, l'intégration récente de l'Espace comme composante officielle de l'armée de l'air (désormais nommée Armée de l'air et de l'espace) illustre la place grandissante du contexte spatial dans les opérations militaires.

Les centres de contrôle sont des organes essentiels des opérations spatiales, nécessitant une communication entre le sol et l'espace. Ils sont depuis longtemps intégrés aux opérations spatiales militaires (Morton & Roberts, 2011). Le NASA Mission Control Center de Houston en est un exemple des plus célèbres. De plus, les problématiques "facteurs humains" liées au partage des représentations sont également bien étudiées dans le domaine spatial. Pour ne citer que deux exemples :

- les difficultés survenues durant la mission Skylab 4, lors de laquelle le centre de mission n'avait pas perçu la fatigue des astronautes et leur besoin de faire une pause, résultant en un arrêt volontaire des communications de la part des astronautes (Kanas, 2005).
- la mésaventure de Luca Parmitano, qui a failli mourir noyé dans son scaphandre spatial lors d'une sortie extravéhiculaire en dehors de l'ISS, le contrôle de mission ne comprenant pas la cause et sous-estimant les problèmes (NASA, 2013).

Nous proposons d'aborder le partage de représentation au travers d'un expérience utilisant une tâche d'orientation spatiale collaborative (Johansson, Hellgren, Oskarsson, & Svensson, 2013). Une telle situation est monnaie-courante des missions de recherche et sauvetage ou des opérations militaires, et présente l'avantage d'être aisément transférable à de nombreux domaines, tout en étant facilement simulable. Dans de tels environnements collaboratifs, la capacité partagée à se guider mutuellement et à localiser des objets sur des terrains inconnus, en temps réel et notamment en situation dégradée, est essentielle à la performance de l'équipe. Pour le faire efficacement, les coéquipiers doivent partager la représentation de la situation la plus actuelle et exacte possible.

La représentation spatiale de la situation a été utilisée dans une grande variété de tâches militaires et civiles et même de la vie courante. Ainsi, dans une étude de 2010, Johannsdottir et Herdman (2010) ont évalué la conscience de situation de conducteurs automobiles en leur demandant de rappeler la position des véhicules les environnant. Un autre exemple est celui de pompiers à qui l'on a demandé de marquer l'emplacement d'un incendie sur une carte. La conscience de situation est alors mesurée comme l'écart entre la situation du feu réel et l'endroit indiqué sur la carte (Artman, 1999). Cette méthode visant à évaluer la différence entre le réel et la représentation que s'en fait le sujet, assimilable à un « mini-SAGAT[24] », présente l'avantage de produire une mesure quantitative et objective de la conscience de situation tout en reposant sur une méthode classique. Nous nous proposons de procéder similairement dans une expérience de simulation d'exploration spatiale. La thématique particulière de l'expérience (environnement spatial) n'est pas centrale dans notre problématique. Elle permet néanmoins d'avoir une situation « pure », sans parasite, et facilement contrôlable.

Lorsqu'une personne est guidée à distance, sa position actuelle est un élément de connaissance partagée nécessaire (NSKE, cf. 2.2) à l'accomplissement de la tâche collaborative. La position est une donnée relativement simple à acquérir, qu'il s'agisse de la position réelle ou de la représentation mentale que s'en font les personnes collaborant. Les distances calculées à partir de ces positions sont alors des mesures quantitatives et objectives du partage des représentations, offrant un outil de mesure relativement simple à mettre en place.

Ainsi, en travaillant sur le partage de la représentation spatiale de la situation, nous considérons la conscience de situation spatiale (Klippel, Hirtle, & Davies, 2010) comme une partie de la conscience de situation, c'est-à-dire une restriction de la compréhension globale de la situation aux éléments concernant la position dans l'espace du coéquipier ou de soi-même. Nous avons donc besoin de mettre un individu dans une situation d'exploration d'un environnement inconnu. Pour cela, la simulation Explore Mars remplit cet objectif. Cette simulation de pilotage de véhicule d'exploration de surface (*rover*) a été développée par l'équipe de Jean-Marc Salotti dans le cadre des travaux portant sur les problématiques facteurs humains de l'exploration de l'environnement martien (Salotti et al., 2016). Nous proposons de nous placer dans une situation collaborative simple, où un opérateur en centre de contrôle doit vocalement guider un agent (astronaute) sur le terrain sans retour visuel de sa position. Le protocole a d'abord été conçu pour une mise en situation réelle au sein de la Mars Desert Research Station[25] et a

[24] SAGAT ou *Situation Awareness Global Assessement Technique* est une méthode d'évaluation de la conscience de situation consistant à interrompre la tâche afin de poser au sujet les questions relatives à la situation (cf. 1.4.2)

[25] Mars Desert Research Station (MDRS) est un laboratoire de recherche exploité par la Mars Society, simulant ce qui pourrait être une base habitée sur la planète Mars. Elle est située dans un désert de l'Utah (USA), où l'isolement relatif de l'installation permet des études de terrain ainsi que des recherches sur les facteurs humains.

fait l'objet d'un pré-test lors de la mission du Mars Society Crew 185 en Décembre 2017 (Salotti, Prébot, & Cinelli, 2019). Nous l'avons ensuite adapté à l'utilisation d'Explore Mars au laboratoire.

Ces travaux feront prochainement l'objet d'une étude comparative complémentaire lors d'une mise en situation dans un désert israélien en terrain analogue. Cette simulation a été retenue parmi d'autres expériences qui seront menées dans le cadre du programme Amadee-2020[26].

Les études présentées dans ce chapitre font l'objet de deux présentations à des conférences internationales avec publication dans les actes (Prébot et al., 2019; Prebot, Salotti, Vennin, & Claverie, 2020).

3.1.2 Hypothèses de recherche

Nous avons voulu explorer le potentiel de la mesure de la conscience de situation liée à la localisation d'un élément de situation, par une mise en situation appropriée et tout en conservant une méthode de mesure classique. Cette métrique pourrait répondre à certains des critères requis pour une automatisation de la mesure.

L'objectif de cette première expérience est alors d'étudier la manière dont la **similarité** et l'**exactitude** des SA sont liées aux performances de l'équipe, afin d'établir dans quelle mesure elles peuvent être utilisées comme indicateurs quantitatifs objectifs des performances de l'équipe. Pour rappel l'**exactitude** correspond à l'évaluation des réponses au regard de la réalité ; la **similarité** correspond au niveau de concordance entre les réponses fournies par les coéquipiers. Nous formulons ainsi la question de recherche suivante :

(Q1) L'**exactitude** et la **similarité** des SA spatiales peuvent-elles être utilisées comme mesures quantitatives de performance d'équipe ?

Dont découlent trois hypothèses :

> H1 : **Les équipes ayant échoué présentent une conscience de situation significativement plus mauvaise que les équipe ayant réussi.**
>
> H2 : **La performance temporelle de l'équipe est corrélée à l'exactitude des SA des coéquipiers, en particulier plus l'équipe possède une SA partagée exacte, plus elle réussit rapidement la mission**.
>
> H3 : **La performance temporelle de l'équipe est corrélée à la similarité des SA des coéquipiers, en particulier plus l'équipe partage une SA partagée similaire, plus elle réussit rapidement la mission.**

[26] Le programme Amadee-2020 est un programme de recherche mené par l'OEWF, l'agence spatiale autrichienne, en coopération avec l'agence spatiale israélienne, consistant à mener une mission intégrée en terrain analogue à Mars dans le désert du Néguev en Israël. Une petite équipe de terrain composée d'astronautes hautement qualifiés et équipés de simulateurs de combinaisons spatiales mènera des expériences préparatoires aux futures missions d'exploration humaine et robotique de Mars. - https://oewf.org/en/portfolio/amadee-20/

Notre travail se structure ici en deux parties. La première, traitant de l'hypothèse H1, est une analyse intergroupe entre les équipes ayant échoué et celles ayant réussi la tâche. La seconde, relative aux hypothèses H2 et H3, se concentre sur le lien entre le partage de conscience de situation des équipes ayant réussi et leur performance temporelle.

3.2 Plan d'expérimentation

Pour répondre à ces objectifs scientifiques nous avons conçu un protocole expérimental se basant sur la collaboration active de deux individus s'orientant dans une simulation informatique représentant un environnement jusqu'alors inconnu. Nous décrivons ci-après ce protocole expérimental en détaillant la situation collaborative étudiée, les données recueillies ainsi que les variables mesurées.

3.2.1 Idée expérimentale

La mise en situation se veut représentative d'une tâche collaborative simple de guidage à distance nécessitant un partage actif d'informations. Par équipes de deux, les volontaires doivent collaborer pour amener un *rover* (véhicule d'exploration de surface), aussi rapidement qu'ils le peuvent, à un endroit précis dans une simulation informatique d'un environnement martien. Pour ce faire des rôles spécifiques leurs sont attribués : astronaute ou capitaine. L'astronaute (Astro), est chargé de conduire le rover dans l'environnement virtuel et possède une carte vierge de cet environnement. Le capitaine (CapCom[27]) est quant à lui chargé de guider l'astronaute jusqu'à une cible, une pierre blanche, dont lui seul connaît la position. CapCom n'a pas accès au retour visuel de la simulation. La carte sur laquelle la position de la cible est indiquée est son seul outil. En résumé, la tâche de l'Astro est de faire naviguer le rover en suivant les instructions données par le CapCom, afin d'atteindre la roche désignée le plus rapidement possible.

3.2.2 L'outil Explore Mars

La simulation Explore Mars est une application développée sur le moteur informatique multiplateforme Unity®. A la manière d'un simulateur de conduite elle propose une immersion « à la première personne » dans un environnement numérique 3D représentant la surface martienne. L'utilisateur (Astro) est aux commandes d'un véhicule d'exploration de surface, contrôlé à l'aide des flèches directionnelles du clavier.

L'outil autorise le libre déplacement dans l'environnement dans les limites du terrain modélisé. Le conducteur est amené à naviguer dans un milieu désertique rocailleux composé de cratères et chaînes de montagnes facilement identifiables. Le véhicule est capable de passer pratiquement partout, à l'exception des portions montagneuses trop abruptes. En général, il y a peu d'obstacles nécessitant un contournement, hormis les grands cratères et les falaises (voir Figure 24 et Figure 25). Le logiciel a été adapté pour l'expérience de manière à se mettre automatiquement en pause toutes les 45 secondes et enregistrer la position du rover à cet instant.

[27] Ces appellations sont conventionnelles en matière spatiale et nous les avons adoptées pour des raisons de facilité de publication.

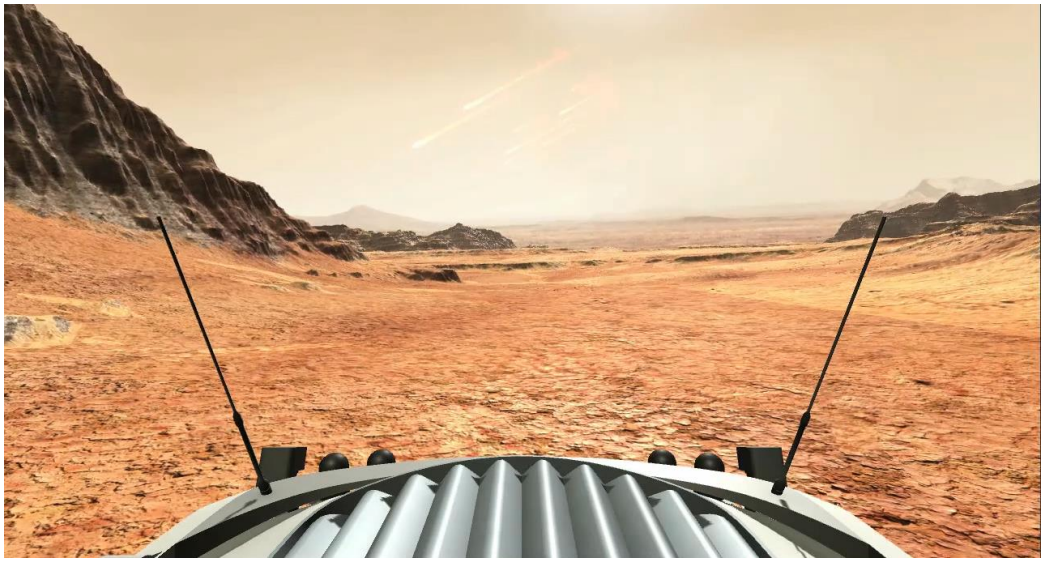


Figure 24. Illustration de la vue à la première personne de la simulation Explore Mars.

3.2.3 **Matériel et participants**

L'échantillon sur lequel porte notre étude est composé de 62 participants, dont 38 femmes et 24 hommes, âgés de 18 à 43 ans (âge moyen : 21,6). Ils ont été recrutés sur le campus de l'Université de Bordeaux. Afin de limiter un possible biais lié à l'habitude de collaboration entre les étudiants recrutés, les équipes ont été formées de binômes issus de groupes, promotions ou parcours différents.

Une carte au format A4 représentant l'environnement simulé (Figure 25) a été remise à chaque participant. Volontairement, aucune échelle n'est indiquée ni aucune grille ou système de coordonnées. Seul le point de départ du rover et son orientation sont présents (en haut à droite). Cette position de départ est la même pour tous les participants. Les cartes ont été données orientées de la même manière à chaque participant. La simulation est affichée sur un moniteur de 24 pouces

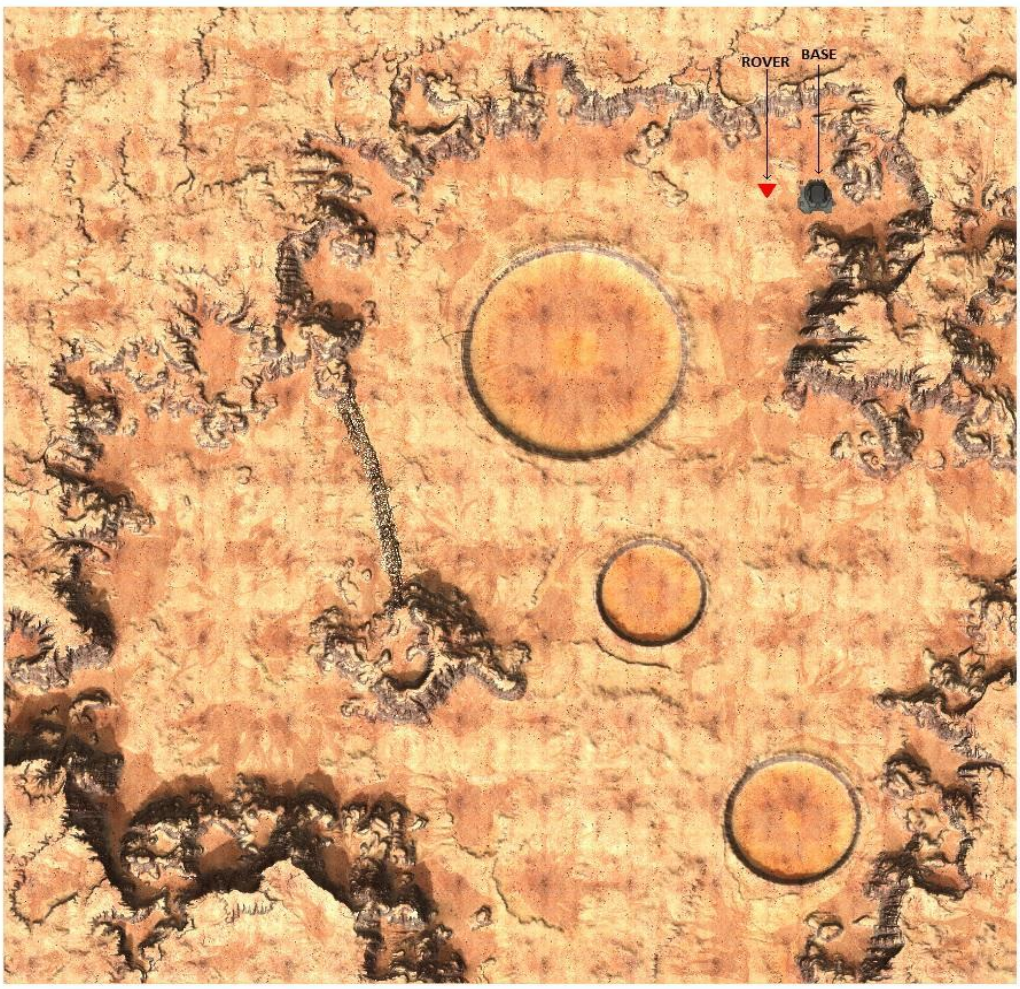


Figure 25. Support de carte représentant une vue aérienne de l'environnement de simulation (version présentée à l'Astro, ne contenant pas la position de la roche cible).

3.2.4 **Déroulement des passations et instructions aux participants**

La durée totale d'une expérience est de 30 et 45 minutes, en tenant compte de la variabilité de performance des participants ainsi que des temps de pause et de remplissage de questionnaire.

Chaque coéquipier se voit aléatoirement attribué le rôle d'Astro ou de CapCom. Ils sont ensuite conduits à leur poste de travail où ils reçoivent les instructions relatives à leur rôle. La carte leur est remise, la position de départ du rover et son orientation initiale y sont indiqués. Chaque participant est au préalable informé des objectifs de l'expérience, des conditions de passation, des données de respect de la RGPD et ils remplissent un formulaire de consentement éclairé.

Les sujets n'ont aucune connaissance préalable de l'environnement. Ils sont postés dans une même pièce mais physiquement séparés par une cloison de façon à ne pouvoir communiquer que de manière orale (Figure 26). CapCom reçoit l'ordre d'orienter l'Astro en temps réel et de ne pas donner directement la position de la roche (par exemple veiller à ne pas dire « la roche est dans le coin supérieur gauche de la carte »).



Figure 26. Configuration expérimentale. Les participants sont séparés physiquement par une cloison ne permettant que la communication orale

Toutes les 45 secondes, la simulation est automatiquement interrompue. Durant cette interruption, la position réelle du rover est enregistrée par le système et les sujets doivent indiquer sur leur carte par une croix où ils estiment que le rover se trouve (point d'évaluation de la position - PEP). Les coéquipiers ne sont pas autorisés à communiquer pendant cette phase de l'expérience, il leur est donc demandé de se taire.

La simulation est relancée lorsque l'expérimentateur a vérifié que les deux coéquipiers ont bien marqué leur estimation. Si après 15 minutes, les participants n'ont pas trouvé la roche, la tâche est considérée comme un échec et la simulation arrêtée.

Immédiatement après la tâche les participants remplissent un questionnaire post-situationnel.

3.2.5 **Recueil des données**

Trois grands types de données sont recueillies : les données temporelles, les positions réelles et estimées du Rover et enfin les données issues d'un questionnaire post-situationnel. Le détail de ces types de données est présenté ci-dessous.

3.2.5.1 ***Données temporelles***

Les binômes ont 15 minutes maximum pour accomplir la tâche demandée. Néanmoins ils sont chargés de la réaliser le plus rapidement possible. Les données temporelles permettent quantifier une performance du groupe. Les équipes n'ayant pas atteint l'objectif durant ce laps de temps sont considérées en échec. Pour les autres, leur performance est évaluée par le temps d'exploration (en secondes) mis par l'équipe pour retrouver la roche.

3.2.5.2 ***Données de position***

Trois positions sont relevées toutes les 45 secondes sur la durée totale de la passation. La position réelle du rover sert de référence pour l'estimation de l'**exactitude** des consciences de situation individuelles des coéquipiers. Cette position est donc automatiquement enregistrée au sein d'un fichier .txt, et exprimée en coordonnées relativement à la limite supérieure gauche de la simulation.

Toutes les 45 secondes il est également demandé aux sujets de marquer sur leur carte personnelle la position estimée du rover. Les cartes sont ensuite scannées et les coordonnées des « Points d'Évaluation de Position » (PEP) sont relevées au moyen d'une application C# développée pour l'occasion. Cette application nous permet également de projeter sur une même carte les trois mesures (Astro, CapCom, Réel) comme illustré sur la Figure 27. Ces données, comparées au réel ou entre-elles, serviront à l'évaluation quantitative de la **similarité** et de l'**exactitude** des consciences de situation.

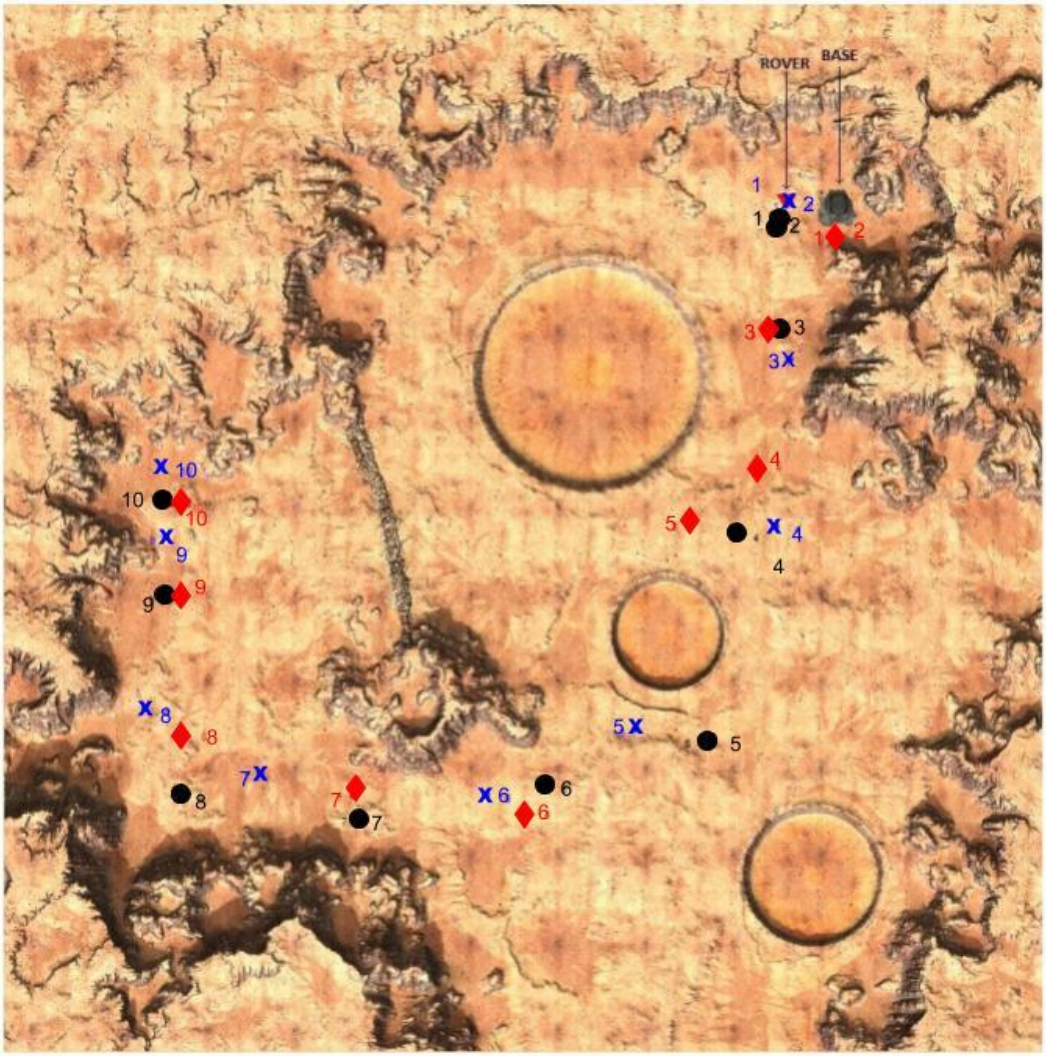


Figure 27. Exemple des positions relevées par Astro (Rouge) et CapCom (Bleu) et la position réelle (Noir). On observe immédiatement un fort écart des représentations pour le point N°5. Dans cet exemple l'équipe utilisée est la N°11. (Points grossis pour faciliter la lecture)

3.2.5.3 *Données du questionnaire post-situationnel*

Un questionnaire post-situationnel (Annexe 1) nous permet de recueillir la représentation que chaque sujet se fait de l'expérience passée et de leur collaboration. Ces évaluations étant personnelles, c'est par la comparaison des réponses au sein d'une équipe que peut être estimée la cohérence et le partage des représentations transmises.

Les questions qui leur sont posées portent sur deux thèmes :

- La performance : il s'agit non seulement de recueillir le ressenti du sujet vis à vis de sa performance individuelle mais également de celle de l'équipe et en particulier de son coéquipier. Trois questions leur sont posées :
    - *Qu. 1 : "Selon vous, avez-vous formé une bonne équipe ?*" (1-2) (Oui/Non)
    - *Qu.2 : "Comment évalueriez-vous votre propre performance ?"* (1-5) (Très faible-Excellente)
    - *Qu. 3 : "Comment évalueriez-vous la performance de votre coéquipier ?"* (1-5) (Très faible-Excellente)
- La communication : la perception générale de la répartition des échanges est évaluée par trois questions :
    - *Qu.4 : "Selon-vous, y'a-t-il eu une bonne communication au sein de l'équipe ?"* (1-2) (Oui/Non)
    - *Qu. 5 : "Qui parlait le plus ?"* (1-2-3) (Moi - Mon Coéquipier - Équilibré)
    - *Qu. 6 : "Qui posait le plus de questions ?"* (1-2-3) (Moi - Mon Coéquipier - Équilibré)

3.2.5.4 *Récapitulatif des données*

Le Tableau 4 résume les catégories de données recueillies.

Catégorie de données	**Données enregistrée**
Données Temporelles	- Temps mis à trouver la roche
Données de Position	
- **Issues de la simulation**	- Timestamp - Position réelle du rover en X - Position réelle du rover en Y
- **Issues des cartes** Pour chaque point d'évaluation de position (PEP)	- N° du Point d'Évaluation de Position - Position estimée du rover en X par l'Astro - Position estimée du rover en Y par l'Astro - Position estimée du rover en X par CapCom - Position estimée du rover en Y par CapCom
Donnée de Questionnaires Généraux	- Âge - Sexe - Habitude de travail avec le coequipier (1-5)
Données de Questionnaires Post-situationnel	- Qualité de la collaboration (1-2) - Performance personnelle (1-5) - Performance du coéquipier (1-5) - Qualité de la communication (1-2) - Prédominance de la parole (A-C) - Prédominance des questions (1-C)

Tableau 4. Résumé des données recueillies par catégories

3.3 Exploitation des données

Les deux métriques spatiales d'exactitude et de similarité de la conscience de situation sont calculées à partir des positions relevées sur les cartes des coéquipiers et des enregistrements réalisés par la simulation. Les paragraphes qui suivent détaillent ce calcul. Les coordonnées de chaque Point d'Évaluation de Position sont exprimées dans le repère relatif à la carte.

3.3.1 Exactitude des consciences de situation

L'exactitude des consciences de situation est mesurée objectivement en comparant la position des marques sur chaque carte avec la position réelle enregistrée par la simulation. Pour chaque Point d'Évaluation de Position, l'exactitude est donc déterminée par la distance euclidienne entre la position estimée et la position réelle. Cette distance est exprimée en métrique de mesure « Unity » (ici 1 unité = 0,083% de la carte). Pour chaque évaluation sont ainsi calculées l'exactitude de l'Astro, l'exactitude de Capcom et, l'exactitude de l'équipe par moyenne des deux autres. Nous en dérivons trois exactitudes moyennes sur la durée de l'expérience, les moyennes individuelles d'Astro et CapCom et l'exactitude moyenne de l'équipe. Si pour l'étude qui suit la moyenne a été privilégiée pour caractériser la performance du groupe, la pertinence d'autres métriques comme la variance ou l'écart-type n'est pas à exclure.

Cette mesure implique que les valeurs les plus faibles correspondent aux meilleures exactitudes et réciproquement. L'exactitude parfaite est ainsi associée à la valeur 0.

3.3.2 Similarité des consciences de situation

La similarité des consciences de situation est mesurée en calculant la distance euclidienne entre les positions marquées par les coéquipiers pour une même Point d'Évaluation. Nous obtenons donc une valeur commune à Astro et CapCom par PEP. Une valeur de similarité moyenne par équipe est également calculée.

Plus la valeur calculée est faible, plus la conscience de situation des coéquipiers est similaire. Pour une parfaite similarité de représentation la valeur mesurée serait de 0. Notons par ailleurs que si l'exactitude tend vers 0, la similarité tend également nécessairement vers 0 par convergence vers le référentiel commun qu'est la réalité.

3.3.3 Questionnaire

Les données du questionnaire sont exploitées comme suit. Les valeurs données par les échelles de Likert[28] sont exploitées telles quelles. En revanche, la formulation des questions relatives à la répartition des communications nécessite de ramener les réponses à un référentiel commun afin de permettre les comparaisons. Les réponses, initialement « Moi », « Mon Coéquipier » ou « Équilibré », ont ainsi été ramenées à 3 valeurs : « Astro », « CapCom », « Équilibré » (par exemple si l'Astro a répondu « Moi » à la question « Qui pose le plus de questions ? » la valeur attribuée est « Astro », s'il a répondu « Mon coéquipier », la valeur attribuée est « CapCom »).

[28] Parfois appelée « échelle de satisfaction », l'échelle de Likert comprend cinq ou sept options de réponse, qui couvrent le spectre d'opinions, d'un extrême à l'autre

3.4 Résultats

Globalement, les conducteurs n'ont eu aucune difficulté à prendre en main les commandes du véhicule et à comprendre l'environnement. Etant données les contraintes géologiques du terrain simulé, le cheminement attendu correspond à peu près à celui qui est observé sur la Figure 27. Parmi les 31 équipes, 3 ont accidenté le *rover* et n'ont donc pas pu mener la tâche à bien. Elles ont été écartées de l'analyse. 8 des 28 binômes restant ont échoué à la tâche, ne réussissant pas à atteindre la cible dans le temps imparti de 15 minutes. Pour les 20 équipes ayant réussi, le temps moyen de complétion est de 450,5s, soit 7 minutes 35s (min = 274 s ; max = 648s, SD = 102,56). Les chemins observés étaient très variés, certains passant d'un côté ou de l'autre d'un cratère, d'autres rebroussant parfois chemin à plusieurs reprises, cherchant un passage entre les montagnes.

Nous exposons ici deux analyses. Pour chaque analyse, un rappel des questions de recherches et des hypothèses associées est effectué. Dans un premier temps nous nous intéressons aux différences entre les deux groupes que nous nommons « Réussite » et « Echec ». Puis nous nous penchons plus précisément sur les équipes ayant réussi. Nous étudions notamment les liens entre la performance temporelle et le partage de conscience de situation. A la suite de ces deux analyses nous présentons brièvement les résultats issus de l'exploitation du questionnaire. Ces résultats ne concernent pas directement l'évaluation du partage de conscience de situation discuté dans ce cadre mais traitent du partage de représentation du travail collaboratif. Ils ont fait l'objet d'une étude particulière détaillée dans (Prébot et al., 2019). Il nous a donc semblé tout de même judicieux d'en faire état à la fin de cette partie. Les analyses et graphiques présentés ont été réalisés avec le langage R. Nous prendrons comme seuil de significativité alpha = 0,05.

3.4.1 Partage de conscience de situation et réussite

Cette première analyse se penche sur la comparaison du groupe constitué des équipes ayant réussi à trouver la roche dans le temps imparti et du groupe comprenant les équipes ayant échoué. Nous étudions l'hypothèse H1.

3.4.1.1 *Exactitude moyenne des SA et Réussite*

Le Tableau 5 décrit la distribution des **exactitudes** moyennes des équipes pour le groupe « Echec » et le groupe « Réussite ».

Exactitude

	Min	**Q1**	**Médiane**	**Q3**	**Max**	**Moyenne**	**Ecart-Type**	**Nb**
Echec	127,1	154,1	213,1	267,1	314,4	213,9	72,3	8
Réussite	46,16	66,70	91,77	106,6	318,87	103,62	65,3	20

Tableau 5. Tableau comparatif des distributions de l'Exactitude moyenne des équipes en échec et en réussite

La Figure 28 présente cette comparaison.

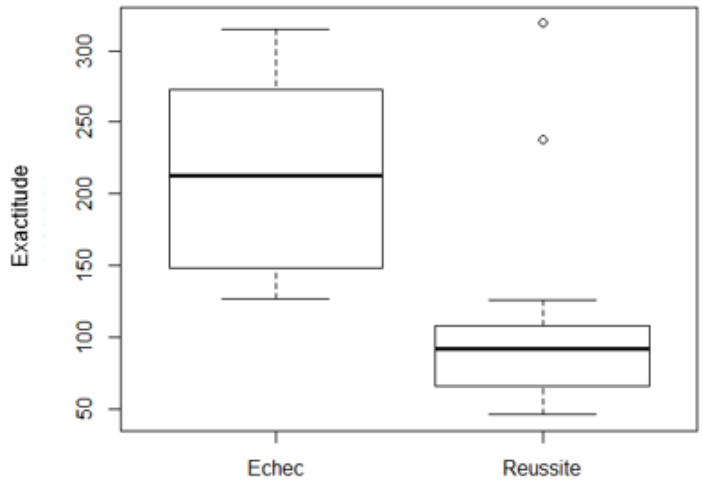


Figure 28. Comparaison Boxplot de l'Exactitude moyenne des SA des équipes appartenant aux groupes Echec et Réussite.

On observe une meilleure **exactitude** moyenne des équipes qui ont réussi, également caractérisée par un écart-type plus faible. La Figure 28 nous autorise à supposer une différence significative entre les deux groupes. Un test de Student nous permet de confirmer cette significativité en terme d'exactitude moyenne du SA entre les équipes qui ont échoué et celles qui ont réussi (p-value = 0,00284 < 0,05). Ce résultat montre que les équipes qui réussissent ont en moyenne une représentation plus précise de la position du Rover que les équipes qui échouent.

3.4.1.2 *Similarité moyenne des SA et Réussite*

Sur le même principe, comparons maintenant les évaluations de **similarité** des équipes ayant échoué et celles ayant réussi. Le Tableau 6 décrit la distribution des similarités moyennes des équipes pour les groupes « Echec » et « Réussite ».

Similarité

	Min	**Q1**	**Médiane**	**Q3**	**Max**	**Moyenne**	**Ecart-Type**	**Effectif**
Echec	94,01	129,99	169,10	194,06	304,72	174,58	69,74	8
Réussite	49,65	73,50	82,87	108,90	201,08	92,94	35,53	20

Tableau 6. Tableau comparatif des distributions de l'Exactitude moyenne des équipes en echec et en réussite.

Cette comparaison est visualisée sur la Figure 29.

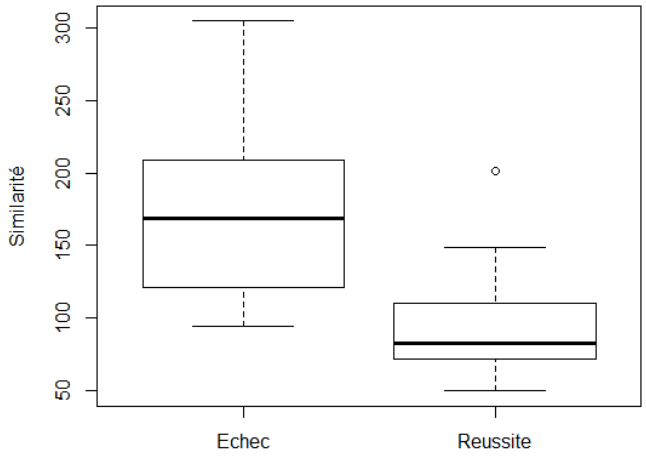


Figure 29. Comparaison boxplot de la Similarité moyenne des SA des équipes appartenant aux groupes Echec et Réussite.

On constate une meilleure similarité moyenne des équipes qui ont réussi, caractérisée également par un écart-type plus faible. En nous appuyant sur la Figure 29 nous pouvons à nouveau supposer une différence significative entre les deux groupes. Cette différence significative est confirmée par un test de Student (p-value = 0,01258 < 0,05). D'après ce résultat les équipes qui réussissent ont en moyenne une représentation plus similaire, c'est à dire commune de la position du *rover* que les équipes qui échouent.

Ce résultat, associé au résultat similaire obtenu pour l'exactitude, permet donc de confirmer l'hypothèse H1 : Les équipes ayant réussi possèdent une conscience de situation à la fois significativement plus précise et plus similaire que les équipes qui échouent.

3.4.2 Partage de conscience de situation et performance temporelle

Nous proposons désormais de nous concentrer sur les 20 équipes ayant réussi. Cette seconde analyse se penche sur les hypothèses H2 et H3 concernant le lien entre les degré d'exactitude et de partage des SA des coéquipiers et la performance temporelle. Les résultats sont présentés dans les paragraphes suivants.

Dans un premier temps l'**exactitude** et la **similarité** moyenne des SA par équipe ont été calculées. Le Tableau 7 résume les variables d'équipe de similarité moyenne, d'exactitude moyenne et de temps.

Réussite

	Min	**Q1**	**Médiane**	**Q3**	**Max**	**Moyenne**	**Ecart-Type**	**Effectif**
Exactitude moyenne	46,16	66,70	91,77	106,67	318,87	103,62	65,28	20
Similarité moyenne	49,65	73,50	82,87	108,90	201,08	92,94	35,53	20
Temps (s)	274	398,5	469,0	514,5	648	461	103	20

Tableau 7. Statistiques descriptives des distributions de l'Exactitude moyenne, Similarité moyenne et performance temporelle du groupe Réussite.

Les scores obtenus par équipe permettent de représenter la performance en terme de SSA de chaque équipe comme illustré par la Figure 30.

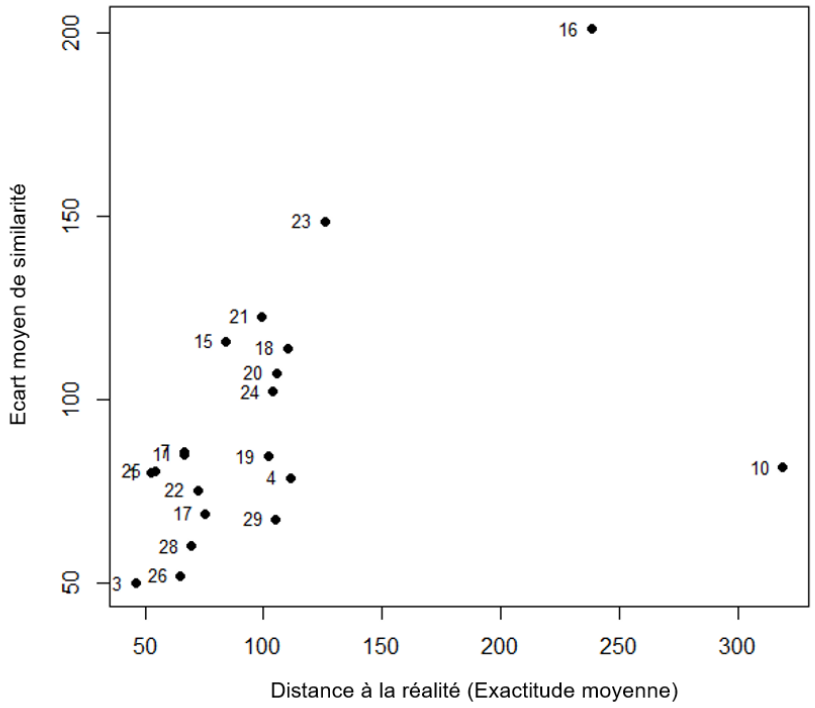


Figure 30. Équipes représentées par leur Similarité et Exactitude moyenne

Deux équipes attirent particulièrement l'attention. L'équipe n°10 (coin inférieur droit) est la moins exacte, avec une **exactitude** moyenne de 318,87, soit plus de trois fois la valeur moyenne, mais possède une **similarité** inférieure à la moyenne (81,39). Elle illustre des coéquipiers qui possèdent une représentation de la situation erronée mais partagée.

L'équipe n°16 (coin supérieur droit du graphique) possède des scores moyens de similarité et d'exactitude tous deux près de deux fois supérieurs aux valeurs moyennes (Similarité = 201,08 ; Exactitude = 238,17). Les deux coéquipiers sont incorrects de manière différente, induisant des représentations de la situation à la fois inexactes et très différentes.

Ces deux cas spécifiques couvrent deux des trois états de SSA possibles tels que décrits par Endsley et Jones (Endsley & Jones, 2001), avec une représentation inexacte mais similaire ou non de la situation. De façon surprenante, la performance temporelle de ces deux équipes est moyenne (G10 = 477s ; G16 = 387s). Ces deux équipes aux comportements particulièrement différents des autres ont néanmoins été écartées de l'analyse. Une étude plus approfondie de leur collaboration serait nécessaire.

L'analyse suivante concerne donc les 18 équipes restantes.

Nous cherchons à expliquer la variabilité des performances temporelles par l'exactitude et la similarité moyenne des équipes.

La Figure 31 représente les équipes positionnées en fonction de leur performance temporelle et de leur exactitude (graphique de gauche) ou de leur similarité (graphique de droite).

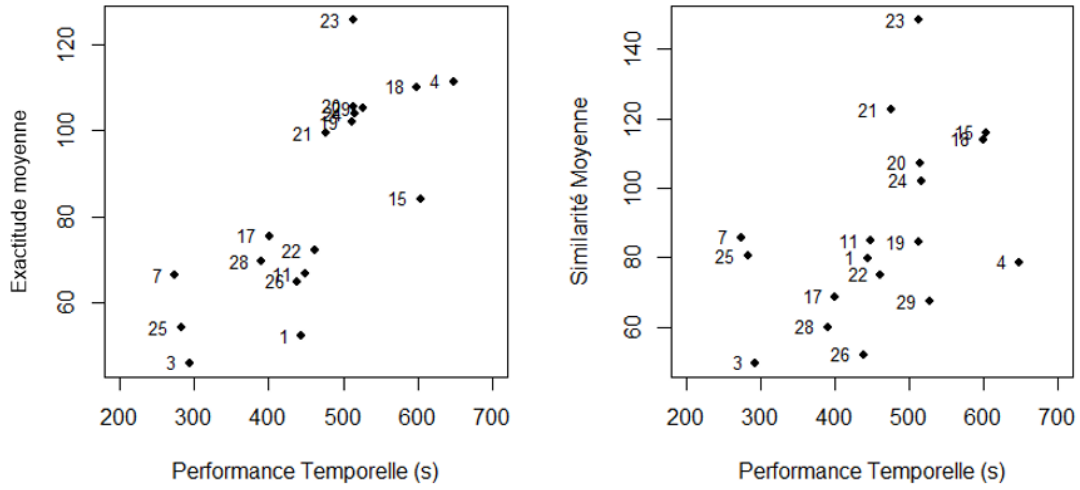


Figure 31. Performance Temporelle des équipes en fonction de leur Exactitude moyenne (Gauche) et Similarité moyenne (Droite)

Des tendances linéaires croissantes sont observables dans les deux graphiques de la Figure 31, la tendance apparaît plus marquée pour l'exactitude (graphique de gauche) que pour la similarité (graphique de droite). Cette tendance est confirmée par modélisation linéaire. Sur les 18 équipes l'exactitude seule explique 58,51% de la variabilité de la performance (p-value : 0,0002175 <0,01), quand la similarité n'en explique que 21,34% (p-value = 0,0536 > 0,05). On constate donc un lien linéaire croissant relativement fort entre l'exactitude de l'équipe et la performance temporelle.

Le test de corrélation de Pearson permet de confirmer la significativité de cette relation. Le résultat confirme la relation linéaire positive entre l'exactitude moyenne du SA partagée et la performance temporelle de l'équipe (r=0,76 ; p<.001). Pour l'équipe, plus la distance moyenne à la réalité est faible au cours de l'expérience, plus le temps mis sera faible et donc la performance meilleure. Ainsi cela valide notre hypothèse H2.

Pour sa part, la similarité seule n'apparaît pas significativement corrélée à la performance temporelle avec une p-value (5,36%) très légèrement supérieure à la valeur seuil (risque de première espèce choisi à 5% pour les interprétations). On est cependant très proche de la significativité et une tendance linéaire croissante reste visuellement observable.

Il va de soi que plus les représentations des coéquipiers sont exactes plus elles sont similaires, puisque convergeant vers la même réalité. Un rôle de la similarité n'est donc à notre sens pas à exclure, mais l'exactitude apparaît comme facteur prédominant pour la performance temporelle.

Par ailleurs, l'analyse de l'exactitude moyenne par sujet révèle une différence intéressante entre les profils Astros et CapCom. En effet la représentation des Astros est en moyenne plus exacte (exactitude moyenne Astro=97.67, écart-type=107.02) que celle des CapCom (exactitude moyenne CapCom=141.06, écart-type=137.03). Cela signifie que l'Astro à en moyenne une représentation de sa position plus exacte que CapCom. Ce résultat est cohérent, l'Astro étant l'individu dont la position est évaluée et celui naviguant la simulation il est de fait plus enclin à posséder une estimation correcte de sa propre position dans l'environnement. Sur la base de ces observations, nous pouvons également supposer qu'en moyenne l'exactitude de l'Astro à plus d'impact sur la performance temporelle de l'équipe que l'exactitude de CapCom.

3.4.3 Résultats complémentaires concernant le Travail Collaboratif

Au-delà des hypothèses principales de ce chapitre, il nous a semblé pertinent de présenter les résultats tirés de l'exploitation des réponses au questionnaire, portant sur le ressenti plus ou moins partagé des coéquipiers vis-à-vis de deux aspects importants du travail collaboratif : la performance et la communication. Ces résultats sont notamment détaillés dans (Prébot et al., 2019).

3.4.3.1 *Partage de représentation de la performance*

La performance perçue de l'équipe est biaisée par le succès ou l'échec à la tâche. Au-delà de l'observation triviale selon laquelle les équipes qui réussissent estiment leur performance meilleure que celles qui échouent, il est intéressant d'explorer l'influence du rôle du sujet dans l'évaluation de ses propres performances et de celles de son coéquipier. Chaque participant a répondu aux questions *Qu.2* et *Qu.3*. Pour rappel, la première consistait à évaluer sa propre performance et la seconde celle de son coéquipier, sur une échelle de 1 à 5 (1 étant la plus faible et 5 la plus élevée).

Une tendance intéressante apparaît lors de l'examen des différences entre l'évaluation de la performance du sujet et celle de son coéquipier. Nous constatons que les équipes ayant échoué présentent un déséquilibre plus important des évaluations que les équipes ayant réussi. Les Astros ont tendance à mieux évaluer les performances de leurs coéquipiers que les CapCom. De même, ils se trouvent généralement moins performants que leur coéquipier. Le résultat de cette analyse intra-groupe suggère que les Astros sont plus enclins à assumer la responsabilité de l'échec que les CapCom.

Une interprétation possible serait celle de l'existence d'un lien hiérarchique implicite perçu par l'Astro. Dans le contexte d'une tâche d'orientation spatiale, les conseils de CapCom peuvent être interprétés comme des ordres. CapCom étant l'instructeur qui sait où aller, Astro pourrait penser que l'échec résulte de sa mauvaise compréhension des instructions de CapCom et de son incapacité à fournir un retour d'information approprié. Il semble donc plus enclin à endosser la responsabilité de l'échec, alors que CapCom ne partagerait pas nécessairement le même sentiment.

Dans le cadre de la future étude comparative (programme Amadee 2020) il sera pertinent de mettre en place des entretiens post-expérimentaux afin d'adresser cette hypothèse d'organisation spontanée d'une hiérarchie.

3.4.3.2 *Equilibre ressenti des communications*

Durant l'expérience nous avons pu observer une diversité intéressante des comportements, notamment au sein des communications. La communication est connue pour être un processus central dans le travail et la performance d'équipe (Salas et al., 1992; Schmidt, 1991). C'est un processus nécessaire à la construction d'une compréhension commune de la situation de l'équipe. Certains groupes ont semblé former très rapidement un langage commun ou mettre en place une structure des échanges d'informations, tandis que d'autres ont fait preuve de plus de liberté et variabilité dans leur comportement. Un enregistrement audio n'ayant pas été inclus au protocole, nous nous concentrons sur le ressenti des participants vis à vis de l'équilibre des communications. Les réponses aux questions *Qu.5* et *Qu.6* ont été étudiées, concernant (1) la répartition perçue des communications générales ("Qui a parlé le plus ?") et (2) la répartition perçue des questions ("Qui a posé le plus grand nombre de questions ?").

Il apparaît que les membres des équipes ayant échoué semblent s'accorder sur le fait que les communications et les questions sont davantage dirigées par CapCom, tandis que les équipes ayant réussi ont tendance à être d'accord sur une répartition davantage équilibrée des communications.

De nouveau, il sera pertinent de prendre en compte ce constat lors de la future étude, notamment par une analyse approfondie et objective des communications, passant par l'enregistrement des échanges.

3.5 Discussion des résultats

3.5.1 Conclusion

Dans ce chapitre nous avons détaillé la mise en pratique d'une méthodologie de mesure de la conscience de situation reposant sur l'interruption de la tâche, pratique classique de ce domaine. L'objectif de cette phase expérimentale a été d'évaluer l'utilisation d'une métrique quantitative objective dont le traitement peut potentiellement être effectué de manière automatique en temps réel, nous permettant de nous rapprocher, avec des techniques de mesure actuelles, des objectifs établis au chapitre 2 en terme de futur de la mesure de conscience de situation. La mise en situation choisie, à savoir un travail collaboratif reposant sur l'orientation spatiale, nous a permis d'évaluer la pertinence des critères d'**exactitude** et de **similarité** des SA.

Par la première étude nous avons pu constater que l'exactitude des consciences de situation individuelles et leur similarité entre coéquipiers permettent a posteriori de distinguer les équipes performantes des équipes en échec. Nous validons ainsi notre hypothèse H1.

Dans la seconde étude, l'analyse approfondie de ce lien a confirmé la corrélation positive existant entre l'exactitude de la conscience de la situation, c'est à dire sa proximité à la réalité, et le temps mis par l'équipe pour mener à bien la mission qui lui était confiée. En particulier il s'est avéré, plus la représentation mentale moyenne de la position est exacte plus l'équipe est performante (rapide) à la tâche, validant l'hypothèse H2. Si cette hypothèse n'avait à notre connaissance jusqu'alors pas été vérifiée pour le critère d'exactitude de la position spatiale, le résultat reste peu surprenant.

En revanche, la similarité moyenne des consciences de situation de l'équipe n'a statistiquement pas pu être corrélée à la performance. L'hypothèse H3 n'a donc pas pu être validée. Cependant le faible rejet statistique de l'hypothèse et la tendance de corrélation linéaire positive malgré tout observée laissent penser qu'une exploration de métriques un peu différentes (par exemple l'écart-type ou la variance moyenne des similarités) peut s'avérer pertinente. Nous mettons l'accent sur le besoin d'étude approfondie de la similarité car nous considérons qu'elle présente un potentiel particulièrement intéressant en tant qu'outil de mesure de la conscience de situation indépendant de la connaissance à priori de l'état réel de l'environnement. En calculant la similarité non pas comme l'écart entre les exactitudes des représentations comme cela peut être fait habituellement, mais comme une distance entre les points symboles de la représentation des individus, nous avons testé une mesure quantitative et objective, potentiellement utilisable dans un environnement où l'état réel est difficilement accessible. Ainsi, en l'absence d'une connaissance absolue de la réalité de la situation, la similarité pourrait néanmoins fournir des informations sur la cohérence de la représentation partagée du groupe.

3.5.2 **Discussion et pistes de recherche**

La démarche entreprise dans ce chapitre suscite de multiples pistes de réflexion que de futurs travaux pourront explorer. Tout d'abord, cette situation expérimentale semble être propice à l'étude approfondie des communications. Bien qu'aucun enregistrement vocal n'ait été réalisé, nous avons pu observer durant l'expérience l'importance de la mise en place d'un langage commun entre les coéquipiers afin de faciliter la construction d'une représentation partagée de l'environnement. Lors d'une expérience similaire, une analyse lexicale des échanges verbaux entre coéquipiers pourrait permettre d'identifier la temporalité de création de ce langage commun et ainsi analyser la dynamique de la conscience de situation partagée au regard de la mise en place de ce langage. Par exemple nous pouvons nous demander si le processus de création d'un langage commun se reflète dans l'évolution de la Similarité des consciences de situation.

Cette approche nous a également permis d'obtenir une mesure quantitative objective du partage de conscience de situation et ouvre la porte à l'étude de sa dynamique en temps réel. En droite ligne des besoins exprimés au chapitre 2, la mesure répétée d'une métrique quantitative permet d'envisager son traitement sous forme de série temporelle. Même s'il ne s'agit pas à proprement parler d'une mesure continue, elle nous donne néanmoins accès à une certaine dynamique d'évolution de la conscience de situation. Elle présente cependant plusieurs limites.

Principalement cette méthode requiert l'intervention du sujet, comme toute collecte de type *Freeze Probe*[29]. Elle conserve donc l'intrusivité reprochée à ce type de méthodes et paraît ainsi peu adaptée à la mesure en situation opérationnelle réelle. Elle reste néanmoins appropriée pour des situations d'entraînement ou de formation et présente un potentiel intéressant en tant qu'indicateur de performance de l'équipe. Par exemple, une version numérique du pointage sur carte permettrait aisément un traitement automatisé et direct des réponses pour fournir un indicateur en temps réel du partage de conscience de situation de l'équipe. Sa dynamique peut alors être évaluée au cours de l'entraînement à la fois relativement aux mesures précédentes pour indiquer si la conscience de situation des coéquipiers est en train de s'accorder ou se désaccorder, et au regard des précédents entraînements pour estimer une progression plus générale de l'équipe. Une analyse contextualisée de ces résultats peut alors fournir des pistes d'amélioration pour le comportement, la formation et l'organisation de l'équipe. Grâce à l'automatisation aisée de son traitement et sa relation avec la performance temporelle, une telle mesure peut également constituer un élément pertinent à intégrer au sein d'un système de prédiction en temps réel de la performance.

Nous pouvons également discuter de la pertinence de l'utilisation de la localisation comme NSKE[30]. Les informations de position et plus généralement les données situationnelles sont centrales dans de nombreuses tâches de C2. Dans une étude menée par la DARPA[31] portant sur le programme RAID[32], Kott et Milks (2009) utilisent notamment les données de localisation des multiples bataillons ennemis pour comparer l'exactitude du système prédictif des déplacements aux prévisions humaines. Cependant elle n'est généralement pas le seul NSKE pertinent. Par exemple la position d'un aéronef doit être

[29] Freeze Probe : ensemble de méthodes de recueil de données de conscience de situation par questionnaires posés durant l'interruption de la tâche

[30] NSKE : Necessary Shared Knowledge Element, élément de la situation dont la connaissance doit être partagée entre les coéquipiers pour remplir leur tâche.

[31] DARPA : Defense Advanced Research Projects Agency, agence américaine chargée de la recherche et du développement de technologies à usage militaire.

[32] RAID: Real-time Adversarial Intelligence and Decision-making. Programme de recherche de la DARPA entre 2004 et 2008 portant sur la mise au point d'outils de fusion des données permettant d'estimer en temps réel les objectifs, les actions, les mouvements et les positions probables de l'ennemi.

associée à son cap, sa vitesse et son identité pour déterminer son intention. Toute mesure de la conscience de situation est relative aux éléments du contexte dans lequel elle s'ancre et une situation plus complexe avec de multiples NSKE de différents types demanderait des traitements séparés et donc de repenser l'estimation globale de la conscience de situation.

De plus, nous n'avons pas cherché à qualifier la conscience de situation. Nous avons privilégié l'estimation de degrés de similarité et d'exactitude, proches des définitions données par Endsley (1995) et Klein (2000). Or une interprétation des valeurs est requise pour permettre la prise de décision ou l'engagement d'actions en réponse à l'état mesuré. Cette qualification des consciences de situation est une étape supplémentaire qui nécessite de déterminer des écarts acceptables pour désigner les représentations comme bonnes ou mauvaises, partagées ou non. Autrement dit il convient de fixer des seuils de qualification de la qualité pour pouvoir agir en fonction et ainsi utiliser les mesures. Cela demande également de prendre en compte l'incertitude dans la précision de la notation par le sujet. Ces sujets seront prochainement abordés par un projet étudiant.

En résumé, la pertinence de cette métrique ne peut être dissociée de la tâche et, même si elle présente un potentiel intéressant pour la mesure de la dynamique du partage de conscience de situation en entraînement, elle ne répond pas aux besoins pour l'évaluation continue en situation écologique. Dans le chapitre suivant nous allons donc aborder l'utilisation de métriques continues se voulant également objectives, issues d'une acquisition non intrusive. Nous tirerons parti du monitoring comportemental et physiologique individuel, lors d'une tâche plus complexe de surveillance d'espace aérien, afin d'étudier la sensibilité de ces outils à la modification de la conscience de situation.

Chapitre 4 Monitoring Individuel - Expérience AMIGOS

Nous abordons dans ce chapitre la question de l'utilisation du monitoring individuel. Nous avons postulé au chapitre 2 qu'une différence de conscience de situation puisse se refléter dans des différences comportementales individuelles. Ce chapitre vise à évaluer la pertinence de la pupillométrie comme outil de mesure continue de la conscience de situation individuelle, dans l'optique par la suite d'une comparaison des comportements au sein d'une équipe.

4.1 Contexte et objectifs de l'expérience AMIGOS

4.1.1 Contexte de collaboration

L'étude du monitoring d'opérateur en situation de travail collaboratif présente deux difficultés majeures. D'une part, un de nos objectifs étant de se rapprocher de situations écologiques, la mise en situation d'un travail collaboratif requiert un outil de simulation à la fois suffisamment complexe pour permettre la réalisation de tâches collectives et suffisamment simple pour être appréhensible sans formation étendue, et ce afin de répondre aux disponibilités du public expérimental (à savoir principalement des chercheurs et des étudiants). Une étude des outils de simulation envisagés est disponible en annexe (Annexe 3). Après quoi la conception d'un outil propre répondant à ces critères et autorisant une induction de différences de conscience de situation (SA) a par ailleurs été engagée, sur la base du logiciel *Warship Commander Task* (St. John et al., 2004) utilisé par la DARPA[33] pour le projet *Augmented Cognition* (voir 6.2.2).

D'autre part, le monitoring est une démarche lourde, coûteuse et complexe d'un point de vue expérimental. Lorsque le matériel et les logiciels spécifiques d'acquisition, de synchronisation et de traitement des données, ainsi que les méthodologies ne sont pas déjà présents dans l'équipe de recherche, la mise en place d'une telle démarche, d'autant plus dans le contexte de travail collaboratif écologique demande le développement d'une expertise et un investissement important.

[33] DARPA : Defense Advanced Research Projects Agency, agence américaine chargée de la recherche et du développement de technologies à usage militaire.

Après une première expérimentation collective dont nous faisons état au Chapitre 4, ces raisons nous ont poussés à entreprendre une collaboration avec l'institut allemand Fraunhofer for Communication, Information Processing and Ergonomics (FKIE).

Cette collaboration a été rendue possible par des contacts établis dans le cadre de ma participation au groupe de recherche IST[34] 157 de la Science and Technology Organization[35] de l'OTAN, portant sur les Considérations Humaines pour l'IA dans le C2. Au sein du FKIE, une équipe de recherche travaille depuis 2014 sur le monitoring multidimensionnel de l'opérateur de frégate de marine militaire, dans l'optique de son utilisation au sein de systèmes adaptatifs. Au vue de leur expertise, de la similarité de nos champs de recherche et de leur contexte d'application, une collaboration a vu le jour, résultant en un séjour de recherche de 4 mois au sein de leur équipe. Nous avons ainsi trouvé un point de convergence de nos intérêts concernant l'applicabilité du monitoring individuel dans le cadre d'une évaluation de travail collaboratif.

Le présent chapitre fait état des travaux que j'ai pu réaliser pendant de ce séjour de 4 mois en Allemagne et de l'exploitation des données que j'ai poursuivie en France.

4.1.2 Expérience AMIGOS

L'institut Fraunhofer pour la Communication, le traitement de l'information et l'Ergonomie (FKIE), développe depuis 2014 un *framework* de diagnostic en temps réel de l'état de l'utilisateur et de sa performance (Schwarz & Fuchs, 2017; Schwarz et al., 2014). Intitulé RASMUS[36], cet outil a pour objectif d'être utilisé au sein de systèmes d'aide adaptatifs. Il doit identifier les moments où l'utilisateur a besoin d'aide puis sélectionner et appliquer la stratégie la plus à même de rétablir la performance de l'utilisateur dans la situation présente.

Alors que de nombreux chercheurs se limitent à l'utilisation de la mesure du *workload* comme déclencheur d'adaptation du système (Kaber, Prinzel, Wright, & Clamann, 2002; Parasuraman, Hockey, Gaillard, & Burov, 2003), certains auteurs considèrent que cette approche unidimensionnelle n'est pas à même d'apporter les adaptations les plus adéquates en situation réelle (e.x., Dorneich, Ververs, Mathan, Whitlow, & Hayes, 2012; S. Fuchs et al., 2006; Steinhauser, Pavlas, & Hancock, 2009).

Ainsi Schwarz et al. (Schwarz et al., 2014) proposent une approche multidimensionnelle et identifient six dimensions interconnectées de l'état de l'utilisateur susceptibles d'impacter la performance humaine[37] (Figure 32)

[34] IST : Information Systems Technology, panel technique de la STO rassemblant des groupes de recherche internationaux collaborant à l'amélioration des systèmes d'informations (C3I) des pays de l'OTAN.

[35] STO : Science and Technology Organisation. Organisme civil de l'OTAN en charge de la recherche scientifique et technologique.

[36] RASMUS : Real-time Assessment of Multidimensional User State

[37] Pour une présentation plus complète de cette approche se référer à Schwarz et al. (2014) "Towards a more holistic view on user state assessment in adaptive human-computer interaction".

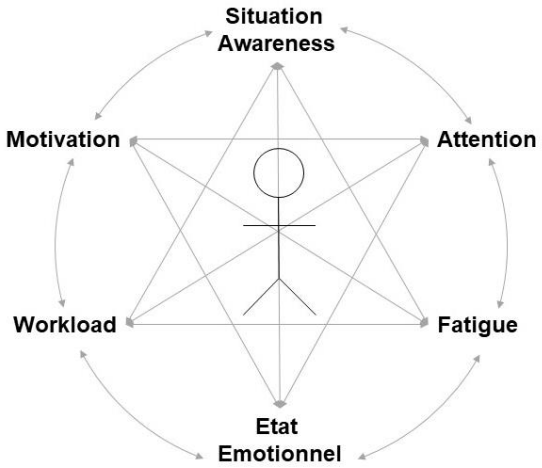


Figure 32. Concept de la multidimensionnalité de l'état de l'utilisateur, d'après Schwarz et al. (Schwarz et al., 2014)

Une preuve de concept de l'outil RASMUS ayant été développée, une première étude de validation de son potentiel de diagnostic a été réalisée en 2016. Cette étude, nommée AMIGOS, s'intéresse particulièrement à trois états de l'utilisateur : *workload* élevé, focus attentionnel incorrect, fatigue passive liée à la tâche.

A chacune de ces trois dimensions correspond une stratégie d'aide particulière, c'est à dire une adaptation de l'interface visant à assister et soulager l'utilisateur. L'objectif de RASMUS est alors de détecter une baisse de performance, définie comme le retard dans le traitement d'une tâche, puis de caractériser l'état de l'utilisateur comme une combinaison de ces trois dimensions. Ce diagnostic lui permet ensuite de déterminer la stratégie d'aide la plus adéquate (Figure 33). Ainsi, lorsque le *workload* est impliqué, le système déclenche une automatisation des tâches de faible importance. Lorsqu'une mauvaise focalisation attentionnelle est diagnostiquée, il présente des indices visuels pour orienter l'attention vers les tâches prioritaires, et des indices auditifs lorsque la fatigue est impliquée.

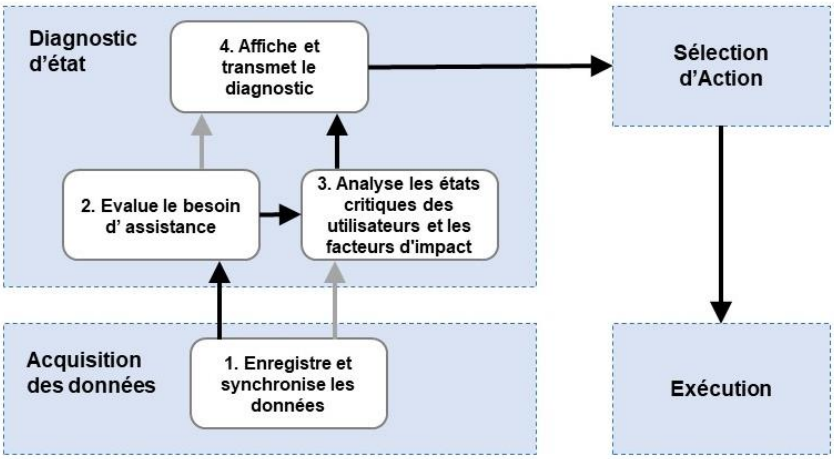


Figure 33. Etapes du processus de diganostic de RASMUS, d'après Schwarz et al. (2014)

Ce système de diagnostic repose sur l'acquisition d'une combinaison de données à la fois physiologiques, comportementales et environnementales. Un électrocardiogramme et un *eye tracker* recueillent des métriques physiologiques d'activité cardiaque, respiratoire et oculaire ; une caméra enregistre les expressions faciales et le système détecte également les clics de souris de l'utilisateur. Le système estime l'état de l'utilisateur en se basant sur des critères et seuils d'activités préalablement fixés. Chaque tâche est associée à un temps imparti à l'utilisateur pour la réaliser. Par exemple, si une minute après l'apparition d'un nouvel évènement, celui-ci n'a pas été traité, le système considère qu'il y a une baisse de performance. Dans le cadre de cette étude, la détection d'une baisse de performance interrompt automatiquement la tâche et le sujet est soumis à une série de questionnaires évaluant son *workload*, sa conscience de situation et sa fatigue (S. Fuchs & Schwarz, 2017).

Afin d'accroître la significativité des résultats, l'expérience a été répétée en Décembre 2018, se concentrant toujours sur la capacité de diagnostic et laissant de côté la partie stratégie d'adaptation. Cette nouvelle expérience, même si elle reprend l'essentiel des conditions et protocoles de la première, a été conçue et menée conjointement lors du séjour de collaboration réalisé au sein de l'équipe. Les travaux présentés dans la suite de ce chapitre résultent de la collaboration avec l'équipe Mensch-Maschine-Interaktion et l'exploitation d'une partie de la grande quantité de données recueillies lors de ces deux phases de test.

4.1.3 Problématique et hypothèses de recherche

Comme nous l'avons précisé au Chapitre 2, nous nous intéressons à l'applicabilité d'une détection automatique de la mise à jour de conscience de situation. La conscience de situation est traitée par RASMUS en assimilant un focus attentionnel incorrect à une erreur de SA de niveau 1. Une tâche dont le traitement n'est pas encore engagé est considérée comme non perçue. Or il se peut qu'un opérateur ait perçu une tâche à réaliser mais qu'il ait délibérément choisi de ne pas la traiter ou de différer son traitement. En d'autres termes, une erreur de focalisation attentionnelle peut également provenir d'une mauvaise interprétation du stimulus (erreur de SA de niveau 2 et 3) ou d'une prise de décision inadéquate. RASMUS ne permet pas de distinguer de tels cas d'une absence de perception.

Conscients des erreurs d'interprétation dues à cette simplification, nous souhaitons explorer l'utilisation de différentes sources de monitoring pour redéfinir la focalisation attentionnelle mesurée par le système, notamment via le biais de la focalisation visuelle.

Afin d'être traité, un stimulus doit d'abord être perçu et intégré à la conscience de situation du sujet. En identifiant les phases de perception et de traitement cognitif il serait possible de qualifier plus sûrement l'intégration (ou non) d'un nouvel événement à la conscience de situation et d'affiner ainsi la mesure de la focalisation attentionnelle. Pour ce faire, nous proposons de compléter la mesure de l'activité comportementale (clic de souris) par les données non encore exploitées de position du regard et de diamètre pupillaire du sujet.

Focalisation attentionnelle et focalisation visuelle sont régulièrement associées (pour une revue voir Carrasco, 2011). Posner et al. (1980) définissent la focalisation attentionnelle visuelle comme la focalisation fovéale sur un élément, c'est-à-dire son passage en vision centrale. L'étude que nous proposons reprend une structure relativement commune en psychophysiologie de l'attention. Nous nous intéressons à la réaction pupillaire induite par une tâche découlant d'un stimulus, s'apparentant ainsi à ce qui est communément référé en tant que *Task Evoked Pupillary Response* (TEPR) (Beatty, 1982; Claverie & Paty, 1999; Klingner et al., 2008; Peysakhovich, Causse, Scannella, & Dehais, 2015).

Lors de l'étude de TEPRs, assez classiques en psychologie cognitive, le sujet est souvent préparé à l'arrivée d'un stimulus particulier (par exemple par l'apparition d'une flèche ou d'un mot) auquel correspond une tâche particulière (par exemple cliquer à droite si la flèche est orientée vers la droite). Cette analyse binominale Stimulus-Réaction est caractéristique des études sur le temps de réaction. Dans ce contexte de recherche, en « situation de laboratoire », la perception du stimulus est considérée comme concomitante à son apparition. De fait, l'étude Perception-Réaction est assimilée à l'étude Stimulus-Réaction et l'intérêt est porté sur l'influence de différents facteurs sur le temps de réaction à la tâche et la réaction pupillaire associée.

Cependant dans le cas de la présente étude, la situation quasi-écologique diffère des situations standard de laboratoire. Elle laisse plus de liberté à l'individu dans le déroulement de la tâche, celui-ci n'est par ailleurs pas préparé à l'arrivée d'un stimulus particulier. Ainsi, comme nous l'avons exposé au Chapitre 2, il existe nécessairement une latence entre l'apparition du stimulus et sa perception par le sujet. La perception du stimulus, puis la focalisation attentionnelle sur ce dernier, ne peuvent pas être confondus avec son apparition. Cela nécessite donc d'effectuer la distinction Stimulus-Détection-Réaction.

Comme nous le détaillons plus bas (4.2.2) nous différencions deux temps de latences à étudier : le temps de perception du stimulus impliquant la tâche (latence entre l'apparition du stimulus et sa détection visuelle) et le temps de réaction (latence entre la détection du stimulus et l'action réalisée dessus). C'est notamment sur cette différence majeure découlant du réalisme accru de la tâche, que s'axe notre travail.

D'après Endsley (Endsley, 1995), la perception marque l'initialisation du processus d'actualisation de la conscience de situation. La réaction marque quant à elle l'initialisation du traitement effectif de la tâche. Entre ces deux temps un traitement cognitif a nécessairement été effectué, notamment l'identification du stimulus, la mise à jour des représentations et le processus de décision. Nous y reviendrons en conclusion de ce chapitre. Dans leurs travaux menés en situation de laboratoire, Privitera et al. (2010) avancent que la détection visuelle d'un stimulus recherché (pour eux une cible), provoque une réaction pupillaire marquée et identifiable. Nous formulons ainsi les questions de recherche suivante :

(Q1) En situation de tâche écologique pouvons-nous distinguer la réaction pupillaire liée à la détection visuelle d'un nouvel élément de situation important dans le cadre de la tâche ?

(Q2) Le traitement cognitif découlant de la détection visuelle, soit le changement de représentation traduisant la prise en compte du stimulus et ses implications, provoque-t-il une réaction pupillaire identifiable ?

Desquelles nous tirons l'hypothèse suivante :

H1 : **La détection visuelle d'un nouvel élément situationnel présentant une importance pour la tâche se traduit par une réaction pupillaire majeure identifiable, témoignant de l'actualisation immédiate de la SA.**

Que nous décomposons en deux hypothèses secondaires :

H1-1 : **Il existe, entre l'apparition d'un stimulus important pour la tâche et l'événement de traitement par un clic, un pic de dilatation pupillaire significatif.**

H1-2 : **Ce pic de dilatation étant provoqué par le traitement cognitif du NKE[38] initialisé lors de sa perception visuelle, il est atteint dès les premières secondes suivant la première fixation sur le NKE.**

Cette étude s'ancre dans une approche multi-monitoring. Nous nous servons conjointement d'enregistrements oculométriques, pupillométriques et comportementaux.

Nous décrivons dans un premier temps le plan d'expérimentation et la tâche de surveillance d'espace aérien utilisée. Puis nous présentons le processus de préparation et de prétraitement des données pupillaires. Enfin nous présentons les résultats de nos analyses, s'articulant autour de deux approches, une première quantitative et une seconde qualitative. Finalement nous discutons des résultats et observations issues de ce travail, au regard de leur applicabilité à la situation écologique.

4.2 Plan d'expérimentation

4.2.1 La tâche de surveillance d'espace aérien

L'expérience s'appuie sur une simulation réaliste de défense d'espace aérien maritime développée au sein du FKIE lors d'une précédente étude (Kaster, Tappert, Ruckert, & Becker, 2010). L'interface de cet outil est similaire à celle de l'outil de gestion de combat implémenté dans les postes de travail des Frégates Allemandes F125 (Figure 34).

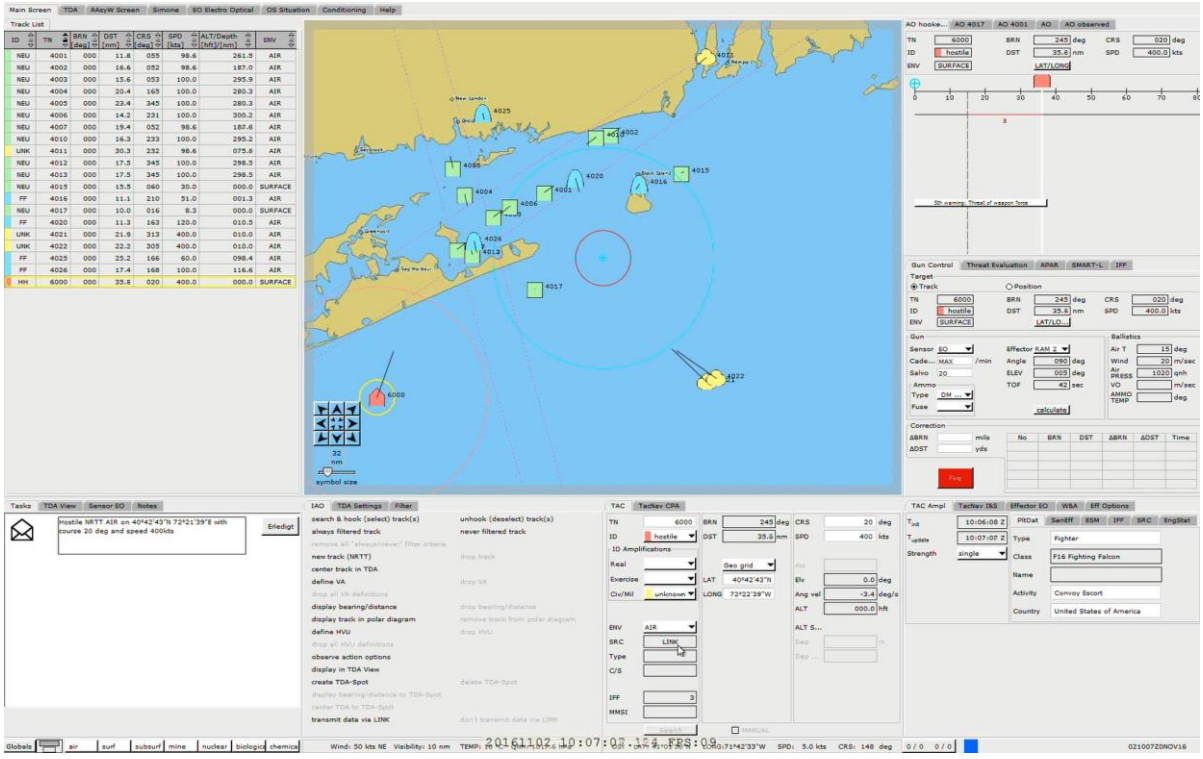


Figure 34. Interface de l'outil de simulation de monitoring d'espace aérien.

[38] NKE : Necessary Knowledge Element. Element d'information dont la connaissance est nécessaire au sujet pour réaliser sa tâche.

Sur la Figure 34, la zone de carte centrale représente la zone d'affichage tactique (TDA[39]). Les aéronefs que doit contrôler la frégate y sont représentés par des pictogrammes jaunes, verts, bleus ou rouges, selon leur identité. Le centre de cette zone (point bleu) simule la position de la frégate. Le cercle rouge bordeaux qui l'entoure symbolise la zone d'engagement (WR[40]). Le cercle bleu de plus grand rayon symbolise quant à lui la limite de la zone d'identification sécurisée (ISR[41]).

La situation simplifiée de surveillance d'espace aérien permise sur cet outil comporte 4 tâches.

Tout d'abord la tâche principale de l'opérateur est l'Identification (ID). Celle-ci consiste à attribuer à tout nouveau contact (un aéronef quel qu'il soit) une identification « ami », « neutre » ou « ennemi ». Cette attribution se fait par modification de paramètres présents sous la zone d'écran radar centrale, sur la base de critères d'identification.

Il se peut également que des pistes doivent être créées manuellement (NRTT[42]) sur la zone d'affichage tactique. L'instruction est transmise visuellement par un message dans la zone basse gauche de l'outil.
La troisième tâche consiste à avertir tout aéronef hostile entrant dans la zone d'ISR de la frégate, en cliquant sur un bouton *Warning* de la zone *Timeline* (haut droit de l'écran sur la figure X).
Enfin, les contacts ennemis se dirigeant vers la frégate malgré l'avertissement doivent être abattus dès qu'ils entrent dans le rayon d'action des armes (WR). Il s'agit de la tâche d'Engagement.

Ces tâches et leur importance sont résumées dans le Tableau 8.

Tache	Description	Priorité
Identifier	Tout contact non identifié doit être identifié comme amical, neutre ou hostile sur la base de critères prédéfinis. Les contacts identifiés peuvent modifier leur comportement d'une manière qui nécessite la réassignation de leur identité	100 (si Hors ISR) 300 (si dans ISR)
NRTT	Lorsqu'un message apparaît dans le panneau de message, un contact (NRTT) doit être ajouté manuellement au TDA. Les informations nécessaires à la création d'un NRTT sont affichées dans le message	200
Avertir	Les contacts identifiés comme hostiles doivent être avertis dès qu'ils entrent dans la zone de sécurité d'identification (ISR ; indiquée sur le TDA par un cercle bleu autour de la frégate)	400
Engager	Les contacts identifiés comme hostiles ayant été avertis doivent être engagés dès qu'ils sont à portée de tir de la frégate (Weapon Range WR; indiqué sur le TDA par un cercle rouge autour du navire)	500

Tableau 8. Description des tâches de l'expérience AMIGOS et leur priorité

Le scénario d'une heure fait alterner période de forte activité (tâches à réaliser en parallèles) et de plus faible activité (taches plus séparée dans le temps). Pour les raisons invoquées au 4.3.1, nous concentrons notre présent travail sur la tâche d'Identification, demandant de détecter l'apparition d'un aéronef non identifié (alors jaune) et de lui attribuer une identification (Ami-Bleu, Neutre-Vert ou Hostile-Rouge).

[39] TDA : Tactical Display Area
[40] WR : Weapon Range
[41] ISR : Identification Safety Range
[42] NRTT : Non Real-Time Tracks

Ce choix est fait sur la base de critères préétablis, notamment le code d'identification, la position et le cap de l'aéronef.

4.2.2 Les Variables

Trois variables ont été mesurées lors de cette expérience, deux temps de réaction issus du monitoring comportemental ainsi que le diamètre pupillaire, donnée de monitoring physiologique.

Le temps de réaction est habituellement défini comme le temps mis par l'individu à initier une réponse moteur à la présentation d'un stimulus (Bertelson, 1967). En psychophysiologie, les conditions expérimentales considèrent généralement la présentation du stimulus et sa détection par l'individu comme simultanées.

Dans notre cas, ou comme cela peut l'être lors d'une tâche de recherche visuelle, nous faisons la distinction entre le temps mis à détecter visuellement le stimulus et le temps mis à engager l'action qui en découle (ici sa sélection par un clic de souris). Ainsi nous distinguons et mesurons deux latences, illustrée par la Figure 35 :

- le temps séparant l'apparition du stimulus (l'apparition d'un avion non identifié à l'écran) et la première fixation fovéale sur celui-ci - que nous noterons *Tp ;*
- le temps entre l'apparition du stimulus et la sélection de l'aéronef associé par l'individu (clic sur l'aéronef) - que nous noterons *Tr*.

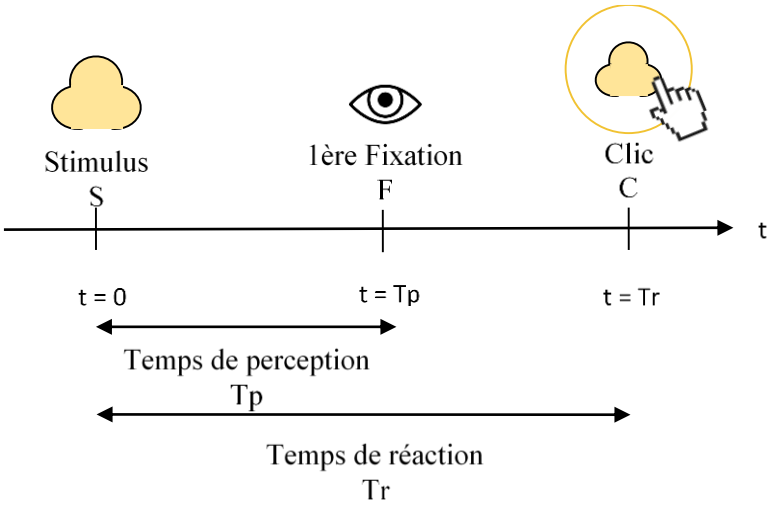


Figure 35. Représentation des deux latences mesurées et des évènements associés.

Outre les données de position du regard qui, comme détaillé au 4.3.2 servent pour déterminer *Tp*, nous extrayons également de l'*eye tracker* des données pupillométriques correspondant à l'amplitude du diamètre pupillaire sous forme de séries temporelles. Celles-là nous permettent de construire la variable dépendante de dilatation pupillaire. Son calcul est détaillé en 4.3.4.

Nous abordons par la suite la sélection des stimuli étudiés ainsi que les méthodes mises en œuvre pour le calcul de **Tp** et **Tr**.

4.2.3 **Participants**

L'expérience AMIGOS, comporte 2 jeux de données issues des 2 séries de passations de 2016 et 2018. Les sujets 1 à 12 font partie du 1er jeu de données, résultant d'une session expérimentale de Décembre 2016. Les sujets 13 à 24 font partie du second jeu de données, récolté lors de la session expérimentale de Décembre 2018. 16 hommes et 8 femmes, âgés de 19 à 39 ans (M= 29.9), ont été recrutés au sein du laboratoire, tous novices à la tâche. 5 sujets ont été exclus de l'étude en raison de mauvaise qualité de l'enregistrement ou de dysfonctionnement du matériel, ramenant ainsi à 19 le nombre total de sujets ayant pu être étudiés. Tous les participants sont en bonne santé et ont une vue normale ou corrigée.

4.2.4 **Matériel**

L'étude a été menée dans une salle calme et sans fenêtre, équipée d'un système d'*eye tracking* déporté. Ce dernier était un SMI® iView X (30 Hz) pour la première session expérimentale et un Tobii® pro X2 60 (40 Hz) pour la seconde. Les données issues du Tobii® ont par la suite été sous-échantillonnées à 30 Hz lors de l'interpolation afin de conserver la cohérence des données (voir 4.3.2 Interpolation). Lors de cette même session expérimentale, d'autres données (électrocardiogramme, posture, clics et expressions faciales) ont été recueillies. L'ensemble des données issues de l'*eye tracker*, de la ceinture cardiaque, des actions de l'utilisateur, du suivi du visage et de l'évolution du scénario, sont enregistrées au sein d'un unique fichier CSV via le logiciel iMotions®. La vidéo de l'écran est également enregistrée. Nous concentrons notre analyse sur l'enregistrement du diamètre pupillaire par l'*eye tracker* tout en exploitant les données de position du regard recueillies par le même dispositif, ainsi que les données comportementales de clic de souris et celles relatives au déroulement du scénario de simulation (par exemple l'apparition d'un avion).

Durant les 45 minutes du scénario, l'expérimentateur n'intervient jamais, astreint le plus possible à la neutralité.



Figure 36. Installation expérimentale.



Figure 37. Installation expérimentale et poste de l'expérimentateur.

4.2.5 Déroulement des passations et instructions aux participants

Les passations ont une durée totale d'environ 1h15. En premier lieu il est demandé au participant de signer un formulaire de consentement éclairé et de remplir un questionnaire démographique. S'en suit une phase d'instruction et de calibration du matériel de mesure, puis le participant se familiarise avec la tâche durant une période d'entraînement de 10 minutes. Le scénario d'expérimentation qui suit dure 45 minutes. Durant cette période, 3 phases expérimentales s'enchainent dans un ordre fixe (Figure 38).



Figure 38. Phases de l'expérience

Ces phases se distinguent par une variation de la fréquence d'apparition des tâches. Le sujet est plus ou moins sollicité selon la phase. La première, dite de « baseline », dure 5 minutes et correspondant à une première période de faible activité. Les tâche sont soumises au sujet à un rythme d'une toutes les 30 secondes, lui laissant suffisamment de temps pour les traiter au fur et à mesure. Vient ensuite une phase de « forte charge induite » d'une durée de 10 minutes. Durant cette période les tâches s'enchainent à une fréquence plus soutenue (en moyenne une tâche toutes les 8s). Il arrive que le sujet ait à traiter plusieurs tâches en simultané par exemple un nouvel aéronef à identifier (ID) et un *Warning* à effectuer. Le sujet est donc très fortement sollicité. La dernière phase, intitulée « Monotonie », dure 30 minutes durant lesquelles les tâches apparaissent à nouveau chacune leur tour, espacées de 60 à 90 secondes.

4.3 Exploitation des données

4.3.1 Sélection des tâches

Afin de minimiser le possible impact du recouvrement du traitement d'une tâche et de l'apparition d'une nouvelle, les tâches de la période « chargée » ont été écartées. Les tâches sélectionnées sont donc séparées à minima de 30s de toute autre tâche.

Sur l'intégralité du scénario, un total de 100 tâches sont à réaliser par les sujets, de 4 natures différentes (voir 4.2.1). La tâche d'identification hors zone d'engagement (ID OOR) est la plus commune et rapide à réaliser, mais également la moins pressante ou dangereuse. Elle consiste à attribuer une identification à un aéronef non identifié (symbolisé par un trèfle jaune sur la Figure 39) venant d'apparaître dans un couloir aérien, ne présentant donc pas de danger immédiat pour la frégate. Tâche ayant la plus faible priorité, elle est considérée comme tâche de fond. Pour l'ensemble de ces raisons ce type de tâche a été retenu pour l'étude.

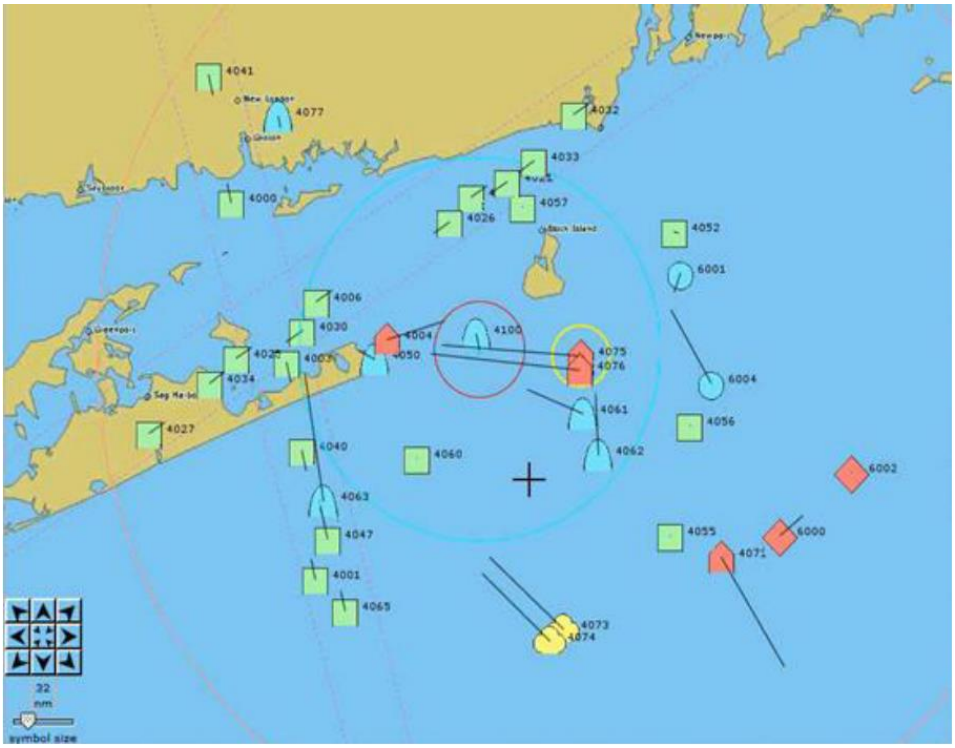


Figure 39. Zone radar de l'outil de simulation de tâche de surveillance d'espace aérien

En résumé, sur les 100 périodes proposées, 18 tâches d'identification ont été sélectionnées de manière à limiter les éventuelles interférences qui pourraient résulter de la pression temporelle ou du traitement des autres tâches.

4.3.2 Méthode de détection de la focalisation visuelle

Le temps à la première fixation est un indicateur couramment utilisé en marketing afin d'évaluer la saillance d'un élément d'interface, d'un slogan, d'une image ou d'un produit (Duchowski, 2002; Goldberg & Kotval, 1999; Jacob & Karn, 2004). Les données oculométriques de position du regard permettent de déterminer quand l'attention du sujet se porte sur une zone spécifique de l'écran. En

définissant une zone d'intérêt, ou AOI pour *Area Of Interest*, autour d'un aéronef, il est donc possible de mesurer le temps de latence entre l'apparition de l'aéronef et la première focalisation fovéale sur celle-ci. Nous définissons cette AOI dynamiquement de manière à suivre le déplacement de l'objet ciblé à l'écran, l'aéronef se déplaçant continuellement le long d'une trajectoire. La taille de cette zone prend en compte l'étendue du champ de vision fovéale[43] avec lequel l'opérateur regarde directement l'objet, ainsi que la distance séparant les yeux de l'opérateur et le sujet regardé, comme présenté dans la Figure 40 suivante.

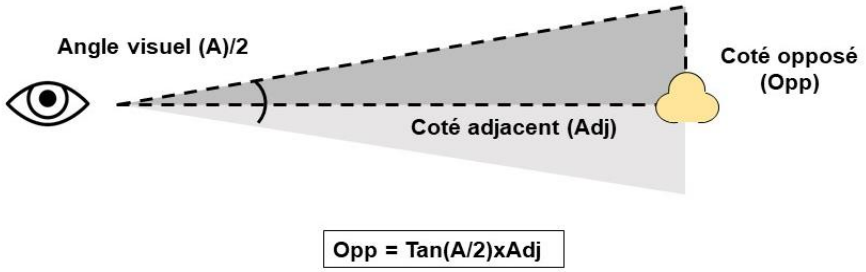


Figure 40. Représentation du calcul de rayon de l'AOI

La précision du matériel de mesure est également à prendre en compte. Elle est définie comme étant la déviation en degrés entre la direction réelle du regard du sujet et la position mesurée. Les deux *eye tracker*s utilisés dans le cadre de cette expérience, SMI® iView X et Tobii® pro X2 60, sont réputés posséder une précision de l'ordre de 0,5 à 1° d'angle visuel. La vision fovéale est pour sa part communément définie comme un angle de 1 à 2° (Scheidt & Schultz, 2011; Shomstein & Behrmann, 2006). Les limites supérieures de ces deux angles sont retenues pour estimer ainsi un angle de vision centrale total de 3°. La distance entre l'œil et la cible est également mesurée en millimètres par l'*eye tracker*. Le rayon de l'AOI autour du centre de la cible est donc calculé de la manière suivante :

$$R = D * \tan\left(\frac{V_a}{2}\right) * Cp + Cr$$

avec :

V_a : angle visuel (V_a = 3) ; D : distance œil – écran ; Cp : constante de conversion de centimètres en pixels de l'écran (HP ZR2440w 1920*1200 px, 52*32cm, Cp = 37,2 pixels/cm) et Cr : constante de rayon de la cible (ici Cr = 14px).

[43] Vision centrale

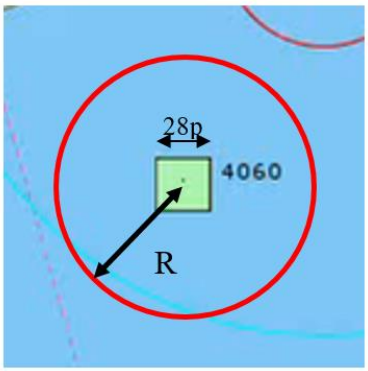


Figure 41. Projection de l'aire de focalisation visuelle sur la cible. Toute fixation ayant lieu dans ce cercle est considérée comme une fixation sur la cible.

Pour chaque échantillon mesuré à partir de l'apparition de la cible, la distance entre le point de regard et le centre de la cible (coordonnées enregistrés par la simulation) est calculé et comparée au rayon de l'AOI, déterminant ainsi si le regard est dans la zone d'intérêt ou pas. L'opération est renouvelée jusqu'à ce que le point de regard soit compris dans l'AOI, définissant l'instant de première fixation sur la zone. Par souci de concision, nous ferons référence à cette mesure comme « temps de perception de la tâche » (noté Tp). Nous gardons cependant à l'esprit que la perception visuelle au sens de la psychophysiologie est un évènement pouvant prendre naissance dans la vision périphérique et déclenchant la focalisation visuelle que nous mesurons. (Eckstein, Guerra-Carrillo, Miller Singley, & Bunge, 2017; Hafed & Clark, 2002; Posner et al., 1980; Shomstein & Behrmann, 2006).

Tp est donc défini comme le temps écoulé entre l'apparition du stimulus et la première rencontre du regard avec la zone d'intérêt définie autour de celui-ci.

4.3.3 Extraction des données pupillaires

La combinaison de 19 sujets et 18 tâches permet de définir 342 cas d'étude, c'est-à-dire 342 couples (sujet, tâche). Dans un premier temps, pour chaque cas, les *timestamp*[44] d'apparition de la tâche et d'action sur celle-ci (sa sélection au clic) ont été extraits du log de l'expérience pour chaque sujet. Cette phase a été réalisée sous Microsoft Excel®.

Les données oculométriques de chaque sujet sont regroupées dans un fichier spécifique. Les données pupillométriques sur la durée spécifique à chaque cas en sont ensuite extraites : de 5 secondes avant l'apparition de la tâche à 10s après l'action.

Pour chaque échantillon, le diamètre pupillaire de l'œil droit et de l'œil gauche sont moyennés afin d'obtenir un unique indicateur de diamètre pupillaire.

4.3.4 Processus de préparation des données pupillaires : sélection des cas exploitables et nettoyage

Le processus de prétraitement des données issues de l'*eye tracking* peut être divisé en 3 phases principales (Figure 42). La première, la phase d'extraction des données pour les tâches qui nous

[44] Valeur représentant l'instant auquel une opération a été effectuée. Ici le *timestamp* correspond au nombre de nanosecondes écoulées depuis le début du scénario d'expérimentation.

intéressent depuis le jeu de données initial, est la démarche manuelle spécifique au formatage des données de l'expérience, abordée au paragraphe précédent qu'il ne nous a pas semblé pertinent de détailler ici. Ce qui suit présente donc les étapes de Sélection des cas utilisés et de Nettoyage des données. Cette dernière étape doit permettre d'améliorer la qualité des données en identifiant et corrigeant ou supprimant les artéfacts et autres comportements aberrants créés à l'enregistrement. Elle se compose de trois sous-étapes d'Interpolation, de Filtrage et de Vérification.

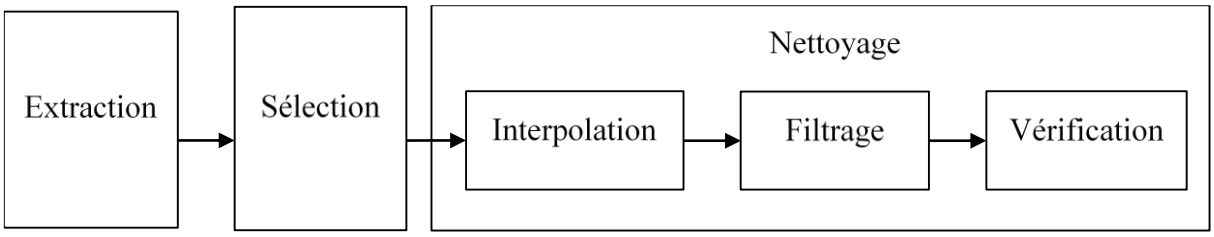


Figure 42. Processus de préparation des données d'eye tracking

4.3.4.1 *Sélection des cas exploitables*

Quelle que soit la qualité du matériel d'acquisition utilisé, il est très délicat de collecter des données de suivi oculaire, en particulier en utilisant un système d'*eye tracking* déporté. La qualité des données recueillies est notamment liée à la perte de données. Bien que cela puisse être intuitivement attribué à un participant ne regardant pas l'écran, une donnée manquante peut également être due à l'instabilité de l'appareil de mesure (Hessels, Niehorster, Kemner, & Hooge, 2017; Wass, Forssman, & Leppänen, 2014). Les raisons d'une absence de détection peuvent donc être multiples. L'orientation de la tête et la position des yeux par rapport au dispositif d'acquisition sont les principales causes de données manquantes. Évidemment, lors de clignements le même phénomène se produit et les clignements sont également traités par l'*eye tracker* comme une absence de donnée. Il est théoriquement possible de différencier les clignements d'une simple donnée manquante.

Cette étape de sélection vise à identifier et exclure les cas dont la quantité de données manquantes nuit à l'exploitation du cas. Dans le cadre de l'analyse effectuée ici, par mesure de précaution, nous ne faisons pas de distinction quant à l'origine des données manquantes, que ce soit des clignements répétés ou une mauvaise orientation de la tête.

Précédemment, la taille moyenne de la pupille pour chaque échantillon a été calculée en moyennant les diamètres mesurés de la pupille droite et gauche. Si l'un des deux yeux n'est pas capté par l'*eye tracker*, la moyenne pour cet échantillon est considérée comme la valeur du seul œil détecté. En effet, la corrélation entre le diamètre de la pupille droite et celui de la gauche est connue comme étant importante (> 0.9) (Jackson & Sirois, 2009). Si une donnée n'est manquante que pour un œil, il est possible d'estimer l'échantillon par la valeur de l'autre œil. En revanche si aucun des deux yeux n'est détecté, l'échantillon est laissé vide.

Sur cette base le nombre d'échantillons vides est alors compté. En le ramenant au nombre total d'échantillons du cas étudié, nous obtenons le pourcentage de données manquantes pour ce cas.

Cette procédure répétée sur l'ensemble des cas permet d'obtenir un tableau dont un extrait est donné ci-après :

	Task1	Task2	TAsk3	Task5	Task7	Task8	Task10	Task11	Task12	Task13	Task14	Task15	Task17	Task18	Task19	Task20	Task21	Task23
1	0,119	0,203	0,057	0,055	0,053	0,051	0,038	0,066	0,115	0,082	0,057	0,157	0,060	0,075	0,038	0,031	0,024	0,062
3	0,062	0,026	0,066	0,044	0,033	0,258	0,062	0,020	0,082	0,082	0,040	0,413	0,015	0,073	0,155	0,104	0,150	0,132
5	0,009	0,082	0,002	0,013	0,018	0,022	0,020	0,018	0,007	0,091	0,060	0,057	0,013	0,046	0,055	0,013	0,188	0,053
6	0,119	0,062	0,040	0,002	0,015	0,001	0,002	0,055	0,035	0,026	0,002	0,042	0,011	0,026	0,013	0,018	0,203	0,020
8	0,086	0,219	0,155	0,060	0,055	0,029	0,024	0,143	0,020	0,042	0,130	0,082	0,011	0,013	0,046	0,053	0,108	0,060
9	0,095	0,159	0,044	0,071	0,024	0,029	0,042	0,064	0,046	0,104	0,029	0,115	0,996	0,077	0,340	0,148	0,185	0,130
12	0,084	0,225	0,066	0,051	0,110	0,110	0,055	0,071	0,042	0,086	0,044	0,040	0,055	0,042	0,071	0,150	0,181	0,389
13	0,091	0,169	0,108	0,149	0,156	0,280	0,349	0,528	0,225	0,295	0,997	0,997	0,116	0,093	0,694	0,111	0,555	0,470
14	0,192	0,091	0,220	0,321	0,197	0,121	0,114	0,161	0,265	0,273	0,517	0,048	0,212	0,240	0,118	0,243	0,237	0,111
15	0,530	0,651	0,227	0,997	0,977	0,278	0,454	0,563	0,295	0,422	0,434	0,525	0,997	0,644	0,887	0,131	0,970	0,578
16	0,071	0,119	0,227	0,151	0,111	0,056	0,126	0,136	0,022	0,036	0,078	0,997	0,124	0,997	0,121	0,104	0,126	0,997
17	0,482	0,341	0,758	0,982	0,210	0,513	0,581	0,993	0,667	0,397	0,927	0,265	0,561	0,974	0,846	0,719	0,487	0,793
18	0,096	0,086	0,126	0,111	0,099	0,162	0,129	0,139	0,184	0,113	0,116	0,184	0,161	0,070	0,220	0,108	0,190	0,652
19	0,167	0,215	0,129	0,076	0,305	0,144	0,200	0,245	0,987	0,396	0,200	0,265	0,207	0,260	0,647	0,126	0,099	0,189
20	0,280	0,497	0,435	0,997	0,581	0,886	0,260	0,258	0,997	0,288	0,308	0,474	0,997	0,220	0,248	0,811	0,217	0,363
21	0,091	0,096	0,033	0,843	0,215	0,334	0,073	0,199	0,296	0,137	0,043	0,803	0,071	0,465	0,326	0,513	0,452	0,280
22	0,230	0,265	0,343	0,354	0,358	0,571	0,242	0,334	0,459	0,311	0,190	0,182	0,283	0,212	0,243	0,512	0,334	0,411
23	0,043	0,276	0,058	0,141	0,164	0,171	0,146	0,175	0,190	0,214	0,185	0,321	0,225	0,278	0,219	0,182	0,197	0,315
24	0,050	0,036	0,111	0,303	0,075	0,993	0,040	0,046	0,068	0,334	0,036	0,055	0,051	0,063	0,071	0,051	0,078	0,141

Figure 43. Extrait du tableau résumant le pourcentage de données manquantes par cas étudié. Chaque ligne représente un sujet, chaque colonne une tâche.

Dans le cas de ces deux expériences, les deux séries (nommées XP1 et XP2) ont été réalisées avec des *eye trackers* possédant des fréquences d'échantillonnage différentes. Les données ont donc été prétraitées de manière indépendante.

Lemercier et al. (2015) proposent la méthode suivante pour effectuer la sélection des cas à conserver. Un graphique (Figure 44) est construit à partir des effectifs cumulés par tranche de 1% de données manquantes.

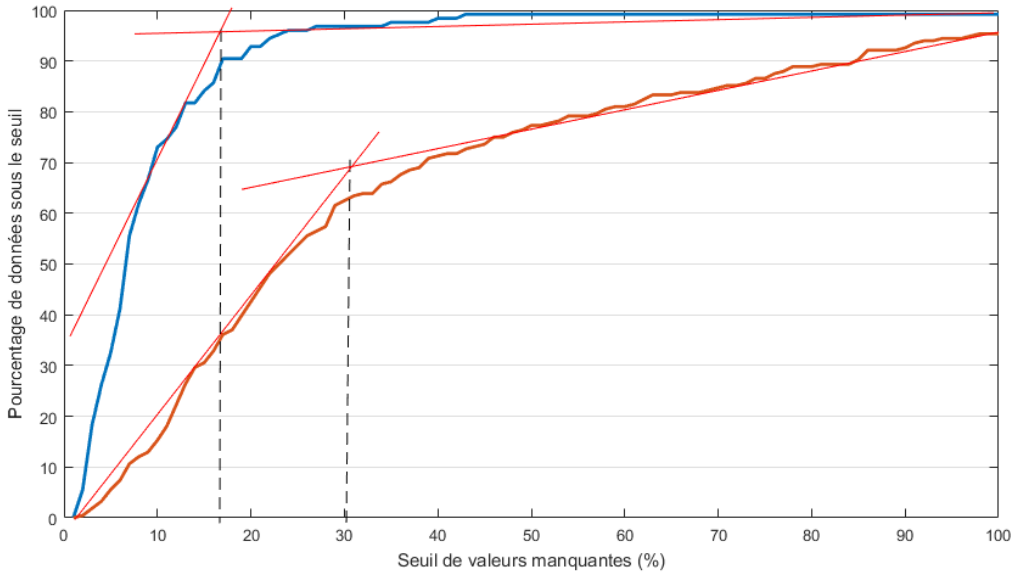


Figure 44. Effectifs cumulés et pourcentages de données manquantes pour XP1(bleue) et XP2 (rouge)

L'observation d'un point d'inflexion de la courbe permet de déterminer la limite entre un pourcentage habituel (pour l'expérience en question) et moins habituel. En nous basant sur ces observations nous pouvons fixer un seuil à 30% de données manquantes pour l'XP2 et 18% pour l'XP1. Au-delà de ce pourcentage de données manquantes, la quantité est considérée comme inhabituelle et le cas est exclu.

Cependant le seuil estimé de 30% pour l'XP2 nous a semblé légèrement élevé au regard de la littérature (Lemercier et al., 2015; Winn, Wendt, Koelewijn, & Kuchinsky, 2018). Nous avons donc choisi de mutualiser les deux seuils à 20%.

Ainsi est obtenue une liste de 171 cas sur lesquelles nous basons la suite des travaux.

4.3.4.2 *Interpolation des données manquantes*

Le traitement des données manquantes est principalement une question d'opinion. Certains chercheurs préfèrent exclure les cas avec trop de données manquantes (Koelewijn, Zekveld, Festen, & Kramer, 2012), d'autres préfèrent interpoler les données (Knapen et al., 2016). Ici nous choisissons d'adopter un traitement mixte de sorte à maximiser la quantité de données disponibles tout en minimisant la création d'artefacts. En effet l'interpolation présente le risque de diverger en cas de trop nombreuses données manquantes successives. Nous reviendrons au traitement lié à cette création d'artefacts au 4.3.4.4.

Lors de l'étape précédente nous avons exclus les cas où trop de données étaient absentes. Nous effectuons désormais une interpolation sur les cas restants. Cela permet d'obtenir des jeux de données complets, nécessaires pour procéder à l'étape ultérieure de filtrage.

L'interpolation consiste à remplacer une donnée manquante par une approximation de sa valeur en se basant sur les valeurs existantes.

Une méthode privilégiée dans le cadre de la préparation des données pupillo-métriques est l'interpolation par une spline cubique (Frank, Vul, & Johnson, 2009; Mathôt et al., 2013; Mathôt, Fabius, Van Heusden, & Van der Stigchel, 2018). Contrairement à une interpolation linéaire classique qui se contente de relier linéairement deux points distants, les splines approximent la courbe de manière polynomiale afin de prédire l'évolution de celle-ci entre les points (nœuds). L'interpolation par une spline cubique est préférée à d'autres formes d'interpolation polynomiale (e.x., courbes de Bézier) car le polynôme d'interpolation de degré 3 généré est plus précis.

Le principe de cette méthode est de calculer un polynôme unique pour chaque intervalle entre les points donnés (les nœuds). La condition étant que les polynômes consécutifs partageant un nœud aient les mêmes dérivées première et seconde afin de conserver une cohérence globale. Une fois que les paramètres des polynômes qui relient chacun des nœuds ont été déterminés, on peut alors calculer l'ordonnée d'un point situé entre deux nœuds : c'est la valeur du polynôme pour l'abscisse correspondante. La spline permet ainsi d'interpoler les valeurs comprises entre les nœuds.

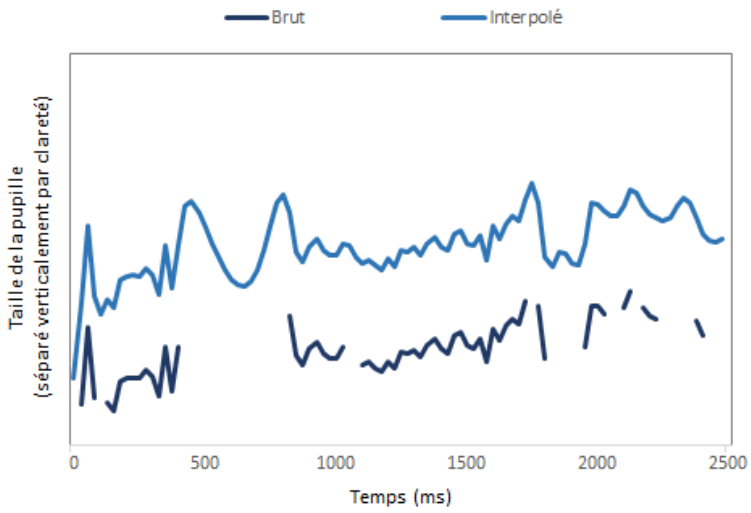


Figure 45. Représentations graphiques de l'évolution du diamètre pupillaire sur la phase étudiée, données brutes en bas et interpolées en haut

L'interpolation est également régulièrement utilisée comme méthode de rééchantillonnage. Ainsi nous avons profité de cette étape pour uniformiser les fréquences d'échantillonnage des deux sessions expérimentales en sous-échantillonner à 30Hz les données de l'XP2, jusqu'alors échantillonnées à 40 Hz.

Cependant, en cas de pente entre deux nœuds beaucoup plus grande que 1, l'évaluation des paramètres de la spline peut diverger, par exemple si les nœuds sont très éloignés (beaucoup de données manquantes). Une vérification de l'interpolation et des divergences est réalisée et les données aberrantes dues à une trop grande succession de données manquantes sont alors supprimées. Cette étape de vérification est détaillée au 4.3.4.4.

4.3.4.3 *Filtrage*

Afin d'éliminer les derniers artefacts et le bruit résultant de l'acquisition et de l'instabilité instrumentale, le signal est filtré. Ce bruit instrumental étant composé de hautes fréquences et la fréquence de réponse de la pupille à une stimulation cognitive étant basse (McLaren, Erie, & Brubaker, 1992) le signal est filtré à l'aide d'un filtre passe-bas[45].

D'abord envisagé, le filtrage par moyenne glissante s'avère faire perdre trop d'information en début et fin de série. Notre choix s'est donc porté sur un filtre Butterworth de 4ème ordre, suivant la méthode proposée par Klingner et al (2008).

La réponse pupillaire à une stimulation cognitive se caractérise par une fréquence de coupure autour de 2 Hz (Claudio M. Privitera & Stark, 2006) et oscille plus généralement entre 0 et 4Hz (Nakayama & Shimizu, 2004). La fréquence de coupure du filtre Butterworth est ainsi fixée à 4 Hz.

L'algorithme utilisé est celui développé par Sam Van Wassenberg, basé sur les travaux d'Erer (2007). Il possède l'avantage de corriger le décalage temporel habituellement créé lors de l'utilisation d'un tel filtre.

[45] Un filtre passe-bas est une transformation mathématique qui laisse passer les basses fréquences et qui atténue les hautes fréquences, c'est-à-dire les fréquences supérieures à la fréquence de coupure.

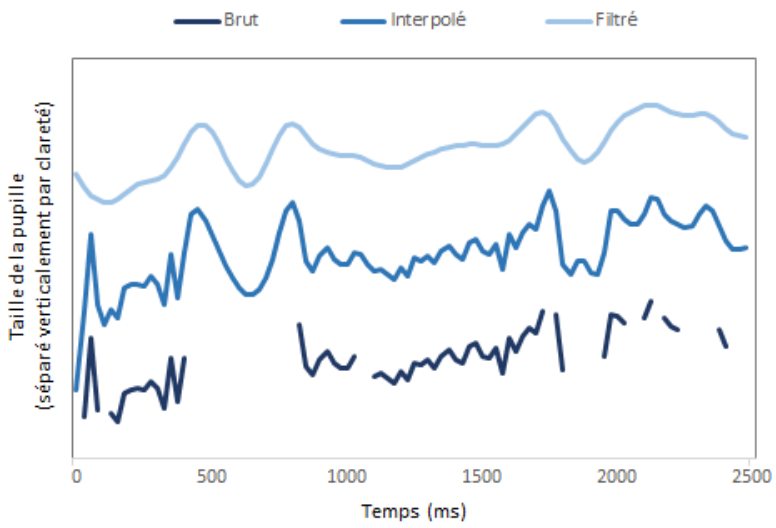


Figure 46. Représentations graphiques de l'évolution du diamètre pupillaire sur la phase étudiée, données brutes en bas, interpolées au milieu et filtrées (courbe du haut)

4.3.4.4 *Vérification*

Une vérification est nécessaire afin d'éliminer les possibles artefacts créés par la divergence de l'interpolation, en cas d'un trop grand nombre de données manquantes successives. Les cas de divergence sont identifiés et les données divergentes sont retirées. De manière générale les données d'interpolation sont retirées (l'échantillon est laissé vide) dès que plus de 5 valeurs successives sont manquantes dans le signal d'origine.

Nous obtenons ainsi une courbe lissée, où les courtes périodes de données manquantes sont complétées et dont l'allure et l'évolution sont visuellement plus facilement interprétables.

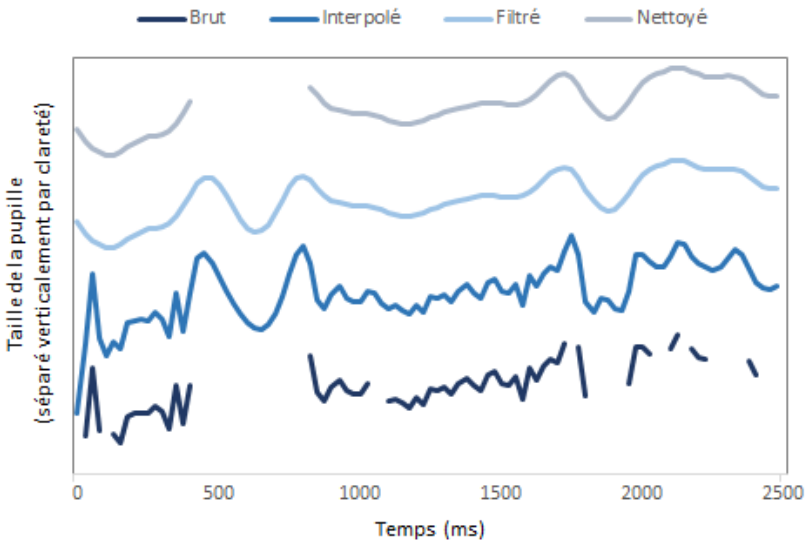


Figure 47. Représentations graphiques de l'évolution du diamètre pupillaire sur la phase étudiée, données brutes (courbe du bas), interpolées (deuxième courbe en partant du bas), filtrées (troisième courbe en partant du bas) et nettoyées (courbe du haut)

4.3.5 Dilatation Pupillaire

Le diamètre pupillaire[46] est rarement exploité tel quel. La valeur absolue mesurée après un stimulus dépend nécessairement d'un état initial dans lequel se trouvait la pupille à la survenue de l'événement.

Ainsi afin d'étudier la réponse pupillaire, une correction mathématique est nécessaire pour faire apparaître la réaction. Cette correction doit faciliter les comparaisons inter et intra-individuelles en standardisant les changements observés. Ceux-ci ne sont alors plus exprimés en diamètre absolu mais en variation relative à un état basal. Si un consensus méthodologique reste encore atteindre (Mathôt et al., 2018), la pratique informelle la plus courante consiste à utiliser une période pre-stimulus pour déterminer une ligne de base ou baseline.

La dilatation pupillaire est donc ainsi exprimée comme l'écart relatif entre le diamètre pupillaire à un instant t et la valeur de référence (baseline) pré-stimulus. Il est courant dans les études de pupillométrie cognitive de retrouver la dilatation pupillaire exprimée en pourcentage de proportion de la baseline (Hayashi, Someya, & Fukuba, 2010; Janisse, 1977; Mathôt, Grainger, & Strijkers, 2017; Sirois & Brisson, 2014; Weiss, Trehub, Schellenberg, & Habashi, 2016) d'après la formule suivante :

$$dilatation = \frac{diamètre\ -\ baseline}{baseline}$$

Ainsi une variation d'un diamètre de 6,0 mm à un diamètre de 6,5 mm représente une dilatation proportionnelle de 8,3%, quand la même variation de 3,0 à 3,5 mm reflète une différence de presque 17%.
Or la non-linéarité de la réponse pupillaire sur laquelle repose cette approche, fait encore débat (Reilly, Kelly, Kim, Jett, & Zuckerman, 2019). Par précaution et sur recommandation du Dr Peysakhovich, spécialiste de l'oculométrie au laboratoire de neuroergonomie et facteurs humains de l'ISAE-SUPAERO, nous avons choisi de nous limiter à une expression de la dilatation en différence à la baseline *(dilatation = diamètre – baseline*) exprimée en millimètres. Nous utilisons ici comme baseline la valeur moyenne du diamètre pupillaire sur une période de 500 ms (15 échantillons) précédant directement l'apparition du stimulus caractérisant le début de l'analyse.

4.3.6 Résumé du processus de prétraitement des données pupillaires

La Figure 48 résume les étapes et choix effectués pour le prétraitement des données. Au final nous obtenons 171 vecteurs de dilatation pupillaire répartis sur 11 sujets et 18 tâches.

[46] Le diamètre pupillaire est la mesure de la taille de l'ouverture de la pupille de l'œil. Celui-ci-varie varie en moyenne entre 2 mm (constriction maximale : myosis) et 8 mm (dilatation maximale : mydriase)

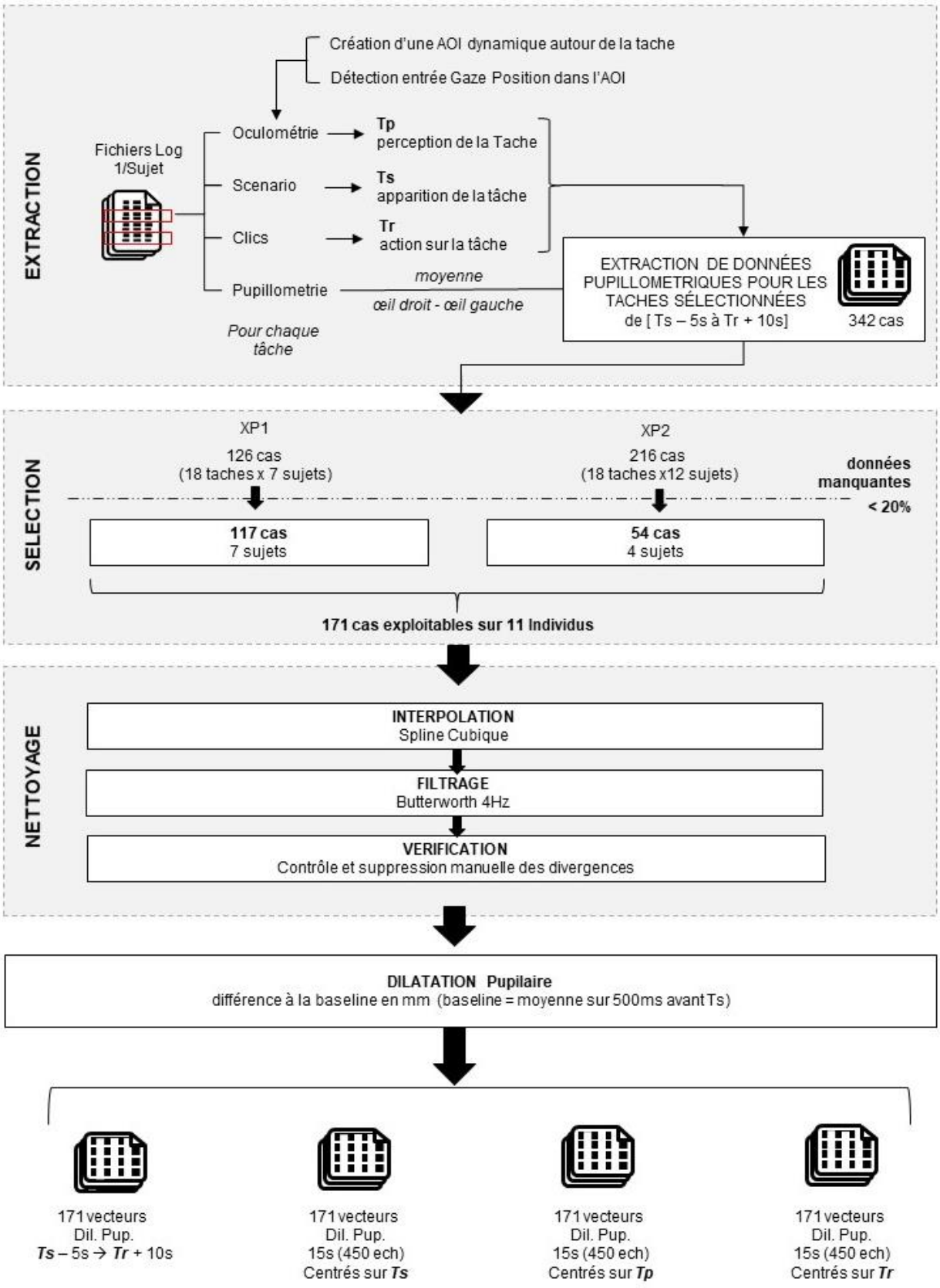


Figure 48. Schéma illustrant le processus de prétraitement des données pupillaires.

4.4 Résultats

Pour rappel, nous cherchons à répondre aux questions de recherche suivantes :

> (Q1) En situation de tâche écologique pouvons-nous distinguer la réaction pupillaire liée à la détection visuelle d'un nouvel élément de situation important dans le cadre de la tâche ?
>
> (Q2) Le traitement cognitif découlant de la détection visuelle, c'est-à-dire le changement de représentation traduisant la prise en compte du stimulus et ses implications, provoque-t-il une réaction pupillaire identifiable

Le traitement de ces questions se structure en 3 sections. Dans un premier temps, nous présentons une description des variables de latences Tp et Tr issues des mesures comportementales. Nous engageons dans une seconde section une analyse quantitative axée autour de la détection automatique d'un pic de dilatation pupillaire. Puis dans un troisième temps nous modérons ces résultats par une approche d'analyse qualitative des courbes de dilatation pupillaire par classification par tri de carte pour faire ressortir des similarités de réaction. Au sein de cette section la diversité des réactions observée, nous contraint de mettre en œuvre une analyse par moyennes, en deux temps, d'abord sur l'ensemble des cas conservés, puis par individu pour tester la sensibilité et la "significativité" des observations.

Pour résumer, nous cherchons tout d'abord à observer un pic de dilatation pupillaire entre le stimulus et le clic, puis à déterminer si la focalisation visuelle en est à l'origine. Chacune des deux approches présentées cherchera donc à répondre à ces deux étapes. Les analyses qui suivent ont été réalisées sous les logiciels Matlab 2016a® et RStudio®.

4.4.1 Présentation des données

4.4.1.1 *Pupillométrie*

Pour chaque vecteurs de dilatation pupillaire issus du processus de prétraitement (4.3.4) le principal pic de dilatation entre l'apparition du stimulus et le clic est identifié à l'aide de la fonction Matlab *findpeaks* (The MathWorks Inc., 2012). L'amplitude des pics identifiés et leur latence respective sont résumées dans le Tableau 9. La latence de pic correspond au temps mis par la pupille depuis l'apparition du stimulus pour atteindre le pic. Le tableau présente également la dilatation moyenne pré et post-stimulus.

	Moyenne	**Ecart type**	**Médian**
Dilatation pré-stimulus	-0,0052 mm	0,1186	0,0016 mm
Dilatation post-stimulus	0,0186 mm	0,1664	0,0130 mm
Amplitude du pic	0.1216 mm	0.2086	0.0770 mm
Latence de pic (Tmax)	2814 ms	5103	1766 ms

Tableau 9. Résumé des caractéristiques, moyenne, écart type et médiane, des pics de dilatation pupillaires.

4.4.1.2 *Latences*

Les données relatives aux latences Tp et Tr sont issues du monitoring comportemental. Pour rappel, Tp correspond au temps écoulé entre l'apparition du stimulus et la première fixation visuelle sur celui-ci. Tr correspond au temps de réaction, entre la fixation visuelle et le clic. Le Tableau 10 résume ces deux latences.

	Moyen (Ecart type)	**Médian**	**min / max**
Tp	2773 ms (6281)	634 ms	33 ms / 50 028 ms
Tr	4767 ms (6824)	2492 ms	1113 ms / 54 562 ms

Tableau 10. Moyennes, Écarts types et extremums pour les deux temps de latence étudiés.

Toutes les latences étudiées présentent un écart-type supérieur à leur moyenne. Les variations dans les réponses peuvent donc être considérées comme très importantes. Nous observons en effet un Tp maximal de plus de 50 secondes quand le minimum est de 33 ms (le sujet regardait déjà la zone d'apparition de la cible avant son arrivée). Cette grande dispersion se retrouve, dans des proportions légèrement moindres mais toujours importantes, pour le temps de réaction Tr. Le temps de traitement s'étend de 500 ms à plus de 22 secondes.

Cependant, malgré cette forte variabilité, la Figure 49 permet d'observer une densité très concentrée. Les deux variables présentent une distribution fortement asymétrique, possédant une très forte concentration de données en dessous de 5s, et une queue de distribution droite très allongée.

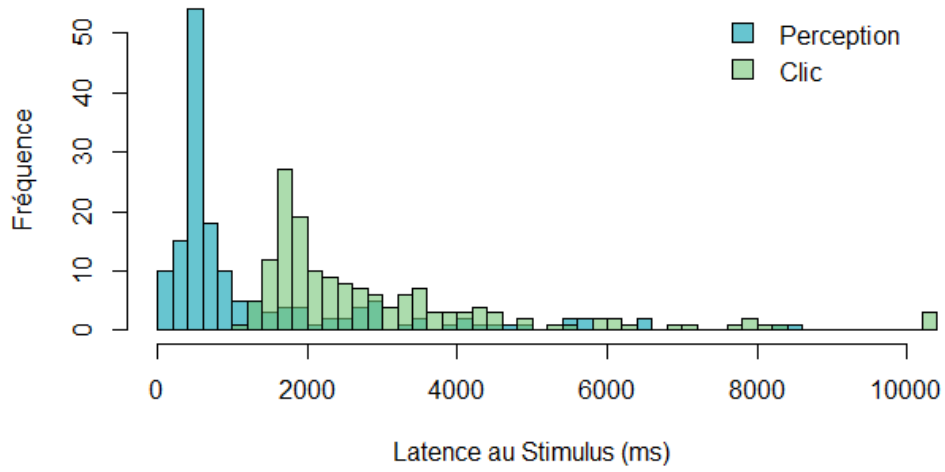


Figure 49. Distributions de l'ensemble des temps de perception (rouge) et temps de réaction (vert). Le graphique est limité à 10000 ms pour des raisons de lisibilité (max 52000).

Cette forte diversité des comportements est intrinsèque au réalisme de la situation, et présente le plus gros challenge à relever en terme d'analyse des données de monitoring.

4.4.2 **Approche quantitative**

Cette première approche vise à examiner de manière statistique les données de dilatation pupillaire au regard des deux latences mesurées Tp et Tr. Pour chaque cas le pic maximal de dilatation pupillaire entre le Stimulus et le Clic a été détecté (4.4.1). Cette méthode automatique identifiant nécessairement un pic, ne garantit cependant pas sa significativité ni même sa pertinence. A notre sens l'hypothèse H1-1, ne peut donc pas être traitée par cette méthode. Cependant, afin de déterminer si ce pic présente néanmoins une corrélation avec l'un des deux événements mesurés (Hypothèse H1-2), sa latence d'apparition est comparée aux latences Tp et Tr. La Figure 50 illustre leur relation.

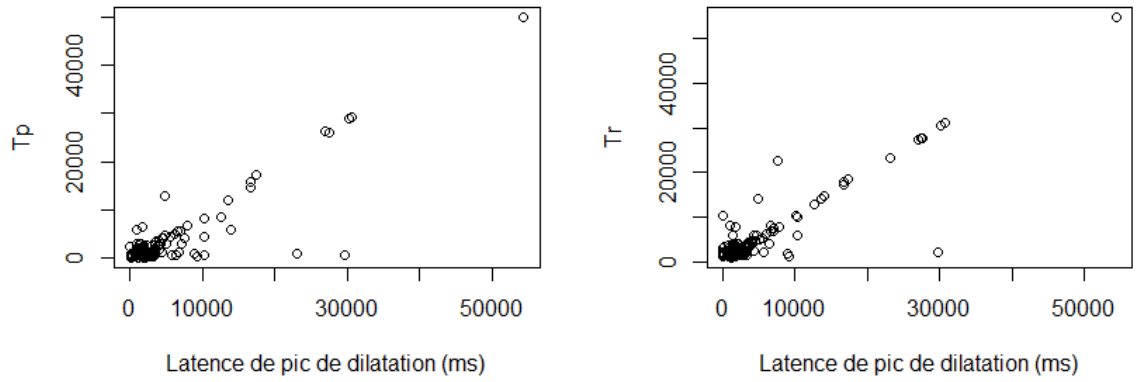


Figure 50. Représentation graphique du temps de perception Tp (à gauche) et du temps de réaction Tr (à droite), en fonction de la latence de pic de dilatation pupillaire

Il semble exister une relation linéaire positive relativement forte entre le moment ou le pic de dilatation est atteint et les deux latences étudiées. La relation semble plus marquée entre le pic et le clic. Le test de corrélation de Pearson nous confirme cette forte corrélation positive et cette différence (cor=0,87 p<2.2e-16, et cor=0.91 p<2.2e-16). Cependant, la relation semble être accentuée par les fortes latences et n'est pas aussi flagrante pour les faibles, comme l'illustre la Figure 51.

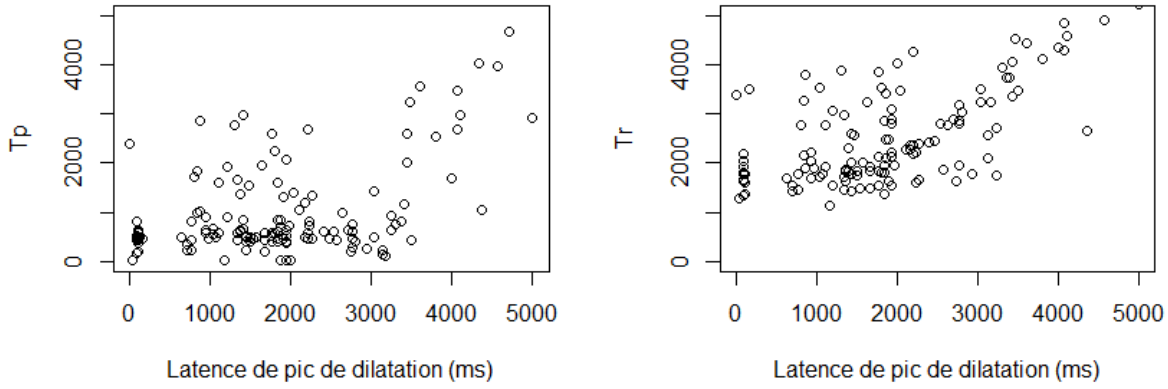


Figure 51. Représentation graphique du temps de perception Tp (à gauche) et du temps de réaction Tr (à droite), en fonction de la latence de pic de dilatation pupillaire, restreinte à 5000 ms.

Au regard de la densité de dispersion des temps de réaction (Figure 49) ces faibles valeurs représentent pourtant 90% de la population. Le test statistique nous confirme cette plus faible corrélation, bien que toujours significative, pour des latences inférieures à 5 secondes (cor=0.24 p=0.0033, cor = 0.34 p=3.33e-5). Le caractère fortement asymétrique de la dispersion semble donc avoir une influence non négligeable sur l'apparente corrélation

La dispersion des données et la proximité temporelle des deux événements déclencheurs, nous obligent à prendre ce résultat avec précaution. Ils ne nous permettent pas de conclure quant à l'origine de la réaction. Une approche qualitative est donc nécessaire.

4.4.3 Approche qualitative

4.4.3.1 *Préambule à l'analyse qualitative*

Si les études sur le temps de réaction et la détection visuelle sont courantes en psychologie cognitive, il est apparu, sans surprise, que mener une étude similaire dans des conditions de réalisme plus avancées, nécessite une méthodologie d'analyse adaptée, tournée dans un premier temps sur la description de l'événement recherché.

L'analyse quantitative des signaux pupillaires ne nous ayant pas permis de caractériser avec suffisamment de certitude cette réaction pupillaire, une analyse qualitative s'est avérée nécessaire.

La recherche sur les comportements humains est confrontée au problème de grande diversité du contenu du matériel de recherche recueilli, comme par exemple la diversité des réponses à une même question d'un entretien semi-directif. Il en va de même lorsque l'on aborde une analyse de grande ampleur de signaux physiologiques, ou une étude en situation d'applicabilité. En s'éloignant des situations extrêmement maîtrisées de laboratoire pour se rapprocher du réalisme de la situation écologique, les contraintes habituellement imposées par la structure de la situation de recherche, garanties de conformité et comparabilité des réponses étudiées, laissent place à une diversité des réactions, tant inter qu'intra-individuelle, certes plus réalistes mais également plus difficilement analysables d'un point de vue purement quantitatif. Afin d'analyser et interpréter cela tout en représentant cette diversité, l'analyse qualitative se pose en méthodologie complémentaire afin de qualifier, caractériser et donner du sens au contenu. C'est notamment ce qu'avance Dumez (2016), pour qui la nature même de l'analyse qualitative est d'être compréhensive, c'est-à-dire de « donner à voir » par la description et la narration. L'analyse qualitative repose sur une approche inductive, voire abductive, plutôt que déductive (Claverie, 2019). Nous cherchons à appréhender la singularité des situations en formulant des hypothèses plausibles dont la vérification pourra être faite ultérieurement (Catellin, 2004). Tout comme l'analyse clinique, c'est une démarche de qualification de la donnée par des critères subjectifs issus d'un travail d'observation, de regroupements intuitifs et de catégorisation. Les rapprochements et la confrontation des données ou encore les analogies et l'analyse des récurrences, doivent permettre l'émergence du sens. L'analyse qualitative revient à augmenter la démarche quantitative, numérique et statistique par la puissance de l'intuition, de l'observation et de la cognition du scientifique.

Dans notre cas, le travail de catégorisation est centré sur la comparaison des phénomènes (dilatation pupillaire), à la recherche de motifs communs au sein de la population. La description de ces motifs permet par la suite d'établir des critères de classification.

Notre analyse comporte trois temps. Un premier où nous comparons tous les comportements pupillaires à la recherche de similitudes. Un deuxième où nous cherchons à faire émerger une stabilité

comportementale via une méthode de moyennage. Enfin un troisième, où nous cherchons à vérifier à l'échelle de l'individu les observations effectuées sur l'ensemble des cas.

4.4.3.2 *Analyse Visuelle*

Notre principale question de recherche concerne l'identification d'une réaction pupillaire propre à l'actualisation de la conscience de situation, suite à la détection visuelle d'un nouvel élément situationnel. Dans cette analyse nous nous intéressons donc à l'évolution temporelle de la dilatation pupillaire suite à la première fixation sur l'élément, afin de déterminer si la réaction attendue peut à la fois être observée et caractérisée.

Nous réalisons une première observation du comportement de la dilatation pupillaire après la fixation visuelle sur le stimulus sur les 171 cas obtenus après prétraitement (voir 4.3.6.)

Afin d'illustrer cela, les Figure 52 et Figure 53 présentent respectivement les courbes de dilatation pupillaires pour un individu et pour une tâche.

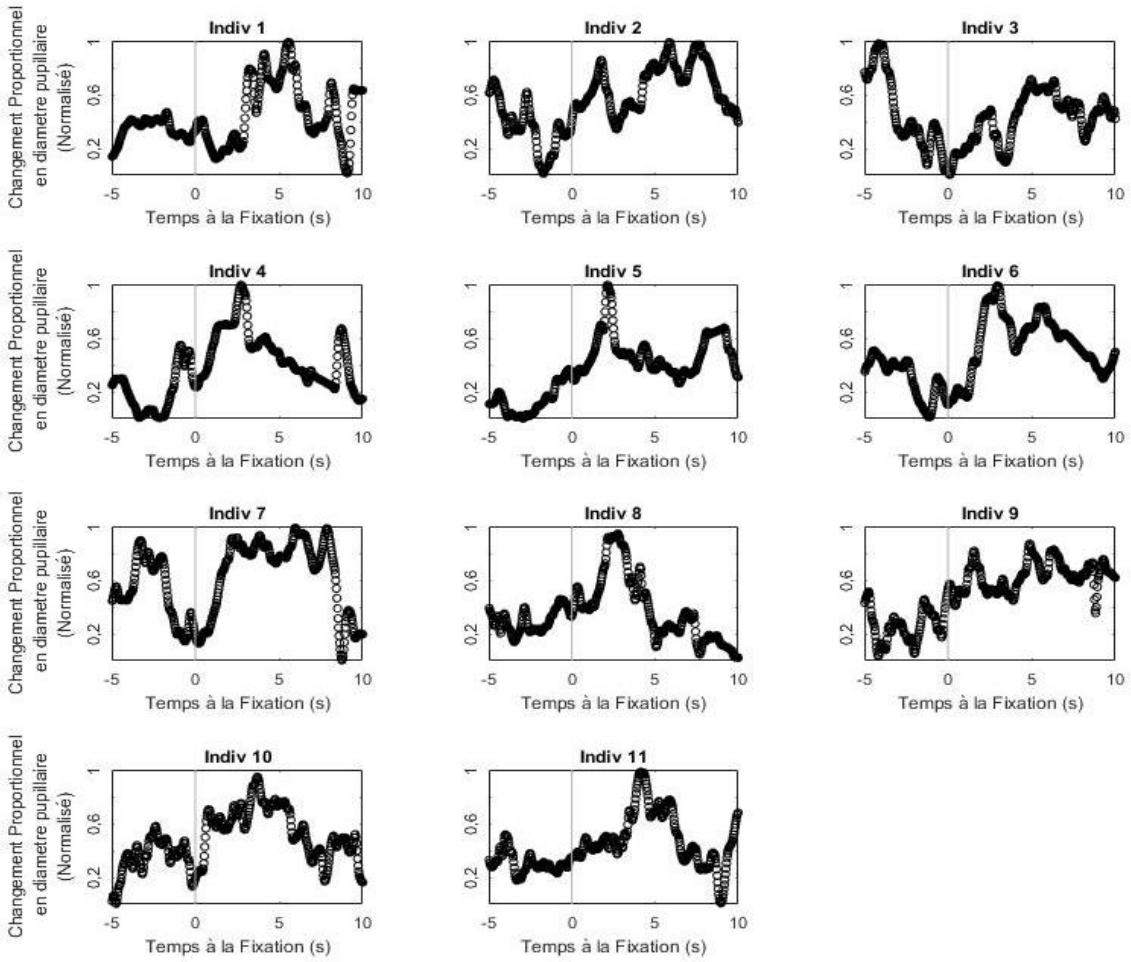


Figure 52. Evolution de la dilatation pupillaire pour la tâche N°1. 11 individus sur une fenêtre de 15s synchronisée sur la première fixation visuelle sur le stimulus.

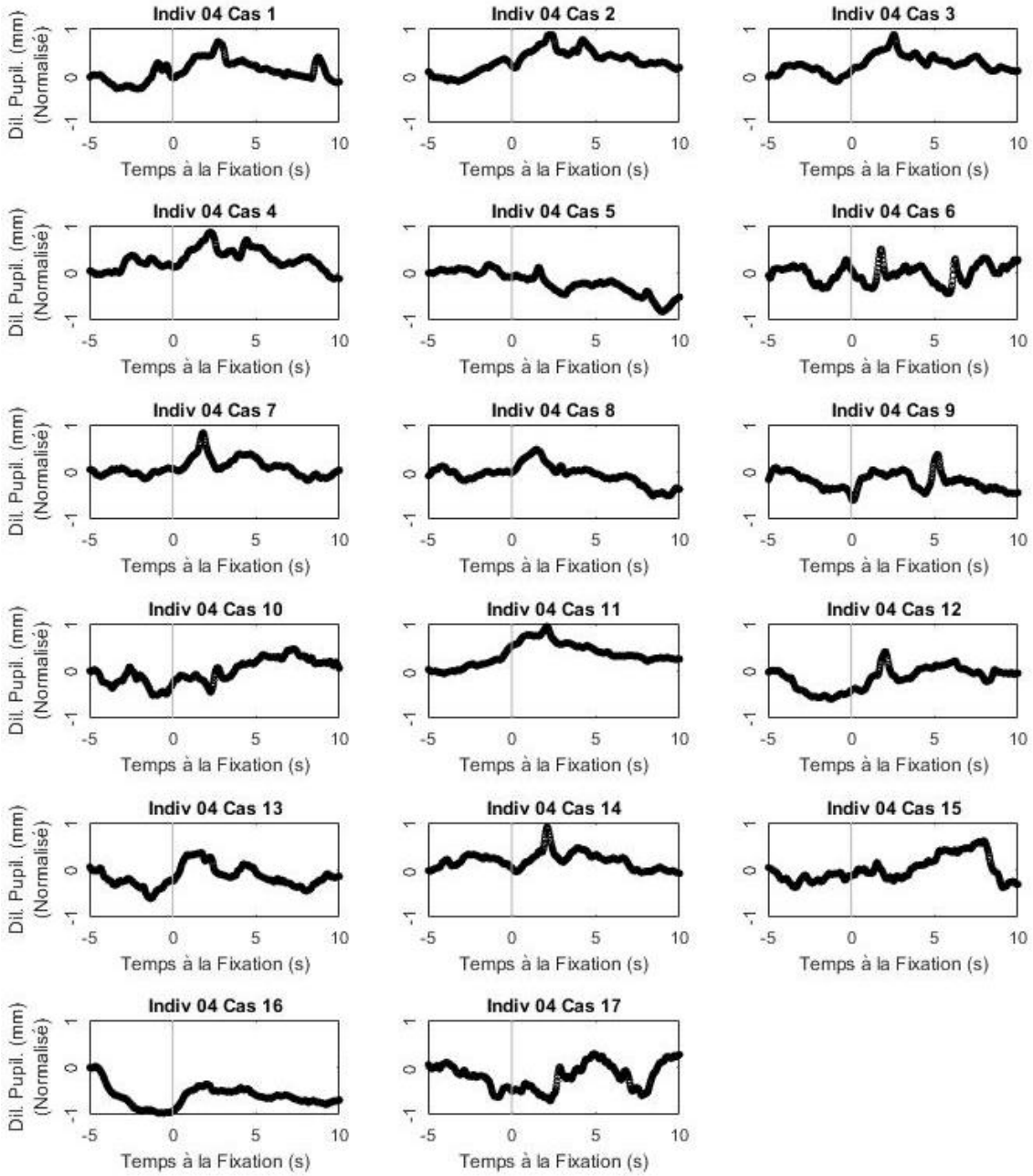


Figure 53. Evolution de la dilatation pupillaire de l'individu N°4 - 17 tâches sur une fenêtre de 15s synchronisée sur la première fixation visuelle sur le stimulus

Dans ans une faible majorité des cas (68,42%), la dilatation pupillaire moyenne est plus élevée dans les 5s après la fixation que dans les 5s avant (Mavant = -0,63, Std = 2,81 ; Mapres = 1,90, Std = 5,15). Cette tendance est accompagnée le plus souvent d'un pic de dilatation atteint durant cette période. Cependant il est difficile d'établir une nette similarité dans les réactions observées. La latence du pic de dilatation, témoin de réaction à un stimulus couramment utilisé, est très variable voire négative (pic précédant la fixation) et dans certains cas n'est tout simplement pas distinguable du reste du comportement pupillaire. La diversité dans la forme, l'origine et la durée de cette réaction empêche de la caractériser rigoureusement. La Figure 52 illustre la variabilité intra et la Figure 53 illustre la variabilité inter-individuelle.

Afin d'approfondir notre compréhension du comportement observé, un tri de carte ouvert a été organisé. Le tri de carte est une méthodologie récurrente en analyse qualitative (Fastrez, Campion, & Collard, 2009). Appliqué à de nombreux domaines (conception centrée utilisateur, architecture de l'information, capitalisation des connaissances…), il permet entre autre d'identifier les catégories pertinentes aux yeux de l'expert. Ainsi les 171 courbes ont été imprimées séparément puis regroupées par similarité visuelle. Cette catégorisation qualitative a finalement permis une première segmentation des cas en 2 catégories.

La catégorie 1, notée CAT1, regroupe 77 courbes à l'allure jugée similaire, comme illustrées par la figure 54. Cette forme de courbe présentant un pic de dilatation clairement identifiable nous est apparue comme la plus répandue et la plus distinctive parmi les réactions observées.

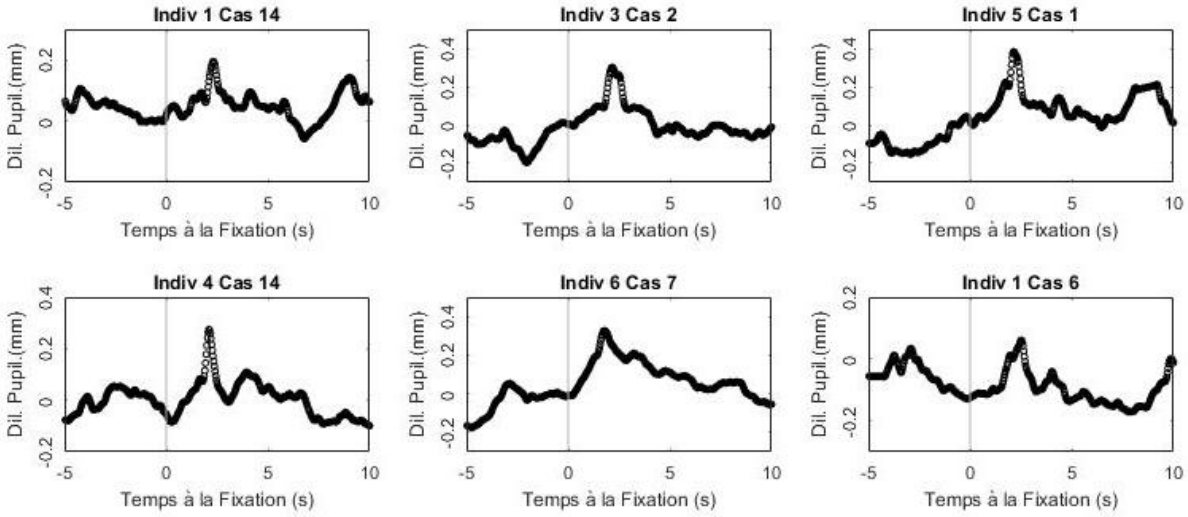


Figure 54. Illustration de cas compris dans la Catégorie 1

La seconde catégorie, CAT0, est composée des 94 courbes restantes dont le comportement, même si présentant parfois une réaction identifiable, est éloigné de la CAT1.

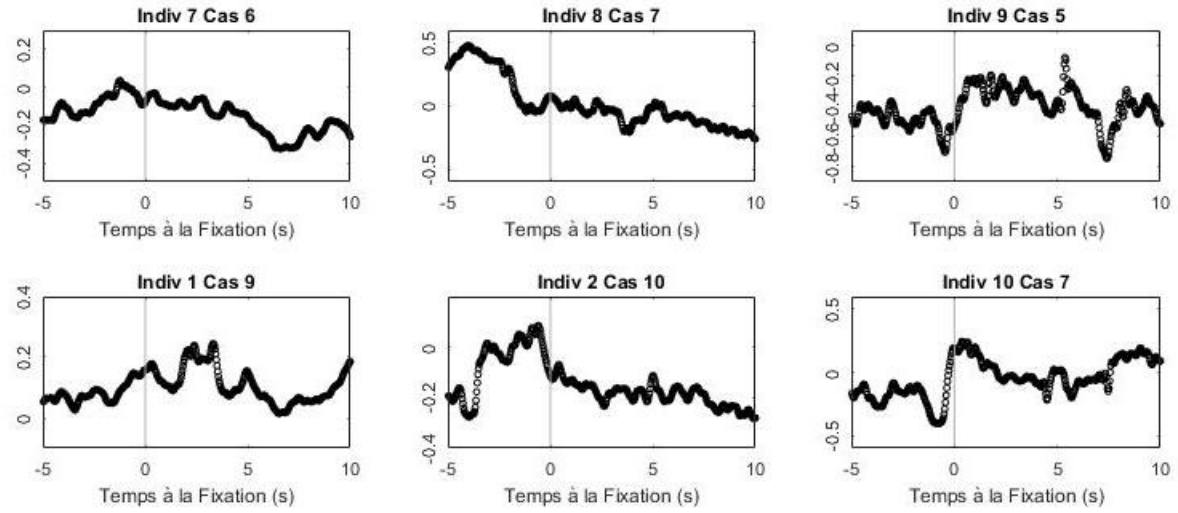


Figure 55. Illustration de cas compris dans la Catégorie 0

Les cas présentés dans la Figure 55 illustrent la diversité de forme du signal pupillaire étudié. Dans certains cas où l'on observe une dilatation après la fixation, donc une possible réaction (graphiques 4 et 6), celles-ci ne sont pas aussi marquées que les réactions de la CAT1. Il arrive également que certaines variations positives soient prises dans une tendance plus générale de baisse ou d'augmentation provenant de bien avant la fixation. Nous sommes conscients que le travail de catégorisation peut être poussé plus loin et que de nombreuses distinctions et regroupements sont encore réalisables au sein des deux catégories établies. Cela traduit d'ailleurs de l'extrême hétérogénéité encore présente au sein des catégories. Dans ces conditions, il est impossible d'établir avec suffisamment de certitude un lien entre la fixation et les variations, quand celles-ci sont effectivement observables.

Au final, ce ne sont que 77 cas qui ont pu être qualifiés par cette analyse visuelle des données pupillométriques. Cela ne représente que 22,6 % des cas initiaux. Nous mettrons par la suite ce résultat en perspective quant à l'applicabilité de la pupillométrie à l'exploitation en temps réel.

La réaction pupillaire attendue présente donc une variabilité très importante, tant inter qu'intra individuelle, frein majeur à sa caractérisation. La question de l'existence d'une réaction reste primordiale. Si celle-ci ne semble pas clairement accessible au cas par cas, l'observation nous laisse néanmoins l'impression qu'une modification, principalement positive à bien lieu mais ne nous permet pas d'accéder précisément à son origine ni de la qualifier.

L'analyse suivante cherche donc à mettre en évidence l'existence d'un comportement moyen, plus aisément descriptible.

4.4.3.3 *Analyse par Méthode de moyennage*

La diversité des réactions et la sensibilité des mesures au parasitage est un problème courant en étude psychophysiologique. L'approximation du signal initial par méthode de moyennage est alors un outil privilégié afin de faire apparaître une forme de stabilité et ainsi extraire le signal pertinent du bruit et autres artefacts. En électroencéphalographie (EEG) la moyenne d'essais successifs synchronisés sur une évènement extérieur (*event*) est utilisée afin de faire apparaître les potentiels évoqués recherchés (ERP[47]). En pupillométrie, Klingner et al. (2008) proposent une méthodologie de moyennage de signaux pupillaires alignés sur la fixation (*Fixation-Aligned Pupillary Response*) afin d'étudier les modifications de charge au moment de découverte visuelle d'une cible. Nous proposons d'appliquer une démarche similaire et de l'étendre à l'étude de la réponse motrice afin de (1) vérifier l'existence d'une réponse pupillaire comme le prévoit la littérature et nous le laissent penser les observations précédentes (Hypothèse H1-1), et (2) déterminer l'origine de la réponse (Hypothèse H1-2). Cette méthode demande au préalable l'alignement temporel des signaux sur le potentiel évènement déclencheur (*trigger*)**.** Afin déterminer l'origine de la réaction, trois événements potentiellement déclencheurs sont comparés : le Stimulus, la Fixation et la Réaction (clic de souris).

[47] Les ERPs, ou Event Related Potentials sont de très petites tensions générées dans les structures du cerveau en réponse à des événements ou des stimuli spécifiques (Claverie 1999, Blackwood et Muir, 1990)

Des fenêtres d'étude de 10 secondes (-3 +7) autour de chacun des trois événements d'intérêt sont définies. L'alignement nécessite une simple translation temporelle de manière à placer l'événement de référence à t=0. Cet alignement est répété pour l'ensemble des cas moyennés et pour les trois événements étudiés. Les événements potentiellement déclencheurs étant alignés, les réponses pupillaires associées le sont aussi. La moyenne des 171 signaux devrait faire apparaître la réponse pupillaire sous-jacente. Nous procédons à trois moyennages qui seront par la suite comparés. Par ce processus, nous nous attendons à ce que l'événement déclencheur principal génère une réponse pupillaire moyenne plus marquée que les autres.
Les graphiques ci-dessous (Figure 56) montrent l'évolution de la dilatation pupillaire moyenne sur une fenêtre de 10 secondes (300 échantillons), alignés respectivement de gauche à droite, sur le Stimulus, la Fixation et la Réaction.

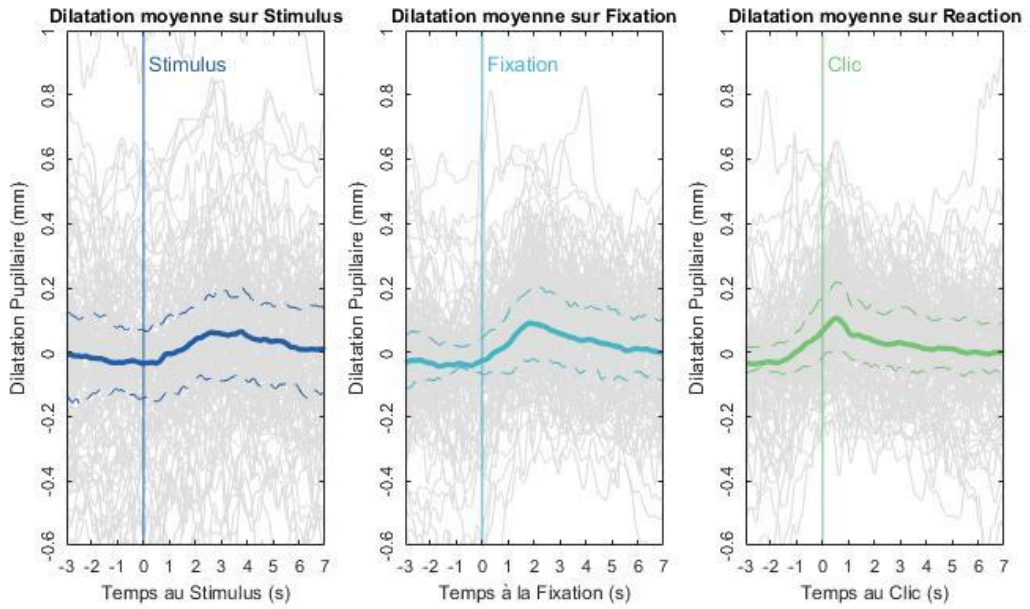


Figure 56. Evolution de la dilatation pupillaire moyenne pour les trois synchronisations étudiées.

Comparativement au constat effectué sur l'analyse individuelle, la moyenne fait ici apparaître une réaction observable. La Figure 56 suggère bel et bien l'existence du phénomène de réaction pupillaire, observable sur chacune des trois synchronisations. Cette observation est renforcée par l'évolution d'allure similaire des quantiles 25 et 75 symbolisés par les lignes fines supérieure et inférieure à la moyenne. Afin de mieux visualiser et comparer les réactions, la Figure 57 les représente alignées sur l'événement potentiellement déclencheur.

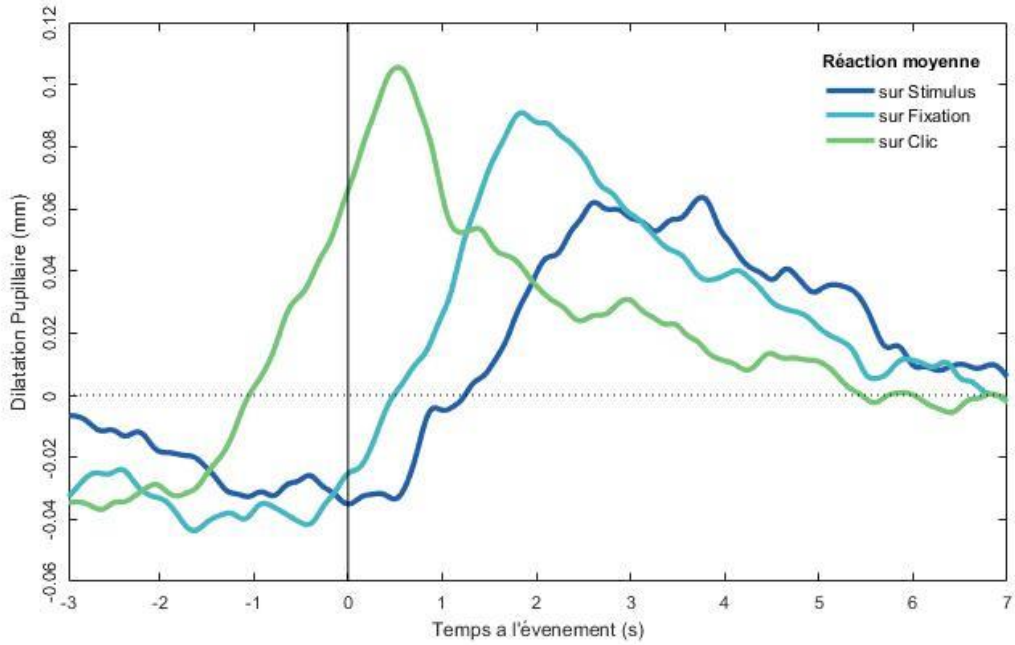


Figure 57. Comparaison des 3 moyennes alignés sur l'événement potentiellement déclencheur. En bleu foncé est représentée la réaction obtenue par moyennage sur le stimulus, en bleu clair celle obtenue par moyennage sur la fixation et en vert est représentée la réaction moyenne par synchronisation sur le clic.

Il est intéressant de noter que bien que nous observons également une augmentation de la dilatation pupillaire après le **stimulus**, cette réaction est cependant nettement moins marquée que lors de la synchronisation sur la **fixation** ou le **clic**. Il semble également raisonnable d'envisager que la plus faible réaction observée après le stimulus puisse être due à la proximité des autres événements, tant ceux-ci sont temporellement proches. Pour rappel, le temps de fixation médian est de 634 ms et celui de réaction de 2492 ms (4.4.1), soit moins de 2 secondes d'intervalle, quand la réaction observée couvre une durée de plus de 2 secondes. Cela laisse donc à penser que ce n'est pas l'apparition du stimulus en lui-même mais plutôt sa perception ou l'action qui causent la réaction observée. Ces résultats confortent donc l'hypothèse H1-1 formulée précédemment, une réaction plus significative étant observée lors de la fixation ou du clic que lors de l'apparition du stimulus.

Outre cette différence d'amplitude entre les courbes, la Figure 57 permet d'observer une similarité certaine dans l'évolution de la réaction. L'écart entre les courbes décroît rapidement dès lors que l'on s'éloigne du pic de dilatation. Il apparaît également que le temps mis pour atteindre le pic soit sensiblement identique dans les 3 cas. Nous observons en effet une initialisation de la dilatation environ 2 secondes avant le pic. Il semblerait donc que la méthode de synchronisation, et par extension l'événement déclencheur n'aie pas d'impact sur la durée du phénomène observé. En revanche cette comparaison apporte de nouveaux éléments de réflexion quant à l'origine de la réaction.

Un pic de dilatation pupillaire est couramment décrit par trois valeurs que sont son amplitude, la latence de pic définie comme le temps mis par la pupille pour atteindre un maximum local de dilatation après un événement déclencheur, et la latence à l'initialisation de la réaction. Le Tableau 11 présente les caractéristiques du pic de dilatation (latences et amplitude) selon les trois synchronisations étudiées.

	Stimulus	**Fixation**	**Clic**
Pic (mm)	0.062	0.091	0.105
Latence de dilatation (ms)	533	-120	-1400
Latence de pic (ms)	2633	1833	533

Tableau 11. Caractéristiques des réactions observées selon l'événement de référence du moyennage.

Relativement à l'événement de référence, la réaction est déclenchée légèrement après le stimulus (+ 533 ms), coïncide pratiquement avec la fixation (-120 ms) et déclenchée avant le clic (- 1400 ms). La coïncidence de l'initialisation de la réaction et de la fixation ne doit cependant pas mener à des conclusions hâtives sur le lien entre celles-ci. De nouveau, la proximité temporelle des événements comparés et la durée des réactions doivent modérer ces observations.

L'observation de la réaction que nous avons pu mettre en évidence par moyennage nous permet de conclure qu'une réponse est initiée à proximité de la fixation, résultant en un pic de dilatation atteint en général rapidement après le clic. En terme de processus, l'interprétation qui peut être faite est la suivante. Le clic apparaît comme étant indicateur d'un processus ayant commencé un peu plus tôt et la Fixation comme déclencheur d'un processus cognitif complexe résultant en une prise de décision d'action motrice sur la cible (déplacement de la souris et clic). Cependant nous ne pouvons distinguer l'influence de la fixation de celle du clic, ne nous permettant donc pas confirmer l'hypothèse H1-2 et donc de répondre à la question.

Ne pouvant donc pas conclure quant à l'origine de la réaction observée il nous est néanmoins permis de vérifier qu'une réaction semble bel et bien identifiable (hypothèse H1-1), témoignant d'un processus résultant de la fixation et/ou d'un processus précédent le clic (initiateur du clic).

4.4.3.4 *Moyennage par sujet*

Afin de consolider les observations effectuées dans les paragraphes précédents, nous vérifions si le comportement révélé par la moyenne est également observable à l'échelle de chaque sujet. Les moyennes sont donc calculées sur un faible nombre de cas, réduisant de fait la force d'amélioration du rapport signal bruit de cette méthode. Entre 9 et 18 cas par sujet ont été moyennés.

Nous reproduisons la méthode de comparaison appliquée plus haut en nous concentrant sur la Fixation (Tp) et le Clic (Tr). Sur la Figure 58 sont représentées les moyennes individuelles calculées sur ces deux événements (en bleu la fixation, en vert le clic).

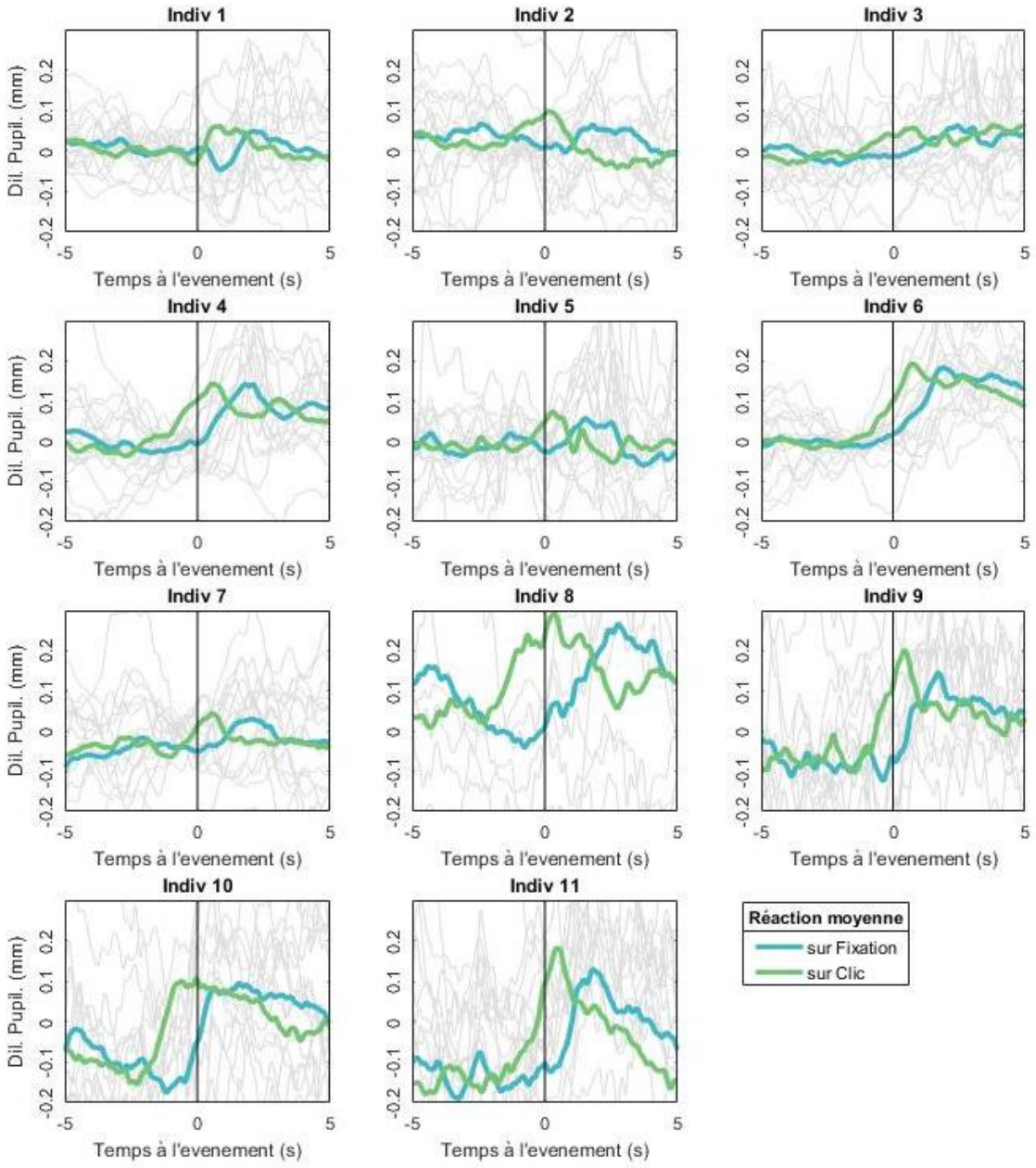


Figure 58. Représentation de la réaction moyenne Individuelle, par synchronisation sur la Fixation (bleu) et le Clic (vert)

Nous constatons tout d'abord que malgré le nombre de cas d'analyse réduit, une réaction est généralement visible pour tous les individus. Similairement aux observations établies en 4.4.3.3, la réaction est amplifiée par le moyennage sur le clic et la dilatation initiée quelques millisecondes avant la fixation. Nous retrouvons également dans la Figure 58 une illustration de la différence inter-individuelle observée au 4.4.3.2, la forme et l'amplitude des réactions moyennes étant propre à chaque individu.

4.5 Discussion des résultats

4.5.1 Conclusion

Dans ce volet expérimental nous nous sommes posé la question de l'existence et de la détectabilité de la réaction pupillaire liée à l'actualisation de la conscience de situation. Nous postulions que si cette réaction existait elle serait observable à la suite de la perception visuelle du nouvel élément situationnel à intégrer à la représentation de la situation. A ce propos et au vu des résultats des analyses présentées dans ce chapitre, nous formulons les conclusions suivantes.

Tout d'abord nous avons vu qu'une réaction pupillaire existe bel et bien, validant l'hypothèse H1-1. Cependant, cette réaction se retrouvant le plus souvent noyée dans un bruit comportemental important, nous n'avons pas pu effectuer la caractérisation générique de celle-ci nécessaire dans l'optique d'une détection automatisée en temps réel.

Concernant notre postulat sur l'origine de la réaction, nous ne pouvons dire si la première détection visuelle du stimulus est bien le déclencheur de la réaction, ou si celle-ci est due à l'initialisation de la réponse motrice. Il apparaît cependant probable que la concomitance des processus cognitifs en jeu puisse en être la cause.

Enfin la variabilité constatée, tant inter qu'intra-individuelle rappelle l'intérêt d'une approche de monitoring multi modale en situation écologique. Cela nous conforte néanmoins dans l'idée qu'il serait théoriquement possible de s'abstraire de la connaissance de l'apparition du stimulus pour mesurer l'actualisation de la conscience de situation ou du moins le traitement d'un nouvel élément situationnel.

Lors de cette étude nous avons tiré parti de l'approche multi-monitoring. Nous nous sommes volontairement positionnés dans un cadre réaliste tout en conservant la connaissance de la temporalité d'induction des stimuli. Cette situation, à mi-chemin entre le laboratoire et le travail réel de l'opérateur, a pour but de nous permettre d'estimer la sensibilité de la pupillométrie à la réalité de la situation opérationnelle. Dans un premier temps le monitoring comportemental, oculométrie et clics, nous a permis de lier le déroulement du scénario au comportement de l'utilisateur. La grande dispersion des Temps de Perception et Temps de Réaction ainsi que l'hétérogénéité des comportements pupillaires, illustrent la réalité de la variabilité comportementale humaine dès lors que l'on se rapproche d'une situation écologique. Nous avons ainsi étudié l'évolution de la dilatation pupillaire au regard de ces comportements.

Nous pouvons tirer 3 grands enseignements de cette analyse :

- les temporalités variables du processus cognitif étudié, et sa complexité, noient la réaction dans un bruit comportemental lié à la situation écologique, rendant ainsi la réaction difficilement détectable ;
- la complexité se traduit par de fortes variabilité tant inter qu'intra-individuelle rendant une caractérisation de la réaction difficilement généralisable ;
- la faible qualité des données initiales (moins de 50% des cas exploitables) questionne quant à l'applicabilité d'un tel monitoring à un traitement en temps réel, en situation opérationnelle.

Tout d'abord, nous avons pu constater la difficulté à identifier suffisamment clairement la réaction attendue. Le processus cognitif que nous cherchons à détecter et la liberté laissée par la situation

écologique n'ont permis d'identifier un semblant de réaction type que dans 77 cas sur les 171 étudiés (soit moins d'un cas sur deux). Seul un moyennage permet de faire apparaître le phénomène, autrement généralement noyé dans la complexité du signal physiologique et le bruit comportemental.

La pertinence du signal moyenné dépend de deux facteurs : la stabilité de la réponse et le nombre de cas moyennés. Ne pouvant pas, *ex-post-facto*, augmenter ce nombre de cas, les résultats obtenus s'avèrent peu marquants.

L'observation des moyennes révèle néanmoins deux phénomènes que nous interprétons comme, d'une part, l'ouverture d'un canal attentionnel lors de la perception, liée à la latence de dilatation, et, d'autre part, le traitement cognitif et la prise de décision, liés à la latence de pic. Cependant la distinction de leur contribution est difficile à faire et il ne nous a pas été permis de différencier la détection visuelle seule de la prise de décision qui en résulte. Cela ne nous permet pas de conclure quant à l'hypothèse H1-2.

De plus, comme mentionné plus haut, sur les 18 tâches isolées parmi les 100 du scénario d'origine, seuls 50% des cas ont présenté suffisamment de données pour pouvoir être traités correctement. La bonne qualité des données d'entrée est un critère indispensable à la fiabilité d'un système de détection automatique. S'il est certain que la qualité du matériel peut toujours être améliorée, il reste des raisons de perte de données inhérentes à la pupillométrie et au comportement humain en situation naturelle de travail ainsi qu'au bruit dû aux caractéristiques mêmes de cette situation. Bien que des pistes informatiques et mathématiques soient exploitables, nous avons pu voir que la réalité de la variabilité inter et intra-individuelle ainsi que du bruit comportemental et de l'omniprésence de réactions "parasites", inhérents à la situation écologique, sont autant de freins à l'applicabilité d'une détection automatique en temps réel d'une réaction physiologique liée à la perception d'un nouvel élément de situation. Un tel système de détection se doit donc d'être couplé à d'autres outils de monitoring pour une estimation dynamique plus fiable de l'état de l'opérateur. Une interprétation basée sur le croisement de signaux continus d'origine différentes (clics, mouvements oculaires, dilatation pupillaire, rythme cardiaque et autres) au regard de la proximité observée des événements, permettrait de renforcer la fiabilité du diagnostic.

Si plusieurs marqueurs issus du monitoring se manifestent dans un laps de temps réduit et dans un certain ordre, une interprétation plus robuste de l'activité de l'opérateur pourrait ainsi être formulée. Dans le cas présenté ici cela pourrait par exemple se traduire par une fixation, suivi d'un clic au même endroit, très rapidement suivi d'un pic de dilatation. Une telle méthode présente également un intérêt certain pour s'abstraire des schémas classiques de stimuli et scénarios contrôlés par l'expérimentateur et proposer des études de réactions pupillaire liées à des patterns attentionnels déduits des comportements visuels et moteurs, plus adaptées à la mesure en situation écologique.

4.5.2 **Discussion : monitoring, variabilité et interprétabilité.**

La pupillométrie a été employée avec succès pour le monitoring de l'activité cognitive dans de nombreuses expériences de perception. Nous pouvons citer entre autres la reconnaissance de la parole ou de mots, la lecture ainsi que différents tests auditifs (O'Neill & Zimmerman, 2000; Steinhauer & Hakerem, 1992). Nous avons documenté la manière dont un sujet réagit à la détection et la reconnaissance d'une nouvelle cible visuelle devant faire l'objet d'une action de sa part.

L'amplitude de la réponse pupillaire à une stimulation cognitive est généralement moindre que celle des variations induites par la luminance, la dilatation culminant à environ 0,1 mm au-dessus de la *baseline* (Beatty & Lucero-Wagoner, 2000; Peysakhovich, Vachon, & Dehais, 2017). L'objet de la pupillométrie

cognitive est précisément la mesure de ces fines réactions, au regard de l'évolution de la cognition et du comportement du sujet. Cependant nous avons pu constater que cet exercice se trouve complexifié par la variabilité de la situation écologique.

La grande variabilité observée à la fois dans le signal pupillaire (physiologique) et dans les temps de réaction (comportementale), illustre la réalité du monitoring dans une telle situation. En effet, chaque sujet étant libre quant à l'organisation de ses comportements à mettre en œuvre et donc de ses réponses, il peut décider de retarder le traitement de la tache où d'y accorder, par exemple, peu d'importance sur l'instant, ou au contraire de la traiter immédiatement. Or cette décision temporelle résulte néanmoins d'un traitement cognitif lié à la perception du stimulus et à son intégration à la conscience de situation. Nous pouvons donc penser que **la grande variabilité inter et intra-individuelle observée traduit la liberté décisionnelle et comportementale laissée par la situation écologique.** Bien que cette hypothèse corresponde à un postulat fort, il est probable qu'une nouvelle expérimentation puisse amener des éléments de réponse quantitatifs quant à la relation entre variabilité mesurée des variables issues du monitoring et degrés de liberté décisionnelle et comportementale laissés par la situation expérimentale, à partir d'analyses fines des comportements et interview des sujets. Néanmoins, nous voyons mal comment ces deux dimensions pourraient être mises en œuvre sans intervenir sur la tâche et donc influencer l'organisation spontanée du travail de contrôle, notamment dans les situations sollicitant beaucoup d'attention.

Nous souhaitons également profiter de ce constat afin d'approfondir la réflexion sur les limites du monitoring découlant de cette variabilité inhérente à la réalité de l'activité humaine en situation de travail, en nous penchant à la fois sur la difficulté d'interprétation des signaux à laquelle nous nous sommes heurtés, et à la contrainte de non intervention de l'expérimentateur en cours de tâche.

Cette variabilité soulève à notre sens deux questions principales. D'un part la question de la lecture et de l'identification de la réaction. D'autre part, quand ces réactions physiologiques ou comportementales sont effectivement observables ou détectables, se pose la question de leur interprétation.

L'identification claire d'une réaction, pourtant recherchée et attendue, fut l'une des difficultés majeures de notre travail. Il n'est pas apparu de réelle constance dans la forme de la réaction pupillaire, quand celle-ci a été identifiable. À la manière d'un signal EEG, le signal pupillaire recueilli en situation écologique semble se faire le reflet de la complexité et de la dynamique de la cognition de l'Homme dans son environnement. L'inévitable enchevêtrement des processus cognitifs en jeu, conscients et inconscients, est une source de différence de lisibilité flagrante entre des signaux pupillaires d'étude en laboratoire et notre étude se voulant plus écologique.

Outre l'influence de la complexité psychologique, les facteurs d'influence physiologiques sont également multiples. Si pour de telles études de *Cognitively Evoked Pupillary Responses* la luminosité ambiante est le premier facteur d'influence à contrôler, de nombreux autres facteurs persistent, dont les variations sont plus difficilement maîtrisables. Le comportement visuel en lui-même peut influencer le diamètre pupillaire dans une proportion difficilement contrôlable. Accommodation, saccades ou encore luminosité de la zone regardée créent des variations locales d'amplitude se cumulant à la réaction cognitive recherchée (Binda, Pereverzeva, & Murray, 2014; Mays & Gamlin, 1995; McDougal & Gamlin, 2010). Cette multitude de facteurs provoque autant de réactions de fréquences, d'amplitude et durées différentes se cumulant, générant in-fine un « bouillon » de réactions. De fait, le signal issu d'une activité écologique semble traduire ce tumulte d'activité. La réaction recherchée se trouve noyée dans un bruit comportemental écologique. De même, dû aux différentes temporalités des réactions, celle

recherchée se retrouver prise dans un processus sous-jacent plus lent comme cela peut sembler être le cas dans la figure ci-dessous.

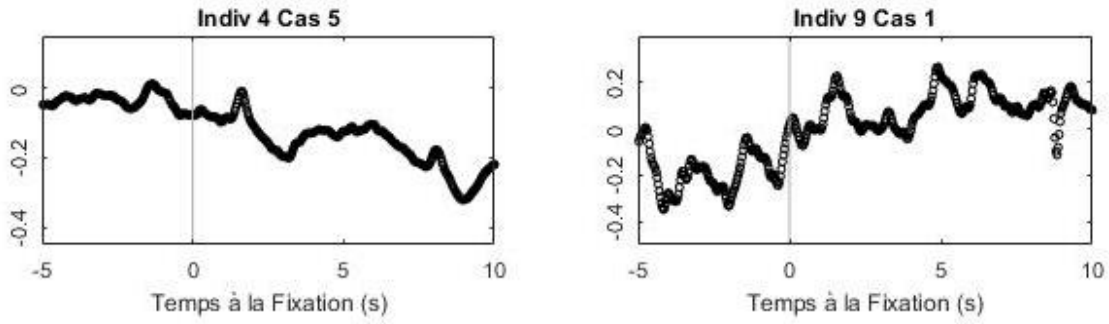


Figure 59. Exemples de cas où une réaction semble être prise dans un processus sous-jacent descendant (à gauche) et ascendant (à droite).

En psychophysiologie, la question de l'interopérabilité de la réponse physiologique est au cœur de plusieurs concepts et lois concernant l'organisation générale de l'organisme[48]. Nous mentionnons ici quelques lois nous ayant semblé avoir une pertinence pour l'interprétation de la complexité du signal physiologique.

En s'appuyant sur le principe de régulation homéostatique proposé par Lacey (1956), Wilder formule en 1957 la « loi du niveau initial » (LIV pour "*Law of Initial Values*")(Wilder, 1957). Selon Lacey, l'organisme est doté de propriétés homéostatiques régulatrices qui neutralisent les perturbations auxquelles sont soumis les différents phénomènes physiologiques et ainsi les stabilise. Le concept proposé par Wilder stipule ainsi que le niveau d'activité pré-stimulus d'une variable physiologique détermine la magnitude de la réponse. Selon la LIV, un niveau initial élevé prémobiliserait les mécanismes homéostatiques compensatoires et diminuerait ainsi l'amplitude de la réponse.

Cependant la majorité des études sur le sujet concernent principalement des signaux d'activité cardiaque, vasculaire ou électrodermale. Plus récemment cette loi appliquée à la pupillométrie a amené à une précision de l'interprétation du postulat de Wilder. La loi voudrait qu'un diamètre pupillaire initialement large implique une réponse plus faible. Cependant la loi d'opposition et d'équilibre homéostatique énoncée par Sokolov (Loi des processus opposants, Sokolov, 1963) voudrait qu'un diamètre pupillaire tonic large initial amplifie la réponse phasique, créant une réponse plus importante. Au vu de cette apparente contradiction, Jin (1992) propose, suite à l'examen de données expérimentales, de préciser la loi du niveau initial. Selon lui, celle-ci agit comme une restriction de la plage d'évolution de la pupille. Ainsi lorsqu' étant largement dilatée, la pupille ne peut se dilater plus (plafond de dilatation). La réaction observée est alors moindre que pour une valeur initiale plus faible, où la pupille aura plus de marge de dilatation. La détection de la réaction s'en trouve donc compliquée.

Tentons ici une analogie extrêmement simplificatrice en imaginant que la régulation du diamètre pupillaire soit assimilable à l'évolution du flux d'un torrent. La taille du torrent varie en fonction du flux apporté par les affluents situés en amont. Nous pouvons alors distinguer deux affluents principaux. L'un lié aux processus cognitifs liés à la tâche, et l'autre aux autres nombreux facteurs pouvant jouer sur le diamètre pupillaire, luminosité, émotion, stress.... Or le volume de ces affluents varie constamment. Estimer la part de dilatation pupillaire due à la sollicitation cognitive sur la tâche revient à estimer la

[48] Pour une synthèse, voir *Les méthodes en psychophysiologie*, Claverie & Paty, 1999.

quantité exacte d'eau apportée par l'affluent cognition quand tout ce qui est mesuré est le niveau du torrent.

L'observation d'une réaction nette au moment du traitement de la tâche peut être illustrée comme un augmentation du débit venant de l'affluent « cognition », mais nécessitant un débit constant ou réduit de l'autre affluent. Il faudrait donc posséder une évaluation/qualification précise de l'ensemble des facteurs de modification pour identifier l'origine et interpréter la variation avec précision. Cela demande une modélisation extrême, non seulement de l'état de l'utilisateur mais également de l'environnement dans lequel il évolue.

Concernant l'interprétation de la réaction

En dépassant le cadre expérimental, la question de l'interprétabilité d'une réaction est cruciale pour l'application de dispositifs de monitoring en situation réelle. L'utilisation du monitoring comme outil de détection de désynchronisation temporelle des consciences de situation (Chapitre 2) nécessite non seulement la détection de réactions, mais également de s'assurer que celles-ci ont bien la même origine.

Pour être comparables les mesures des temps de réaction, physiologiques et comportementaux, doivent pouvoir être rattachés à leur source. Or nous avons pu constater dans ce chapitre que relier la réaction pupillaire à une partie du raisonnement (ensemble coordonné de stratégies cognitives) lié à la tâche est extrêmement complexe en situation non maîtrisée. Nous souhaitons illustrer ici la difficulté de remonter à la source de la modification dans un environnement complexe où les facteurs de modification sont multiples et hétéroclites.

Le temps de réaction simple, défini comme le temps d'accomplissement de la boucle visuo-motrice (perception visuelle, prise de décision et réponse moteur) est, pour Fabre-Thorpe et al. (2001) décomposable au regard du temps requis par les étapes de traitement de l'information. Comme illustré par la Figure 60, 6 étapes relevant des traitements sensoriels, perceptifs, décisionnels et moteurs se succèdent pour déterminer un Temps de réaction minimal estimé à 250–290 ms dans le cas d'une tâche de catégorisation visuelle simple.

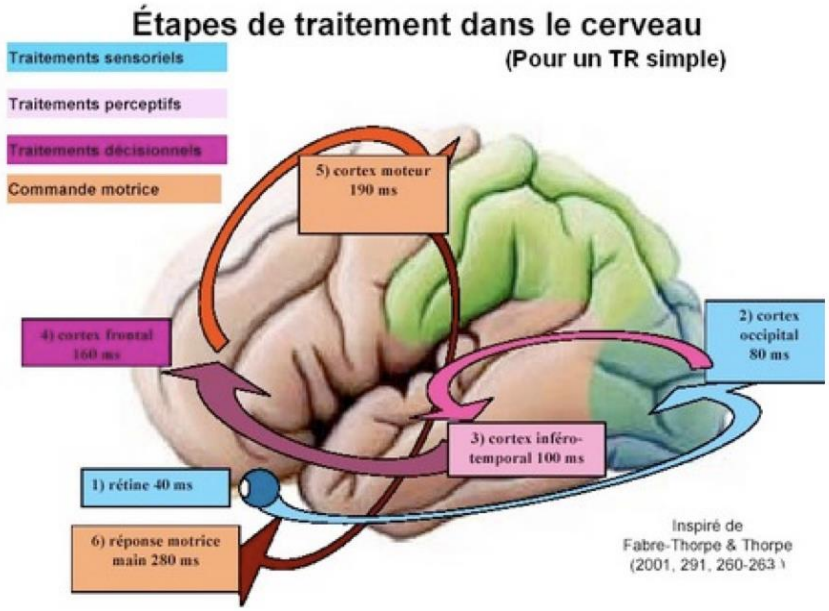


Figure 60. Etapes de traitement pour un temps de réaction simple, théorie hypothétique de chronométries des opérations mentales obtenues à partir de composantes tardives des potentiels évoqués. Inspiré de Fabre-Thorpe et Thopre (2001)

Des variantes de ces étapes du processus de traitement de l'information se retrouvent dans de très nombreux modèles (e.x., M. R. Endsley, 1995a; Gordon & Gill, 1997; Wickens & McCarley, 2007). Notamment avec le modèle COSMODRIVE, Bellet (2003) introduit pour l'activité de conduite une boucle de régulation dynamique Perception - Cognition - Action (Figure 61), dans laquelle intervient la conscience de situation. Le point intéressant est l'approche fractale de la boucle, où la perception et la cognition sont elles-mêmes représentées comme des cycles. La conscience de situation, au sens de Endsley, se situe alors à cheval entre ces deux étapes.

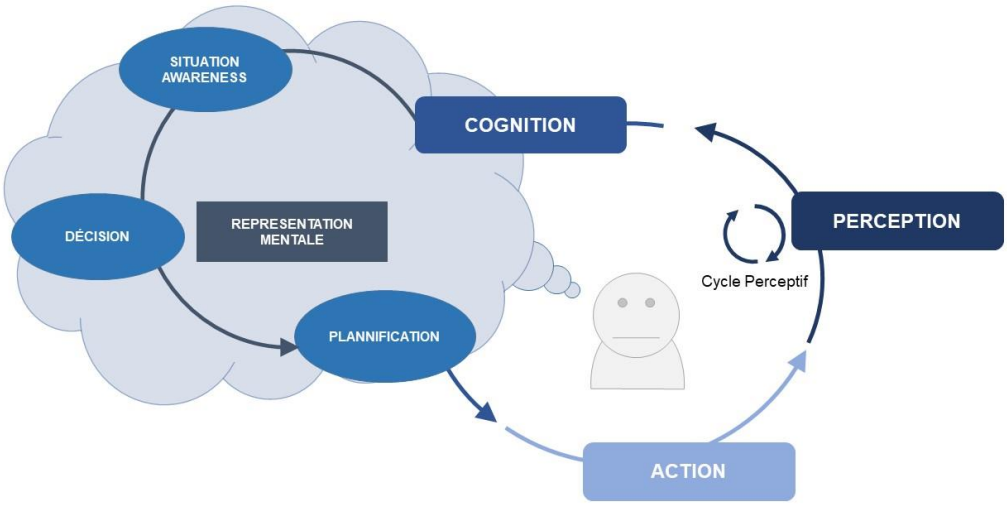


Figure 61. Modèle COSMODRIVE, d'après Bellet et al. 2010

Cependant de telles représentation donnent l'impression qu'étudier l'action revient à étudier la résultante du processus complet de traitement de l'information. Or, comme le présente Chalandon (Chalandon, 2007) "Les modèles linéaires de la cognition, où la compréhension intervient en aval de la perception et en amont de la prise de décision et de l'action, sont en fait des artefacts de méthodes de description". Dans cet esprit Hollnagel (1993) associe les étapes de traitement de l'information à de très nombreux court-circuits et boucles de rétroactions (Figure 62), cherchant à représenter les raccourcis et parallélisme pouvant exister entre les étapes.

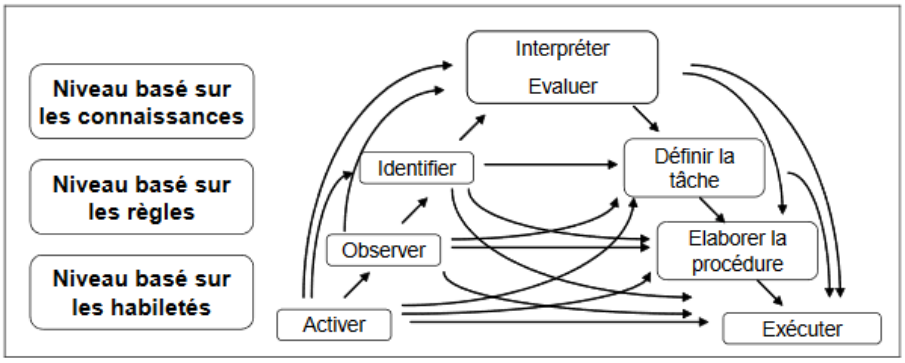


Figure 62. Echelle de décision et modèle SRK de Rasmussen (d'après Cacciabue & Hollnagel, 1995)

Regardons alors les étapes requises pour la réalisation de la tâche d'identification de l'expérience AMIGOS. Celle-ci demande dans un premier temps la catégorisation de l'aéronef en ami ou ennemi en se basant sur sa position et son code. Il s'agit là du traitement cognitif ou raisonnement lié à la tâche. Doit alors s'en suivre l'assignation de cette catégorie dans le logiciel, c'est à dire la réalisation de la tâche. Cette réalisation nécessite trois clics dont le premier est la sélection de l'aéronef, événement sur lequel nous avons basé notre analyse.

En détaillant la phase de traitement cognitif nous pouvons distinguer les 3 étapes suivantes :

- Etape 1 : Le sujet doit reconnaître le stimulus comme un nouvel élément situationnel à traiter et lui associer la tâche correspondante. Cela revient à répondre à la question "Quoi ?" ou "Dois-je le traiter ?". C'est l'étape de Compréhension et Interprétation de l'élément perçu, correspondant au Lvl 2 SA de Endsley.

- Etape 2 : Il doit ensuite déterminer si la tâche doit ou peut être traité maintenant, c'est à dire répondre à la question "Quand ?". Cela demande au sujet une priorisation des tâches en cours basée sur une Projection à court terme de l'évolution de la situation. On retrouve ici le Lvl 3 SA de Endsley.

- Etape 3 : Enfin, le sujet doit décider de la catégorisation de l'aéronef et élaborer la procédure à mettre en œuvre. Cela revient à répondre à la question "Comment ?", phase de recherche de la solution à la tâche et récupération des modèles mentaux associés à l'action à accomplir.

Ces étapes représentent un raisonnement lié à la tâche plus ou moins aboutit. Cette profondeur de traitement est à mettre en relation avec les niveaux SRK (pour *Skills-Rules-Knowledge*) de Rasmussen

(1983). Modèle de référence, SRK présente l'avantage de rendre accessible des niveaux de complexité des raisonnements en les illustrant par des boucles plus ou moins grandes, justifiant un coût cognitif plus ou moins élevé. Nous entendons donc part là que la même action observée peut résulter de niveaux de traitement plus ou moins profond et ne clôt pas nécessairement la phase de cognition, comme illustré par l'adaptation de Cacciabue et Hollnagel des niveaux SRK (Figure 62).

Dans le cas étudié dans ce chapitre, le clic peut aussi bien traduire l'initiation d'un plan d'actions pour la réalisation de la tâche (Etape 3 - je sais ce que j'ai à faire et je le mets en œuvre - Décision d'initiation de la réalisation), que la première action d'un traitement en cours d'élaboration (Etape 2 : Je sais que je dois la traiter maintenant mais ne sais pas encore comment - Décision d'initiation du traitement) ou bien que l'exécution mécanique d'une action liée à l'identification de l'élément à traiter (Etape 1 : Je sais que c'est à traiter mais ne sais pas encore quand - Décision de l'ordre du réflexe).

Cette distinction est illustrée par la Figure 63 dans laquelle nous intégrons les niveaux de SA de Endsley dans un processus Perception-Cognition-Action mettant en avant l'action comme une sortie pouvant résulter de n'importe quelle étape du processus de traitement cognitif.

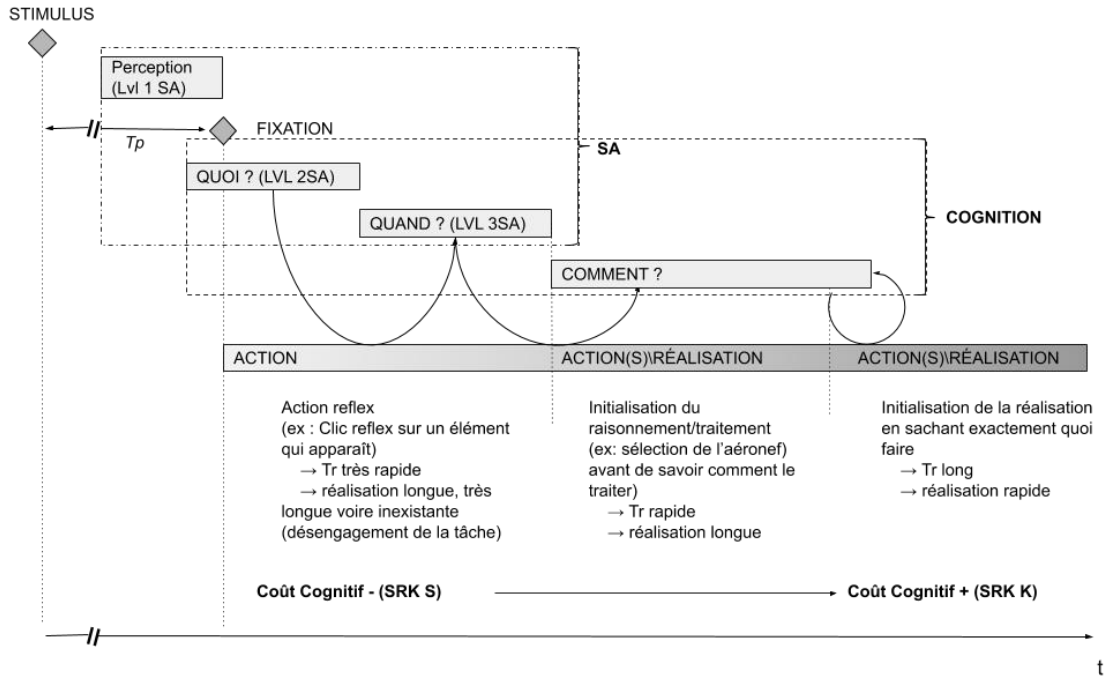


Figure 63. Model de traitement de la situation. La perception du stimulus, premier niveau de conscience de situation d'après Endsley, est suivie d'une fixation visuelle (ou focalisation attentionnelle) qui initialise le processus de traitement cognitif (Cognition) en 3 phases : « Quoi » (niveau 2 de SA), « Quand » (niveau 3 de SA) et « Comment ». L'action peut alors être déclenchée à n'importe laquelle de ces étapes, traduisant d'un traitement plus ou moins aboutit. Cette représentation du processus Perception-Cognition-Action, incluant les 3 niveau de SA de Endsley, traduit la variabilité des temps de réaction observée lors de l'expérience AMIGOS.

Nous illustrons ici la difficulté de lier un monitoring comportemental, comme le clic de souris, à l'évaluation d'un niveau de conscience de situation ou de traitement cognitif d'une tâche. En situation écologique, où l'individu est laissé libre dans la perception, le traitement et la réalisation de ses tâches, une telle méthode d'acquisition ne semble pas permettre l'interprétation correcte d'un temps de réaction.

Le temps de réaction Tr mesuré peut varier non seulement en fonction de la durée de chaque étape du processus comme le présentent Fabre-Thorpe *et al*., mais également en fonction de l'étape à laquelle intervient le clic. La latence Tr ne traduit donc pas nécessairement le temps de traitement cognitif de la tâche, pouvant ainsi expliquer en partie la variabilité de la réaction pupillaire individuelle relativement au clic et à la fixation. Cela illustre la difficulté de lier d'une part une action à une réaction pupillaire et d'autre par une action à une partie précise du traitement cognitif.

Les deux questions adressées par ces réflexions, de détectabilité et d'interprétabilité de réaction issues du monitoring, nous amènent à reconsidérer l'applicabilité du monitoring à la situation écologique.

La démarche de monitoring, et multi monitoring, qui est prônée par de nombreux auteurs apparaît comme l'opportunité majeure pour développer le paradigme jusque-là balbutiant, de la Machine comprenant l'Humain s'adaptant dynamiquement à celui-ci. Cependant pour être applicable au traitement en temps réel requis, il ne s'agit pas seulement détecter une réaction, comme cela peut l'être en situation de laboratoire, mais de détecter la bonne réaction parmi un ensemble parasite d'activités, de réactions plus ou moins importantes dans laquelle la réaction d'intérêt peut se retrouver noyée. Le monitoring de l'opérateur, physiologique ou comportemental, doit être dans ce contexte envisagé comme un outil d'inférence d'un état sans chercher à en interpréter la cause.

Plus concrètement pour nous, la question de la granularité de l'évolution étudiée se pose également. Si à l'échelle de quelques secondes le signal est noyé sous la masse des micro fluctuations parasites provenant de la multitude de facteurs influent, une approche sur période temporelle plus longue (quelques minutes) autoriserait peut-être l'étude de tendances et le lissage du bruit. Cette approche est celle adoptée par l'étude MOCOM présentée en Chapitre 5.

Chapitre 5 Monitoring collectif - Expérience MOCOM

Le chapitre 3 a été consacré à la présentations d'une situation de travail collaboratif en l'absence de monitoring. Dans le chapitre 4 nous avons appliqué un monitoring physiologique (pupillométrie) et comportemental à une situation de travail individuel. Ce chapitre concerne l'étude du monitoring dans un contexte de tâche de C2 collaborative.

5.1 Contexte et objectifs de l'expérience

5.1.1 **Contexte**

Nous nous intéressons ici à la première expérience réalisée durant le doctorat. Celle-ci a été effectuée dans le cadre du co-encadrement d'un groupe d'étudiants en charge de la poursuite d'une expérimentation sur le développement de stratégies de groupe et le monitoring cardiaque par électrocardiographie. Cet encadrement m'a amené à compléter le protocole de cette expérience, pour mettre en pratique un premier monitoring pupillaire en situation de travail collaboratif. Cela a permis d'initier la mise en place d'un cadre expérimental pour l'étude du partage de conscience de situation et de me familiariser avec la pupillométrie et les procédures de traitement du signal associées.

Nous avons rapidement identifié les fortes contraintes relatives à l'expérimentation du monitoring en situation de travail collaboratif ainsi qu'à l'étude du partage de conscience de situation dans ce contexte. Comme nous l'avons évoqué précédemment[49], la démarche de monitoring est coûteuse et d'autant plus complexe à mettre en place dans le cadre du monitoring d'un groupe. Les systèmes d'acquisition et appareils de mesure sont multipliés, posant également la difficile question de leur synchronisation et de la mise en cohérence de leurs mesures.

Afin de répondre à ce nouveau besoin du laboratoire, nous nous sommes procuré deux dispositifs d'*eye tracking* déportés compatibles avec celui déjà présent. La production du modèle étant officiellement abandonnée nous nous sommes rapprochés du constructeur pour obtenir un modèle d'occasion. Le second appareil a été prêté par Thales dans le cadre du partenariat avec l'ENSC. Ces trois appareils ont pu être utilisés grâce à un outil de calibration et d'enregistrement que j'ai dû développer avec le Professeur Legrand, proche du projet. En revanche aucune solution d'acquisition centralisée et

[49] Voir chapitre 4, pp. 77

synchronisée n'a pu être mise en place. Nous revenons sur la méthode de synchronisation post-expérimentale des données au 5.3.1.

Nous mentionnons également au Chapitre 3 la difficulté d'accès à un outil pertinent de simulation de tâche collaborative. L'un de nos objectifs étant de se rapprocher de situations écologiques, la simulation d'une situation de travail collaboratif requiert un outil à la fois suffisamment complexe pour permettre la réalisation de tâches de C2 collectives réalistes et suffisamment simple pour être appréhensible sans formation étendue afin de répondre aux disponibilités du public expérimental (à savoir principalement des étudiants et quelques chercheurs). De plus l'étude du partage de conscience de situation par monitoring nécessite que les tâches et les outils de mise en situation répondent à des besoins particuliers. C'est pourquoi il est principalement nécessaire d'avoir la main sur le scénario de simulation, en particulier en ce qui concerne l'apparition et la présentation des éléments de la situation. Afin d'étudier les latences de réaction à l'intégration d'un nouvel élément de situation, il est primordial de posséder la maîtrise de ces évènements.

L'outil de mise en situation de travail collaboratif prévu dans le protocole de cette expérience se présente sous la forme d'un jeu vidéo de commandement d'opérations militaires. Toutefois, cet outil ne permet pas d'avoir la maîtrise du scénario nécessaire à l'étude des latences de synchronisation de la conscience de situation comme nous l'avions souhaité. Si la mise en situation non contrôlée se rapproche un peu plus de la réalité du terrain, l'étude de la dynamique du partage de conscience de situation demande un lourd travail d'identification post-expérimentale des marqueurs de modification de la conscience de situation.

Cependant une étude comparative d'une sélection d'outils de simulation auxquels nous pouvions avoir accès a été réalisée. Trois outils proposant une tâche réaliste de commandement et contrôle (C2) d'opérations ont été comparés. Aucun d'eux ne répond à la fois au critère de tâche collaborative simple et réaliste et au besoin d'accès au scénario. La conception d'un outil répondant à ces critères et autorisant une induction de différences de conscience de situation (SA) a été engagée, sur la base du logiciel *Warship Commander Task* (St. John & Kobus, 2003) utilisé par la DARPA pour le projet *Augmented Cognition* (St. John et al., 2004). Le programme *Augmented Cognition* visait à évaluer la sensibilité d'outils de monitoring aux modification de l'état cognitif d'un opérateur. Les scénarios proposés par ce logiciel sont calibrés pour induire de manière précise différents niveaux de charges sur un opérateur individuel lors d'une tâche de surveillance d'espace aérien (version similaire mais simplifié de la tâche utilisée au chapitre 4). Il répond également aux mêmes contraintes de réalisme et simplicité que nous rencontrons. Cet outil est présenté au 6.2.2.

Malgré ces multiples difficultés nous avons pu mener à bien cette expérience et nous confronter à un manque de maturité technologique, nous amenant à effectuer les études des chapitres 3 et 4.

5.1.2 Hypothèses de recherche

La situation expérimentale – monitoring en situation de travail collaboratif réaliste - est la plus complète qu'il nous a été possible de réaliser par rapport à nos hypothèses de recherches principales (voir 2.4). Nous reprenons donc les grandes lignes théoriques concernant la synchronisation comportementale et des indices physiologiques des membres d'une équipe en situation de collaboration. Nous précisions ici ces questions de recherches et ces hypothèses au regard des conditions expérimentales à notre disposition.

Historiquement, les études portant sur la cognition collective se sont majoritairement axées autour du partage d'information, de connaissances et de l'organisation du groupe. Cependant depuis quelques années ce courant d'étude de la cognition d'équipe est enrichi par des méthodes venues de la psychophysiologie et de la neuroergonomie. Dans ce contexte, la synchronisation comportementale mais également physiologique a des implications importantes pour les équipes en termes de gestion de la charge de travail, de comportement coordonné et de fonctionnement global. Pour quantifier le degré de synchronisation entre les membres d'une équipe, la synchronisation des réponses physiologiques de ses membres a été observée dans de nombreux contextes et à travers divers indicateurs dont nous pouvons donner plusieurs exemples.

Dans une étude portant sur un groupe de quatre étudiants (Henning et al., 2009), au cours de 18 réunions, il a été constaté que la synchronisation des variabilités cardiaque (HRV) était liée à une plus mauvaise évaluation du travail en équipe et de la productivité. Selon une autre étude (Elkins et al., 2009) portant sur 10 équipes de quatre personnes ayant effectué une tâche d'entraînement militaire, une meilleure performance de l'équipe était associée à une meilleure synchronisation des rythmes respiratoires. Par électroencéphalographie (EEG) Stevens et al. (2014) observent également un lien entre la synchronisation de l'activité du système nerveux autonome au sein de six équipes de sous-mariniers de l'US Navy et la dynamique du groupe. L'impact de la collaboration sur la synchronisation des réponses électrodermales (GSR) a également été mis en évidence dans le contexte de 4 coéquipiers participant à un jeu de d'intervention d'urgence (Guastello et al., 2016).

Si l'*eye tracking* est également un outil plébiscité, c'est principalement le cas dans le cadre précis de la synchronisation des comportements visuels et l'analyse des fixations que se concentrent la majorité des études (e.x., Bornard, Bellet, Mayenobe, Gruyer, & Claverie, 2011; Dehais, Causse, Vachon, & Tremblay, 2012; Lini, 2013; Peysakhovich & Hurter, 2018). À notre connaissance, la synchronisation de dilatation des pupilles n'a fait l'objet que de très peu d'études, principalement orientées autour du couplage émotionnel. En particulier Kang et Wheatley (2017) ont observé une synchronisation pupillaire collective entre les locuteurs et les auditeurs, maximale pendant les pics émotionnels d'un récit, et diminuant à mesure que les récits devenaient moins engageants. Nous proposons donc une première ébauche de l'utilisation de la pupillométrie comme outil de mesure du partage de conscience de situation. En particulier au regard du concept de synchronie des consciences de situation développé au Chapitre 2, nous adressons la question de recherche suivante :

(Q1) Le décalage temporel (latences de synchronie) dans l'actualisation des consciences de situation de coéquipiers est-il détectable par l'étude de la dynamique de dilatation pupillaire des individus du groupe ?

L'hypothèse qui découle rejoint celles formulées au Chapitre 1 :

H1 : **L'actualisation de la conscience de situation relative à un élément de connaissance partagée de la situation (NSKE) produit un pic de dilatation pupillaire permettant, par comparaison des dynamiques individuelles relativement à l'événement déclencheur, d'observer la désynchronisation temporelle et de déduire les latences.**

Cependant, conscients des limites liées à l'absence d'accès au scénario de la simulation, nous posons également la question de recherche suivante, à notre sens plus adaptée au contexte expérimental :

(Q2) La collaboration entre deux coéquipiers se traduit-elle par la synchronisation de la dynamique des dilatations pupillaires ?

Le traitement de cette question passe par deux approches complémentaires, l'une "bottom-up" dont l'hypothèse de recherche peut être formulée comme suit :

H2-1 : **Des dynamiques de dilatations pupillaires similaires entre coéquipiers indiquent une collaboration active entre eux.**

La seconde, "top-down" reposant sur l'hypothèse

H2-2 : **Les coéquipiers collaborant présentent une similarité accrue de leur dynamique de dilatation pupillaire.**

Nous commençons par décrire la situation de collaboration utilisée et le plan d'expérimentation. Ensuite, nous présentons les méthodes et les développements visant à garantir l'exploitabilité des données. Puis, nous exposons les résultats de nos analyses, principalement axées sur deux approches qualitatives complémentaires et les grandes lignes d'une méthode quantitative pour traiter ces mêmes hypothèses de synchronisation des dilatations pupillaires. Pour terminer, nous discutons des résultats et des observations issues de ces travaux.

5.2 Plan d'expérimentation

Afin de traiter ces objectifs scientifiques nous avons adapté le protocole expérimental initialement centré sur le monitoring cardiaque en y intégrant les étapes et conditions requises pour les mesures pupillométriques. Ce protocole expérimental est décrit ci-dessous. Nous détaillons la situation de collaboration étudiée, les données collectées et les variables évaluées.

5.2.1 La tâche collaborative : Wargame Red Dragon

Wargame Red Dragon est un jeu vidéo de stratégie en temps réel dans lequel des joueurs collaborent afin de vaincre une équipe virtuelle contrôlée par l'ordinateur. Publié en 2014, il est développé par Eugen Systems. Il s'agit d'une simulation relativement réaliste de guerre moderne dans lequel un sujet déploie et supervise des unités militaires de différentes sortes. Les joueurs contrôlent une armée contemporaine, composée d'infanteries, de véhicules terrestres et aériens issues des forces de l'OTAN. L'objectif est alors de conquérir et défendre plus de zones géographiques que l'adversaire dans un temps imparti. Les deux équipes s'affrontent sur un terrain découpé en 8 zones (Figure 64) nommées selon l'alphabet OTAN (Alpha, Bravo, Charlie, Delta, Echo, Foxtrot, Golf, Hotel).



Figure 64. Aperçu du champ de bataille virtuel ou les équipes s'affrontent. Les zones à conquérir sont nommées d'Alpha à Hotel. Ici la zone Alpha (en bleu) est la zone de départ occupée par les coéquipiers, la zone Hotel (en rouge) est celle occupée par l'adversaire (ordinateur)

Les zones de départ des deux équipes, Alpha ou Hotel, sont diamétralement opposées sur la carte. Petit à petit les coéquipiers doivent conquérir les zones intermédiaires en y déployant des troupes, occupant des immeubles et repoussant l'équipe adverse. Pour occuper une zone, une unité de commandement doit y être positionnée. Les coéquipiers doivent ensuite défendre cette position contre les attaques ennemies. L'occupation d'une zone est récompensée par l'obtention de "Points de conquête", qui s'accumulent avec le temps. L'équipe atteignant 500 points de conquêtes remporte la partie. Si au bout de 15 minutes de jeu aucun des deux adversaires n'a atteint ce score, celui possédant le plus de points est considéré vainqueur. Ces points de conquêtes servent donc à la fois de score pour le jeu et d'indicateur expérimental de performance à l'équipe.

Au sein de l'équipe chaque joueur contrôle ses propres troupes mais à visuellement accès à la situation dans son ensemble. Il peut donc à tout moment observer les mouvements de troupes de ses coéquipiers comme de son adversaire. Les joueurs sont libres de déplacer la caméra dans l'environnement. Plusieurs niveaux de zoom sont disponibles allant d'une visualisation globale de la carte à une très précise (Figure 65).



Figure 65. Illustration des niveaux de zoom sur l'environnement. L'images de gauche représente le niveau le plus dézommé disponible, l'image de droite le niveau le plus rapproché

Les joueurs ont à leur disposition un catalogue d'unités militaires comprenant 8 catégories : Logistiques, Infanterie, Support, Blindés, Reconnaissance, Véhicules, Hélicoptères et Air.

Deux grands types d'actions sont réalisables :

- **déploiement** : cette action consiste à sélectionner une nouvelle unité à déployer sur la carte. Chaque unité à un coût, variant selon son type et son niveau d'entraînement. Le joueur dispose alors d'une réserve personnelle de "Points de commandement" lui permettant "d'acheter" des unités. Il revient donc au joueur de gérer ses propres dépenses et achats ;
- **commandement** : il s'agit de gérer les troupes présentes sur la carte. Le joueur doit réaliser plusieurs actions telles que : ordonner des déplacements, faire monter ou descendre des troupes des véhicules, attaquer, cibler des positions ou troupes ennemies, positionner des défenses...

Les sujets sont amenés à prendre des décisions à la fois sur le plan stratégique (par exemple décider quelle prochaine zone attaquer), tactique (décider quel type d'unité déployer ou quelle attitude une unité doit adopter face à l'ennemi - agressive, défensive, dissuasive...) et opérationnel (déplacer un bataillon). Cette diversité couvre un panel de tâches représentatif du commandement et contrôle d'opérations. Pour l'expérimentation, le jeu à au préalable été paramétré de la manière suivante pour toutes les équipes :

- Mode : Conquête.
- Nombre de points de conquêtes à atteindre : 500
- Revenus : Normaux (+7)
- Carte : Wonsan Harbor (version terrestre)
- Difficulté : Facile (parties 1 à 6) à Normal (parties 7 à 11)
- Type de Partie : Multi-joueurs
- Mode : 3 joueurs VS IA
- Temps de positionnement : Pas de limite
- Durée d'une partie : 15 minutes (hors temps de positionnement).
- Deck des participants : Deck de départ Bluefor
- Deck de l'ennemi : Deck de départ Redfor

5.2.2 Participants

Pour cette étude 9 participants ont été recrutés sur le campus de l'Université de Bordeaux, 2 femmes et 7 hommes âgés de 20 à 26 ans (M= 22,9), possédant tous une vue bonne ou corrigée. Les participants ont été groupés en 3 équipes (A, B et C) de 3 joueurs[50]. Les passations ont eu lieu entre le mois de décembre 2017 et le mois de janvier 2018. Les sujets n'avaient qu'une infime ou inexistante connaissance préalable du jeu et du domaine militaire mais présentent tous une consommation à minima mensuelle de jeux vidéo.

5.2.3 Matériel

L'expérience s'est déroulée au sein de la plateforme technologique de l'ENSC dans une salle optimisée pour les situations de travail collaboratif. Chaque participant a été installé à un poste de travail équipé d'un PC et d'un moniteur Dell® de 24 pouces. L'organisation de la salle est présentée sur la figure X. Concernant le matériel d'acquisition, le rythme cardiaque a été enregistré par des cardiofréquencemètres Polar montés sur ceintures ventrales portés à même la peau. Les données sont récoltées via l'application *HRV logger* installés sur 3 *ipods* connectés via bluetooth avec le cardiofréquencemètre.

L'acquisition du diamètre pupillaire est effectuée via 3 *eye trackers* déportés Tobii® Pro X2-30 (30Hz) montés sous l'écran. Un programme de calibration et d'acquisition des données issues des *eye trackers* a été développé sous MATLAB® pour l'occasion sous la supervision de Pierrick Legrand (Institut de Mathématiques de Bordeaux). Les données ont été enregistrées séparément pour chaque partie dans un fichier .csv.

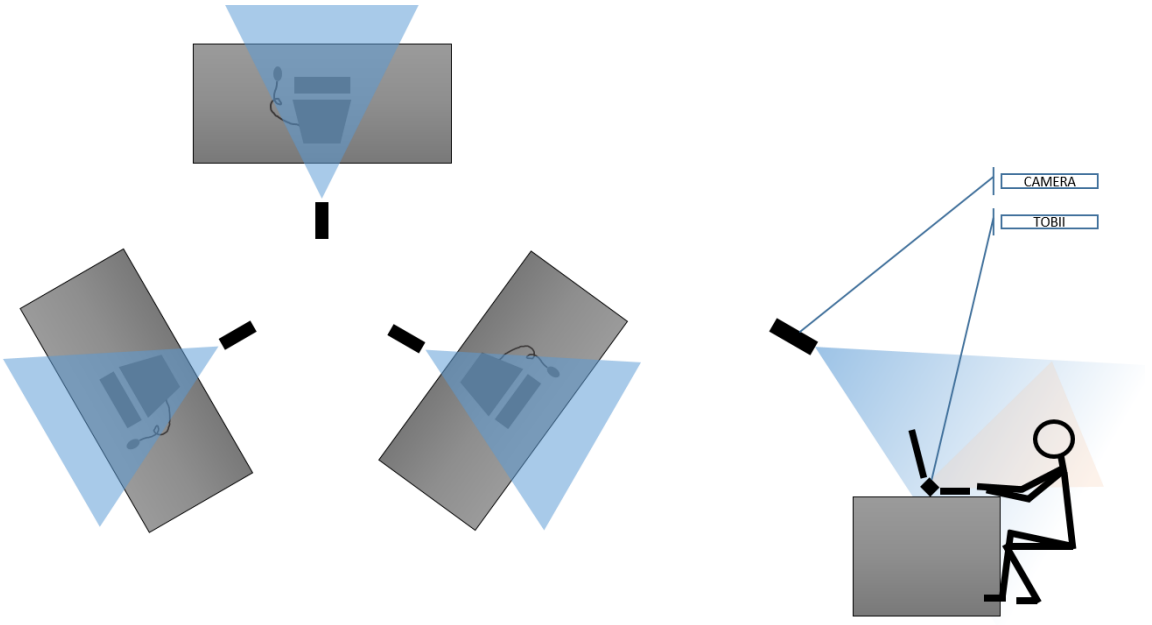


Figure 66. Représentation de l'organisation spatiale des postes.

Chaque poste de travail est également équipé d'une caméra connectée en réseau. L'acquisition synchronisée des vidéos est effectuée par le logiciel iSpy. Une capture d'écran vidéo de chaque partie

[50] Nous avons constitué des groupes de 3 du fait de notre limitation expérimentale à 3 *eye trackers*.

est également faite via le logiciel libre OBS Studio. L'absence d'un logiciel de centralisation des acquisitions a demandé un travail conséquent de synchronisation post-expérimentale des données de sources et types différents[51].

Bien que l'analyse conjointe des métriques cardiaques et pupillaires présente un intérêt scientifique potentiel, nous nous concentrons sur la partie oculométrie du protocole et sur l'analyse des données recueillies. En ce qui concerne les données cardiaques, elles ont fait l'objet d'une analyse séparée par les étudiants.

5.2.4 Déroulement des passations et instruction aux participants

Afin d'adresser les questions concernant les stratégies de groupes, le plan expérimental a été conçu de la manière suivante :

Séance 1 (4 parties)	Séance 2 (4 parties)	Séance 3 (3 parties)
Participant **A1**	Participant **A1**	Participant **A1**
Participant **A2**	Participant **A2**	Participant **B3**
Participant **A3**	Participant **A3**	Participant **C3**
Participant **B1**	Participant **B1**	Participant **A2**
Participant **B2**	Participant **B2**	Participant **B1**
Participant **B3**	Participant **B3**	Participant **C1**
Participant **C1**	Participant **C1**	Participant **A3**
Participant **C2**	Participant **C2**	Participant **B2**
Participant **C3**	Participant **C2**	Participant **C2**

Tableau 12. Plan expérimental

Pour chaque équipe, établies à l'avance, l'expérience s'est décomposée en 2 sessions de 4 parties, puis les équipes ont été mélangées pour une dernière session de 3 parties.

Ce mélange des équipes n'étant pas pertinent au regard de nos questions de recherche, nous nous penchons uniquement sur les 2 premières sessions expérimentales dans la suite de ce chapitre, soit 24 parties.

Lors de la première séance le contexte de l'expérience et déroulement des diverses séances est présenté aux sujets. Ce briefing initial comporte également une phase de présentation du jeu et de la tâche, sous forme d'un tutoriel présenté par l'expérimentateur. Ainsi les grands principes du jeu, l'interface, les

[51] Ce travail est détaillé au 5.3.1

catégories d'unités ainsi que les phases de déploiement et de combat sont présentées. Cette initiation se termine par une démonstration de 5 minutes où sont détaillées les principales interactions de sélection et déploiement de troupes ainsi que la prise d'un nouveau territoire par une unité de commandement.

Les participants remplissent ensuite un formulaire de consentement éclairé.

En suivant les coéquipiers sont équipés du cardiofréquencemètre dont le fonctionnement correct est vérifié. Ils sont ensuite amenés à leur poste de travail où la calibration de l'*eye tracker* est effectuée. Il est demandé aux sujets d'adopter une position de travail la plus naturelle possible. Celle-ci est alors ajustée (variation de la hauteur de chaise) afin de répondre aux critères de qualité de détection des pupilles requis par l'*eye tracker*.

Ces dernières étapes d'appareillage et de calibration son bien entendu répétées à chaque séance.

L'enregistrement des données cardiaques, pupillaires et vidéos est ensuite lancé par l'expérimentateur sur chaque poste ou outil correspondant dans l'ordre suivant : caméra, vidéo écran, *eye tracker*, cardiofréquencemètre.

Afin de pallier le manque d'outils d'acquisition synchronisée et compte tenue de l'hétérogénéité des sources d'enregistrement il a été nécessaire de donner un élément de synchronisation temporel entre tous les enregistrements. Pour ce faire nous avons demandé aux sujets de fermer les yeux pendants 5 secondes avant de lancer la partie. La durée est contrôlée par l'expérimentateur, chargé de leur signaler cette durée par un bruit sonore, provoqué par un clap des mains au début et à la fin. Les claps s'entendant également sur les enregistrements vidéos et audio, cette méthode avait pour but de permettre de posséder un repère de synchronisation commun à tous les enregistrements. Cependant il s'est avéré que ces 5 secondes n'ont pu être clairement identifiées sur les enregistrements pupillaires. Les enregistrements vidéo ont en effet montré des comportements très variables entre les sujets, notamment certains entre-ouvrants les yeux pendant cette période de synchronisation ou les ré-ouvrants trop tôt ou trop tard (certains semblant compter mentalement), générant de fait un manque de précision gênant pour l'utilisation de cette méthode. De plus, une défaillance des *eye trackers* nous a poussé à développer une méthode de synchronisation des signaux pupillaires spécifique dont nous faisons état au 5.3.1.

Au deuxième clap les sujets rouvrent donc les yeux et lancent la partie préalablement configurée par l'expérimentateur.

Il est alors possible de décomposer une partie en deux phases distinctes. Au début aucune troupe n'est présente sur la carte. Dans une première phase les coéquipiers effectuent un déploiement initial des unités qu'ils souhaitent. C'est également l'occasion pour eux de se mettre d'accord sur une première stratégie. Les coéquipiers ont autant de temps qu'ils le souhaitent pour cette phase de positionnement. Dès que les trois coéquipiers sont prêts, ils démarrent la phase de jeu réelle. Cette phase dure un maximum de 15 minutes, le jeu peut se terminer plus tôt si l'équipe ou l'adversaire atteint les 500 points de conquête nécessaires pour remporter la victoire totale. Nous concentrons notre travail sur cette phase de jeu réel, la phase de positionnement étant trop variable et trop courte.

Une fois la partie finie, il est à nouveau demandé aux sujets de fermer les yeux durant 5 secondes, dirigés une fois de plus par l'expérimentateur. Les enregistrements sont ensuite éteints dans l'ordre inverse duquel ils ont été lancés. Nous avons accordé une pause de 5 minutes aux participants de l'expérience, qui avaient pour consigne de rester à leurs postes durant cette période afin de ne pas avoir à recalibrer le matériel. Ce protocole est répété pour les 4 parties de chaque session.

La Figure 67 montre cette mise en situation expérimentale.

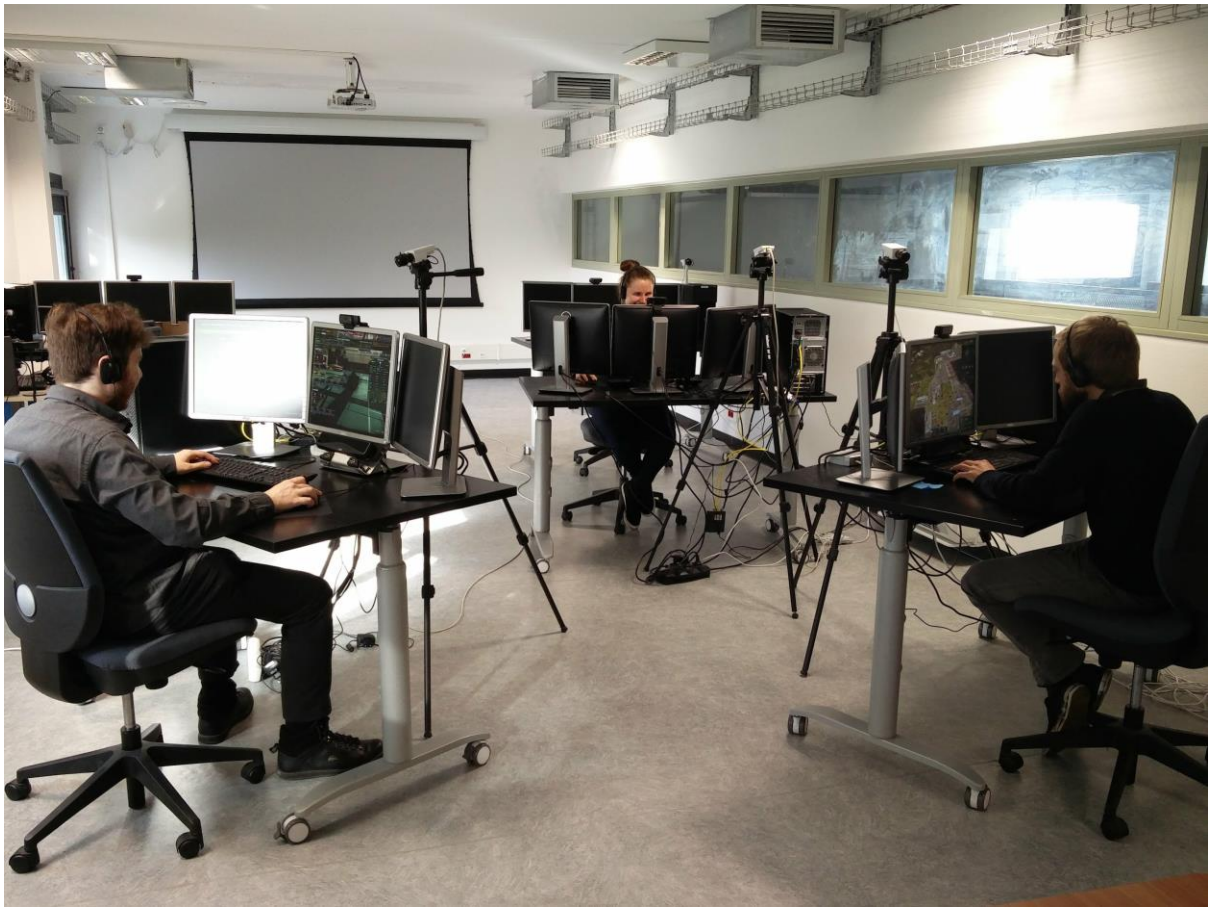


Figure 67. Mise en situation expérimentale

5.2.5 **Les variables**

Nous avons concentré notre analyse sur les données vidéo issues de l'enregistrement de l'écran des participants ainsi que les données pupillaires et temporelles récoltées par les *eye trackers*. Les enregistrements vidéo sont utilisés pour qualifier des situations de collaboration et non-collaboration entre les coéquipiers. Ces enregistrements sont indépendants les uns des autres et leur synchronisation ne peut se faire qu'a posteriori. Afin d'adresser l'hypothèse H2-1, nous utilisons la qualification de la collaboration comme variable indépendante et la dilatation pupillaire comme variable dépendante. L'hypothèse H2-2 étant le complémentaire de celle-ci, la dilatation pupillaire est alors notre variable indépendante et la qualification de la collaboration la variable dépendante.

5.3 Exploitation des données

5.3.1 **Synchronisation des données pupillométriques**

À plusieurs reprises nous avons mentionné le besoin et déploré l'absence d'un outil d'acquisition synchronisée des données. La méthode de synchronisation externe (par clap) utilisée durant l'expérience nous a permis de synchroniser entre eux les flux audio et vidéo issus des captures d'écrans et des caméras, mais ne nous a pas permis à procéder à la synchronisation des données pupillaires. Les

acquisitions étant lancées et arrêtées les unes après les autres par l'expérimentateur les fichiers de données sont de fait de longueur variables. Une synchronisation est donc primordiale afin d'extraire les données correspondantes à la phase de jeu.

La défaillance de l'horloge interne des 3 *eye trackers* nous a poussé à procéder à une synchronisation post expérimentale des données. Dans un *eye-tracker*, l'horloge interne sert de référence temporelle pour l'horodatage. Or, nous avons observé que l'horloge est réinitialisée au 01/01/2001 à chaque nouvelle mise sous tension de l'appareil. Les timestamps récupérés lors de l'acquisition de chaque échantillon sont donc erronés, ne correspondant ni à l'heure et à la date réelle, ni à ceux des autres *eye trackers*. Nous avons corrigé chaque timestamp à partir de l'heure d'enregistrement du fichier (celle-ci étant correcte car basée sur l'heure synchronisée de l'ordinateur), censée être le dernier timestamp enregistré dans le fichier, et en rétro-propageant les pas d'échantillonnages. Le procédé complet de re-synchronisation est détaillé en Annexe 4.

Pour chaque partie un nouveau jeu de données est ainsi créé, formé des données pupillaires synchronisées des trois coéquipiers.

5.3.2 Prétraitement des données pupillaires

Les données pupillaires ont été nettoyées et préparées en suivant le même processus et en utilisant les mêmes algorithmes que les données du chapitre 4[52]. La dilatation pupillaire a été dérivée des données de diamètre pupillaire par soustraction d'une valeur baseline, moyenne des diamètres pupillaires calculée sur la première minute de la phase de jeu. Un filtre Butterworth d'ordre 3 avec une fréquence de coupure à 4 hertz a été appliqué. Les données manquantes ont été interpolées grâce à une spline cubique de second ordre.

5.3.3 Méthodologie d'observation comportementales : accord inter-juges

Le propre de la démarche qualitative est de faire appel à un expert pour comparer, discriminer, trier ou encore qualifier des situations. Cette méthodologie peut nécessiter de faire appel à de multiples experts afin de confronter leurs qualifications. Ainsi, nous avons fait appel à trois experts (ingénieurs et docteurs en cognitique, habitués des observations comportementales) dans le but de contrôler la qualification des situations de collaboration que nous avions initialement effectuée. Chaque expert s'est soumis au même travail d'observation et un dispositif d'accord inter-juges a été mis en place pour vérifier la concordance de nos qualifications. Cet accord inter-juges chiffre, sous la forme d'un coefficient, la qualité de l'accord réel entre des jugements qualitatifs appariés, dont le calcul est basé sur la table de contingence des évaluations.

Pour deux évaluateurs le coefficient utilisé est le Kappa de Cohen (Cohen, 1960), exprimé par une différence relative entre la proportion d'accord observée $\boldsymbol{P_o}$ et la proportion d'accord aléatoire $\boldsymbol{P_e}$ qui est la valeur espérée sous l'hypothèse nulle d'indépendance des jugements, divisée par la quantité disponible au-delà de l'accord aléatoire, soit :

$$k = \frac{\boldsymbol{P_o} - \boldsymbol{P_e}}{1 - \boldsymbol{P_e}}$$

[52] Voir Chapitre 4.3.4

Un accord parfait vaut $\kappa = 1$. Pour plus de deux évaluateurs le Kappa de Light est privilégié (Conger, 1980). Il s'agit de la moyenne de tous les kappas possibles.

5.3.3 Synchronisation des dilatations pupillaires

Lors de notre analyse nous utilisons deux méthodes d'évaluation de la synchronisation des dilatations pupillaires des coéquipiers. D'une part la synchronisation sera estimée par inspection visuelle de la similarité des courbes. Implicitement cette inspection visuelle intègre l'existence de latences, retard dans la similarité d'évolution observée.

D'autre part, et de manière quantitative, nous pouvons mesurer la corrélation entre une série temporelle et une autre, c'est à dire le degré de similarité, en décalant leur alignement par des intervalles croissants. Ce procédé de corrélation croisée (*cross-correlation*) est particulièrement privilégié pour des analyses exploratoires et permet de trouver la similarité maximale entre deux courbes ainsi que le décalage associé. Cet indicateur est calculé en utilisant la fonction *xcorr* de MATLAB® appliquée sur le signal d'évolution du diamètre pupillaire sur la période voulue. Ainsi il est possible d'établir un coefficient de corrélation par dyade (binôme) : J1-J2, J2-J3 et J1-J3. Pour chaque phase de jeu étudiée, 3 coefficients de corrélation des activités pupillaires sont donc calculés.

5.4 Résultats

Pour rappel, nous traitons les questions de recherche suivantes :

(Q1) Le décalage temporel (latences de synchronie) dans l'actualisation des consciences de situation des coéquipiers est-il détectable par l'étude de la dynamique de dilatation pupillaire des individus du groupe ?
(Q2) La collaboration entre deux coéquipiers se traduit-elle par la synchronisation de la dynamique des dilatations pupillaires ?

L'analyse qui en découle se structure en 3 parties. Dans un premier temps nous adressons rapidement la première question de recherche et l'hypothèse H1 en faisant état de la difficulté rencontrée, et attendue, à identifier post-expérimentalement de manière suffisamment claire des situations de partage de perception d'un même élément situationnel.

Dans un deuxième temps nous abordons la seconde question de recherche par les deux approches qualitatives traitant respectivement des hypothèses H2-1 et H2-2 par qualification visuelle de la collaboration et des similarités d'évolution des dilatations pupillaires. Nous effectuons également une analyse approfondie de la collaboration dans ce contexte.

Enfin, dans une troisième et dernière partie, nous présentons une première analyse quantitative via l'utilisation du coefficient de corrélation croisé pour adresser ces mêmes hypothèses de synchronisation des dilatations pupillaires. Nous en profitons pour discuter des perspectives pour une analyse plus approfondie et adéquate de la désynchronisation.

Seules trois des vingt-quatre parties se sont terminées avant la fin du temps réglementaire, l'équipe ayant concédé 500 points de conquête en moins de 15 minutes.

Deux parties souffrent de l'absence de données pupillaires pour un joueur de l'équipe suite à une panne d'enregistrement. Ces parties sont tout de même conservées afin d'exploiter les données des deux autres joueurs.

5.4.1 Latence de synchronie des SA.

Dans le chapitre 2 nous faisons état de l'existence de trois latences d'intérêt pour l'évaluation de la dynamique de partage de la conscience de situation. Notre première question de recherche et l'hypothèse H1 cherchent à adresser l'observation de ces latences. Cependant, ne possédant pas la maîtrise du scénario du logiciel, nous n'avons pu induire et contrôler la perturbation de l'environnement. Nous avons donc cherché à identifier post-expérimentalement des événements marquant l'apparition dans l'environnement d'informations nécessaires aux coéquipiers (NSKE). Par observation des enregistrements vidéos des parties, nous ne sommes cependant pas parvenus à identifier une situation d'étude présentant à la fois l'apparition d'un NSKE et des latences comportementales suffisamment observables. Si quelques marqueurs de perception initiale d'un élément de la situation sont identifiables dans les vidéos, nous n'avons pas pu identifier suffisamment clairement sa prise en compte par un second coéquipier. Par exemple, lors de la partie 12, le joueur 1 remarquant que des ennemis entrent dans la zone delta informe clairement ses coéquipiers. Mais rien dans leur comportement n'a signifié suffisamment lisiblement l'intégration de cette information à leur SA (pas de changement de zone, pas de retour vocal…). Il s'agit pourtant là d'un des exemples de détection d'un NSKE les plus explicites que nous ayons pu identifier.

Nous ne pouvons donc conclure sur l'hypothèse H1 et la question associée.

5.4.2 Similarité des dynamiques de dilatation pupillaire et qualification de la collaboration

Nous nous concentrons maintenant sur la question de recherche Q2 portant sur la qualification de la collaboration et la similarité des dilatations pupillaires. Pour ce faire nous présentons deux approches.

5.4.2.1 *Approche N°1 : de la collaboration à la similarité des dilatations pupillaires*

Par cette approche nous adressons l'hypothèse H2-1 d'après laquelle une collaboration active de deux coéquipiers durant une période donnée se traduit par une similarité d'évolution des diamètres pupillaires durant cette période. Nous cherchons donc dans un premier temps à observer des situations de collaboration entre deux coéquipiers, puis à comparer l'évolution de leurs courbes de dilatation pupillaire sur les mêmes périodes. Parmi les 12 heures de jeu enregistrées, 21 phases de 30 secondes ou 1 minute ont été sélectionnées à partir de l'analyse visuelle des captures d'écran vidéos. 11 périodes ont été sélectionnées car une collaboration active de deux joueurs peut y être observée. Les 10 périodes restantes ont été choisies car les coéquipiers n'y présentent pas de signe flagrant de collaboration.
Cette sélection a fait l'objet d'un contrôle par un procédé d'accord inter-juges. Il a été demandé à un membre du laboratoire, docteur en cognitique, de visionner les vidéos des 21 phases et de déterminer pour chacune d'elles s'il y avait collaboration ou non entre les deux coéquipiers présentés.

Le Tableau 13 présente les qualifications initiales et celle du juge.

Phase	1	2	3	**4**	5	6	**7**	8	9	10	**11**	12	13	14	**15**	16	17	18	19	20	21
Juge 1	1	1	0	1	1	0	0	1	1	0	1	1	0	0	0	1	1	1	0	0	0
Juge 2	1	1	0	0	1	0	1	1	1	0	0	1	0	0	1	1	1	1	0	0	0

Tableau 13. Qualification des situations présentées aux juges 1 et 2. 1 signifie que les deux coéquipiers dont les vidéos ont été observées collaborent sur la période étudiée, 0 signifie qu'ils ne semblent pas collaborer. Les évaluations divergentes sont indiquées en gris.

Le Kappa de Cohen, traduisant la similarité des qualifications entre les deux juges est de k=0.62, ce qui représente un degré d'accord juste à bon au-delà du hasard selon la classification de Fleiss et al. (1982). Les juges s'accordent donc sur la qualification des périodes de collaboration à l'exception de 4 d'entre-elles. Par ailleurs les juges s'accordent sur une certaine ambiguïté dans les comportements observés sur ces 4 phases, ceux-ci étant moins tranchés que pour les autres phases, la qualification a été plus difficile.

Les courbes de dilatation pupillaire correspondantes à ces phases sont présentées sur la Figure 68. La phase N°10 a été écartée de l'étude en raison de données pupillaires manquantes pour un des deux coéquipiers comparés sur cette période. Les 4 phases où un désaccord est présent ont également été écartées de l'analyse qui suit. Celle-ci porte donc sur 16 phases de jeu dont 9 ont été qualifiées de "collaborative" et 7 de "non collaborative".

Afin de faciliter l'observation et la qualification des tendances sur les périodes étudiées, nous fixons empiriquement un filtre à 1/15 Hz, présentant un compromis correct entre gain de lisibilité et fidélité à l'allure générale de la courbe.

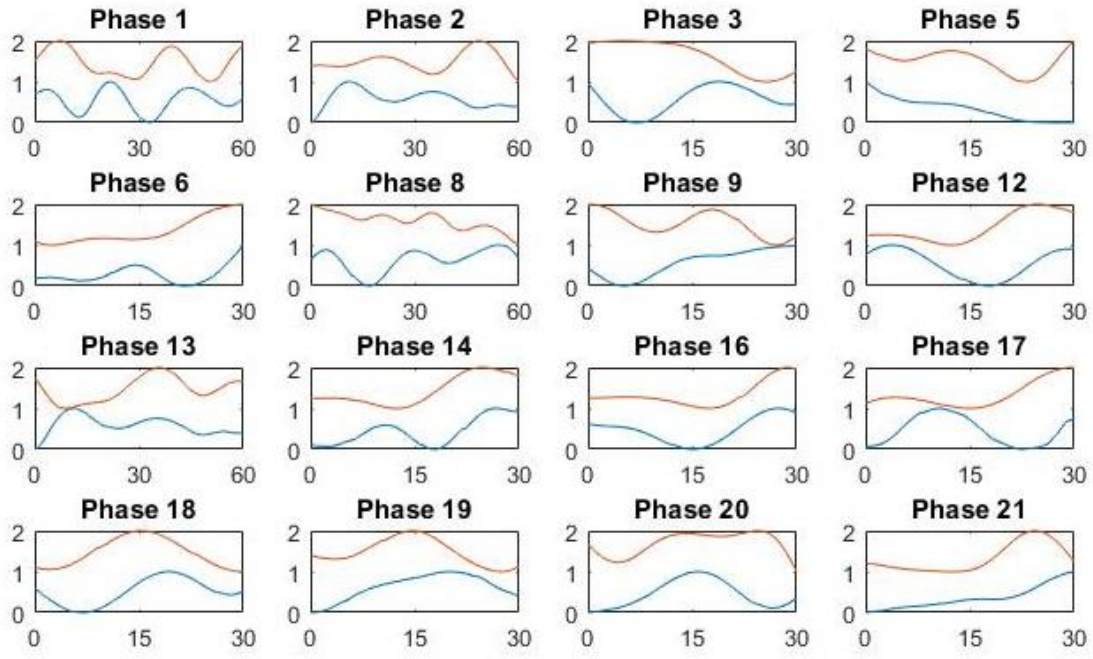


Figure 68. Courbes de dilatation pupillaire des joueurs étudiés sur les 16 phases retenues. Les courbes sont normalisées et espacées verticalement pour plus de lisibilité.

Les résultats montrent une grande disparité des courbes des diamètres pupillaires. Certaines similarités peuvent être mises en évidence. En se basant sur ce critère de similarité visuelle des courbes, les phases ont ainsi été classées en deux catégories : « similaires » ou « non similaires ». Cette classification est alors comparée à la classification précédente de collaboration des coéquipiers afin de déterminer s'il existe une relation entre la similarité d'évolution des activités pupillaires et le niveau de collaboration des coéquipiers évalué par les observateurs. Le Tableau 14 présente la table de contingence de ces classifications.

	Similaires	Non Similaires	Total
Collaboration	4	5	9
Non Collaboration	3	4	7
Total	7	9	16

Tableau 14. Table de contingence des évaluations de Collaboration et de Similarité des courbes

La répartition des qualifications nous permet d'avancer qu'il ne semble pas apparaître de dépendance entre la similarité de la dynamique des dilatations pupillaires et la collaboration entre les individus. Cette observation est corroborée par un test de Fisher, qui ne nous permet pas de rejeter l'indépendance des variables (p-value=1). Nous ne pouvons donc pas conclure quant à l'impact de la collaboration sur la similarité d'activité pupillaire des coéquipiers.

Nous reviendrons sur les limites de cette approche au 5.5.

5.4.2.2 *Approche N°2 : de la similarité des dilatations pupillaires à la collaboration*

Cette seconde approche consiste à effectuer la démarche réciproque de la précédente, en utilisant cette fois-ci la pupillométrie comme variable indépendante et le comportement observé comme variable dépendante. Nous adressons alors l'hypothèse H2-2 selon laquelle une évolution similaire du diamètre pupillaire de deux coéquipiers ou plus durant un certain lapse de temps signifie qu'ils collaborent sur cette période. Dans un premier temps nous sélectionnons 21 périodes par inspection visuelle des courbes de dilatation pupillaire des 24 parties : 7 périodes où deux courbes présentent une évolution similaire, 7 périodes où les courbes des 3 coéquipiers présentent une évolution similaire et enfin 7 autres périodes où aucune des courbes ne se ressemblent visuellement.

Ces courbes sont représentées sur la Figure 69.

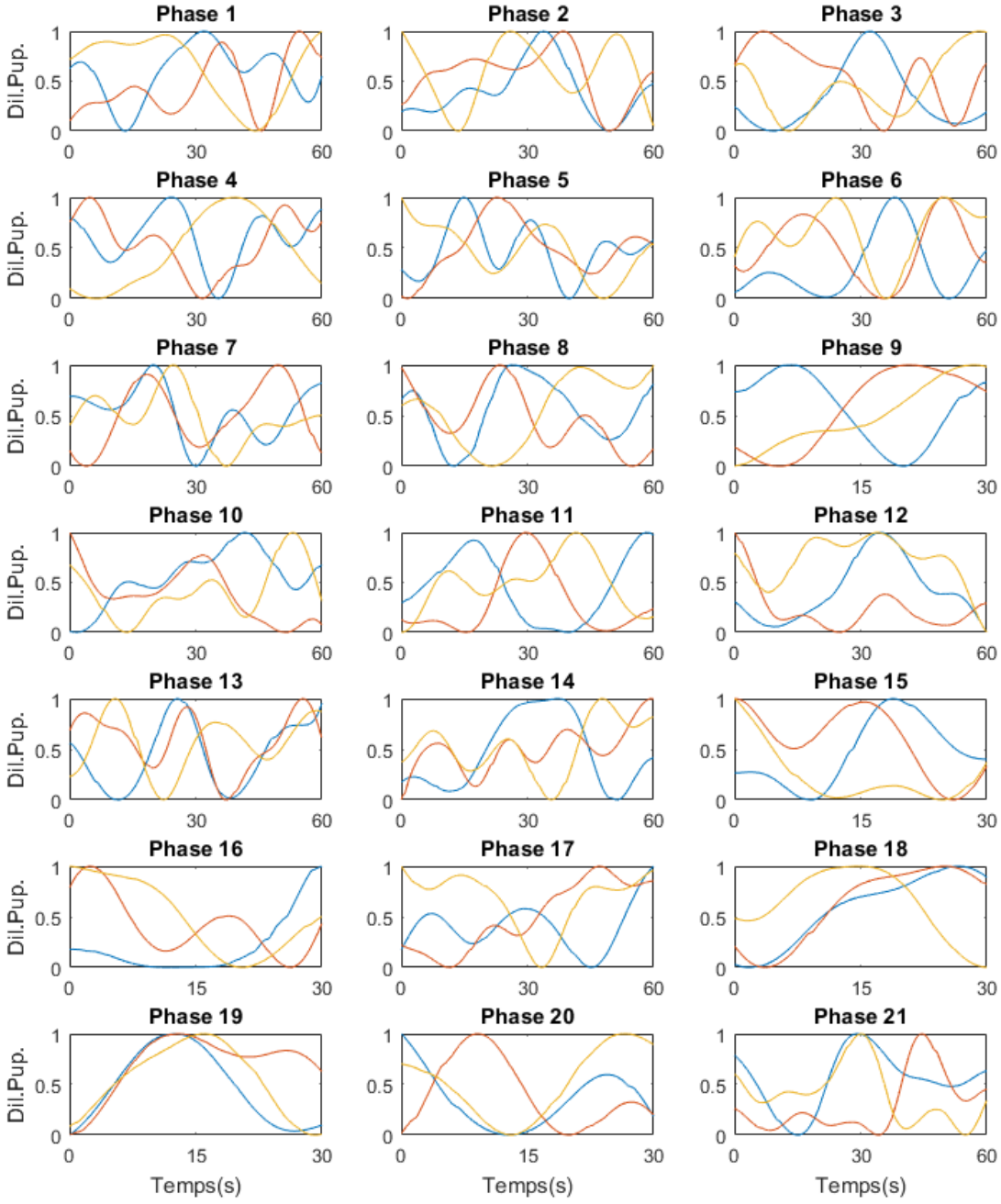


Figure 69. Courbes de dilatation pupillaire pour les 21 phases de 1 minute sélectionnées. Opérateur 1 courbe bleu ; opérateur 2 courbe rouge ; opérateur 3 courbe jaune

Afin de qualifier la collaboration sur ces périodes, nous avons appliqué le même protocole d'accord inter-juges que précédemment. Pour chacune des périodes ainsi sélectionnée 3 experts, ingénieurs et docteurs en cognitique, ont successivement observé les captures d'écran vidéos des 3 coéquipiers. Il leur a été demandé, pour chaque période, d'observer le niveau de partage de représentation de la situation et de qualifier le comportement de l'équipe afin de définir quels coéquipiers semblaient être sur la même longueur d'onde ou collaborer.

Pour cela ils avaient le choix entre 5 qualifications possibles : le joueur 1 collabore avec le 2 ; le 2 collabore avec le 3 ; le joueur 1 collabore avec le joueur 3 ; aucun coéquipier ne collabore ou bien tous collaborent (Annexe 5). Ils disposaient également d'un champ de description libre de leur observation

Cette grille de notation et leurs observations ont été traduites sur une échelle de description de la situation allant de 1 à 5, présentée dans le tableau ci-dessous.

Qualification	Le joueur 1 collabore avec le joueur 2	Le joueur 2 collabore avec le joueur 3	Le joueur 1 collabore avec le joueur 3	Aucun joueur ne semble collaborer	Tous les joueurs semblent collaborer
Notation	1	2	3	4	5

Tableau 15. Table de correspondance des qualifications de la collaboration

Le Tableau 16 présente les résultats des qualifications de la collaboration des trois juges après leur visionnage des situations.

	1	2	**3**	4	5	**6**	7	8	9	10	11	**12**	13	14	15	16	17	18	19	20	21
J1	1	3	**1**	5	5	1	5	1	4	4	5	**5**	3	4	2	4	5	5	5	5	5
J2	1	3	5	5	5	5	5	1	4	4	5	4	3	4	2	4	5	5	5	5	5
J3	1	3	5	**1**	5	5	5	1	4	4	5	4	3	4	2	4	5	5	5	5	5

Tableau 16. Qualification des situations présentées aux 3 juges. La traduction des notations est disponible dans le tableau X. Les 4 périodes où un désaccord est observé sont en gris.

Seules 4 phases présentent des désaccords entre les juges, l'accord semble donc relativement fort. Le Kappa de Light's nous permet de mesurer la concordance entre les trois évaluateurs Il s'agit de la moyenne des Kappa de Cohen pour plus de deux variables catégorielles. Nous utilisons la fonction R *kappam.light()* (Gamer, Lemon, Fellows, & Singh, 2012). Nous observons un excellent accord au-delà du hasard (k = 0.814) selon la classification de (Fleiss et al., 1982).

Malgré cette accord, nous ne constatons une fois de plus pas de corrélation entre cette qualification de la collaboration effectuée par les juges et les synchronisations d'évolution observées sur les courbes des 17 phases pour lesquels l'accord était total. Le Tableau 17 présente la table de contingence associée à ces classifications.

		Similarité des courbes					
		1 J1 avec J2	2 J2 avec J3	3 J1 avec J3	4 Aucun	5 Tous	**Total**
Collaboration	1 - J1 avec J2	2	0	1	0	1	**4**
	2 - J2 avec J3	0	0	0	1	1	**2**
	3 - J1 avec J3	0	0	0	0	0	**0**
	4 - Aucun	0	1	0	2	2	**5**
	5 - Tous	0	0	1	1	4	**6**
	Total	**2**	**1**	**2**	**4**	**8**	**17**

Tableau 17. Table de contingence des évaluations de Collaboration et de Similarité des courbes

Le test de Fisher ne nous permet pas de rejeter l'indépendance des variables (p-value = 0,47 > 0,05). Ainsi d'après cette approche nous ne pouvons pas conclure que la similarité des dilatations pupillaires se traduise par une collaboration accrue des coéquipiers concernés.

5.4.2.3 *Retour sur la qualification de la collaboration*

Nous avons souhaité nous pencher sur l'important accord observé lors de la qualification de la collaboration, cela afin de mettre en perspective les résultats obtenus. Les commentaires recueillis sur les grilles des juges et les échanges post-évaluation nous permettent de proposer les réflexions et remarques suivantes.

Bien que dans la seconde approche les catégories choisies pour la qualification soient plus complètes que celles utilisées lors de la première, et que le nombre de juges soit également plus important, nous ne constatons pas plus de situation de désaccord. Notons que nous n'observons pas non plus de cas de désaccord flagrant où les trois juges auraient une évaluation radicalement différente (par exemple une évaluation de type 1-2-4[53]), ou même de désaccord sur les collaborations deux à deux (aucun juge n'a estimé une collaboration uniquement entre les joueurs 1 et 3 quand les autres juges auraient estimé une collaboration uniquement entre les joueurs 1 et 2, par exemple). Les désaccords observés sont seulement de deux types.

(1) Dans trois cas sur quatre une partie des juges a estimé que seuls les joueurs 1 et 2 semblaient collaborer tandis que le ou les autres juges ont estimé que le troisième coéquipier collaborait également avec eux (désaccord de type 1-5). Dans cette configuration, les juges s'accordent à minima sur une collaboration entre les joueurs 1 et 2, seul le comportement du joueur 3 est sujet à débat.

[53] D'après le grille de notation cela revient à un expert estimant une collaboration entre les joueurs 1 et 2 (1), un deuxième expert estimant lui une collaboration entre les joueurs 2 et 3 (2), et un troisième expert estimant que les joueurs ne collaborent pas (4)

(2) Le dernier cas de désaccord quant à lui concerne une divergence plus large entre une collaboration de tous les coéquipiers versus aucun collaboration détectée (désaccord de type 4-5).

Il est intéressant de noter que les juges s'accordent sur le caractère ambigu de ces cas, illustrant leurs doutes par des points d'interrogation sur la grille ou des commentaires équivoques. Selon eux il s'agit de cas plus difficiles à qualifier car les comportements sont plus variés et moins lisibles que dans d'autres cas. Ils constatent également que les actes de collaboration sont bien moins fréquents ou flagrants mais existent néanmoins. Cela laisse alors plus de place à leur liberté d'interprétation et leur demande une réflexion plus approfondie sur "quel(s) comportement(s) dénote(nt) d'une collaboration ?". Nous avons donc demandé aux trois experts ayant été sollicités de décrire leurs critères de qualification afin d'obtenir une forme d'objectivation de leur évaluation. Nous nous sommes également soumis au même exercice. Les critères verbalisés par les 4 juges sont ainsi consignés dans le tableau ci-dessous.

	Collaboration	Non Collaboration
Juge N°1	Sont-ils en train de communiquer verbalement sur un événement spécifique dans le jeu ? Sont-ils en train de se répartir des rôles ou actions à faire ? Suivent-ils la répartition des tâches établies entre eux (à l'oral) ? S'ils ne communiquent pas, agissent-ils sur la même zone ? S'ils ne communiquent pas, observent-ils le comportement de l'autre (coéquipier) et réagissent en conséquence ? S'ils ne communiquent pas, observent-ils le comportement de l'autre sans forcément agir ?	Pas de communication, concernant les tâches et/ou évènements se produisant à l'écran Chacun réalisant des actions pour gérer sa propre zone, sans communiquer sur celle-ci. Chacun réalisant des actions pour traiter une menace directe sur les troupes qu'eux-mêmes ont déployées dans leur zone, sans communiquer sur celle-ci. Aucun apport d'information oral ou visuel (mouvement de la souris pour indiquer une menace par exemple).
Juge N°2	S'il y a un partage des tâches explicite ou implicite : - Répartition spatiale - chacun son côté - Répartition des rôles S'il y a un partage de l'information, ils donnent de	S'il n'y a aucun des signes précédant, Si les actions individuelles sont contradictoires, tout le monde dans la même zone et/ou avec le même rôle Si la communication n'est pas

	l'information ou ils en demandent S'il y a des prises de décision collectives : - soit parce qu'il y a consensus - soit par qu'il y a un leader et que les autres acceptent de suivre le leader	sur la tâche : ils s'énervent, critiquent, s'insultent, ou parlent de sujets sans lien.
Juge N°3	Il y a un échange d'informations entre les coéquipiers et que cette communication porte sur : - leurs actions actuelles ou celles des coéquipiers - un événement du jeu et qu'une réaction (verbale ou comportementale) à cette communication est observable. - leur organisation, ils se répartissent les tâches et discutent des stratégies et actions à entreprendre dans un futur proche Leur attention est portée sur le même zone à une échelle à peu près similaire Ils semblent partager une compréhension des urgences liées à la situation et agir de façon cohérente en fonction Ils s'occupent de situation ou d'événements communs, s'apportent du soutien	Le joueur est isolé sur une même situation ou zone géographique durant longtemps sans prise de recul, communications, ou prise d'information au niveau du groupe. Les communications divergent de la tâche et dénotent d'un désengagement. Il y a un déséquilibre dans leurs niveaux d'activité, ne semblent pas répondre aux même urgences. De grandes actions (par exemple capturer une zone ou attaquer un ennemi) sont prises individuellement et sans en informer le groupe

	Ils suivent une stratégie préalablement discutée et consentie par l'équipe	
Juge N°4	Visualisation de la même zone ou du même ennemi	Visualisation de deux zones ou ennemis bien distincts
	Actions coordonnées (déplacement de troupes, attaques, etc.) dans le but de mener à bien une tâche commune	Actions « solitaires » : suite à une répartition organisée ou tacite, les joueurs ont chacun leur tâche
	Communication orale d'objectifs/d'informations directement en lien avec la tâche du coéquipier (qui est en cours ou doit être stoppée pour une autre tâche prioritaire)	Pas de communication orale concernant la tâche du coéquipier ou celle qu'il devrait/devra effectuer

Tableau 18. Critères de qualification de la collaboration utilisés par les experts.

La façon dont les opérateurs traitent l'information, c'est à dire la manière dont ils la perçoivent, l'interprètent et la partagent ainsi que les actions individuelles et collectives qui en découlent, apparaît comme le principal point commun à tous ces critères. Par l'analyse conjointe des communications et des comportements, les juges estiment un niveau de cohérence et de cohésion du travail de l'équipe. Par exemple, la fréquence, le contenu et le sens des échanges permettent de déterminer la dynamique de partage d'information dans laquelle se trouvent les individus. L'analyse des comportements permet quant à elle d'estimer la perception des éléments de la situation et l'interprétation qui a pu en résulter. En observant les actions menées par les joueurs ainsi que les zones et les événements traités ou ignorés, les juges appréhendent les stratégies cognitives individuelles et identifient la dynamique attentionnelle de chacun. On retrouve ici une évaluation implicite du partage de conscience de situation des joueurs, par ailleurs illustrée par des commentaires comme « ne se comprennent pas », « n'a pas relevé l'incident » ou « ne réagit pas » lors de l'évaluation, ou encore des critères tels que « Ils semblent partager une compréhension des urgences liées à la situation ».

Les opérateurs sont plongés dans une situation collaborative par essence. Il peut donc paraître étrange de qualifier des situations où aucun coéquipier ne collabore. Certes la dynamique de l'activité varie entre des tâches communes et des tâches individuelles, néanmoins ces dernières sont réalisées avec une perspective et un objectif commun au groupe, ce qui rejoint toujours la définition de la collaboration comme forme intégrative de la coopération (Schmidt, 1991).

Ceci peut cependant expliquer pourquoi les juges ont été déstabilisés par la catégorisation tous/aucun. Les individus travaillant à un but commun dans un espace virtuel partagé, il existe donc toujours un certain niveau de collaboration entre eux. En revanche cette collaboration est ponctuellement plus ou moins accentuée, concentrée par un événement ou une stratégie. C'est cette concentration temporaire de l'activité commune qui a été évaluée, et qualifiée comme un niveau plus ou moins « actif » de

collaboration. La collaboration qualifie alors une période durant laquelle les individus présentent des marqueurs comportementaux et verbaux structurants un partage accru d'information et traduisant une compréhension de la situation similaire et cohérente de la part des coéquipiers.

Les qualifications réalisées lors de nos deux approches sont de fait sensibles à cette définition.

5.4.3 Ébauche d'approche quantitative de la similarité.

Le travail des juges nous fournit une qualification de la collaboration pour chaque dyade (couple de joueur ; J1-J2, J2-J3, J1-J3) de chaque période étudiée lors des deux approches qualitatives. Pour chacun de ces couples de joueurs il a donc été estimé qu'ils collaboraient ou non. Nous construisons ainsi deux modalités : « Collaboration » et « Non collaboration ».

Nous effectuons cette étude sur les 32 périodes où tous les experts s'accordent sur la qualification de la collaboration (16 périodes pour l'approche N°1 et 17 périodes pour l'approche N°2).

Nous souhaitons comparer la similarité des dilatations pupillaires en fonction de la collaboration ou non des coéquipiers. Ainsi pour chacune des 67 dyades de joueurs, le coefficient de corrélation croisée maximum de leurs dilatations pupillaires sur la période donnée a été calculé. Ce coefficient exprime une valeur de similarité des formes des courbes de dilatation.

Le Tableau 19 décrit la distribution de la corrélation croisée des dyades pour les deux modalités : « Collaboration » et « Non Collaboration »

Cross-correlation

	Min	Q1	Médiane	Q3	Max	Moyenne	Ecart-Type	Nb
Collaboration	158,0	229,4	334,4	482,9	798,9	376,5	169,0	38
Non Collaboration	110,0	206,1	305,8	414,3	849,3	347,8	184,3	29

Tableau 19. Tableau comparatif des distributions de la Précision moyenne des équipes en échec et en réussite.

Cette distribution est présentée sur la Figure 70.

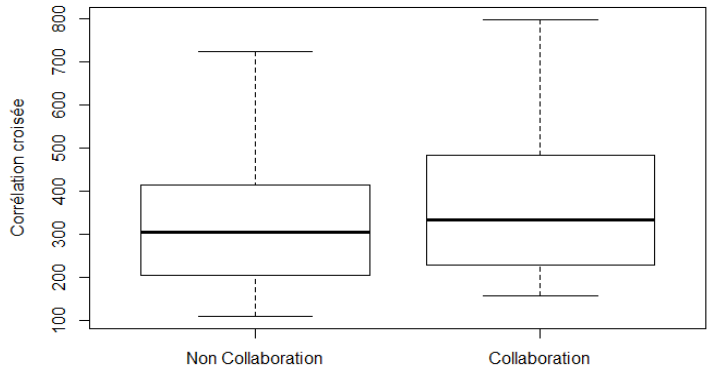


Figure 70. Comparaison boxplot de la corrélation croisée des dilatations pupillaires des dyades présentant une Collaboration et celle ne collaborant pas.

Bien qu'en situation de collaboration les dyades présentent une corrélation croisée moyenne légèrement supérieure, ainsi qu'une dispersion plus importante dans les grandes valeurs, aucune différence significative n'est mise en évidence. Cette méthode ne nous permet donc pas non plus de conclure quant à la dépendance de la similarité des dilatations pupillaires des coéquipiers à leur situation de collaboration.

5.5 Discussion des résultats

Pour ce chapitre expérimental nous avons choisi de présenter la première étude menée durant ce doctorat. Nous expliquons cela par la situation expérimentale particulière, plus complète au regard de notre sujet. En effet nous avons pu mettre trois coéquipiers en situation de travail collaboratif suffisamment réaliste, sur une tâche relevant du domaine du commandement et contrôle d'opérations militaires, tout en leur appliquant un monitoring physiologique.

Cependant des contraintes principalement matérielles et logicielles ont compliqué l'étude de la dynamique du partage de conscience de situation au sein de l'équipe. Nos résultats ne permettent pas de répondre à la question Q1.

La question du partage de conscience de situation a pu néanmoins être abordée par le biais de la qualification de la collaboration lors des deux approches qualitatives présentées. Nous y examinons les hypothèses H2-1 et H2-2, qui supposent une relation linéaire entre la similitude de la forme des courbes de dilatation pupillaire des coéquipiers et leur niveau de collaboration. Cependant aucune des deux approches n'a permis de mettre en évidence une telle relation, et cela malgré tous les calculs et traitements opérés sur le signal. Nous n'avons pas eu plus de succès en utilisant le coefficient de corrélation croisée. Ces résultats ne nous permettent donc pas de valider nos hypothèses. Il conviendra probablement d'utiliser des scénarios plus simples, ou du moins maitrisés, mais alors plus déconnectés de la réalité opérationnelle.

Néanmoins la qualification commune de la collaboration ouvre un domaine de réflexion intéressant. Il est apparu que les critères intuitivement utilisés par les experts s'apparentent à des critères d'évaluation de la cognition collective, relevant de la quantité de partage d'information, du partage de conscience de situation et de synchronisation des actions.

En situation de travail collaboratif, l'état de l'opérateur est apparu avant tout sensible aux événements, actions et intentions entreprises dans son espace d'action au sein de sa stratégie personnelle. Le partage d'information par un coéquipier, ce que nous avons nommé la méthode indirecte de synchronisation des consciences de situation (voir 2.2.2), doit néanmoins être soumis à son interprétation. La situation de collaboration demande un prétraitement de chaque nouvelle information pour déterminer s'il s'agit d'une information nécessaire aux coéquipiers et si oui, décider d'un moment et d'une méthode de partage de cette information. De même chaque information reçue d'un coéquipier doit être analysée au regard de son importance par rapport à la tâche individuelle en cours et à la situation collective pour décider d'un moment de réaction opportun.

Comme lors d'une tâche individuelle, les absences de réaction que nous avons pu observer peuvent être dues à une absence perception (erreur de SA de niveau 1), à une décision erronée provenant d'une mauvaise interprétation de l'information (erreur de SA de niveau 2 et 3), ou bien d'une décision délibérée due à la priorisation personnelle de la tâche. Néanmoins la provenance de l'information en elle-même - un coéquipier - peut être un indicateur d'importance de l'information. Entre alors en jeu la question de la confiance envers le coéquipier et des habitudes de collaboration.

Inversement nous avons constaté des situations de collaboration spontanées issues de l'observation active de l'environnement par les coéquipiers, illustrant à nouveau la diversité des comportements collaboratifs au sein d'une équipe. Le traitement de l'information et l'importance qui lui est accordée varie nécessairement d'un coéquipier à l'autre. Le même événement peut donc tout à fait résulter en deux traitements différents et donc deux réactions physiologiques ou comportementales différentes. Il s'agit là d'un point critique pour la recherche de similarité des réactions et l'interprétation des comportements d'équipe.

Notons par ailleurs qu'une critique majeure peut être formulée envers cette étude, résultant de l'absence de matériel nécessaire. La synchronisation des données a dû être faite manuellement après l'expérience. Tout résultat relatif à la comparaison des évolutions et corrélations des dilatations pupillaires dépend donc nécessairement de la qualité de cette synchronisation manuelle. Ce risque ne serait pas présent avec un système adéquat de centralisation des acquisitions, mais hors de portée des marges financières du laboratoire.

Nous avons pu observer au Chapitre 4 que la réaction à la détection visuelle d'un nouvel élément situationnel se retrouvait souvent noyée dans le bruit comportemental. En situation de collaboration, ce bruit se trouve exacerbé et la prévalence des éléments situationnels communs s'en trouve réduite.

Dans ce contexte, il paraît de moins en moins probable que la perception seule d'un NSKE soit observable via la dilatation pupillaire. Dans un environnement si complexe, où les sollicitations sont très nombreuses et de nature variées, il paraît d'autant plus compliqué de s'attendre à pouvoir détecter, en tout cas visuellement, cette réaction. Il convient donc de réfléchir à des méthodes plus appropriées. Nous considérons la qualification de la collaboration assez fiable et adéquate car validée par des experts indépendants, ayant procédé séparément. La question se pose donc de la méthode de mesure de la synchronisation des dilatations pupillaires.

Il apparaît dans les recherches récentes, que des méthodes aussi simples que la corrélation croisée ou la corrélation de Pearson ne sont pas non plus les plus adaptées à la complexité du signal physiologique et trop sensibles aux caractéristiques intrinsèques du signal (Dean & Dunsmuir, 2016). Afin de discuter nos résultats nous proposons maintenant des pistes d'analyses susceptibles d'être plus performantes.

Tout d'abord, nous souhaitons mentionner une méthode actuellement en développement au laboratoire LOMA (Laboratoire Ondes et Matière d'Aquitaine) de l'Université de Bordeaux, visant à détecter les corrélations entre les modifications de comportement de signaux physiologiques. Cette méthode repose sur une analyse de séries à la fois dans le domaine temporel et dans le domaine fréquentiel en combinant une transformée en ondelette croisées (*Cross Wavelet Transform* - XWT) et une analyse des cohérences (Cazelles et al., 2008; Grinsted, Moore, & Jevrejeva, 2004). L'analyse en ondelette s'avère être un outil utile pour analyser des séries temporelles présentant de nombreuses échelles de temps différentes ou des fluctuations de la variance (Torrence & Compo, 1998). Le but de cette analyse est d'identifier les périodes et localisations pour lesquelles deux séries auraient des variations communes (produit positif) ou opposées (produit négatif). Le calcul de la cohérence s'apparente ensuite à au calcul d'un "coefficient de corrélation" à la fois temporel (un déphasage) et en amplitude. L'utilisation de cette méthode n'est malheureusement pas compatible avec les contraintes temporelles de cette thèse. Néanmoins, le jeu de données sera utilisé par les scientifiques concernés. Nous avons tout de même réalisé un test sur les données issues de la partie n°2 de l'équipe A (Figure 71), avec l'aide d'Alexandre Guillet, doctorant travaillant sur cette méthode au sein de l'équipe Matière Molle et Biophysique du LOMA.

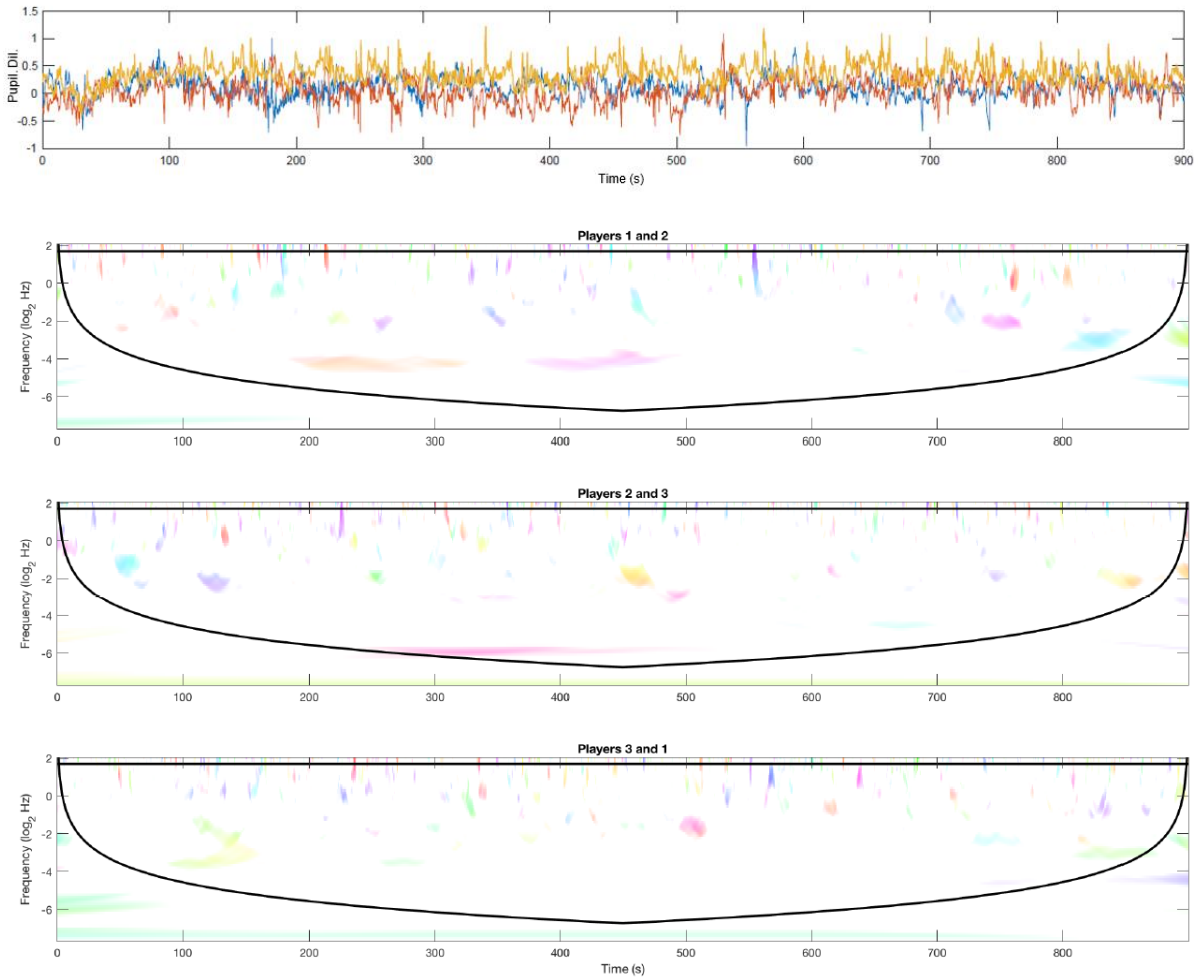


Figure 71. Exemple d'analyse des synchronisations des dyades par transformée en ondelette croisées et analyse des cohérences. Sur la première ligne est représentée l'évolution des dilatations pupillaires des 3 coéquipiers (Joueur 1 en bleu, joueur 2 en rouge, joueur 3 en jaune) de l'équipe A durant la partie n°2. Les trois dernières lignes représentent le résultat de l'analyse pour chaque dyade. Les taches colorées correspondent à une cohérence (localisée à un moment et une échelle de fréquence donnée) qui excède les 95% de signifiance. La couleur correspond au déphasage de ces composantes cohérentes (vert = 0, rouge = +/- 3s)

Sur les trois dernières lignes de la Figure 71, chaque tache de couleur correspond à une période de synchronisation détectée entre deux signaux (ici deux individus). De nombreuses petites tâches peuvent être observées dans les hautes fréquences, traduisant de périodes de synchronisation ponctuelles, possiblement dues à la gestion d'un évènement en commun. Les tâches plus importantes et étalées, observées dans les basses fréquences, représentent des synchronisations sous-jacente traduisant, par exemple, une collaboration passive des individus sur ces périodes.

Ce test nous amène à penser que, si elle s'avère fiable, une telle méthode pourrai permettre d'identifier des périodes de cohérence dans l'activité pupillaire des coéquipiers et de quantifier l'intensité de la collaboration entre les individus. Appliquée à différents signaux physiologiques individuels et interindividuels elle peut représenter un axe intéressant pour l'étude de la synchronie des consciences de situation et des latences associées dans un contexte de monitoring multi modal.

Cependant, ces analyses dans le domaine du temps et des fréquences reposent sur le postulat d'une relation linéaire entre les variables. Or, des preuves de plus en plus nombreuses soulignent l'existence d'une relation non-linéaire. Une telle relation n'est pas visible avec de tels outils et requiert l'utilisation d'outils mathématiques et de transformation du signal plus complexes. Pour cela les systèmes dynamiques non linéaires (SND) sont une approche théorique et méthodologique visant à comprendre les systèmes complexes et les relations dans et entre les sous-systèmes. Ils comportent un ensemble de formalismes mathématiques qui peuvent être utilisés pour comprendre l'évolution dans le temps des systèmes physiques, comportementaux et cognitifs (Stevens, Gorman, Amazeen, Likens, & Galloway, 2013). Entre autres, des indicateurs tels que l'entropie (Dias, Zenati, Stevens, Gabany, & Yule, 2019), ou des outils tels que la *Detrended fluctuation analysis* ou l'exposant de Lyapunov s'avèrent être adaptés à l'étude de la complexité chaotique de certains signaux physiologique autorisant, par exemple d'étudier le rôle du système nerveux autonome sur le contrôle du rythme cardiaque (Costa, Goldberger, & Peng, 2002; Goldberger et al., 2002). Citons également l'analyse de quantification croisée (*Cross-Recurrence Quantification Analysis* - CRQA) , technique d'analyse non linéaire qui quantifie le couplage entre deux signaux, et qui a été largement utilisée pour quantifier la dynamique comportementale et physiologique dans des contextes interpersonnels (Mønster, Håkonsson, Eskildsen, & Wallot, 2016). Ainsi, le CRQA fournit des mesures de la synchronisation entre les interactions. Il paraît alors pertinent de réfléchir à l'applicabilité de ces outils à l'étude de la dynamique pupillaire et à la détection de synchronisation des consciences de situation.

Pour terminer je souhaite revenir sur la place qu'a occupée cette expérience et son impact sur le travail de thèse. Cette expérience nous a beaucoup appris. Elle nous a tout d'abord permis de faire l'état des lieux des besoins pour expérimenter sur la conscience de situation, sur la cognition collective et via le monitoring des opérateurs, ainsi que des outils à notre disposition. Les limites rencontrées au cours de cette étude ont structuré la suite de nos travaux et ont, en particulier, motivé les contextes de recherche des chapitres 3 et 4. Cette expérience, au sens vécu et scientifique, nous a permis de repenser l'approche de notre problématique en nous a poussant à développer la collaboration avec le laboratoire allemand FKIE et l'expérience Explore Mars. Tout cela a indéniablement nourri les réflexions présentées dans le prochain chapitre, notamment concernant les besoins expérimentaux pour l'étude de la cognition collective par monitoring.

Enfin, cette expérience fut l'occasion de nous perfectionner à l'acquisition et au traitement des données pupillaires. Malgré les difficultés techniques rencontrées, les données ont pu être nettoyées, synchronisées et traitées. Elles constituent un jeu de données propre, prêtes pour une analyse plus approfondie.

Nous avons utilisé deux types de méthodes d'analyse, numériques et comportementales, qui ne nous ont pas permis d'obtenir des résultats satisfaisants. Nous avons cependant identifié des méthodes nouvelles dont l'application fera éventuellement l'objet d'un post-doctorat.

Chapitre 6 Contextualisation et perspectives

À partir de notre démarche, nous devons contextualiser et évaluer l'applicabilité de nos recherches au C2 en tant qu'environnement de travail collaboratif. Pour cela, nous nous appuierons sur la doctrine de l'Armée de l'air et de l'espace française ainsi que sur les formations et les observations réalisées au CASPOA et au CEAM. De nos travaux de thèse, nous avons tiré des enseignements concernant l'étude de la cognition collective en situation de travail collaboratif réaliste. Cette expérience nous amène à formuler des recommandations et à proposer les prochaines étapes de cette recherche. Il nous semble enfin important d'aborder la place de la cognition collective et de son étude dans le futur de la défense et de l'augmentation du C2.

6.1 L'opérateur en centre de commandement et contrôle

Au Chapitre 1 nous avons présenté le C2 en tant que machine conceptuelle visant à traduire des ambitions stratégiques en actions et effets. Au cours des chapitres 3 à 5, nous avons inclus ce contexte à nos situations expérimentales par le choix de tâches relevant de la conduite d'opérations ou du contrôle aérien.

L'immersion dans un C2 opérationnel a été une étape incontournable afin de mieux appréhender les problématiques et processus mis en place dans le cadre du travail collaboratif en C2. Celle-ci a été rendue possible grâce au soutien de le Direction générale de l'armement, qui a notamment permis l'obtention de l'habilitation requise pour avoir accès aux bases aériennes. Compte tenu des partenaires de l'École Nationale Supérieure de Cognitique et de cette thèse[54], nous avons privilégié le contexte de l'Armée de l'air et de l'espace (AAE). Ainsi, deux structures ont pu être impliquées : la salle de simulation JFAC du CASPOA Lyon Mont Verdun et le Centre de détection et de contrôle (CDC) de la base aérienne 118 de Mont-de-Marsan.

[54] Principalement par le biais des conventions avec l'Armée de l'air et l'espace (AAE) et avec l'*Allied Command Transformation* de l'OTAN.

Les connaissances théoriques concernant l'organisation des C2 sont issues de la formation qualifiante JFAC OC-B que j'ai suivie au CASPOA, ainsi que de la doctrine[55] de l'AAE et de l'OTAN[56].

Les deux structures C2 dans lesquelles nous avons effectué des observations et des entretiens sont propres aux deux principales missions confiées par l'Etat à l'AAE. Le Commandement de la défense aérienne et des opérations aériennes[57] (CDAOA) s'appui d'une part sur les centres de détections et de contrôle (CDC) répartis sur le territoire. Ceux-ci sont chargés de la surveillance et de la défense de l'espace aérien français. Des observations et des entretiens avec les opérateurs du CDC de la BA-118 de Mont-de-Marsan nous ont permis d'appréhender l'organisation et la place de l'opérateur dans une telle structure.

D'autre part, le commandement et la conduite des opérations aériennes extérieures (hors du territoire) dans le cadre de missions, sont assurées par les JFAC (*Joint Force Air Command*). Ces structures sont implantées dans plusieurs pays membre de l'OTAN, dont la France, le Royaume-Uni et l'Allemagne. Le JFAC opérationnel français est situé sous la BA-942 Lyon Mont Verdun. Au-delà des salles opérationnelles, la salle de JFAC à laquelle nous avons eu un accès facilité est un espace d'entrainement du CASPOA destiné à la formation des futurs opérateurs. Cette salle reprend la structure habituelle d'un JFAC à une échelle réduite, adaptée à sa vocation de salle de formation. Elle comporte une vingtaine de places (pour comparaison le JFAC opérationnel français dispose d'environ 100 postes).

6.1.1 Le poste de l'opérateur en JFAC

La doctrine DAA-3.3_COA (2017) de l'armée de l'air concernant le Commandement des opérations aériennes donne la définition suivante des différentes composantes du C2 Air :

« Le C2 Air constitue l'une des parties terminales du commandement des opérations, lequel s'étend du niveau stratégique aux niveaux tactiques. Dans la continuité du C2 opératif qui conçoit et planifie les effets, le C2 Air planifie et réalise les actions nécessaires à l'atteinte de ces effets. »

[55] Dans le domaine de la défense, une doctrine est un document cadre dont le contenu sert de référence commune, donne à la réflexion un cadre analytique rigoureux et contribue à définir un langage et des méthodes partagées par tous ceux qui ont pour tâche d'élaborer ou d'exécuter des plans, des missions ou des ordres.

[56] Cette dernière m'a été présentée par le lieutenant-colonel Emmanuel Latasse, commandant en second du CEAM ayant co-dirigé la rédaction de la version de 2017.

[57] Le CDAOA est l'un des quatre commandements par nature de forces de l'armée de l'air français. Il planifie, conduit et coordonne tout type de missions aériennes sur le territoire national, depuis le territoire national et en opérations extérieures. Le CDAOA est situé sur la base aérienne BA-942 Lyon Mont Verdun.



Figure 72. Les différentes structures de la conduite des opérations aériennes en France (*Doctrine Armée de l 'air DAA-3.3 « commandement des opérations aériennes »*, 2017)

Le commandement des opérations aériennes nécessite d'abord une phase de planification opérationnelle, puis une phase de conduite des opérations aériennes. Celle-là désigne l'ensemble des processus C2 de cette structure, c'est-à-dire la planification de mise en œuvre, la programmation, l'exécution et l'évaluation des actions militaires.

Si les étapes de planification et de programmation sont également réalisées par des opérateurs de C2, elles se déroulent dans un contexte temporel moins cointraint. Le monitoring des opérateurs présente donc un intérêt plus particulier dans la phase de *current ops* (opérations en cours), au cours de laquelle les opérateurs doivent réagir et gérer en temps réel des situations de crise sous forte contrainte temporelle. Il s'agit là de la partie « contrôle » de la conduite des opérations aériennes, comme défini par la doctrine citée.

Lors d'opérations extérieures (hors du territoire français), la conduite des opérations aériennes est assurée par le JFAC dont la division *combat ops* s'occupe du commandement et de la conduite des opérations en cours. Celle-ci est structurée en 7 cellules dirigées par le chef de salle. La salle de formation JFAC du CASPOA reprend cette organisation (Figure 73). Autour du chef de salle et de son assistant, généralement situés au centre, se trouvent les 7 cellules suivantes :

- la cellule « Offensive » s'occupe des actions visant à perturber ou à dégrader les capacités de l'adversaire, voire à les détruire ;
- la cellule « Défensive » a pour mission de coordonner la protection des forces alliées et des infrastructures vitales pour la mission ;
- la cellule « Support » est en charge, notamment, de la logistique, de la gestion du ravitaillement en vol des vecteurs opérationnels (avions, hélicoptères…) et de l'organisation des évacuations médicales ;
- la cellule « Renseignements », à proximité du chef de salle, s'occupe de la collecte et du traitement d'informations concernant l'environnement, permettant d'identifier les menaces et d'offrir des possibilités d'exploitation aux décideurs (NATO Standardization Agency, 2016) ;

- la cellule « Legad » (Legal Adviser[58]) est composée de personnel expert en droit des conflits armés, et auquel se réfère le commandeur pour assurer son action dans la stricte limite du droit international ou, le cas échéant, national ;
- la cellule « Joint Personnel Recovery[59] » (JPR sur la Figure 73) coordonne la récupération et la réintégration du personnes qui se retrouveraient isolées, militaires ou civiles. Elle entretient des liens étroits avec la cellule Support, notamment pour gérer les évacuations médicales ;
- la cellule « Time Sensitive Targetting[60] » (TST sur la Figure 73) est en charge d'identifier les opportunités pour les cibles à fort potentiel.

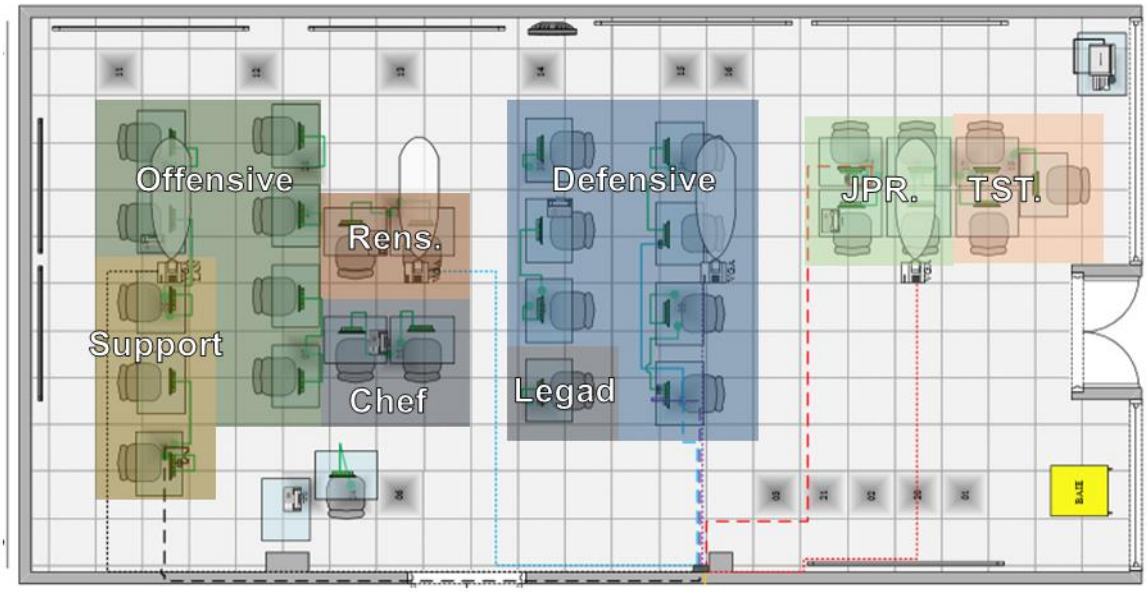


Figure 73. Organisation spatiale de la salle JFAC du CASPOA. Le nombre de poste attribué à chaque cellule est donné de manière indicative, le propre de cette salle de formation étant de pouvoir adapter sa structure. Dans le cadre des exercices de formation du CASPOA, les cellules JPR et TST sont tenues par des experts, d'où leur position excentrée sur le plan qui les sépare du personnel en formation.

Un JFAC est traditionnellement organisé en travées ou ilots identifiés. Il s'agit de postes de travail individuels voisins, sur une ou plusieurs rangées. Chacune des cellules décrites dispose d'une position dédiée. Il est d'ailleurs courant de se référer à la « travée Offensive » plus qu'à la « cellule Offensive ».

Le poste de travail de l'opérateur se compose de deux écrans disposés face à l'opérateur sur un bureau personnel. S'y ajoutent de nombreux écrans muraux. Ces dispositifs partagés de visualisation sont sensés servir de support et à faciliter la construction d'une représentation commune de la situation. La salle de formation JFAC du CASPOA en dispose de 4, sur lesquels sont projetées d'une part la situation aérienne (une carte, écran mural de droite sur la Figure 74), d'autre part les tâches en cours et à venir (écran mural de gauche sur la Figure 74).

[58] Conseiller juridique
[59] Récupération du personnel interarmées
[60] Ciblage dynamique



Figure 74. Salle de formation JFAC du CASPOA. Source CASPOA

Dans la majorité des cas, les communications entre opérateurs ou travées reposent sur un système électronique de discussion instantanée (tchat). Il existe un groupe de discussion par travée, ouvert à tous les opérateurs du JFAC, ce qui représente 8 à 10 fenêtres de tchat différentes. Le tchat assure un rôle de communication mais aussi de traçabilité des échanges d'information et permet aux opérateurs d'y accéder au moment qu'ils jugent le plus opportun.

L'environnement de travail, plongé dans la pénombre, est généralement silencieux. Les communications verbales par téléphone sont généralement réservées au contact avec l'extérieur (autres C2 ou autres structures telles que les centre de contrôle aérien civil, les commandements opérationnels d'autorités extérieures, etc.). Les échanges entre opérateurs d'une même travée demeurent cependant assez spontanés et la promiscuité induit certains comportements de partage. Les opérateurs peuvent ainsi avoir tendance à regarder l'écran du voisin afin de récupérer une information sans le déranger ou à se déplacer pour aller consulter directement un autre opérateur ou le chef de salle. Ces échanges oraux et ces déplacements sont beaucoup plus importants entre voisins de travée qu'avec le reste de la salle. Néanmoins, si un événement exceptionnel le requiert, par sa gravité ou son urgence, il arrive qu'un individu se lève et s'adresse oralement à sa travée ou à l'ensemble de la salle. Ce comportement favorise le transfert direct et immédiat de l'information, améliorant ainsi le partage de conscience de situation concernant l'élément concerné.

Notons que la salle observée est une salle de formation et qu'aucun opérateur de JFAC n'a pu être interviewé pour des raisons de confidentialité. De fait, dans une salle plus grande avec des opérateurs ayant un niveau d'expertise plus élevé, les comportements décrits ici peuvent être différents, et on observe d'ailleurs des comportements auto-organisés, variant d'ailleurs à la marge avec la culture nationale des opérateurs.

6.1.2 Le poste de l'opérateur en CDC

Pour mener à bien ses missions de contrôle de l'espace aérien, l'AAE dispose en France de trois centres de détection et de contrôle (CDC) opérationnels (Mont-de-Marsan, Cinq-Mars-la-Pile et Lyon Mont Verdun) et d'un mis en sommeil depuis 2015 (CDC de Drachenbronn) mais pouvant être réactivé. Le contrôle de l'espace aérien de la métropole, découpé en 4 zones, est réparti entre ces centres (Figure 75). Le CDC de Mont-de-Marsan est ainsi chargé du contrôle de la zone aérienne Sud-Ouest. Le CDC de Cinq-Mars-la-Pile est en charge des espaces aériens Nord-Ouest et Nord-Est (initialement à la charge du CDC de Drachenbronn). Enfin le CDC de Lyon Mont Verdun contrôle l'espace aérien Sud-Est.

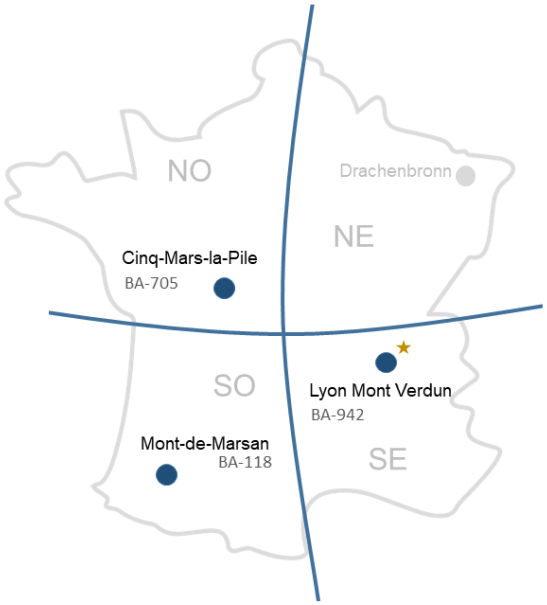


Figure 75. Centres de détection et de contrôle (CDC) du territoire français et répartition des zones aériennes

Lorsqu'un aéronef pénètre dans la zone de responsabilité du CDC, un opérateur lui associe un indice qualifiant son comportement (normal, anormal ou dangereux). La mission classique d'un CDC consiste à s'assurer qu'aucun aéronef dans sa zone ne présente de comportement non conforme (par exemple une déviation de son plan de vol, une absence d'identification correcte ou la pénétration dans une zone interdite).

Généralement les opérateurs du centre et les membres de l'aéronef collaborent afin de comprendre et de résoudre rapidement la situation. Cependant il peut arriver que les opérateurs du CDC ne puissent gérer seuls la situation (par exemple si l'aéronef ne coopère pas ou a un problème de communication directe). Ceci peut être le cas pour un aéronef tentant de pénétrer dans une zone interdite, par exemple volant à basse altitude vers une centrale nucléaire. Lorsqu'un opérateur constate qu'un appareil se dirige vers une telle zone interdite à un niveau inférieur au plafond minimum autorisé, il prévient le chef de travée et lui fait part de la situation. Celui-ci sera en charge d'appeler le Centre national des opérations

aériennes[61] (CNOA). Une classification particulière sera accordée à l'aéronef ayant un comportement anormal. En fonction du déroulé de l'opération l'appareil changera sa route, sortira de la zone en question et l'alerte cessera. Dans le cas contraire le CNOA pourra donner l'ordre de déclencher la PO[62] et d'intercepter l'appareil en question, comprendre ses intentions (détresse, panne, méconnaissance de la zone, pénétration volontaire…) et agir s'il le faut. L'intervention de la PO peut aller jusqu'à la destruction de l'appareil délinquant sur décision des hautes autorités[63]. Pour effectuer ce type de mission, une communication active est nécessaire entre les opérateurs de différents centres (CDC et CNOA) qui collaborent afin de réaliser une tâche donnée.

Similairement à un JFAC, un CDC militaire est composé de plusieurs travées divisées en plusieurs postes de travail. L'organisation concentrique y est cependant plus marquée. La travée extérieure comprend les postes des opérateurs et des contrôleurs : les « cabines ». La travée intermédiaire est celle des chefs contrôleurs et des chefs opérateurs. Enfin, la dernière travée est celle du chef de salle. Cette organisation spatiale traduit la structure hiérarchique de l'équipe.

Les opérateurs de CDC en charge du contrôle aérien, aussi appelés « contrôleurs »[64], s'occupent de la gestion des missions d'interception et des entraînements dans l'espace aérien du CDC. En l'absence de Liaison 16[65], le champs de détection visuelle des pilotes ne couvre que 60°. Les contrôleurs sont donc généralement en lien direct avec les différents pilotes (chasseurs, ravitailleurs, AWACS[66]…) pour leur décrire la situation et coordonner leurs déplacements, le tout également en fonction de l'activité aérienne civile.

Dans sa cabine, le contrôleur est souvent épaulé par un assistant contrôleur. Ce dernier est notamment en contact téléphonique avec les autres CDC et les contrôleurs militaires présents au sein des tours de contrôle civiles ou militaires pour coordonner les transferts de zone et effectuer le suivi de la mission. Il fait également le relais entre le chef de travée et le contrôleur, ce dernier étant le plus souvent engagé dans un échange avec un pilote.

Les binômes contrôleur-assistant ne sont pas fixes. Des appariements sont prévus selon le niveau d'expérience requis pour la mission. Cependant, quel que soit le binôme, l'assistant est

[61] Le Centre National des Opérations Aériennes (CNOA) est la partie opérationnelle du CDAOA, commandant les 3 CDC nationaux. Il est également situé sur la BA-942 Lyon Mont-Verdun.

[62] La Permanence opérationnelle est le dispositif d'interception de l'armée de l'air et de l'espace. En permanence des hélicoptères et des avions de chasse répartis sur l'ensemble de l'hexagone sont prêts à décoller à tout moment, afin d'assurer des missions de sûreté aérienne. En moyenne les aéronefs d'alerte décollent deux fois par jour pour reconnaître et intercepter notamment les appareils ne se conformant pas aux règles de l'air ou nécessitant une assistance en vol.

[63] En France, le Premier Ministre

[64] Ce terme est distinct des fonctions de contrôleurs de la circulation aérienne civile des centres de contrôle en route.

[65] La Liaison 16 (ou L16) est un standard de liaison de données tactiques de l'OTAN pour l'échange d'informations tactiques entre des unités militaires aériennes, terrestres et maritimes. Elle a pour but l'élaboration d'une image « Opérative » et « Tactique » unique et cohérente entre les unités.

[66] L'AWACS (pour Airborne Warning and Control System) est un système de commandement et de détection aéroporté. Il s'agit d'avions ou d'hélicoptères équipés de capteurs passifs et d'un radar à longue portée capables de détecter des contacts aériens et au sol à de grandes distances.

conventionnellement assis à la gauche du contrôleur. Cette configuration est décrite dans l'organisation du poste de travail de la Figure 76.

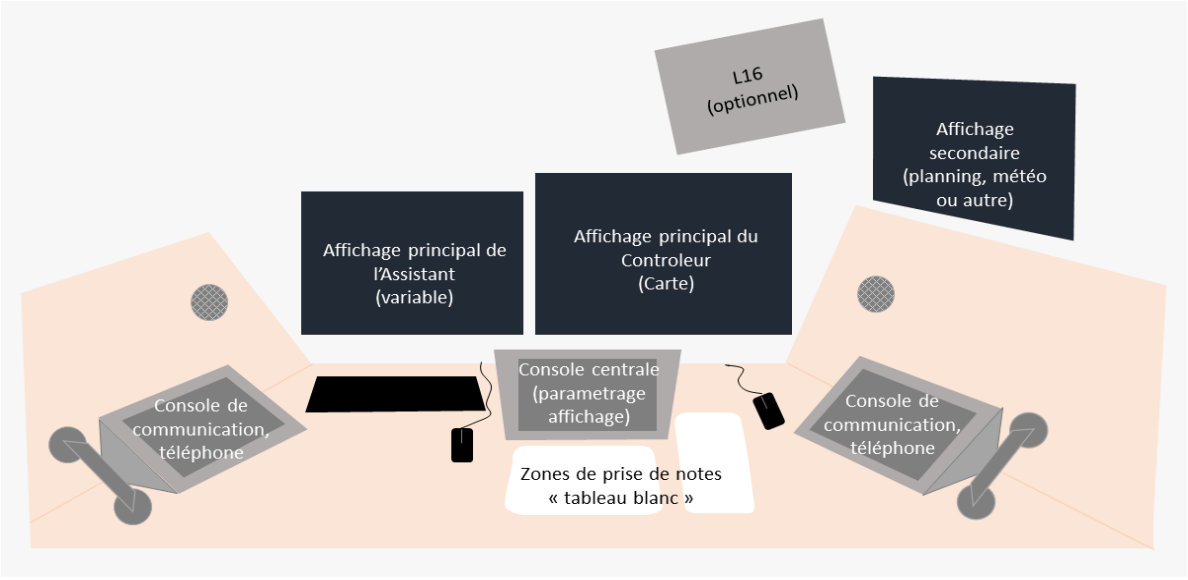


Figure 76. Représentation schématique de la cabine d'un contrôleur de CDC.

Une cabine est composée de 3 à 4 écrans et de 3 consoles tactiles. En incluant la zone de prise de notes sur le bureau, cela ne représente pas moins de 6 zones potentielles d'interaction ou de récupération d'information pour un contrôleur. Le nombre et la disposition de ces zones demandent au contrôleur une grande attention et mobilité visuelles.

Selon la mission ou le scénario de l'exercice pour lequel il est sollicité, une configuration différente de la station de travail peut être nécessaire. La console centrale permet de modifier le contenu des affichages afin de les faire correspondre au besoin de la mission et/ou aux habitudes de travail du contrôleur. L'interaction avec les 3 écrans principaux se fait par le biais d'une souris, la plupart des tâches du contrôleur se réalisant « au clic ». Seul l'assistant contrôleur dispose d'un clavier si nécessaire. Chacun à accès à une console latérale qui sert à gérer les communications et les flux audio transmis par les casques.

Contrairement au JFAC, où les échanges se font majoritairement par tchat, les communications au sein de la salle de CDC sont principalement orales. Seuls les chefs de travées et le chef de salle disposent d'un tchat pour échanger avec le CNOA.

Le travail en binôme favorise les échanges directs entre le contrôleur et son assistant, auxquels s'ajoute la liaison audio quasi constante avec des pilotes. On observe régulièrement l'assistant se pencher sur l'écran du contrôleur, soit pour y récupérer des informations non disponibles sur le sien, soit pour lui indiquer directement une position ou un élément sur la carte.

Les échanges d'information avec la structure hiérarchique se fait également oralement. Les discussions avec les autres opérateurs, les chefs de travée et le chef de salle créent ainsi des aller-retours visuels, les opérateurs se retournant naturellement pour s'adresser directement à leurs collègues.

Bien que l'environnement de travail soit assez calme du fait de la faible pression temporelle, celle-ci laisse également une liberté d'action, de comportement, d'attitude et de déplacement plus ample que dans un JFAC. Les opérateurs ne sont pas toujours sollicités en continu et certaines tâches peuvent s'avérer routinières. Néanmoins, ils doivent se tenir prêts à tout instant à supporter une augmentation subite de charge de travail. Alternent donc des périodes d'attention très soutenues et des périodes plus calmes. Cela se traduit par des variations de comportement et d'attitude des individus, parfois concentrés et penchés en avant sur le bureau, ou au contraire parfois adossés dans le fauteuil, prenant de la distance avec le poste de travail pour discuter avec des collègues.



Figure 77. Une cabine du CDC de Mont-de-Marsan. Source Armée de l'air.

Cette méthode de travail et l'organisation notamment physique du poste seront amenées à évoluer par le déploiement progressif de l'ACCS (*NATO Air Command and Control System*), système ayant l'ambition d'uniformiser le commandement et contrôle pour les pays européens de l'OTAN, afin de faciliter la collaboration des nations[67]. Il y a quelques années, le déploiement de la liaison 16 (L16) s'est traduit par l'ajout d'écrans dédiés. On aura de même des modifications similaires et une modernisation des espaces de travail qui auront nécessairement un impact sur l'organisation des tâches des opérateurs.

6.1.3 Applicabilité du monitoring aux opérations et au contrôle aérien : de l'observation à la recommandation

Le monitoring d'opérateur est une solution d'étude et de connaissance en temps réel de l'évolution de l'état d'un opérateur. Bien qu'elle réponde aux contraintes de continuité, de non-interférence avec la tâche et d'objectivité des méthodes de mesure, il convient de réfléchir à son adéquation aux opérateurs de C2, à leur environnement de travail et leurs tâches de contrôle. Une première réflexion concerne

[67] L'ACCS est système unique et intégré de commandement et de contrôle aériens de l'OTAN pour la gestion des opérations aériennes à l'intérieur et à l'extérieur de la zone euro-atlantique. Il permet, aux pays en disposant, de posséder une approche commune en termes de planification, d'attribution des tâches, de suivi et d'exécution de missions aérospatiales. Le programme comporte un hypersystème mettant en réseau des sous-systèmes de surveillance aérienne de 17 sites partageant en feedback des données opérationnelles sur le même réseau de communication à ultra haut débit.

l'applicabilité de ce type de méthodes à l'environnement de travail d'un CDC et d'un JFAC, en nous appuyant sur les entretiens et les observations présentées au 6.1.1 et 6.1.2 ainsi que sur notre expérience des différents outils de monitoring rencontrés durant cette thèse.

Les structures étudiées, CDC et JFAC, relèvent toutes deux de la conduite des opérations aériennes. Cependant, nous avons pu voir que les situations de travail collaboratif y sont différentes. Principalement, le nombre d'opérateurs mobilisés pour une même mission est plus faible en CDC qu'en JFAC et la pression temporelle pesant sur les tâches habituelles y est également plus faible. Il est alors probable que la pertinence des outils de monitoring diffère entre ces deux structures.

La tâche d'un opérateur dans une salle de contrôle consiste essentiellement à être vigilant. Cela implique l'interprétation, le diagnostic et l'intervention afin d'éviter de potentiels problèmes qui devraient être détectés avant qu'ils n'aient des conséquences graves. Ainsi, les opérateurs traitent en permanence des informations afin de diagnostiquer et de résoudre les problèmes qui surviennent en temps réel.

Nous avons pu observer en CDC que toutes les cabines ne sont pas actives et ne travaillent pas nécessairement sur la même mission simultanément. La collaboration la plus importante à lieu entre le contrôleur et son assistant. En nous référant à la définition d'équipe de Salas et al. (1992, voir 1.1), ce binôme peut alors être vu comme le cœur de l'équipe de travail, possédant le niveau de collaboration le plus élevé. Pour la mission qui leur est confiée, c'est avec son assistant que le contrôleur partage naturellement les tâches les plus interdépendantes.

Ainsi, le cercle de collaboration active d'un contrôleur de CDC est relativement réduit, comparé à celui d'un opérateur de JFAC pour lequel toute la salle peut être mobilisée sur la même opération en cours (Figure 78).

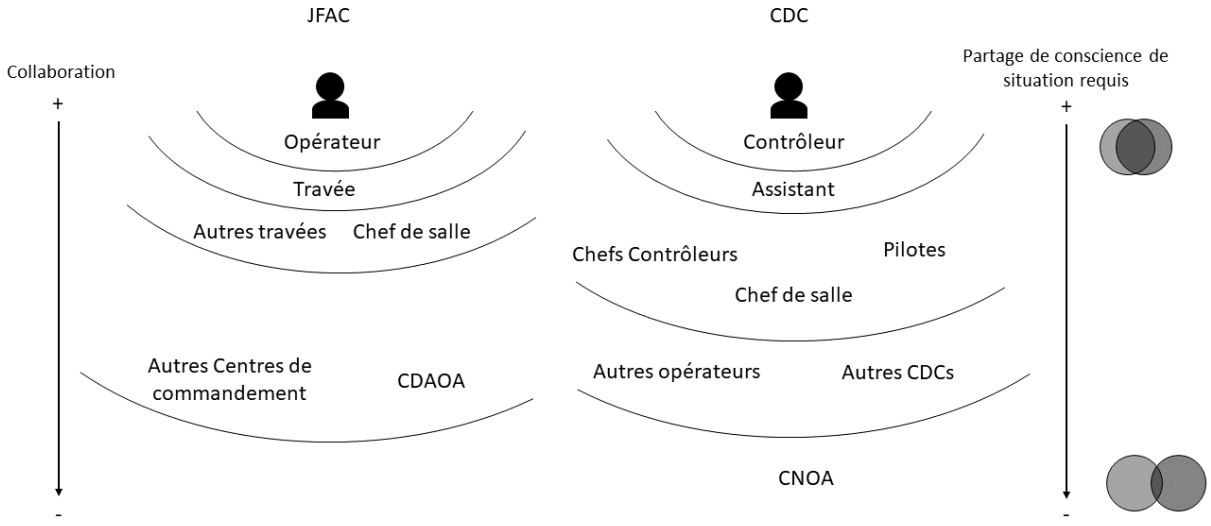


Figure 78. Comparaison des « cercles de collaboration » entre JFAC et CDC

Plus les coéquipiers ont besoin de collaborer de manière proche, plus il est nécessaire qu'ils partagent une grande quantité d'informations, donc qu'ils aient une conscience de situation commune.

Au regard de ces observations, il nous semble que le périmètre d'application du monitoring en CDC puisse être tout d'abord restreint au binôme contrôleur-assistant. Dans le cas du JFAC, la plus grande

complexité et proximité informationnelle requise par la situation de collaboration, fait de l'évaluation du partage de conscience de situation de toute l'équipe un enjeu plus important.

Après ce premier constat concernant les périmètres d'application du monitoring d'opérateur en CDC et JFAC, nous pouvons discuter de l'impact de l'organisation et des comportements observés, d'une part sur l'acquisition des données, c'est-à-dire les contraintes liées au matériel de monitoring, d'autre part sur les dispositifs requis pour leur analyse en temps réel.

Quelles contraintes pour le monitoring ?

En situation de travail collaboratif écologique ou simulée, il est important que le sujet reste libre dans ses déplacements et ses mouvements. Pour une intégration optimisée, il est indispensable que l'opérateur soit le moins impacté possible par les outils de monitoring. La question de l'appareillage exclut nécessairement un ensemble d'outils de mesure très pertinents en laboratoire mais trop contraignants en situation de travail réel. Une grande partie des systèmes de mesure psychophysiologique demandent notamment un équipement assez lourd, qui plus est souvent très sensible aux perturbations extérieures.

Ainsi pour maximiser l'acceptabilité du système en minimisant son intrusivité, l'utilisation de dispositifs de monitoring déportés (sans appareillage de l'opérateur) est théoriquement à privilégier.

De tels systèmes sont très sensibles aux déplacements et aux mouvements des sujets. Un exemple est certainement celui de l'utilisation de l'*eye tracking*. Pour un *eye tracker* déporté, c'est-à-dire reposant sur une ou plusieurs caméras montées sur l'espace de travail, les mouvements du corps et de la tête perturbent l'enregistrement de données précises et fiables, les yeux pouvant ne plus être détectés en cas de mouvements importants de la tête ou du corps.

Or, nous avons pu observer de nombreux comportements susceptibles de parasiter les mesures de tels outils. En effet, les opérateurs sont libres de se déplacer dans les travées et d'échanger directement avec leurs coéquipiers. En JFAC, ces échanges s'effectuent souvent au sein de la même travée mais il peut arriver qu'un individu se lève pour signaler oralement au groupe une information majeure, par exemple un comportement dangereux inattendu de l'ennemi. En CDC, l'échange inter-rangées est beaucoup plus courant, les opérateurs faisant remonter l'information à leur chef de travée et au chef de salle pour coordonner la prise de décision avec le CNOA. Il en résulte des comportements naturels dictés par l'organisation spatiale de la salle, comme le fait de se retourner pour échanger directement avec le chef de travée ou le chef de salle, ou de se pencher sur l'écran du coéquipier. Même si ces comportements restent ponctuels, ils génèrent du bruit, des réactions anormales ou des pertes de données problématiques pour l'interprétation automatique par un système.

L'organisation du poste de travail en lui-même a également son importance. Dans un contexte de travail collaboratif et multi-écrans, l'information est récupérée sur différents supports, incitant l'opérateur à d'amples mouvements de la tête. C'est particulièrement le cas pour les opérateurs de CDC, dont le poste de travail est chargé et complexe. En ce sens, le poste d'un opérateur de JFAC semble, par exemple, plus adapté à l'utilisation d'*eye tracker* déporté car, moins complexe, il requiert théoriquement des mouvements moins amples.

Les outils de monitoring des interactions et des communications sont moins sensibles à cette variabilité comportementale. Il s'agit principalement d'outils numériques, tels que des programmes informatiques d'enregistrement d'activité de la souris ou du clavier, ou des communications vocales. Ces mesures se concentrant sur l'utilisation du système, c'est-à-dire l'action (ou interaction), le comportement physique de l'opérateur n'est que peu ou pas impactant.

Cependant, des comportements inhérents à la collaboration restant présents, il convient d'en tirer parti. En JFAC, la circulation de l'information se fait à 90% par tchat, la majeure partie des comportements collaboratifs se font donc dans l'espace virtuel commun via des conversations à plusieurs. Le monitoring des communications au sein de l'équipe doit donc être abordé à la fois par l'enregistrement et l'analyse des échanges oraux, mais aussi par celle des interactions entre opérateurs via les systèmes numériques. Ces tchats sont des espaces de partage d'information plus que de conversation active, néanmoins une étude de leur utilisation réelle permettrait de déterminer les métriques quantitatives (fréquences d'échanges, niveau d'activité, etc.) ou qualitatives (analyse du contenu) pertinentes à mesurer en temps réel.

La principale contrainte, relative aux outils de monitoring de l'activité numérique et aux communications, relève des limitations techniques d'intégration aux outils et logiciels en place. Il est en effet nécessaire que le système de monitoring ait accès au contenu, afin de contextualiser les actions de l'opérateur. Il s'agit de recueillir des métriques provenant de l'utilisation des outils (par exemple le nombre de fenêtres de tchat actives pour un opérateur de JFAC ou les changements d'affichage sur l'écran de l'assistant opérateur en CDC). Ces *analytics*[68] représentent une source de données comportementales supplémentaires pouvant notamment aider à la contextualisation et à l'interprétation des données issues des autres sources.

Si les comportements et le contexte dictent le type de capteurs utilisables, ils génèrent également des contraintes pour l'acquisition et l'utilisation des données. Les données issues des capteurs doivent être traitées et interprétées en temps réel, afin d'être par exemple utilisées par un système de diagnostic de l'état de l'utilisateur et de l'équipe.

Nous avons vu au chapitre 4 et au chapitre 5 la nécessité de contextualiser les comportements pour palier la difficulté d'interprétation des réactions provenant des variabilités inter et intra individuelles. Cette interprétation de l'évolution des métriques se fait au niveau individuel, notamment par l'utilisation de seuils personnalisés.

Or les tâches relevant d'un CDC ou d'un JFAC *current-ops* ont à être assurées 24h/24. Cela signifie qu'un poste n'est pas attribué à un opérateur spécifique, et que par le jeu des rotations et des relèves différents opérateurs travaillent successivement sur le même poste. Cette observation illustre le besoin de prise en compte du changement d'opérateur dans la conception du système de monitoring. Cela a plusieurs impacts, principalement pour les outils de monitoring physiologique. Si le capteur est porté par l'opérateur, par exemple un bracelet ou une ceinture avec cardiofréquencemètre, l'appareil devra être couplé au système d'acquisition lors de la prise de poste et associé aux autres outils présents pour ce poste, puis découplé lors de la relève.

De plus, tout appareil nécessitant une calibration avant usage, tel qu'un *eye tracker*, est peu adapté à cette dynamique de travail. Pour cela des outils de monitoring comportemental ou des outils physiologiques portés semblent être plus adaptés.

Le changement de poste à également un impact sur l'analyse et l'interprétation des données. Le système doit développer une connaissance de l'état de l'utilisateur pour en effectuer un diagnostic. Pour cela, il est nécessaire de posséder une connaissance préalable de certaines valeurs, *baselines* et seuils, propres à l'individu. Il doit donc être à même d'identifier l'opérateur monitoré pour appliquer les analyses et les

[68]Dans le domaine du web, les *analytics* sont un ensemble de métriques et de statistiques permettant d'étudier précisément les comportements et les habitudes des visiteurs d'un site afin de l'optimiser, principalement dans un but commercial.

seuils personnalisés nécessaires à une interprétation adéquate de son activité personnelle. Une telle contrainte peut aisément être adressée, par exemple, par la mise en place d'un système de connexion chargeant le « profil » de l'opérateur lors de la prise de poste.

Quelles solutions pour l'utilisation des données en temps réel ?

Les paragraphes précédents abordent l'impact majeur de la variabilité comportementale sur le choix des outils de monitoring et sur l'analyse des données. Or cette liberté comportementale est un élément intrinsèque à la situation de travail collaboratif. Elle représente donc le principal défi à relever pour permettre l'utilisation d'un monitoring fiable et adapté à l'environnement écologique.

Nous identifions 4 leviers permettant d'adresser cette contrainte.

Premièrement, cela passe nécessairement par l'amélioration continue des algorithmes d'acquisition, de nettoyage et de traitements de données, qui permettent de compenser, de limiter, de comprendre ou d'interpréter les comportements particuliers ou la perte de données.

Deuxièmement, il est possible de jouer sur le degré d'invasivité des outils. Les récents progrès technologiques ont permis l'apparition de capteurs physiologiques portables, légers et peu invasifs. Des bracelets ou montres connectés incorporent par exemple des capteurs de plus en plus sophistiqués du rythme cardiaque, de l'oxygénation du sang ou de la sudation. Ils rendent ainsi progressivement possible l'accès à un plus grand nombre de marqueurs psychophysiologiques exploitables en situation écologique par un système de monitoring.

Troisièmement, l'utilisation de plusieurs capteurs de même type peut également s'avérer une solution. Ainsi des dispositifs d'*eye tracking* reposant sur un réseau de caméra synchronisées permettent par de contextualiser le suivi du regard dans un environnement 3D multi-écrans (Chennamma & Yuan, 2013; Kocejko & Wtorek, 2013).

Une quatrième solution consiste à combiner différents types de capteurs. C'est l'approche défendue par le monitoring multimodal. La complémentarité des données issues de sources différentes doit permettre dans une certaine mesure d'interpréter des comportements particuliers observés sur les autres (comme la perte de données). L'ajout d'un capteur d'orientation de la tête (Marion, 2018) ou d'un système de suivi dans l'espace peut simplifier l'interprétation de données manquantes, ne serait-ce que pour s'assurer de la présence de l'opérateur à son poste et être en mesure de stopper le monitoring en cas de déplacement dans la salle. Ainsi, afin de prendre en compte la complexité du contexte, ce n'est pas seulement le poste de travail qui doit être équipé mais aussi la salle, en tant que lieu d'expression de comportements collaboratifs.

Quels risques concernant l'acceptabilité et la cyber-sécurité ?

Nous souhaitons relever deux derniers points à considérer pour la conception d'un tel système : les risques liés à l'acceptabilité du système et ceux liés aux menaces cyber.

Connaître l'état de ses opérateurs en temps réel peut s'avérer être un atout pour le commandeur. Il est néanmoins légitime de se poser la question de l'impact et des déviances possibles d'une telle connaissance de l'activité de l'individu et de son suivi intensif en situation de travail.

L'évaluation de la performance collective par le biais de mesures de performances individuelles repose essentiellement sur le diagnostic de l'état de l'opérateur. L'application à des systèmes adaptatifs autonomes n'est qu'une utilisation parmi d'autres de cette évaluation.

Elle peut servir à fournir en temps réel au commandeur un aperçu de l'état de chacun des membres de son équipe, ou être utilisée a posteriori pour l'analyse des performances et le RETEX[69].

La perception de la performance individuelle, jusqu'alors une estimation subjective provenant des coéquipiers ou de la hiérarchie, est objectivée par un tel outil. Si l'identification claire des éléments faibles est un avantage pour l'amélioration continue des performances de l'équipe, elle risque cependant d'accroître l'individualisation de la responsabilité collective et éventuellement d'impacter la cohésion du groupe.

Quelle que soit l'application, les questions « Qui a accès à l'information ? Quand est-elle utilisée et dans quel but ? » ont un impact sur la perception de l'utilité de l'outil par les opérateurs.

Dans tout autre contexte que celui de la gestion de crise, qui plus est dans le domaine de la défense, l'utilisation d'un tel système serait éthiquement discutable et sujet à débats.

La structure militaire pourrait « forcer » son acceptation, en imposant un tel outil à des fins d'amélioration de la performance collective. Pour fonctionner, il doit être utilisé par tous les opérateurs d'une équipe. Le défi de l'intégration homme-machine est alors de faire en sorte que le système ne fasse pas peser de nouvelles contraintes temporelles, organisationnelles, psychologiques ou sociales sur les opérateurs. Son utilisation doit passer « inaperçue » du point de vue de la tâche tout en étant bénéfique du point de vue de la performance.

Comme pour tout système automatisé, il se pose également la question de la confiance envers l'estimation effectuée par le système. Cela vaut d'autant plus si l'opérateur est confronté à l'estimation de son propre état. Etre « qualifié » de manière automatique par un système n'a rien d'attrayant et il est fort à parier que cette qualification soit remise en question. Le risque est alors de voir apparaître des comportements et des stratégies dictées par un biais d'opposition au système.

Néanmoins, les opérateurs et commandeurs avec lesquels nous avons pu échanger ont semblé plutôt enjoués et curieux du potentiel de ces outils technologiques.

L'ajout à un système de commandement, d'un outil composé de nombreux capteurs reliés en réseau représente nécessairement une source de faiblesse cyber. Celle-ci est d'autant plus importante que cet outil génère une grande quantité de données sensibles sur l'état physique, physiologique et cognitif des membres d'une équipe.

Un système estimant « l'état » d'un groupe et capable de l'influencer en proposant en temps réel des stratégies d'adaptation/d'optimisation, devient un élément de faiblesse. Une faille de ce type peut être l'objet de manipulations malveillantes compromettant les prises de décision des opérateurs et la performance du groupe. Cela resitue alors le problème dans le contexte nouveau des guerres d'influences cognitives.[70]

[69] Le RETEX est une démarche d'analyse *a posteriori* de la gestion d'un événement réel ou fictif (exercices). Il a pour fonction de rechercher des informations émanant de la gestion de ces évènements et de les exploiter pour les traduire en enseignements qui conduiront à des adaptations.

[70] Le c*ognitive warfare* est un domaine récent qui théorise les moyens d'action qu'un état ou un groupe d'influence utilise pour manipuler les mécanismes spontanés de la cognition d'un ennemi ou de son peuple, pour l'affaiblir, le pénétrer, voire le soumettre ou le détruire

En résumé

L'étude de l'applicabilité du monitoring à la situation écologique en C2 soulève principalement la question de la sensibilité des outils à la variabilité comportementale des opérateurs inhérente à la situation réaliste de travail, qui plus est collaboratif. Le monitoring n'étant pas une fin en soi, se pose également la question de son utilisation par un système de diagnostic de l'état de l'opérateur, première étape vers un système adaptatif. Nous proposons ci-dessous quelques recommandations visant à guider les futurs travaux d'applicabilité :

- Les capteurs et outils d'acquisition de l'activité utilisés doivent être le moins intrusif et invasif possible, afin de laisser à l'opérateur la liberté comportementale propre à la situation écologique. Ils ne doivent parasiter ni la tâche ni l'opérateur. Ainsi il est préférable de privilégier l'utilisation d'outils non portés et capables de s'abstraire du besoin de calibration répétée. À ce titre, les outils de mesure de l'activité IHM et des communications présentent un avantage certain sur les outils de mesure psychophysiologique.

- Les contraintes techniques pour l'intégration des outils de monitoring aux systèmes en place doivent être étudiées.

- L'interprétation correcte des activités et comportements individuels et collectifs nécessite une personnalisation des systèmes de diagnostic de l'état de l'opérateur. Afin de prendre en compte la variabilité interpersonnelle observée, il est de notre avis que le diagnostic ne peut se faire sur la base de règles et de seuils généraux. De plus, si les opérateurs changent de poste, le système, lui, reste à la même place. Celui-ci devra donc être capable d'identifier un opérateur afin de lui faire correspondre ses critères de diagnostics, *baselines* et seuils personnalisés.

- Le système doit également acquérir une connaissance du contexte et de l'opérateur pour pouvoir déterminer si un comportement déviant observé sur une des métriques résulte d'une situation à risque ou d'un comportement normal de l'opérateur. Il s'agit, à notre avis, du principal défi à relever pour l'application et l'utilisation de tels systèmes.

- Le monitoring de l'opérateur ne doit pas être pensé comme issu d'un outil unique. Aucun marqueur seul ne saurait rendre compte de manière suffisamment fiable et interprétable de la complexité de l'environnement dynamique étudié. L'approche par monitoring multi-modal semble proposer une piste de réponse à la complexité de la situation. La mise en relation des différents types de monitoring (physiologique et comportemental) peut permettre d'accéder à la compréhension plus approfondie de la situation, nécessaire pour l'interprétation correcte des comportements mesurés. Il faut désormais chercher à unifier et à combiner efficacement ces domaines pour exploiter leur complémentarité.

- L'avancée technologique du matériel d'acquisition et des techniques d'analyse, de plus en plus tournées vers l'utilisation appliquée du monitoring, présente une opportunité majeure pour bénéficier de la multiplicité des sources nécessaires pour augmenter la fiabilité des interprétations.

- Les futures modifications de l'environnement de travail, avec l'automatisation accrue de certaines tâches et le déploiement d'outils tels que l'ACCS, vont nécessairement faire évoluer

la façon de travailler et de collaborer dans ces situations. Il convient de tenir compte de l'impact de ces futures évolutions.

- Le risque cyber engendré par l'utilisation d'un outil de diagnostic de l'état des opérateurs au sein d'un C2 est à la mesure du potentiel que peut représenter un tel outil.

Il s'agit dès-lors de trouver le meilleur compromis entre la fiabilité des données recueillies et l'acceptabilité des systèmes de monitoring. Selon nous, un tel compromis ne peut être trouvé que par l'expérimentation progressive de différentes combinaisons. Dépendant fortement de la singularité de la situation d'usage (salle, opérateurs…), ce sont avant tout les contraintes environnementales qui définissent les outils et métriques utilisables. Une solution générique ne peut donc convenir. Ce travail d'adaptation du système à son environnement relève autant de la recherche que de l'ingénierie cognitique.

Le travail d'observation et d'entretiens que nous avons menés est une ébauche de la démarche à mettre en place pour la conception d'un tel système dans un environnement aussi complexe.

Il reste beaucoup à faire, mais le C2 s'avère être un terrain d'étude exemplaire de la situation typique de travail collaboratif. C'est un support expérimental privilégié pour étudier la cognition collective.

6.2 Expérimenter pour l'évaluation de la cognition collective en C2

L'étude de la cognition de groupe dans une situation simulée de C2 s'est avérée complexe. Le développement de recherches sur ce sujet au sein d'un laboratoire nécessite la mise en place d'une structure de travail particulière qui doit être préalablement définie. Notre travail exploratoire a permis de défricher pour le laboratoire les conditions expérimentales qu'impose l'étude de la cognition collective par le monitoring des opérateurs. Dans les chapitres précédents nous avons mentionné les difficultés rencontrées pour la mise en place d'expériences et les stratégies adoptées pour pallier certaines des contraintes. Nous en avons tiré des enseignements et identifié des besoins, notamment pour une expérimentation orientée vers l'évaluation du partage de la conscience de la situation.

Ainsi, notre question est la suivante : « Que signifie expérimenter dans le cadre de l'évaluation de la cognition collective par le monitoring des individus ? ». Nos réponses n'auront d'autre ambition que de guider, dans ce contexte, le développement de situations expérimentales plus pertinentes.

Dans la même optique nous aborderons également la conception d'un logiciel de simulation.

A l'issue de ce travail d'expérimentation, d'observation et d'analyse, nous sommes à même de soumettre des préconisations pour les futures recherches liées à la conception d'un système de monitoring des opérateurs en C2.

6.2.1 Monitorer la cognition collective : quelles améliorations pour l'expérimentation ?

Lors de nos travaux, nous avons appliqué les techniques de monitoring dans deux situations expérimentales.

La situation de monitoring collectif (Chapitre 5), nous a permis d'aborder la complexité de la mise en place d'un tel environnement et les contraintes techniques, matérielles et organisationnelles qu'impliquent l'enregistrement coordonné de l'activité des membres d'une équipe.

La seconde situation, de monitoring individuel (Chapitre 4), nous a permis de nous situer dans un cadre de travail où une grande partie de ces contraintes étaient maîtrisées. L'environnement expérimental (le matériel, les outils et les logiciels) a permis le développement d'expériences rigoureuses.

Nous avons ainsi dégagé deux éléments majeurs permettant d'améliorer l'expérimentation : la synchronisation et la simulation.

La synchronisation (entre capteurs, entre individus et avec la simulation)

Le but du monitoring collectif est double. Il permet, d'une part, d'étudier de manière objective et non intrusive les réactions d'un sujet (et du groupe) face à une modification de l'environnement ; d'autre d'étudier la synchronisation des activités des coéquipiers entre eux.

Cela signifie que les signaux issus des capteurs doivent posséder un référentiel commun, de manière à être temporellement alignables.

Nous avons pu voir que des méthodes de synchronisations externes sont trop peu fiables pour être utilisées dans un contexte aussi complexe que le monitoring collectif. La synchronisation par signal externe consiste à imposer simultanément aux sujets un stimulus suffisamment important pour déclencher une réaction lisible dans les signaux enregistrés par les capteurs. Cela paraît peu adapté à la synchronisation de capteurs de types aussi différents qu'un cardiofréquencemètre et un système de suivi de l'activité de la souris.

Toutefois la synchronisation des enregistrements est indispensable, d'autant plus dans l'optique d'une utilisation des données en temps réel.

La solution la plus pertinente est la synchronisation des acquisitions, lancées de manière centralisée par l'expérimentateur, le commandeur, ou toute personne en charge du système. Cette acquisition synchronisée doit également intégrer le déroulement de la simulation afin de faciliter et améliorer l'interprétabilité des activités mesurées dans un cadre expérimental.

Lors de l'expérience AMIGOS (Chapitre 4), nous disposions d'un tel système d'acquisition centralisé. Cela nous a permis de mettre en relation les évolutions du diamètre pupillaire, de la position du regard et des clics de souris sur un élément d'interface, avec les évènements de la simulation (apparition d'un nouvel élément).

Il est essentiel que le système enregistre simultanément les données provenant de tous les capteurs individuels ainsi que de la simulation.

Ainsi les mesures sont à la fois synchronisées (1) entre elles pour chaque individu (permettant de comparer les données issues de différents capteurs au niveau individuel), (2) entre individus (permettant de comparer les individus entre eux) et (3) avec l'environnement, permettant ainsi d'analyser la réaction d'un individu ou du groupe à un évènement de l'environnement.

Toutefois, un système intégré d'acquisition synchronisée des activités individuelles, d'équipe et de l'évolution de l'environnement demande la mise en place d'une structure technique importante mais indispensable.

Cependant c'est un défi technique que certains logiciels spécialisés permettent de surmonter, comme nous avons pu le constater lors de l'expérience AMIGOS.

La simulation

Balancer maîtrise expérimentale et validité écologique est un exercice difficile, nécessaire à la recherche en sciences cognitives appliquées. Il s'agit de soumettre le sujet à une tâche représentative de la complexité de la situation de travail réel, tout en maîtrisant différents paramètres expérimentaux.

Ainsi la simulation est un outil potentiellement puissant pour évaluer la cognition d'une équipe dans ce contexte. Elle permet de concevoir un « monde synthétique » qui reproduit les principaux aspects d'un environnement de travail réel, évoque les processus psychologiques et comportementaux liés à la tâche critique et permet d'évaluer une série de résultats possibles en termes de performances (Bell, Kanar, & Kozlowski, 2008). Les simulations sont régulièrement employées pour la formation, où différents scénarios permettent d'entraîner l'individu ou l'équipe à des situations particulières.

L'utilisation de simulations dans un environnement expérimental demande cependant une grande vigilance et un contrôle accru des paramètres de recherche. Cela est obtenu principalement par la maîtrise du scénario de la simulation, qui a deux objectifs.

Elle permet, d'une part, d'étudier la réaction spécifique d'un individu ou du groupe à un évènement particulier, d'autre part de manipuler sciemment l'état de l'utilisateur pour le mettre dans des situations contraintes ou extrêmes que l'on souhaite étudier.

Rappelons que la mesure de la conscience de situation repose sur l'évaluation de la connaissance ou non d'éléments de la situation (évènements) par le sujet. Nous avons pu nous apercevoir, dans le Chapitre 5, que l'absence de maîtrise du scénario demande un lourd travail post-expérimental de caractérisation et recensement de ces évènements critiques. Afin de poursuivre l'étude par monitoring de la cognition collective et particulièrement du partage de conscience de situation, les scénarios doivent être accessibles, paramétrables, permettre le choix de la tâche, du type d'évènement, et le moment de leur apparition. À notre sens, il s'agit de la contrainte majeure reposant sur l'outil de simulation destiné à ce type d'expérimentation.

Ces besoins ne sont certainement pas nouveaux. Ils reposent sur la nécessité de maîtrise des conditions expérimentales propre à la recherche en laboratoire.

La simulation compense la difficulté d'accès à l'expérimentation en environnement réel tout en fournissant un support de tâches réalistes mais maitrisées(-ables).

Le dispositif d'acquisition centralisé et simultané des différentes sources et capteurs assure, en plus de faciliter leur exploitation, une meilleure fiabilité des données et des interprétations grâce à la synchronisation temporelle.

Ces deux éléments sont essentiels à la mise en place d'un environnement expérimental convenable pour la poursuite de toute étude portant sur l'évaluation de la cognition collective par monitoring.

6.2.2 Collaborative Warship Commander Task : proposition d'un outil collaboratif pour l'expérimentation sur le partage de conscience de situation

Du manque naît la création et l'adaptation. Notre besoin d'une simulation de tâche de C2 collaborative, réaliste et au scénario maitrisé, est à l'origine de la conception d'un nouvel outil. Ce dernier trouve ses racines dans le programme *Augmented Cognition* de la DARPA (St. John et al., 2004).

Au début des années 2000, les progrès technologiques de l'automatisation et de la gestion de l'information ont conduit la DARPA à explorer le potentiel de leur application à la mesure de la cognition humaine. Leur objectif était proche du nôtre : évaluer la dynamique de la cognition humaine pendant l'exécution des tâches, appliquées aux systèmes de C2.

La sensibilité de 20 mesures psychophysiologiques aux variations de la charge cognitive des participants a été testée. Le programme devait déterminer la validité et l'utilité des métriques en tant qu'identificateurs de la dynamique de la cognition humaine. Pour ce faire, St John et al. ont conçu une tâche simple, analogue à une tâche de surveillance d'espace aérien maritime comme celle décrite et utilisée dans le Chapitre 4. Appelée *Warship Commander Task* (WCT), sa conception est détaillée dans (St. John & Kobus, 2003). La WCT consiste en une tâche primaire de surveillance de l'espace aérien (Figure 79). Le rôle du participant est de protéger un convoi militaire situé en bas de l'écran. Pour y parvenir il doit empêcher tout aéronef menaçant, arrivant du nord, de passer au sud de la « ligne d'engagement » (LOE). Cette ligne symbolise une proximité critique avec le convoi à protéger. Tout avion se trouvant dans cet espace pourrait potentiellement nuire au convoi. Les sous-tâches comprennent l'identification de chaque avion entrant dans l'espace aérien (c'est-à-dire apparaissant à l'écran), l'avertissement des avions potentiellement dangereux qui pénètrent après la ligne d'engagement et l'abattage de tout avion qui ignore les instructions.

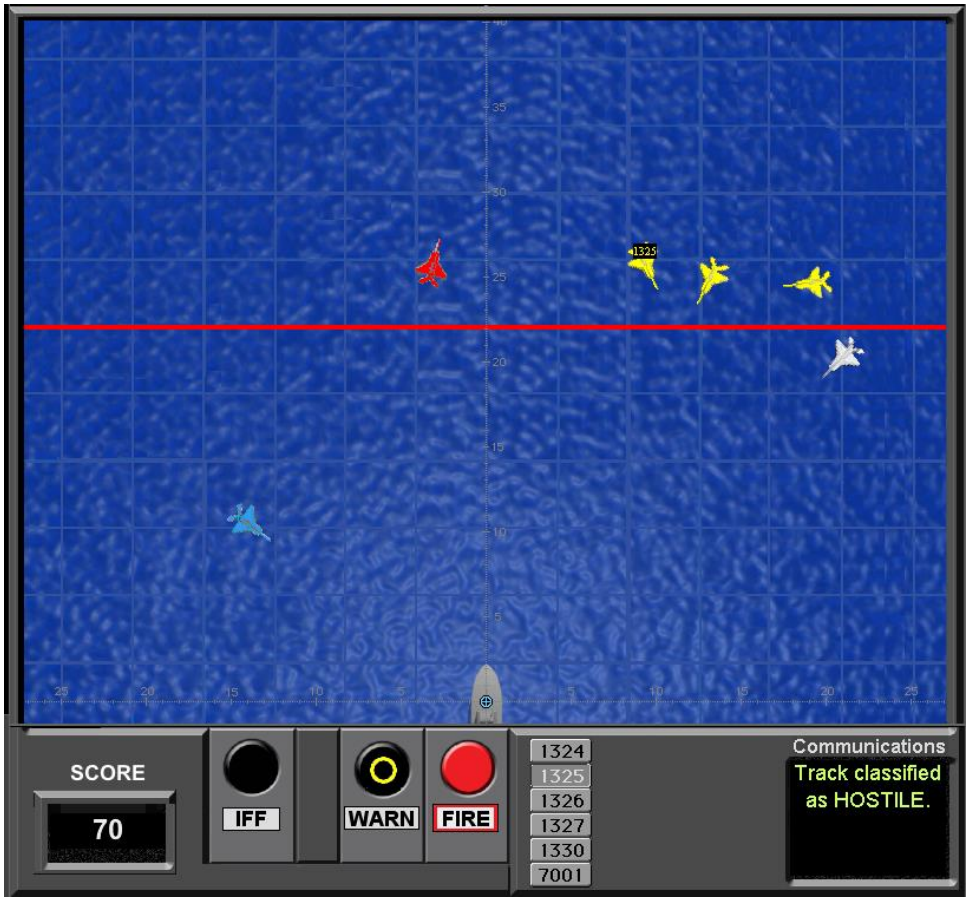


Figure 79. Capture d'écran du Warship Comander Task. En blanc la trace d'un avion non encore identifié, en bleu celle d'un avion identifié comme ami, en rouge celle d'un avion identifié comme ennemi. Les traces jaunes représentent des avions dont l'identité est ambigüe.

Cette étude a permis d'identifier 11 métriques psychophysiologiques présentant une sensibilité pertinente aux changements dans l'activité cognitive durant une tâche relativement complexe de type C2, parmi lesquelles on trouve la pupillométrie.

La WCT présente plusieurs avantages justifiant notre intérêt. La tâche a été conçue pour (1) être facilement utilisée avec des participants de premier cycle (non-experts), (2) être suffisamment complète pour être représentative du processus complexe de prise de décision auquel sont confrontés les opérateurs des environnements C2, (3) manipuler la cognition de l'opérateur en fonction de la charge de travail demandée par le système.

Au regard de la similarité de ces critères avec nos objectifs et nos contraintes, la conception d'un outil de simulation de situation de travail collaboratif basé sur la WCT a donc été entrepris, nommé *Collaborative Warship Commander Task* (CWCT).

Notre travail initial s'est appuyé sur les scénarios et la procédure conçue et validée pour la WCT en apportant cependant trois modifications majeures :

- transformer la tâche en une tâche collaborative demandant un partage d'informations ;

- proposer des paramètres permettant de manipuler le niveau de partage de conscience de situation ;

- permettre l'enregistrement des événements et des interactions des sujets avec le système.

La principale modification à apporter à été de transformer cette tâche en une tâche collaborative. Il s'est donc agit de scinder, partager les sous-tâches entre deux opérateurs, tout en conservant une difficulté suffisante. Le prototype que nous décrivons ici a été conçu pour être utilisé par une équipe de deux personnes, en simulant le rôle de Contrôleur et d'Assistant contrôleur d'un CDC. Les tâches de la WCT ont été réparties entre les deux postes de manière à ce que chacun d'eux dispose à la fois de tâches individuelles et de tâches nécessitant une collaboration.

Rôle du premier opérateur : Contrôleur

L'interface du contrôleur est une représentation visuelle de l'espace aérien autour du convoi. Sur une carte simplifiée, les aéronefs apparaissent sous la forme d'icônes (trace) représentant des avions blancs. Sa première tâche consiste à les identifier en les sélectionnant et en cliquant sur le bouton "IFF". La trace change alors de couleur en fonction de son identité (prédéfinie dans le scénario). Les traces des avions amis, en bleu, peuvent être ignorées car ils ne représentent pas une menace. Les avions ennemis sont signalés en par une trace rouge et doivent être engagés par le contrôleur dès qu'ils franchissent le LOE.

Les traces jaunes demandent un traitement plus complexe. Elles correspondent à une identification « incertaine », nécessitant que l'identité soit établie par une seconde étape. Celles-ci peuvent alors être qualifiés d " ami potentiel " ou de " menace potentielle ". Les amis potentiels peuvent être ignorés tandis que les avions qualifiés de " menaces potentielles " doivent être surveillés et avertis s'ils franchissent la LOE. Si, dans les 3 secondes qui suivent leur avertissement, les avions n'ont pas fait demi-tour, le contrôleur doit engager leur destruction.

Rôle du second opérateur : Assistant contrôleur

Ce second niveau d'identification requis pour les traces jaunes n'est pas directement accessible au contrôleur. Le rôle de l'assistant contrôleur est alors de récupérer l'identification de 2e niveau des avions jaunes sur son système, à la demande du contrôleur pour les lui fournir.

Un numéro à quatre chiffres (ID) est attribué à chaque avion. A la place d'une représentation visuelle, l'assistant dispose d'une liste non ordonnée d'ID d'avions, présents ou non sur l'écran du contrôleur. Pour répondre à la sollicitation du contrôleur, l'assistant doit d'abord retrouver le numéro de l'avion dans la liste à sa disposition pour en demander au système l'identité de niveau 2. Celui-là fournit son évaluation de l'identité de l'avion dans une zone d'affichage temporaire dédiée, après un temps de traitement simulé de 2 secondes. L'assistant peut alors la communiquer au contrôleur.
La seconde responsabilité de l'assistant est la gestion du système d'alerte. A la demande du contrôleur, l'assistant doit sélectionner l'avion suspect et l'avertir si celui-ci peut l'être (s'il s'agit bien d'un avion jaune « menace potentielle » ayant franchi la LOE).

Collaboration et répartition des tâches

Cette répartition s'appuie sur une *Cognitive Task Analysis*[71] effectuée sur la WCT. Cette analyse utilise une décomposition standard de la cognition en étapes de traitement de l'information (Parasuraman, Sheridan, & Wickens, 2000) afin de déterminer quels processus cognitifs, impliqués dans la réalisation de la tâche, peuvent être manipulés. Grâce à cette méthode, les informations utilisées par les individus pour prendre des décisions sont identifiables. Le concepteur à la possibilité de les intégrer à la simulation.
Nous forçons donc volontairement les coéquipiers à échanger, notamment en privant l'opérateur 2 de toute connaissance spatiale de la position des avions (utilisation d'une liste) et en lui attribuant des tâches qui en dépendent. Le contrôleur, quant à lui, a besoin d'informations (identité de 2nd niveau) et de fonctions (avertissement) uniquement disponibles pour l'assistant. Ainsi, chaque coéquipier a besoin d'éléments détenus par l'autre pour prendre des décisions et accomplir efficacement deux des sous-tâches les plus importantes : (1) l'identification de 2e niveau des aéronefs critiques, et (2) l'avertissement des aéronefs qui traversent la LOE.

Scénarios

Nous avons conservé les 24 scénarios utilisés par la WCT, ceux-ci ayant fait l'objet d'une conception et analyse approfondie (St. John & Kobus, 2003) pour garantir la maîtrise de la difficulté de chacun d'entre eux. Ils se présentent sous la forme de 12 vagues successives de 75 secondes chacune, soit 15 minutes par scénario. À chaque vague, un nombre différent d'aéronefs à gérer et surveiller est présenté à l'équipe (entre 6 et 24).

Performance

La performance de l'équipe est évaluée par un score cumulé. Pour chaque action, un certain nombre de points est attribué. Les tâches collaboratives, plus difficiles, valent plus que les actions individuelles simples. Chaque erreur est comptée comme un point négatif. Les points sont attribués comme indiqué dans le Tableau 20. Un taux de performance peut également être calculé sur la base d'un score théorique maximum propre à chaque scénario.

[71] La *Cognitive Task Analysis* (CTA), est un type d'analyse de tâche qui se concentre sur la description des processus cognitifs et des connaissances requises par utilisateur pour accomplir un objectif de travail particulier (Chipman, Schraagen, & Shalin, 2000)

Action	Points	Responsabilité
"Piste identifiée pour la première fois"	+10	Indiv
"Première identification de niveau 2 d'une piste jaune"	+10	Collab
"Piste oubliée"	-5	Indiv
"Tentative d'action sur une Piste n'appartenant pas à la vague actuelle"	-5	Indiv/Collab
"Piste présumée ennemie correctement avertie "	+20	Collab
"Tentative d'avertissement d'une Piste appartenant à un niveau de menace n'autorisant pas l'avertissement "	-10	Collab
" Piste ennemie ou présumée ennemie correctement détruite (après LOE) "	+75	Indiv/Collab
"Tentative de destruction d'une Piste appartenant à un niveau de menace n'autorisant pas la destruction "	-150	Indiv/Collab
"Tentative d'action avant la LOE "	-10	Indiv/Collab
"Tentative d'action avant Avertissement "	-10	Collab
"Tentative d'action avant identification "	-10	Indiv
"Une piste dangereuse a atteint le convoi "	-100	Indiv/Collab

Tableau 20. Actions évaluées et le nombre de points qui leur est associé. Une erreur est sanctionnée par un retrait de points.

Manipulation de la charge cognitive

Les 24 scénarios sont conçus de manière à induire différentes séquences de niveaux de charge à l'opérateur. Pour ce faire, nous jouons sur 2 principaux facteurs :

- le nombre d'aéronefs à identifier par vague, variant de 6 pour une vague à faible charge cognitive induite, à 24 pour une vague à forte charge induite ;
- la difficulté des aéronefs, en faisant varier la proportion de aéronefs "jaunes" menaçants apparaissant au sein de chaque vague (Haute : 67 % des aéronefs sont jaunes, Faible : 33% d'avions jaune). Ces aéronefs ambigus nécessitent une surveillance plus étroite et un traitement plus complexe que les autres.

Manipulation du partage de conscience de situation

Ce sont le plus souvent les facteurs de fréquence et de difficulté des tâches qui sont utilisés pour manipuler la charge cognitive d'un sujet. Il est moins courant de chercher à manipuler la conscience de situation, car il s'agit avant tout d'une représentation traduisant la compréhension individuelle de la situation. Les études sur la conscience de situation évaluent le plus souvent l'impact de la manipulation de la quantité d'informations ou de la difficulté des tâches, c'est-à-dire des manipulations de la charge cognitive induite.

Nous avons souhaité profiter de ce projet pour proposer un outil permettant de tester différentes modalités de partage de conscience de situation. Nous avons donc inclus à la simulation trois options visant à le manipuler, résumées dans le Tableau 21.

La première option permet de faire varier la disponibilité d'une information nécessaire à l'accomplissement de la tâche collaborative d'avertissement d'avions. Une analyse de tâche a mis en exergue trois principaux NSKE :

- la position des avions (ex : en dessous de la LOE, le plus bas, proche de…) ;
- l'identité des avions (couleur) ;
- le comportement des avions (cap, demi-tours…) ;

Notre choix s'est porté sur l'information de position des avions. Nous proposons aux expérimentateurs de faire apparaître ou non, sur l'écran de l'assistant, une indication de position pour les avions ayant franchi la LOE (par exemple sous la forme d'une mise en gras de son numéro d'identification dans la liste).

La deuxième option consiste à injecter dans les scénarios des événements imprévus qui n'apparaissent que pour l'un des coéquipiers à la fois. Deux types d'événements ont été choisis :

- pour le contrôleur, l'apparition soudaine d'un avion non identifié en dessous de la LOE ;
- pour l'assistant un avion déjà identifié ami potentiel change d'identification en menace potentielle.

Deux modalités sont envisagées, une modalité « urgence faible » (l'avion concerné est au-delà de la LOE) et une « urgence forte » (l'avion concerné est entre la LOE et le convoi). Ce type d'événements a été pensé dans le but d'étudier les latences de synchronisation des consciences de situation.

La troisième option relève de la manipulation de la conscience de situation d'équipe (Team SA). Nous proposons cela par le biais d'ajout d'information sur l'écran de l'**assistant.** Ainsi disposera-t-il, ou non, du nombre d'avions présents sur l'écran du **contrôleur** ainsi que du nombre d'avions restant à identifier

par ce dernier. Ces ajouts procurent à l'assistant des informations supplémentaires afin de construire une meilleure estimation du niveau de charge de son coéquipier, ainsi qu'une représentation plus précise de l'avancement de la tâche.

Modalités	**Option de la simulation**	**Description**	**Conscience de situation impliquée**
Avec ou sans **Assistance de franchissement**	Afficher pour l'assistant un indicateur de dépassement de LOE dans la liste des avions	Manipule la quantité de NSKE directement disponible pour la construction d'une représentation plus commune de la situation	Shared SA
Avec ou sans **Perturbations**	Apparition d'événements imprévus visibles uniquement par un des coéquipiers : - Avion non identifié apparaissant sous la LOE (pour le contrôleur) - Changement d'ID d'un avion jaune-ami en jaune-ennemi (pour l'assistant)	Requiert la mise à jour de la conscience de situation individuelle et le partage de l'information. Permet l'étude des latences de synchronie des SA.	Individuelle et Shared SA
Avec ou sans **Assistance d'équipe**	Afficher pour l'assistant, le nombre d'avions présents sur la carte du contrôleur et le nombre d'avions non identifiés.	Manipule la présence d'un indicateur de charge cognitive possible du coéquipier ou d'avancement de la tâche.	Team SA

Tableau 21. Présentation des trois paramètres de l'outil permettant la manipulation d'une partie de la conscience de situation de l'équipe

Architecture technique

L'outil se présente sous la forme d'une application web, développée en .php, s'exécutant sur un serveur local. Sur son poste, l'expérimentateur lance le serveur de simulation auquel sont ensuite connectés les postes des sujets. Tous les évènements du scénario et d'interaction des sujets sont centralisés et enregistrés au sein d'un unique fichier.

Le poste de l'expérimentateur autorise une maîtrise de l'expérience. La simulation peut ainsi être mise en pause au besoin, s'adaptant par là à des méthodes de mesure plus classiques de la conscience de situation. Structurée ainsi, la solution est portable ce qui peut permettre de varier les situations expérimentales et convient aussi bien à l'étude de situation de collaboration en présentiel (les sujets sont localisés dans la même salle) ou à distance. A terme l'outil pourrait être déployé sur une serveur internet et ainsi être utilisé à distance.

Avancement

La conception de cet outil est encore en cours. Son développement a débuté mais demande encore un travail conséquent, notamment pour l'implémentation des 24 scénarios et des paramètres destinés à l'étude de la conscience de situation. A ce jour la structure technique générale est fonctionnelle.

Doivent également suivre de nombreux tests concernant la validité de la répartition des tâches et des modalités de manipulation de la conscience de situation. Ce travail fera l'objet de projets étudiants pour les élèves de l'ENSC et éventuellement ceux de l'Ecole de l'Air et de l'Espace, à la rentrée 2020.

6.2.3 **Proposition de recherche**

L'intérêt pour les systèmes autonomes et adaptatifs s'est accru au cours des dernières décennies, les progrès technologiques posant les premières briques de l'optimisation continue des performances du système Homme-Machine. Dans ce contexte, le monitoring de l'humain est central.

La qualification dynamique de la cognition humaine *in-situ*, orientée vers l'application à des systèmes C2, pourrait conduire à l'attribution automatisée des tâches aux opérateurs ayant la capacité cognitive la plus inutilisée ou à des interfaces adaptatives avec une présentation optimisée des informations.

Si l'évaluation de l'état de l'utilisateur est un sujet de recherche courant, l'efficacité des méthodes reste à démontrer dans un environnement collectif et moins contrôlé que le laboratoire.
C'est dans cette optique que les courants du *Naturalistic Decision Makin*g (NDM) et du *Cognitive System Engineering* (CSE), vers lesquels s'orientent nos travaux, se concentrent sur la mesure de la cognition en situation réelle (Chauvin, 2003; Endsley, Hoffman, Kaber, & Roth, 2007; Hollnagel & Woods, 1999; Raymond, Prun, & Cegarra, 2016). Pour cela, ces cadres de travail suggèrent de procéder par une succession d'étapes possédant différents équilibres entre la validité écologique de la situation et le contrôle expérimental. Cela doit permettre de passer progressivement de la recherche fondamentale sur un concept, à son utilisation dans une situation réelle de travail (Rasmussen, Brehmer, & Leplat, 1991).

Dans notre cas, nous avons identifié 3 étapes, correspondant à trois contextes expérimentaux de plus en plus fidèles, permettant de poursuivre le travail de recherche entrepris par le biais d'une évolution progressive vers l'expérimentation en situation de moins en moins contrôlée de C2. Les étapes qui suivent déroulent une complexification graduelle de la situation d'étude, pour se rapprocher de la situation écologique finale, dans le but de sélectionner progressivement les outils, les métriques et les traitements les plus appropriés.

Nous estimons que la prochaine étape doit être de tester la sensibilité des différents capteurs aux variations de charge et leur complémentarité, en utilisant des méthodes de laboratoire dans un contexte C2 de base. Les capteurs testés doivent néanmoins d'ors et déjà répondre aux contraintes d'intégration exprimées aux 6.1.

Cette étape doit reposer sur une tâche de C2 simple et contrôlée, ainsi que de petites équipes (2 ou 3 coéquipiers).

Pour cela un outil de simulation simplifié, comme le CWCT, est pertinent. Les différentes modalités conçues dans le CWCT doivent notamment permettre de tester les hypothèses concernant la sensibilité des capteurs à la modification de la conscience de situation ainsi que l'étude des latences de synchronisation. Les sujets des expériences ne sont pas nécessairement des experts. Il peut s'agir d'étudiants ou du personnel en formation, plus disponibles.

Dans un deuxième temps, l'expérimentation doit porter sur de plus grandes équipes, composées d'opérateurs de C2 en formation ou experts. Il s'agit d'utiliser une simulation avancée, comme celle

proposée dans le cadre de la formation de personnel de C2 au CASPOA. Pour cette étape, les « CAX[72] » et « CPX[73] » qui y sont joués, représentent de potentiels environnement expérimentaux particulièrement intéressants.

Une telle étude peut être pensée de manière modulaire et porter dans un premier temps sur une seule travée de la salle de C2, puis sur l'ensemble du JFAC simulé.

Cette mise situation doit également permettre de se confronter à une première intégration d'outils de monitoring aux systèmes utilisés hors-laboratoire.

Enfin, en suivant la logique de complexification de la situation d'étude, il faudrait intégrer un cadre expérimental opérationnel. Il peut s'agir de situations d'exercice à plus grande échelle, comprenant de grandes équipes d'experts. C'est le cas, par exemple, des nombreux exercices ouverts à l'expérimentation menés chaque année par l'OTAN (CWIX[74], Trident, …), ou des exercices de type Volfa[75] en CDC.

La structure que nous proposons est une base de travail ayant principalement pour but d'accompagner le développement d'outils de monitoring pertinents et adaptés aux situations de travail complexes en C2. Cela représente un projet d'ampleur rappelant, à l'échelle du collectif et dans une approche multi-capteurs, le projet *Augmented Cognition* de la DARPA, qui mobilisa 11 groupes de recherche et autant d'universités au début des années 2000. L'ambition portée par ce projet et sa nature profondément transdisciplinaire, nécessite le développement d'une collaboration entre laboratoires.

Une approche d'ingénierie cognitique de l'étude de la cognition collective en C2 doit permettre de tirer pleinement parti des opportunités que représentent les progrès technologiques en matière d'intelligence artificielle et augmentée, d'automatisation, de traitement de données massives et de *cognitive computing*.

Au-delà de ce projet, la compréhension de la cognition collective est primordiale pour le perfectionnement de l'organisation et l'adaptation du système homme-machine aux évolutions des situations de conflits. C'est une clé de la compétitivité du C2 du futur.

[72] *Computer Assisted eXercice*. Les exercices assistés par ordinateur (CAX) utilisent un environnement opérationnel simulé par ordinateur en temps réel pour entraîner le personnel à la planification, la conduite des opérations, et la prise de décision militaire.

[73] *Command Post eXercice.* Les exercices de poste de commandement (CPX) sont des exercices faisant intervenir les commandants et leurs états-majors, mettant à l'épreuve les communications entre les quartiers généraux participants.

[74] Le *Coalition Warrior Interoperability eXercise* fournit chaque année aux pays membres de l'OTAN et partenaires un environnement contrôlé pour tester, identifier et résoudre leurs problèmes d'interopérabilité afin d'améliorer continuellement l'interopérabilité et la résilience de l'OTAN lors de crises futures. C'est un environnement privilégié pour l'expérimentation scientifique et technologique appliquée.

[75] Les exercices « Volfa » sont des entraînements majeurs de préparation au combat, mobilisant l'ensemble des composantes des forces aériennes. Ils s'adressent à des équipages qualifiés et permettent la réalisation de scénarios complexes et réalistes mettant en œuvre un grand nombre de moyens aériens.

6.3 Enjeux de la recherche pour le futur du C2

L'histoire de l'augmentation artificielle de la pensée se confond avec celle du calcul, des nombres et de leur utilisation (Ifrah, 1994). L'épopée du calcul automatique a ainsi structuré l'histoire et la politique, les religions et les mythes. Elle a également permis l'avènement de la science et le développement des technologies. Aujourd'hui, nous sommes confrontés à une percée dans les outils et les méthodes, ainsi que dans le domaine de la pensée humaine assistée. Cette avancée promeut de nouvelles capacités et de nouvelles perspectives pour la compréhension et la maîtrise de l'organisation du monde. Les développements récents du calcul intensif, du parallélisme et de l'intelligence artificielle permettent d'envisager une véritable assistance technologique dans la résolution des problèmes complexes, et notamment la gestion des crises sociotechniques, avec celle de conflits d'un nouveau type. Le C2 est précisément, à la fois pour sa structure et ses fonctions, au cœur de cette évolution.

6.3.1 De l'intelligence naturelle à la cognition augmentée en C2

Le développement de l'informatique puis de l'intelligence artificielle a parallèlement intensifié la recherche sur l'humain (Claverie, 2005). La promesse, portée par l'analogie fondatrice machine-cerveau, de reproduire artificiellement la cognition humaine, a guidé l'expansion des recherches dans ce domaine, à la fois structurales (neurones) et fonctionnelles (processus cognitifs).

Si l'on conçoit le cerveau et la machine comme lieux de traitement de l'information cognitive, l'intelligence (naturelle ou artificielle) se traduit alors par un degré d'adaptation dynamique à l'environnement, reposant sur une capacité de traitement plus ou moins complexe et complet de la donnée d'entrée (signal).

En s'appuyant sur les bases théoriques générales issues des travaux de Simon[76], de nombreux modèles ont été développés pour rendre compte de l'adaptation dynamique de l'individu à son environnement. Ces modèles explicatifs centrés sur la prise de décision en tant qu'acte de pensée prototypique du choix d'adaptation, ont permis de perfectionner l'approche en prenant en compte des variables d'ajustement issues de l'environnement (théorie de la perception écologique de Gibson), conduisant notamment à l'apparition de la théorie « écologique » de la maîtrise des risques (Amalberti) et à des méthodes de conception « écologiques » (Norman).

Mac Lean[77] propose une approche structuro-fonctionnelle du cerveau entre plusieurs niveaux de compétences qui s'installent progressivement et qui expriment des fonctions de plus en plus

[76] Le système STI (pour système de traitement de l'information) dû à Simon est un modèle explicatif du traitement de l'information pour la décision. Il prend en considération et fait des variables d'ajustement organisées en niveaux. Il permet de hiérarchiser des niveaux de décision, avec certains que l'on peut confier à des systèmes automatiques, voire des machines spécialisées, d'autres qui nécessitent un traitement complètement humain ou d'autres, hybrides, dont une partie peut être confiée à un ou des automates et d'autres parties nécessitent l'intervention d'un ou plusieurs opérateurs. Les décisions sont alors programmables, non programmables ou partiellement programmables (Newell & Simon, 1972).

[77] Le modèle du « cerveau tri-unique » de Mac Lean (1967), théorise une hiérarchie structure-fonctionnelle de niveaux de développement du cerveau se traduisant par des fonctions cérébrales accumulées, de plus en plus sophistiquées. Ainsi, sur un système nerveux de base (proto-kinétique), se construit une organisation à trois niveaux. Le premier niveau (archéo-kinétique) dote l'animal (et cela dès les premiers animaux terrestres) d'une intelligence essentielle à la survie, ainsi qu'à la meilleure adaptation à l'environnement, de l'individu et de l'espèce. Cette intelligence continue son évolution et son perfectionnement alors que se développent de nouvelles structures (paléo-kinétiques) qu'elle colonise. Ces structures sont amenées à compléter et contrôler les fonctions sous-jacentes

sophistiquées. À l'instar de ce modèle développemental, certains auteurs ont proposé des modèles descriptifs de la cognition selon des niveaux opératoires organisés en hiérarchie.

Ainsi, pour Rasmussen (1983), la perte de contrôle dans un système complexe est le résultat d'une mauvaise adaptation dynamique de l'opérateur aux contraintes de l'environnement. Il définit trois façons dont l'humain se représente les contraintes (signaux, signes et symboles)[78], décrivant alors l'appareil cognitif selon trois niveaux d'activité basés sur les compétences (*skills*), les règles (*rules*) et les connaissances (*knowledge*) (Figure 80).

Les comportements basés sur les compétences (*skill-based*) résultent de tâches sensori-motrices sans contrôle conscient (automatismes). Le risque est alors l'erreur « de routine » ou d'inattention due à l'habitude.

Lors de comportements basés sur les règles (*rule-based*), l'individu possède un ensemble de solutions prédéfinies activées si nécessaire (il est coûteux en ressources) par des signes ou conditions spécifiques sur lesquels il raisonne par inférence. Le risque associé est d'activer une règle inadaptée à la situation.

En présence d'une situation inattendue pour laquelle aucune procédure ou règle n'est connue, l'individu doit lui-même trouver une solution à partir de ses connaissances (*knowledge-based*). Le risque est évidemment de ne pas trouver de solution (sans compter la panique créée par la situation de raisonnement déductif dans l'incertitude).

En empruntant un exemple à Behymer et al. (2016), on peut imaginer un chef cuisinier préparant un repas. Hacher des légumes ou ajuster continuellement la flamme d'un brûleur pour faire frire un œuf à la perfection sont des comportements basés sur des compétences (*skill-based*), et suivre une recette est un comportement basé sur des règles (*rule-based*). Imaginez maintenant que la recette demande un extrait de vanille, mais que lorsque le chef regarde dans le garde-manger, la bouteille soit vide. Le chef examine ses options : (1) délaisser le processus de cuisson, qui est actuellement dans une phase critique, pour obtenir plus d'extrait de vanille ; (2) se passer de l'extrait de vanille ; ou (3) trouver un substitut approprié. Lorsque le chef décide d'essayer l'extrait d'amande comme substitut, il fait alors preuve d'un comportement basé sur la connaissance (*knowledge-based*).

en fonction des enjeux de second niveau, notamment grâce à des aptitudes d'intelligence mnémoniques, émotionnelles et affectives (fonctions apparues plus tardivement). Une dernière phase de céphalisation (néo-kinétique) correspond à l'explosion néo-corticale (remarquable notamment chez les mammifères) et voit émerger des capacités intellectuelles et représentationnelles ouvrant la porte au raisonnement déductifs, facultés qui seront largement mises en œuvre chez les primates pour une intelligence collective complexe menant à des représentations extracérébrales et aujourd'hui à des éléments cognitifs extra corporel et d'intelligence artificielle. (Claverie, 2018)

[78] Dans la classification de Rasmussen reprise notamment par Tosquelles :
Tosquelles, F. (2003). II. Signal, signe et symbole. In F. Tosquelles, Cours aux éducateurs (pp. 31-44). Nîmes, France: Champ social.

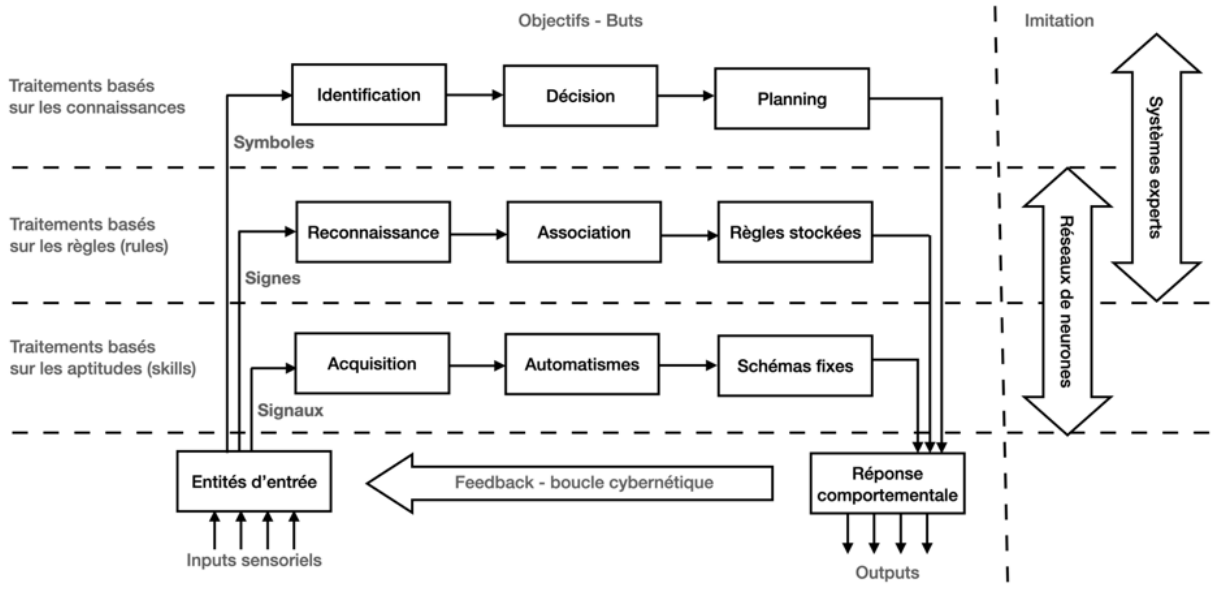


Figure 80. Organisation hiérarchique fonctionnelle des niveaux d'intelligence. Chaque niveau d'information traité selon des compétences qui mobilisent des aptitudes de base, des règles cognitives ou des connaissances de haut niveau. Chaque ensemble informationnel est décomposé et traité en parallèle par les différents étages, selon les objectifs et buts implicites ou explicites de l'individu. L'imitation machinique en IA permet de rattacher les niveaux inférieurs aux performances des réseaux connexionnistes, alors que les étapes supérieures sont référées aux performances des systèmes symboliques (Claverie 2018, adapté de Rasmussen)

Ainsi a-t-on pu hiérarchiser des niveaux de décision. Certains peuvent être confiés à des systèmes automatiques. Certains nécessitent un traitement complètement humain. D'autres, hybrides, peuvent être en partie confiés à un ou des automates et nécessiter l'intervention d'un ou plusieurs opérateurs pour l'(les) autre(s) partie(s). Les décisions sont alors programmables, non programmables ou partiellement programmables (cf. supra : Simon).

En considérant le cycle de décision d'un commandeur d'opérations militaires par exemple, l'ensemble des processus (management de l'information, planification, décision, programmation, évaluation) est géré de manière cyclique avec des outils traditionnels (fiches, tableurs et diaporamas). Ces derniers mobilisent de nombreuses personnes dans des tâches chronophages et à faible valeur ajoutée. Le niveau global de performance devient de plus en plus fragile face à la complexité et à la surcharge d'informations dues à la capacité technologique, sans précédent, à capter, transmettre et recevoir des données, à les combiner de manière significative (signes) pour leur donner une valeur décisionnelle (symboles)

Ainsi, il devient fondamental de pouvoir fusionner intelligemment et rapidement des sources d'information multi-domaines et d'augmenter les capacités humaines par des technologies numériques pour pouvoir à la fois donner un sens à l'ensemble informationnel et aider à comprendre la dynamique des événements.

Dans cette perspective, le projet ANTICIPE[79], porté par le laboratoire HEAL[80] et la société Thales-Raytheon Systems dans le cadre du groupe de recherche IST-157[81] (intitulé *Human Considerations for Artificial Intelligence in C2*), travaille à l'augmentation du processus de CCIR[82] (*Commander's Critical Information Requirements*), comme une illustration de l'apport de l'approche *man machine teaming* pour le C2.

Il s'agit d'un processus central du commandement des opérations. Les CCIR sont des informations requises identifiées par les commandants au cours des phases de planification comme étant essentielles pour faciliter les décisions clés et pour garantir la stratégie souhaitée. Ce sont des éléments de connaissances nécessaires de haut niveau. C'est donc un outil d'anticipation ; il dirige la collecte, l'analyse et la diffusion d'informations clés appuyant les principales décisions.

L'anticipation consiste à construire des règles d'action pour des situations hypothétiques (choisies d'après la probabilité d'occurrence ou le risque leur étant associé) afin d'éviter la prise de décision en situation inconnue sous forte contrainte temporelle, extrêmement coûteuse en temps et en ressources. L'objectif d'un tel outil de commandement est donc de pouvoir aisément ramener le système dans une boucle de traitement de l'information de niveau *rules*, pour lequel le temps de réaction sera plus rapide.

Pour ce faire, ANTICIPE repose sur la notion simple mais innovante de décomposition d'une information critique (CCIR) en deux sous-niveaux d'information, appelés *triggers* et *cues* (ou signaux faibles), qui sont liés par des règles définies au cours du processus de planification.

Cet arbre informationnel (indices + déclencheurs + CCIRs), parallèle analogique aux « signaux-signes-symboles » de Rasmussen, constitue les ontologies qui seront utilisées pour rechercher les informations dans les documents transitant dans le Centre de commandement ou de gestion de crise.
Le soutient au commandeur et aux opérateurs porte alors sur trois types de tâches macrocognitives propres à la plupart des systèmes sociotechniques (Parasuraman et al., 2000) : l'analyse des données, le développement d'une représentation partagée et la prise de décision. Chacune de ces tâches représente un niveau de collaboration potentielle entre un/des humains et un/des agents autonomes.

Les *cues* (signaux, pistes informationnelles issues d'un premier niveau d'intelligence artificielle) sont extraites par analyse automatisée des signaux provenant de toutes les sources disponibles (documents, chat, courriel, communications vocales, notifications, sources ouvertes). Cette assistance à la recherche d'indices est le premier niveau de collaboration où l'IA se substitue en grande partie à l'humain pour analyser la masse de données issues du C4ISR[83].
L'accumulation et l'organisation intelligente (second niveau d'intelligence artificielle) de ces indices permet l'émergence d'une interprétation de l'environnement. Les triggers sont obtenus par agrégation

[79] ANTICIPE signifie *Augmented Near real time Instrument for Critical Information Processing and Evaluation* ou « Instrument amélioré en temps quasi réel pour le traitement et l'évaluation des informations critiques ». Pour plus de détails sur le projet, voir (Desclaux, Marion, & Claverie, 2019).
[80] HEAL (*Human Engineering for Aerospace Lab*) est un laboratoire commun THALES/ENSC
[81] IST est un panel de la STO (Science and Technology Organization : branche scientifique de l'OTAN) qui traite des « *Information Systems Technology* » ou « Technologies des systèmes d'information », dans lesquelles on trouve celles consacrées aux dispositifs hybrides ou impliquant des humains dans le traitement de l'information.
[82] CCIR-AICC : « *Commander's Critical Information Requirements* » ou « Analyse de besoin en Informations Critiques pour le Commandeur ». Les besoins d'information cruciaux du commandeur d'une opération militaire comprennent les informations identifiées par le commandeur comme étant essentielles pour faciliter la gestion de l'information et la décision en lieu et en temps opportuns pour contribuer à la réussite de l'opération.
[83] C4ISR : « *Command, Control, Communications, Computers, Intelligence, Surveillance, Reconnaissance* », quelquefois complété par TAR « *Target Acquisition and Reconnaissance* » - partie technologique et d'usage du C2.

de signaux faibles. Ils représentent un niveau de création de sens en référence à des registres mémoriels (bases de données, comparaisons multiples, etc.). Cette interprétation, dictée par des règles d'association automatique ou manuelle d'indices, est l'objet d'une collaboration active entre automates et opérateurs. Les triggers (signes, dans la classification de Rasmussen) permettent d'activer différents états des CCIR. Le système est à même de proposer instantanément des mesures de prévention, d'atténuation ou majeures selon le niveau critique de chaque CCIR. Ces modes d'action présentés à la décision humaine sont le fruit d'une IA hybride reposant sur des accords de collaboration. Le suivi de l'évolution des informations, s'effectue alors par le biais d'interfaces adaptatives, supports au partage de conscience de situation entre les opérateurs et le commandeur.

Le processus « CCIR ainsi augmenté » doit permettre d'offrir un niveau de performance robuste pour assurer une prise de décision rapide et des réponses pertinentes et sécurisées en temps voulu dans le contexte de complexité et de déluge de données caractéristiques des guerres modernes.
Ainsi doté de propriétés d'adaptation dynamique à l'environnement, le C2 « augmenté » se rapproche des objectifs d'évolution formulés par Alberts & Hayes (Alberts & Hayes, 2006).

L'augmentation, représente alors la contextualisation de la recherche technique à l'Humain et à son environnement (Figure 81).

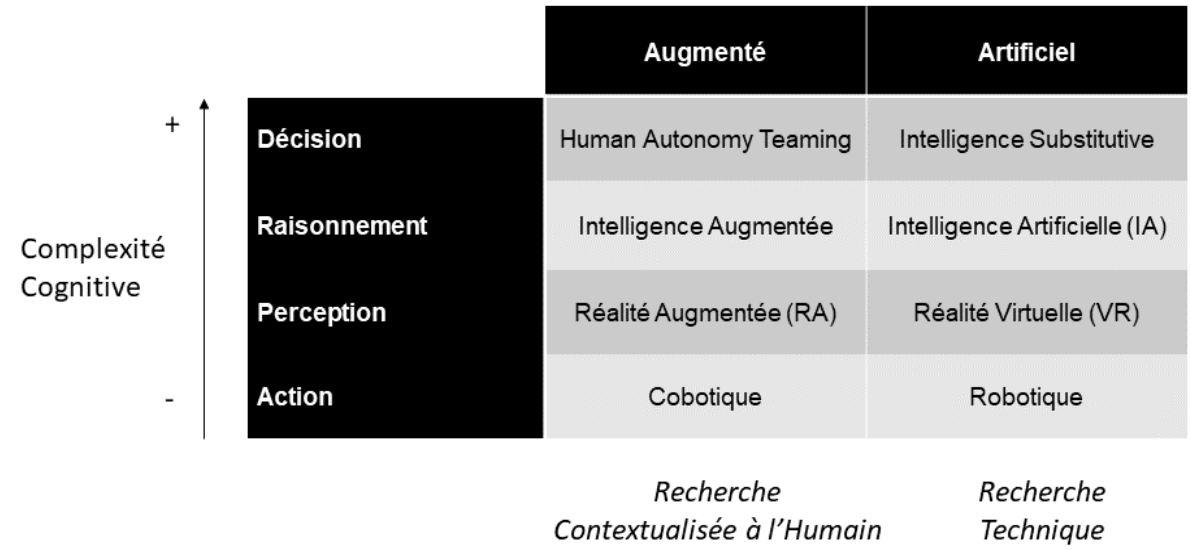


Figure 81. Les recherches technologiques peuvent être séparées selon le niveau de complexité de la boucle cognitive/cybernétique auquel elles s'intéressent. On peut également effectuer une distinction selon l'objectif d'artificialité ou d'augmentation auquel elles répondent ; l'artificiel découlant de recherches purement techniques, tandis que des recherches contextualisées tendent vers l'augmentation de l'Humain dans son environnement.

Le système ANTICIPE est conçu comme une agrégation de tâches de collaboration dynamique entre agents autonomes et agents humains, formant une équipe (HAT[84]) au service de l'amélioration de la représentation de la situation en soutient à la prise de décision du commandeur. L'autonomie n'arbore pas un format unique mais celle d'une multitude d'agents adoptant différents niveaux et formes de collaboration. L'humain peut intervenir à chaque niveau, pour vérifier, confirmer, infirmer ou compléter les informations. Il n'est pas cantonné à un rôle de supervision, puisqu'il peut nourrir et augmenter le système tout en étant alimenté d'informations nouvelles par lui.

Dès lors, il ne s'agit pas d'adopter une approche substitutive aveugle de l'automatisation qui tenterait de remplacer l'humain à chacun des trois niveaux de compétences, de règles ou de connaissances.

La conception des systèmes autonomes et des interfaces associées doit être orientée vers le soutien de l'humain dans la perception et l'interprétation des contraintes (signaux, signes et symboles). Cela permettra d'aboutir à la construction d'une représentation éclairée et exacte de la situation de l'équipe, aidant l'utilisateur à explorer l'espace des possibles.

De même, dans une approche de travail d'équipe (HAT), la machine doit bénéficier du support de l'humain. Cela peut passer par des feedbacks sur ce qu'elle fournit (action, information, suggestion), mais aussi par l'accès à une représentation de l'état, cognitif, physiologique ou physique de son coéquipier humain.

Le défi, aujourd'hui, est de dépasser l'approche classique basée sur une description dichotomique des compétences : « l'humain est meilleur à… » / « la machine est meilleure à… ». L'approche classique tend à réfléchir aux limitations et à l'optimisation de l'un et de l'autre. Elle résulte généralement d'une conviction qu'une approche techno-centrée est meilleure, alors qu'elle handicape la qualité de conception des interfaces et des processus d'équipe.

Il convient d'adopter, dans un esprit de qualité, une perspective holistique et de penser la conception en termes d'intégration des capacités humaines et technologiques dans les systèmes sociotechniques. La réussite dans des domaines complexes dépendra en fin de compte de la capacité des humains ET des technologies à travailler en équipe.

Le *human-autonomy teaming* représente un réel changement de paradigme pour penser les organisations de demain de manière systémique. Il s'agit d'œuvrer dans le sens de l'hybridation de l'un et l'autre et non plus du remplacement de l'un par l'autre. Cette approche demande de porter attention et de mobiliser une réflexion anthropotechnique sur la structure même du C2. Elle doit définir les complémentarités et redéfinir les rôles de chacun, la répartition des tâches et l'organisation des processus. Cela représente probablement une évolution significative de l'organisation du C2 actuel, à tous les niveaux hiérarchiques, conduisant, *in-fine*, à un C2 du futur.

6.3.2 La complexité compréhensible, flagship de l'Europe de la défense

Jusqu'à récemment, l'automatisation a principalement été pensée pour épauler ou remplacer l'humain dans sa tâche principale (*taskwork*). Or, en considérant le système homme-machine comme une équipe

[84] HAT : *Human Autonomy Teaming* - ou MMT pour *Man-Machine Teaming* plus fréquemment utilisé en France. Le HAT caractérise une collaboration entre un ou des personnels humains et un ou des agents autonomes, au-delà de la simple aide à la décision. Le HAT est considéré comme essentiel pour interagir avec une technologie autonome et partager des objectifs, des informations et des représentations mobilisant un langage commun. Le concept a été initié par la NASA, puis par la DARPA pour être appliqué à l'ensemble des forces armées US.
Voir par exemple : (Ho et al., 2017).

à part entière, l'approche HAT requiert également d'étudier les processus de collaboration propres au travail collectif (*teamwork*).

Si l'on se réfère au modèle de conscience de situation d'équipe de Wickens *et al.*(2009), la compréhension de l'environnement en situation de travail collaboratif inclut nécessairement un niveau de représentation des coéquipiers. En plus d'interpréter les signaux provenant de l'environnement extérieur à l'équipe, l'individu doit se tenir au fait de l'état cognitif de ses coéquipiers afin d'initier en temps voulu les communications, les échanges et les stratégies adaptées.

La performance du groupe repose donc en grande partie sur cet aspect métacognitif, inhérent à la collaboration. Dès lors, il s'agit de fournir à des coéquipiers de nature différente (humain et machine) les outils pour construire une compréhension mutuelle, support de cette collaboration.

Deux approches complémentaires en découlent naturellement (Figure 82). L'une concerne la compréhension de l'humain par la machine ; l'autre vise à amener l'humain à comprendre la complexité (de fonctionnement, de décision…) de cette dernière.

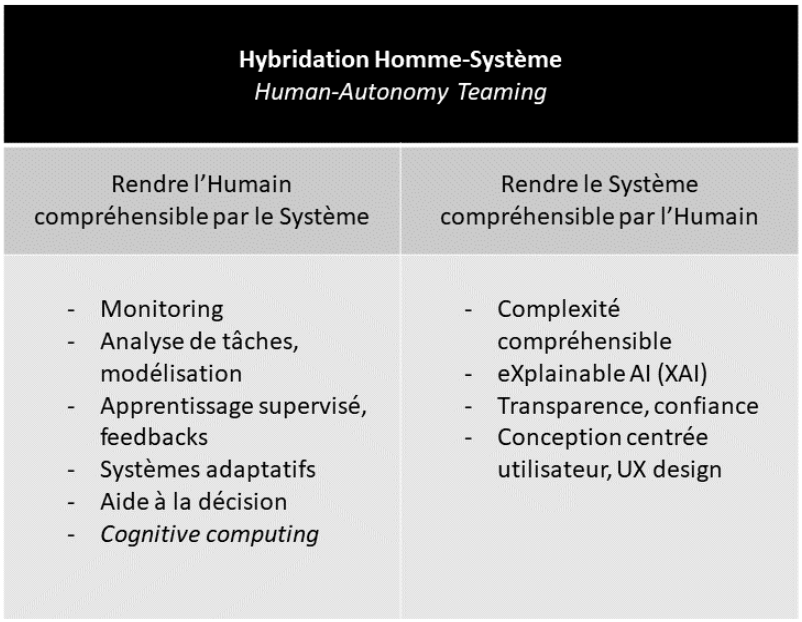


Figure 82. Approches complémentaires dans la conception de systèmes pour le Human-Autonomy Teaming : rendre l'Humain compréhensible par la machine et rendre la machine compréhensible par l'Humain.

La première approche considère que le système doit posséder les outils pour percevoir, interpréter et comprendre la complexité de son environnement humain. Cet aspect est nécessaire à la conception de système adaptatifs, d'allocation dynamique de tâches et plus généralement d'optimisation automatisée continue de la performance collective. Il ancre le besoin d'évaluation en temps réel de l'état cognitif des utilisateurs dans une démarche applicative immédiate qui souligne notamment l'intérêt du monitoring dans ce contexte. L'analyse et la modélisation des tâches et de l'organisation de l'environnement de travail sont également nécessaires pour fournir au système les modèles « mentaux » lui permettant l'interprétation des comportements humains.

Dans le cas où le système se retrouverait privé de certaines informations, que ce soit en raison d'une panne, d'une défaillance d'un capteur ou du comportement de l'humain, il serait passagèrement amputé d'une partie de sa capacité à percevoir ou interpréter correctement la situation.

Cette cécité partielle et temporaire, assimilable à un syndrome de sortie de boucle (*Out Of The Loop*), pourrait s'avérer problématique, la machine devant se mettre à jour avec l'intention ou la stratégie humain entreprise dans ce laps de temps.

En ce sens, la relation humain-machine en cas de sortie de boucle de la machine doit être réfléchie. Tout comme des protocoles de reprise en main par l'humain suite à une délégation à la machine sont au centre des réflexions sur le véhicule autonome, des protocoles de reprise en main par la machine et d'interaction accrues avec l'humain dans ces situations doivent être étudiés. Dans cette situation, la forte puissance de calcul du système et sa capacité de traitement instantanée d'une grande quantité d'information sont des avantages incontestables pour sa résilience.

La seconde approche considère qu'il est nécessaire que l'humain puisse comprendre le système. Jusqu'à présent l'autonomie était progressivement « apprivoisée » par la formation et l'utilisation répétée. L'évolution vers des systèmes capables de fournir de l'aide ou soutenir des décisions dans des situations de plus en plus complexes et inattendues, nécessite d'approfondir l'étude de cette relation afin qu'à terme un opérateur humain n'ait pas besoin d'une utilisation prolongée sans incident pour faire confiance au système.

L'automatisation doit être au service de la supériorité informationnelle et décisionnelle recherchée par le C2. L'un de ses principaux objectifs est généralement de supporter l'humain dans sa compréhension de la situation et appréhension correcte de l'environnement ; objectif d'autant plus important dans un environnement hautement collaboratif. Dans ce domaine, l'implantation d'IA et de système automatisés, fonctionnant en boite noire, représente un risque, car sans compréhension du raisonnement de la machine, l'homme à deux options : renier sa décision (ou conseil) ou lui faire une confiance aveugle.

La compréhension claire et immédiate de cette complexité apparait alors comme une nécessité, dans une volonté d'amélioration du lien Homme-machine. En renforçant ce lien on diminue également la vulnérabilité/sensibilité aux attaques extérieures.

Cependant l'Humain n'est pas par nature performant quand il s'agit d'interpréter les statistiques pour la prise de décision. La complexité explicable (ou compréhensible) est une doctrine scientifique (SHS) basée sur le constat du refus et/ou de l'intolérance d'usage des technologies numériques, par des sujets lamda, notamment dans le cadre de l'augmentation humaine et de l'hybridité. Ceci est d'autant plus observé qu'augmente la complexité de cet usage ou celle de l'environnement dans lequel il s'insère ou qu'ils génèrent.

Les systèmes complexes ou induisant de la complexité doivent être, dès leur conception, prévus pour les utilisateurs (anthropologie, médecine et ergonomie). Ils doivent être accessibles comme objet de raisonnement (cognitique), indépendamment des différences culturelles ou nationales des groupes amenés à collaborer grâce à ces dispositifs. Notamment, leur logique et leur degré d'autonomie doivent pouvoir être transparentes et facilement compréhensibles. Ainsi, le domaine de l'*eXplainable AI* (« IA explicable » – XAI), se concentre sur la transparence des modèles d'intelligence artificielle. En somme, il s'agit de développer une IA non seulement techniquement explicable mais surtout interprétable et actionnable (Linkov et al., 2020). Le but est de doter ces systèmes des formes/aptitudes communicationnelles permettant aux sujets concernés de comprendre à quoi ils servent (pour faire quoi

?), en quoi ils servent (comment font-ils cela ?) et pourquoi ils servent (quelle est la justification de leur existence ?).

Cela s'articule autour de l'acceptabilité, de la tolérance et de la confiance. La complexité explicable est habituellement évoquée avec les dimensions de la formation des personnels et de leur collaboration. Elle doit pouvoir être comprise par les sujets qui y sont confrontés, permettant de faciliter à la fois les procédures d'apprentissage, mais aussi celles de l'adaptation continue. La complexité explicable doit être conçue selon un plan d'action conceptuel (design) qui prend en compte :

- les points forts et les faiblesses des dispositifs ;

- ceux des interfaces, notamment dans le cas d'interfaces adaptatives ;

- la justification des démarches de décision impliquant l'humain ou d'autonomie substitutive ;

- la logique de ces choix décisionnels ;

- le niveau de confiance à accorder à chaque niveau explicatif ;

- le type d'erreurs susceptibles d'être produits ;

- les mesures correctives (ou d'échappement) en fonction du but à atteindre.

L'IA interprétable et exploitable doit ainsi comprendre des quantifications et des visualisations explicites de la confiance des utilisateurs dans les recommandations fournies par les systèmes. Ce faisant, elle permettra d'examiner et de tester les prévisions des systèmes d'IA afin d'établir une base de confiance dans la prise de décision.

Il est probable que la manipulation et la corruption des données deviennent progressivement l'une des principales préoccupations des nations. En conséquence, l'IA ne peut être une boite noire pour les leaders de l'OTAN. Les preneurs de décision doivent comprendre comment la donnée est collectée et analysée, pour qu'elle prenne sens. Ainsi, et dans la continuité des recherches menées par les groupes de travail STO-IST 160[85] et 141[86], l'automatisation ne devrait pas être laissée aux experts et aux sous-traitants, mais être rendue accessible aux décideurs et conçues avec les utilisateurs.

Les politiques commencent à prendre la mesure de l'importance de cette question. En 2020, dans son Agenda de Recherche Stratégique[87] en Facteurs Humain, l'Agence Européenne de Défense (EDA) place la complexité compréhensive comme l'un des 6 sujets de recherche d'importance majeure pour le domaine de la défense à l'horizon 2027. Sa transversalité l'établit comme un projet *Flagship* de l'Europe de la défense.[88]

[85] Le *Specialists meeting* du groupe de travail IST-160 portant sur le thème *Big Data and Artificial Intelligence for Military Decision Making* s'est tenu du 30 Mai au 01 Juin 2018 à Bordeaux. Il a été co-organisé par la STO et l'ENSC (Bastiaansen, Smallegange, Venema, & Bronkhorst, 2018).
[86] Le groupe de travail IST-141, dans lequel l'ENSC est impliqué, réuni des spécialistes autour de la thématique de recherche *Acute Situation Awareness and Informed Decision Making.*
[87] *Strategic Research Agenda*
[88] Tel que présenté par M. Shazad Ali, Project officer Protection and Human Factors à l'EDA, joint par e-mail et téléphone (Septembre 2020)

6.3.3 **Cyberpsychologie et conflits du futur**

Si jusqu'à présent nous avons discuté de la nécessité de l'approche HAT pour la conception de systèmes d'augmentation de la prise de décision en C2, nous ne saurions terminer ce chapitre sans resituer cette démarche dans le contexte plus général de l'évolution conjointe des conflits et de la technologie.

Les opérateurs de C2 sont chaque jour plus immergés dans le monde numérique, la représentation virtuelle du réel, la programmation et la gestion coordonnée de moyens au service d'une stratégie. Ils disposent d'outils, de méthodes et d'un champ pratique, participant à la construction d'un véritable domaine d'étude théorique, à visée applicative.

Avec l'avènement des machines intelligentes et autonomes, il est devenu primordial de développer une nouvelle forme de psychologie, qui examine la manière dont les humains et les machines se comportent les uns envers les autres. De nombreux scientifiques ont développé une réflexion à ce propos sous le terme de « cyberpsychologie ».

La cyberpsychologie peut être définie comme l'étude des phénomènes mentaux relatifs aux systèmes cybers et leur contexte (Claverie & Kowalxzuk, 2018). C'est un domaine scientifique en développement qui englobe divers phénomènes et différents sous-thèmes. Elle se situe à l'interface de plusieurs disciplines (psychologie, sociologie, cybernétique, intelligence artificielle) et examine la façon dont les humains et les machines s'impactent mutuellement. Dans ce contexte, elle se doit également d'explorer comment cette relation influence les interactions entre les humain et les machines, mais aussi celles des machines et des humains entre eux (Figure 83).

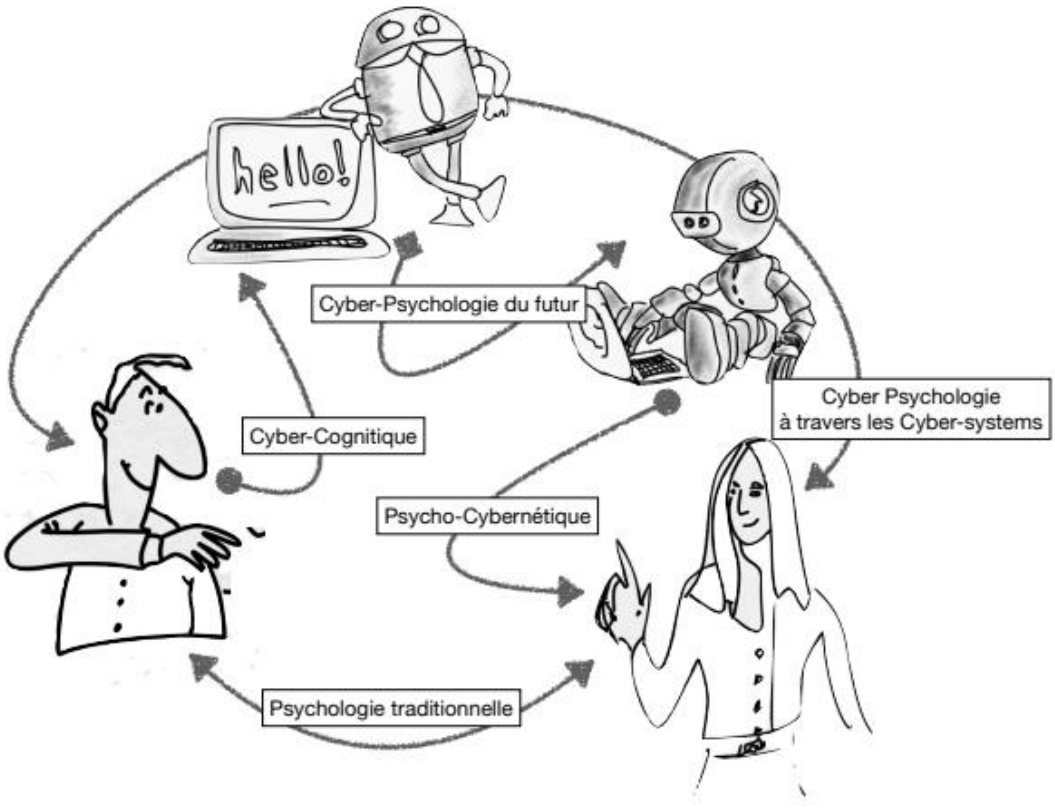


Figure 83. Les champs de la Cyberpsychologie (Claverie & Kowalxzuk, 2018)

L'humanité développe des cyber-outils qui apportent des opportunités mais aussi des risques exceptionnels. Il est nécessaire que des mesures soient prises pour garder ces créations sous contrôle. Alors que la technologie progresse rapidement, certains pourraient oublier que tout dans le cyberespace a d'abord été pensé et créé par l'homme. Si l'on peut tenter de contrôler les processus technologiques et leurs effets, il existe moins de moyens de contrôler la pensée humaine. Cette dernière peut pourtant être utilisée pour produire des outils destructeurs dans le cyberespace. Il faut donc s'efforcer de comprendre les esprits humains qui sont à l'origine du développement de la technologie et d'évaluer les motifs de ces créations.

Pour conclure, la cyberpsychologie s'ancre dans une démarche de responsabilisation des leaders et des technologues, ainsi que de prise de recul concernant l'impact psychologique, comportemental et sociétal des technologies cyber et les risques afférents (NATO Innovation Hub, 2018). En termes de recherche et d'applications, les domaines de la cyberpsychologie sont amenés à prendre une place prédominante dans la transformation de l'OTAN pour adresser les conflits modernes orientés cyber (cyber warfare[89]) et faire face aux adversaires qui sont prompts à utiliser ces technologies émergentes.

Les progrès de la technologie ont augmenté les distances sur lesquelles les individus peuvent rester en contact, la vitesse à laquelle les informations et les instructions peuvent être partagées, la quantité d'informations créées et la charge qui en découle, les possibilités de surveiller le « comportement » des opérateurs et les possibilités d'automatiser les instructions (National Research Council, 1990). La prise de décision, tout comme la représentation partagée sur laquelle elle s'appuie, est désormais « distribuée ». Dans ce contexte de décision décentralisée, une seule erreur d'évaluation de la situation, commise par une machine, un homme ou une équipe réunissant les deux, peut conduire à de graves erreurs de portée mondiale.

Chercher à influencer la représentation de l'adversaire n'est pas nouveau. Depuis des siècles les conflits s'appuient sur des stratégies visant à induire l'adversaire en erreur en brouillant, cachant et manipulant l'information, pour faciliter l'obtention de la supériorité escomptée. Cependant, les outils technologiques actuels et l'orientation prise dans le développement des futurs outils, ouvrent un accès direct à la représentation des individus. Les environnements virtuels deviennent le terrain de jeu privilégié dans lequel la tromperie et la manipulation sont pratiquées à une échelle sans précédent (les récentes manipulations électorales dans les pays de l'OTAN en sont un exemple criant).

Nous pouvons ainsi voir se dessiner les prémices des conflits futurs, ne se jouant pas seulement dans les nouvelles sphères cyber ou spatiales, mais s'appuyant également sur la pleine puissance de manipulation cognitive des outils technologiques.

[89] Terme officiel employé par l'OTAN pour désigner les conflits de type cyber et affrontements ayant lieu dans le cyberespace.

Conclusion

La thématique de la collaboration est un domaine d'étude habituellement abordée en psychologie sociale. Elle a récemment fait l'objet d'un nouvel intérêt de la part des spécialistes de psychologie cognitive, que ce soit à propos de l'étude du partage des représentations, ou de celles des communications et de la répartition des activités entre les membres d'un collectif.

En échangeant des connaissances et en partageant leurs représentations, en manipulant des artefacts ensembles, les membres d'une équipe partagent et coordonnent leurs consciences de situation. Ils deviennent alors conscients des mêmes états et conditions des éléments de l'environnement. Quand la conscience de situation maintenue par certains diverge de celle des autres, le risque de prises de décisions inconsistantes voire incohérentes au sein de l'équipe peut devenir critique et engage le sens même de la collaboration.

L'évaluation de la cognition d'équipe se pose alors comme un nouveau champ d'étude, notamment dans les conditions complexes de collaboration en situation de gestion de crises. Il n'est dans ces cas-là pas question que des représentations erronées soient partagées, que des erreurs ou des incohérences informationnelles interviennent dans les activités partagées. Ce point est particulièrement décisif dans le domaine de la défense, où de telles situations de gestion des grands systèmes critiques sont fréquemment rencontrées. Par exemple, dans le domaine du « Commandement et du Contrôle des opérations de défense » (Command and Control : C2), ensemble sociotechnique par essence collaboratif, elle présente un intérêt tout particulier. La compréhension des mécanismes de la cognition qui sous-tendent la performance des membres de l'équipe, et l'accès à des dispositifs de mesure pertinents, sont décisives pour la conception d'aides technologiques, aussi bien dans le cadre de l'environnement de travail réel des opérateurs militaires que dans celui de la formation.

Cette thèse est une contribution au domaine de la recherche sur la cognition collective et le représentation partagée. Elle situe ses expérimentations dans le domaine du C2 en souhaitant apporter des briques scientifiques au projet du C2 du futur. Il nous semble que celui-là doit s'appuyer sur un travail scientifique qui vient en appui de l'entreprise actuelle, plus du domaine du travail militaire et organisationnel. Cette approche concerne de nombreuses disciplines, dont les sciences cognitives appliquées sont un élément fondamental pour aborder cette problématique.

Le travail n'est pas simple. Il présente deux types de difficultés. La difficulté scientifique est bien entendu celle de toute thèse. Mais, dans le cadre des études à réaliser, nous avons dû faire face aux limitations et aux difficultés de mise en œuvre expérimentale pratique, de l'accès aux informations classifiées ainsi qu'aux convictions motivationnelles des personnes étudiées. C'est une fois ces

problèmes surmontés que nous avons obtenu des résultats qui nous ont incités à approfondir notre réflexion concernant l'impact de la complexité de la situation naturelle sur la mesure de la cognition.

A notre sens, une principale contribution au domaine de l'approche scientifique de la cognition collective concerne une limite à laquelle se heurte l'évaluation de la conscience de situation : son caractère naturellement dynamique. Pour cela nous avons introduit le concept de « synchronie des consciences de situation » (Chapitre 2) comme métrique d'évolution temporelle de leur partage au sein de l'équipe. En proposant un tel concept, nous avons souhaité établir une base de travail et de réflexion sur l'évaluation de la dynamique du partage de conscience de situation, thématique jusqu'alors peu étudiée de manière concrète. Il s'agit de fournir des premiers éléments de réponses sur la nécessité d'une telle évaluation, sur les mesurables associés ainsi que sur les perspectives d'application qui en découlent en environnement simulé comme en environnement réel. Nous retiendrons principalement la mise en évidence de trois latences dans le processus de synchronisation des consciences de situation. Leurs mesures nous semblent pouvoir ouvrir de nouvelles voies pour quantifier, qualifier et répondre en temps réel aux problèmes de partage des consciences de situation et aux risques afférent.

Nous nous sommes confrontés à l'évaluation en temps réel du partage de conscience de situation, à travers 3 chapitres expérimentaux.

Le Chapitre 3 détaille la mise en pratique d'une méthodologie de mesure de partage de conscience de situation entre deux individus, qui repose sur l'interruption de la tâche, procédé courant de ce domaine. L'objectif de cette phase expérimentale est d'évaluer l'utilisation d'une métrique quantitative objective dont le traitement peut potentiellement être effectué de manière automatique, en temps réel, dans un contexte d'entraînement ou de formation. Cette métrique s'étant révélée pertinente, la situation expérimentale mise en place ouvre des perspectives d'étude du partage de cognition et de la dynamique d'équipe pour une éventuelle poursuite de ce travail de recherche.

Nous avons, par la suite, cherché à évaluer l'apport du monitoring pour mesurer le partage de conscience de situation des opérateurs de C2. Dans notre contexte, il offre la possibilité de recueillir des données objectives et quantitatives sur l'opérateur, dont le traitement pourrait être facilement automatisé et continu.

Les contraintes matérielles de l'environnent de recherche, et d'introduction d'une technologie de mesure dans cet environnement particulier et contraint, nous ont amenées à choisir un outil de mesure léger, facilement toléré par les opérateurs. La technique pupillométrique a été adoptée comme méthode de monitoring physiologique, les autres méthodes étant soit trop lourdes, soit incompatibles avec un environnement électromagnétique très parasité et non modifiable. Ce choix oriente nécessairement nos travaux. Ils gagneraient évidemment à être complétés par d'autres technologies plus lourdes, pour une approche multi-modale du monitoring, ce qui ne se fera pas sans une évolution des technologies aujourd'hui disponibles, tant en ce qui concerne le criticisme du recueil des données que l'acceptabilité des outils de mesure par les opérateurs.

À travers une expérience menée conjointement avec l'institut allemand FKIE (Chapitre 4) nous avons pris la mesure de la sensibilité de la pupillométrie et du monitoring comportemental aux modifications de conscience de situation d'un individu seul, impliqué dans une tâche de C2 réaliste. La variabilité comportementale est apparue comme le frein principal à la détection et l'interprétation des réactions psychophysiologiques attendues. Mais l'existence de ces dernières a pu être mise en évidence par d'autres méthodes, cependant moins adaptées à l'utilisation en temps réel. La situation de travail collectif étant naturellement plus complexe et bruitée que la situation individuelle, ces résultats nous ont amenés approfondir l'étude de l'applicabilité du monitoring à la situation de C2.

Nous avons étayé ces observations à l'aide de la situation expérimentale plus complète développée au Chapitre 5. Celle-ci consiste en un monitoring d'opérateurs dans une tâche de C2 collaborative. Dans les conditions expérimentales et d'analyse qui ont été les nôtres, la technique de monitoring pupillaire a été décevante, et elle n'a pas permis de détecter de synchronisation ou désynchronisation des activités des coéquipiers. En situation de collaboration, le bruit comportemental se trouve exacerbé et la prévalence des éléments situationnels communs s'en trouve réduite. L'état psychophysiologique de l'opérateur est apparu avant tout sensible aux événements, actions et intentions entreprises dans son espace d'action, au sein de sa stratégie personnelle. Nous avons néanmoins identifié des pistes de méthodes d'analyse qui ouvrent des perspectives pour l'étude de la dynamique pupillaire et de la synchronisation des phénomènes de consciences de situation.

La transposition des connaissances de laboratoire à une application en situation complexe, dans laquelle s'ancre notre démarche, soulèvent la problématique, à notre sens capitale, de la complexification de la situation d'étude. Les marqueurs de cognition en situation réelle sont loin d'être aussi accessibles qu'en laboratoire mais peuvent être mis en évidence par des méthodes appropriées. Pour les futurs travaux, il ne s'agit donc plus de questionner l'existence de ces réactions, mais de travailler sur leur identification, leur « détectabilité », leur « interprétabilité ». Le développement de méthodes automatiques pertinentes et fonctionnelles que nous entrevoyons, devra être accompagné d'une validation mettant en œuvre des méthodes d'analyses classiques et qualitatives.

Afin de dépasser les limites que nous avons pu mettre en évidence, il est nécessaire d'explorer et identifier les configurations techniques du monitoring et les méthodes d'analyse permettant de faire émerger du sens des données « réelles », c'est-à-dire complexes, plongées dans le bruit et la variabilité.

Le Chapitre 6 rend compte de formations, d'entretiens et d'observations dont nous avons tiré des conclusions sur l'utilisation du monitoring en situation de travail réel en C2. Nous avons formulé des contraintes concernant le matériel pertinent, analysé les risques que la liberté comportementale fait peser sur l'interprétabilité des données, et signalé les risques cyber qui s'en trouvent exacerbés.

Nourris par le vécu et les résultats des expériences, nous proposons alors des recommandations concernant les contraintes et dispositifs de recherche nécessaires à l'expérimentation portant sur l'évaluation de la cognition collective. Notamment nous précisions celles afférentes à l'utilisation de la simulation comme outil de recherche en laboratoire pour l'étude de la prise de décision en situation naturelle.

Enfin, ces travaux ne sauraient se dispenser d'une réflexion sur la place de la cognition, de son étude et de son évaluation, dans les enjeux technologiques et humains pour le C2 du futur.

Le projet du Human-Autonomy Teaming, d'un travail collaboratif effectué en symbiose avec les systèmes technologiques, de la machine comme « coéquipier à part entière », soulève une série de questions. Tenter d'y apporter des réponses représentera la tâche incontournable des futurs travaux de ce domaine.

Comment renforcer les liens entre coéquipiers humains et machines ? Comment un ou des automates peuvent-ils aider, sans interrompre les processus humains ou compromettre les capacités humaines ? Comment assurer le partage des représentations entre humains et machines ? Comment un automate peut-il mériter la confiance d'un opérateur ? L'automate peut-il remettre en question la décision de l'opérateur ?

Ces questions adressent des champs d'étude aussi divers que complémentaires. Pour espérer y répondre, les scientifiques et les industriels doivent, à notre sens et dès maintenant, puiser dans la transdisciplinarité. Pour cela, et dans un cadre cyberpsychologique, les domaines de la cognition collective, de l'interaction homme-système, de la psychologie sociale, de la psychophysiologie, ou encore ceux de l'intelligence artificielle et de la réalité augmentée, sont appelés à se retrouver pour aborder la complexité des défis à relever. Pour cela, le monitoring des opérateurs est porteur d'un potentiel évident pour permettre à la machine de « comprendre » l'humain. Il convient de continuer à l'explorer pour mieux l'utiliser, notamment par l'amélioration des méthodes de traitement et le croisement des sources d'acquisition.

Sur un autre plan, ce travail doit également se poursuivre de manière résolument orientée vers la prise en compte des facteurs inhérents à la situation de travail écologique. Il s'agit d'une ingénierie scientifique, non d'une seule approche instrumentale. La démarche conceptuelle vise avant tout à contribuer à la progression des idées et des connaissances. Nous nous situons résolument, en ce sens, dans la perspective d'une ingénierie cognitique, conçue comme part entière de la constellation interdisciplinaire des sciences cognitives, articulant neurosciences, psychologie et informatique. A ce titre l'ingénierie n'est pas une technicité, c'est une approche scientifique, raisonnée, tournée vers l'application et la résolution de problèmes complexes. Appliquée à l'homme et à la machine, cette ingénierie définie l'humain comme élément du système devant être conçu comme un tout, une vision systémique d'une équipe humain-autonomie (HAT). Cela a d'ailleurs ouverts de nombreux débats au sein du laboratoire, à propos de la valeur scientifique de cette notion, et de sa force explicative, mais cela dépasse les limites formelles de cette thèse.

La décision, en tant qu'objet cognitif, reste d'un grand intérêt pour les militaires. Elle concerne évidemment les alliés comme l'adversaire, et leur connaissance réciproque. Les progrès technologiques ouvrent en effet un accès direct à la manipulation des représentations qui la motive. Dans un milieu où la décision partagée est une des composantes du travail collaboratif, la conscience de situation partagée devient un des nœuds de la cognition augmentée, voire hybride. La cognition devient ainsi progressivement le terrain des batailles de demain, où neurosciences, technologies de l'information et psychologie seront les armes au service des stratégies militaires.

La « cognitive warfare » représente donc la convergence de tous les éléments de l'information warfare (IW) élargie à des notions opérationnelles de psychologie, de psycho-sociologie et de neurosciences. Ce terrain scientifique qui s'ouvre se place à l'interface des opérations psychologiques (PsyOps), des cyberopérations (cyberdéfense) et des opérations d'influence (OMI). Le principe majeur n'est pas uniquement d'accompagner une stratégie et de gagner sans se battre, c'est aussi une guerre contre ce que pense, aime ou croit une communauté. L'objectif affiché est d'attaquer, détériorer voire détruire le tissu de la décision, de la confiance ou même de l'espoir nécessaire au fonctionnement d'un groupe constitué en ennemi.

Ce type d'approche cognitique n'est pas à considérer selon les classifications des instruments de la guerre traditionnelle, mais plutôt comme l'accès à des outils de perturbation d'individus, ciblée ou de masse, pour posséder un impact à plusieurs échelles, allant de l'homme seul au système socio-politique dans son entier.

La construction de représentations communes, que ce soit de la situation, des outils, des concepts, des organismes ou des structures encadrant la collaboration, au sein et entre les forces et les nations, est au centre de ces enjeux. Elle constitue de fait un facteur clé pour l'efficacité du C2. Elle doit permettre à ce dernier de répondre aux mutations à l'œuvre dans le paysage des conflits. Le partage de

représentations requiert l'attention particulière et conjointe d'une collectivité qui ne semble pas encore en prendre toute la conscience : chercheurs, industriels, militaires et citoyens impliqués dans une culture de la défense. Pour cela, il est nécessaire d'orienter la conception des systèmes, l'éducation et la formation, du commandeur au technicien, vers les problématiques humaines en lien avec la technologie. Des recommandations dans ce sens ont déjà été formulées, notamment par les experts du *High Level Group* lors du forum franco-allemand sur la « *Global Security / Civil Security Research* » (2018), ainsi qu'à la suite du *Young Disruptors Forum*[90] qui s'est déroulé lors de la *Chiefs Of Transformation Conference*[91]*2018* (COTC)(2018).

C'est ainsi que pour adresser la question du rôle de plus en plus prépondérant de la cognitique dans le C2, il nous semble indispensable que les différents acteurs de la recherche pour une défense globale (DGA, EDA, AID...) puissent établir un environnement scientifique partagé, interdisciplinaire, propice à l'échange et à la collaboration, favorisant justement l'émergence de représentations partagées.

[90] *Young Disruptor* est le terme utilisé dans le cadre de la COTC pour définir un professionnel, homme ou femme, âgé de 22 à 32 ans, militaire ou civil, spécialisé dans le domaine des relations internationales, de la diplomatie, de la sécurité internationale, de l'économie, des sciences, des technologies, de l'ingénierie et/ou des mathématiques, qui travaille dans une agence gouvernementale, une industrie, un groupe de réflexion ou un institut de recherche. Durant la NATO COTC 2018, chaque nation était invitée à convier un représentant *Young Disruptor* pour participer de manière indépendante aux sessions plénières et contribuer au *Young Disruptors Forum.*
Dans le cadre de la COTC 2018, j'ai été invité à participer en tant que *Young Disruptor* de la délégation française.

[91] Chiefs Of Transformation Conference (COTC), est un événement annuel clé organisé par l'ACT de l'OTAN afin de partager un niveau de compréhension plus élevé sur les défis de sécurité futurs, l'innovation et la transformation. C'est le seul événement qui réunit les chefs nationaux de la transformation, de l'ensemble de l'Alliance et des pays partenaires, avec les hauts responsables de l'OTAN pour une interaction et une discussion ouvertes sur les initiatives nationales de transformation et le développement des capacités à long terme.

Glossaire

AAE	Armée de l'air et de l'espace
ACCS	Air Command and Control System
ACT	Allied Command Transformation (OTAN)
ACT	Allied Command Transformation
AO COE	Air Operations Center Of Excellence (designation anglo-saxonne du CASPOA)
AWACS	Airborne Warning and Control System
AWC	Air Warfare Center (désignation anglo-saxonne du CEAM)
BTH	Between The Heads
C2	Command and Control (Commandement et Conduite des opérations)
C4iSR	Commandement, Contrôle, Communications, Computers, Intelligence (renseignement), Surveillance, Reconnaissance.
CASPOA	Centre d'Analyse et de Simulation pour la Préparation aux Opérations Aériennes
CCIR	Commander's Critical Information Requirements
CDAOA	Commandement de la Défense Aérienne et des Opérations Aériennes (BA-942, Lyon Mont Verdun)
CDC	Centre de Détection et de Contrôle
CEAM	Centre d'Expertise Aérienne Militaire (BA-118, Mont-de-Marsan)
CIH	Equipe « Cognitique et Ingénierie Humaine » de l'IMS
CNOA	Commandement National des Opérations Aériennes
CWCT	Collaborative Warship Commander Task
D2D	Data To Decision
DARPA	Defense Advanced Research Projects Agency

DGA	Direction Générale de l'Armement
ENSC	Ecole Nationale Supérieure de Cognitique
GDTA	Goal Directed Task Analysis
HAT	Human-Autonomy Teaming
HFM	Human Factors and Medicine
IIL	Initial Integration Latency
IMS	Laboratoire « Intégration du Matériau au Système » - UMR 5218 CNRS
IST	Information Science and Technology
ITH	In The Head
JFAC	Joint Force Air Command
LOE	Line of Engagement (Ligne d'engagement)
NATO	North Atlantic Treaty Organization (designation anglo-saxonne de l'OTAN)
NDM	Naturalistic Decision Making
NKE	Necessary Knowledge Element
NSKE	Necessary Shared Knowledge Element
OTAN	Organisation du Traité de l'Atlantique Nord
S2T	Strategy To Task
SA	Situation Awareness
SAGAT	Situation Awareness Global Assessment Technique
SHS	Sciences Humaines et Sociales
SSA	Shared Situation Awareness
STO	Science and Technology Organization
TIL	Team Integration Latency
TSA	Team Situation Awareness
TSL	Team Synchronization Latency
WCT	Warship Commander Task

Productions scientifiques et techniques

Publications et communications avec actes

- **Human Computer Interaction International** (HCII 2020) – Copenhague, DK – (conférencier invité) Communication orale, publication dans les actes - Prébot, B., Schwarz, J., Fuchs, S., & Claverie, B. (2020). From "knowing what" to "knowing when": Exploring a concept of situation awareness synchrony for evaluating sa dynamics in teams. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *12214 LNCS*, 505–518.
- **International Astronautical Congress** (IAC 2019) - Washington D.C., DC, USA – Communication orale, publication dans les actes - Prébot, B., Cavel, C., Calice, L., Mahaut, M., Leduque, A., & Salotti, J. M. (2019). Team performance analysis of a collaborative spatial orientation mission in Mars analogue environment. *Proceedings of the International Astronautical Congress, IAC*, *2019-Octob*(October), 21–25.
- **International Conference on Applied Human Factors and Ergonomics** (AHFE 2019) - Washington D.C., DC, USA – Communication orale, publication dans les actes – Prebot B., Salotti JM., Vennin C., Claverie B. (2020) Shared Spatial Situation Awareness as a Team Performance Indicator in Collaborative Spatial Orientation Task. In: Boring R. (eds) *Advances in Human Error, Reliability, Resilience, and Performance. AHFE 2019. Advances in Intelligent Systems and Computing*, vol 956. Springer, Cham.
- **EGU General Assembly** (2019) (co-auteur) - Viennes, AU - Poster et publication dans les actes - Salotti, J. M., Prébot, B., & Cinelli, I. (2019). Situation awareness issues during outdoor activity. *Geophysical Research Abstracts*, *21*, 2484–2491.
- **International Command and Control Research and Technology Symposium (ICCRTS 2018)** (co-auteur) - Pensacola, FL, USA - Communication orale, publication dans les actes - Desclaux, G., & Prébot, B. (2018). Command and Control at the Autonomy and Cognitive Era : For a decision cycle augmented by the symbiosis between human and systems. *Proceedings of the 23rd International Command and Control Research and Technology Symposium: Multi-Domain C2*.
- **European Conference on Computer Supported Cooperative Work (ECSCW 2018), Doctoral Colloquium** - Communication orale, publication dans les actes - Prébot, B., Claverie, B., & Salotti, J. M. (2018). Real-time teamwork evaluation and C2 crisis management: Overview of doctoral research. *ECSCW 2018 - Proceedings of the 16th European Conference on Computer Supported Cooperative Work*, 1–4.

Communications et conférences

- **Séminaire Annuel équipe AUCTUS Inria/ENSC** (2019), Talence (Communication orale) **–** *Conscience de situation et Cobotique* – Prébot, B.
- **Table ronde ANORAA** « IA et nouvelles approches du C2 » (2019), Talence
- **NATO Chiefs Of Transformation Conference (COTC 18) Young Disruptors Workshops**, Norfolk, US, Décembre 2018.
- **IST-160 Specialists Meeting** « Big Data and Artificial Intelligence for Military Decision Making » (2018), Bordeaux (Communication orale), Young Scientists Elevator Pitch - Prébot, B.
- **IST-160 Specialists Meeting** « Big Data and Artificial Intelligence for Military Decision Making » (2018), Bordeaux (Poster), *Operator monitoring for real-time team cognition assessment in C2 crisis situation.* - Prébot, B., Claverie, B., Salotti, J.M
- **DAS IHS pôle Aerospace Valley** (2017), Agen (Communication orale) – *Représentation partagée, travail collaborative et gestion de crise dans le C2 du futur* – Prébot, B.

Notes

- Note on Data centricity in command and control and military décision making – pour l'Etat-major des Armées (EMA) dans le cadre de la participation française à la COTC18 - Norfolk, US, 2018 – (**Annexe 6**)

Participations

Évènements

- **Colloque de la chaire STAH** (Systèmes Technologiques pour l'Augmentation de l'Humain), *Augmentation de l'humain : vers des systèmes cognitivement augmentés*, Talence, FR, 2019
- **Matinale de l'Augmentation de l'Humain**, Bordeaux, FR, 2019, *L'Expérience Utilisateur : 10 ans plus tard où en est-on ? Quel est l'après ?*
- **NATO Chiefs Of Transformation Conference (COTC 18)**, *Disruptive Advances Shaping Warfare* – en tant que « Young Disruptor » de la délégation française, identifié par l'EMA, Norfolk, US, Décembre 2018.
- **IST-160 Specialists Meeting** « Big Data and Artificial Intelligence for Military Decision Making », Bordeaux, FR, Mai 2018
- **Hackathon Sciences Po Executive Education / EMA Etat-Major des Armées**, Paris, FR, 2018, *Planifier dans l'urgence, décider dans l'incertitude, conduire en agilité à l'ère de l'Intelligence Artificielle ?*
- **NATO Chiefs Of Transformation Conference (COTC 17)**, *Prepare for Tomorrow, Share the Future*, Norfolk, US, Décembre 2017.
- **Lecture Series IST-143**, *Cyber Security Science and Engineering*, Talence, FR, 2017.
- **NATO Command and Control Center Of Excellence (C2COE) Annual Seminar**, Valence, ES, 2017.

Encadrement d'étudiants

- Projet étudiant « Mars Explore V2 » (2020)
- Projet étudiant « ACDC, Méthode d'analyse de la cognition d'équipe » (2020)
- Encadrement de 2 stagiaires ingénieurs (2020) concernant l'applicabilité des processus C2 de gestion de crise aux urgences psychiatriques dans le cadre de la crise sanitaire du COVID-19.
- Projet étudiant « Mars Explore » (2018)

- Encadrement d’1 stagiaire ingénieur (2018) sur le traitement de données des pré-tests de l’expérience MARS Explore
- Co-encadrement d’un Projet de Fin d’Etudes (PFE) (2017) – « MOCOM : Cognition collective et monitoring ».
- Projet étudiant « Miroir connecté » (2017)
- Projet étudiant « Space Invaders Head Tracking » (2017)

Bibliographie

Adams, M. J., Tenney, Y. J., & Pew, R. W. (1995). Situation awareness and the cognitive management of complex systems. *Human factors*, *37*, 85-104.

Alberts, D., & Hayes, R. E. (2003). Power to the Edge: Command and Control in the Information Age. In *Information Age Transformation Series*. CCRP.

Alberts, D., & Hayes, R. E. (2006). Understanding Command and Control. In Assistant secretary of defense (Éd.), *The future of command and control* (Vol. 71). Washington DC: CCRP.

Amalberti, R. (1998). Les Facteurs Humains à l'aube de l'an 2000. *Phoebus*, 5-12.

Artman, H. (1999). Situation awareness and co-operation within and between hierarchical units in dynamic decision making. *Ergonomics*, *42*, 1404-1417.

Bailey, L. L., & Willems, B. F. (2002). The moderator effects of taskload on the interplay between en route intra-sector team communications, situation awareness, and mental workload. *FAA Office of Aviation Medicine Reports, DOT/FAA/AM*, 1-14.

Baroutsi, N. (2015). Measures of Command & Control Performance. *Proceedings of the 20th International Command and Control Research and Technology Symposium: C2, Cyber, and Trust*, 21. Annapolis,MA,USA.

Bastiaansen, H., Smallegange, A., Venema, A., & Bronkhorst, A. (2018). Big Data and Artificial Intelligence for Military Decision Making: Dutch Position Paper. *STO-MP-IST-160*, 11. Bordeaux, France.

Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin*, *91*, 276-292.

Beatty, J., & Lucero-Wagoner, B. (2000). The pupillary system. In *Handbook of psychophysiology (2nd ed.).* (Vol. 1039, p. 142-162). Cambridge University Press, xiii.

Behymer, K. J., & Flach, J. M. (2016). From Autonomous Systems to Sociotechnical Systems: Designing Effective Collaborations. *She Ji: The Journal of Design, Economics, and Innovation, 2*, 105-114.

Bell, B. S., Kanar, A. M., & Kozlowski, S. W. J. (2008). Current issues and future directions in simulation-based training in North America. *The International Journal of Human Resource Management*, *19*, 1416-1434.

Bellet, T., & Tattegrain-Veste, H. (2003). 3. Cosmodrive : un modèle de simulation cognitive du conducteur automobile. In J.-C. Sperandio (Éd.), *Formalismes de modélisation pour l'analyse du travail et l'ergonomie* (p. 77-110). Paris cedex 14: Presses Universitaires de France.

Berggren, P., Kermarrec, Y., Banko, K., Wikberg, P., & Oleksandur, B. (2014). *Measuring and Analyzing Command and Control Performance Effectiveness: Final Report of Task Group HFM-156*. NATO, Science and Technology Organization.

Berggren, P., Prytz, E., Johansson, B., Nahlinder, S., & Nählinder, S. (2011). The relationship between workload, teamwork, situation awareness, and performance in teams: a microworld study. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *55*, 851-855. SAGE Publications Sage CA: Los Angeles, CA.

Bergson, H. (1939). *Matière et mémoire. Essai sur la relation du corps à l'esprit.* Les Presses universitaires de France.

Bertelson, P. (1967). The time course of preparation. *The Quarterly journal of experimental psychology*, *19*, 272-279.

Binda, P., Pereverzeva, M., & Murray, S. O. (2014). Pupil size reflects the focus of feature-based attention. *Journal of neurophysiology*, *112*, 3046-3052.

Bolstad, C. A., Foltz, P. W., Franzke, M., Cuevas, H. M., Rosenstein, M., & Costello, A. M. (2007). Predicting Situation Awareness from Team Communications. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *51*, 789-793. SAGE Publications Sage CA: Los Angeles, CA.

Bornard, J.-C., Bellet, T., Mayenobe, P., Gruyer, D., & Claverie, B. (2011). A perception module for car drivers visual strategies modeling and visual distraction effect simulation. *Proceedings of the 1st international Symposium on Digital Human Modelling (IEA-DHM)*, 1-6. France.

Bowers, C., & Braun, C. C. (1997). Team workload: Its meaning and measurement. *Team performance assessment and measurement: Theory, methods, and applications*, 85-108.

Boyd, J. (1976). *Destruction and Creation* (US Army Co). US Army Co.

Boyd, J. (1987). The Essence of Winning and Losing. *A Discourse on Winning and Losing*.

Brehmer, B. (2005). The Dynamic OODA Loop: Amalgamating Boyd's OODA Loop and the Cybernetic Approach to Command and Control. *Proceedings of the 10th International Command and Control Research and Technology Symposium: The Future of C2*, 365-368.

Brehmer, B. (2007). Understanding the functions of C2 is the key to progress. *The International C2 Journal*, *1*, 211-232.

Breton, R., & Rousseau, R. (2007). The analysis of team decision making architectures. *Decision Making in Complex Environments*, 243.

Cacciabue, P. C., & Hollnagel, E. (1995). Simulation of cognition: Applications. *Expertise and technology: Cognition and human-computer cooperation*, 55-73.

Cain, A. A., Schuster, D., Edwards, T., & Schuster, D. (2016). A Quantitative Measure for Shared and Complementary Situation Awareness. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *60*, 1816-1820. SAGE Publications Sage CA: Los Angeles, CA.

Calcagnì, A., Lombardi, L., & Sulpizio, S. (2017). Analyzing spatial data from mouse tracker methodology: An entropic approach. *Behavior research methods*, *49*, 2012-2030.

Carneiro, D., Pimenta, A., Gonçalves, S., Neves, J., & Novais, P. (2016). Monitoring and improving performance in human-computer interaction. *Concurrency Computation, 28*, 1291-1309.

Carrasco, M. (2011). Visual attention: The past 25 years. *Vision Research, 51*, 1484-1525.

Catellin, S. (2004). L'abduction: une pratique de la découverte scientifique et littéraire. *Hermès, La Revue, 39*, 179-185.

Cazelles, B., Chavez, M., Berteaux, D., Ménard, F., Vik, J. O., Jenouvrier, S., & Stenseth, N. C. (2008). Wavelet analysis of ecological time series. *Oecologia, 156*, 287-304.

Chalandon, X. (2007). *Conscience de la situation : invariants internes et invariants externes*. CNAM.

Chauvin, C. (2003). Gestion des risques lors de la prise de décision en situation d'interaction dynamique : approches systémique et cognitive. *Epique'2003, 3*, 123-178.

Chen, Y., Qian, Z., & Lei, W. (2016). Designing a Situational Awareness Information Display: Adopting an Affordance-Based Framework to Amplify User Experience in Environmental Interaction Design. *Informatics, 3*, 6.

Chennamma, H. R., & Yuan, X. (2013). *A Survey on Eye-Gaze Tracking Techniques*. *4*, 388-393.

Chevalier, A., & Cegarra, J. (2008). Une approche psychologique de la notion de contrainte en résolution de problèmes. *Le Travail Humain, 71*, 173.

Chipman, S., Schraagen, J. M., & Shalin, V. (2000). Introduction to cognitive task analysis. In J. M. Schraagen, S. F. Chipman, & V. L. Shalin (Éds.), *Cognitive task analysis* (p. 3-23). Psychology Press.

Claverie, B. (2005). *Cognitique: science et pratique des relations à la machine à penser*. L'Harmattan.

Claverie, B. (2019). *Introduction à l'épistémologie et à la méthode de recherche : à l'usage des ingénieurs et autres scientifiques de l'industrie*. L'Harmattan.

Claverie, B., Hamacher, H., & Guyader, H. Le. (2018). *Report of Franco-German Experts on Global Security / Civil Security (GS / CS) Research*.

Claverie, B., & Kowalxzuk, B. (2018). *Proposal of topics to be considered in relation to the scientific theme « Cyberpsychology »*. Innovation Hub - Allied Command Transformation - NATO.

Claverie, B., & Paty, J. (1999). Les méthodes en psychophysiologie. In J.-P. Rossi (Éd.), *Les méthodes de recherche en psychologie*. Dunod.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and psychological measurement, 20*, 37-46.

Conger, A. J. (1980). Integration and generalization of kappas for multiple raters. *Psychological Bulletin, 88*, 322.

Cooke, N. J., Gorman, J., & Kiekel, P. A. (2017). Communication as team-level cognitive processing. In M. P. Letsky, N. W. Warner, S. M. Fiore, & C. A. P. Smith (Éds.), *Macrocognition in Teams: Theories and Methodologies* (p. 51-64). Ashgate Publishing.

Cooke, N. J., Salas, E., Cannon-Bowers, J., & Stout, R. J. (2000). Measuring Team Knowledge. *Human Factors, 42*, 151-173.

Cooke, N. J., Stout, R. J. R. J., & Salas, E. (2018). A Knowledge Elicitation Approach to the Measurement of the Team Situation Awareness. In E. Salas (Éd.), *Situational Awareness* (p. 157-182). Routledge.

Cooke, N. J., Stout, R. J., & Salas, E. (1997). Broadening the measurement of situation awareness through cognitive engineering methods. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *41*, 215–219. SAGE Publications Sage CA: Los Angeles, CA.

Costa, M., Goldberger, A. L., & Peng, C.-K. (2002). Multiscale Entropy Analysis of Complex Physiologic Time Series. *Physical Review Letters*, *89*, 68102.

Costello, A. M., Strater, L. D., Bolstad, C. A., Cuevas, H. M., & Endsley, M. R. (2006). Communication and situation awareness in ad hoc teams. *Proceedings of the 25th Army Science Conference*, 27-30.

de Winter, J. C. F., Eisma, Y. B., Cabrall, C. D. D., Hancock, P. A., & Stanton, N. A. (2019). Situation awareness based on eye movements in relation to the task environment. *Cognition, Technology and Work*, *21*, 99-111.

Dean, R. T., & Dunsmuir, W. T. M. (2016). Dangers and uses of cross-correlation in analyzing time series in perception, performance, movement, and neuroscience: The importance of constructing transfer function autoregressive models. *Behavior research methods*, *48*, 783-802.

Dehais, F., Causse, M., Vachon, F., & Tremblay, S. (2012). Cognitive conflict in human–automation interactions: a psychophysiological study. *Applied ergonomics*, *43*, 588-595.

Delaherche, E., Chetouani, M., Mahdhaoui, A., Saint-Georges, C., Viaux, S., & Cohen, D. (2012). Interpersonal Synchrony: A Survey of Evaluation Methods across Disciplines. *IEEE Transactions on Affective Computing*, *3*, 349-365.

Deming, W. E. (2019). Out of the Crisis. In *Out of the Crisis*. MIT Press.

Desclaux, G., & Claverie, B. (2015). C2 et Cyber. *Penser les Ailes françaises*, *32*, 61-68.

Desclaux, G., Claverie, B., Chaudron, L., Le Guyader, H., Salotti, J. M., & Prébot, B. (2016). Cognitive computing and augmented intelligence in man-system integration: Impact on C2 HQ key processes. *COTC'16 : 2016 Chiefs of Transformation Conference « Innovate, Adapt, Transform : Prepare Now for 2030 ».*, 4. Norfolk, VA, USA: NATO Allied Command Transformation.

Desclaux, G., Marion, D., & Claverie, B. (2019). « Reading the mind of the enemy » through an augmented multi-domain Commander's Critical Information Requirements (CCIR) process. *Proceedings of the 24th International Command and Control Research and Technology Symposium*. Laurel, MD, USA.

Dias, R. D., Zenati, M. A., Stevens, R. H., Gabany, J. M., & Yule, S. J. (2019). Physiological synchronization and entropy as measures of team cognitive load. *Journal of Biomedical Informatics*, *96*, 103250.

Doctrine Armée de l 'air DAA-3.3 « commandement des opérations aériennes ». (2017).

Dorneich, M. C., Ververs, P. M., Mathan, S., Whitlow, S., & Hayes, C. C. (2012). Considering Etiquette in the Design of an Adaptive System. *Journal of Cognitive Engineering and Decision Making*, *6*, 243-265.

Duchowski, A. T. (2002). A breadth-first survey of eye-tracking applications. *Behavior Research Methods, Instruments, & Computers*, *34*, 455-470.

Dumez, H. (2016). *Méthodologie de la recherche qualitative: Les questions clés de la démarche compréhensive* (2e édition). Vuibert.

Durso, F. T., Hackworth, C. A., & Truitt, T. R. (1999). Situation Awareness As a Predictor of Performance in En Route Air Traffic Controllers. *Air Traffic Control Quarterly*, *6*, 1-20.

Eckstein, M. K., Guerra-Carrillo, B., Miller Singley, A. T., & Bunge, S. A. (2017). Beyond eye gaze: What else can eyetracking reveal about cognition and cognitive development? *Developmental Cognitive Neuroscience*, *25*, 69-91.

Egeth, H., & Kahneman, D. (1975). Attention and Effort. *The American Journal of Psychology*, *88*, 339.

Elkins, A. N., Muth, E. R., Hoover, A. W., Walker, A. D., Carpenter, T. L., & Switzer, F. S. (2009). Physiological compliance and team performance. *Applied Ergonomics*, *40*, 997-1003.

Endsley, M. R. (1988). Situation awareness global assessment technique (SAGAT). *Proceedings of the IEEE 1988 National Aerospace and Electronics Conference*, 789-795. Dayton, OH, USA: IEEE.

Endsley, M. R. (1993). Situation Awareness and Workload: Flip Sides of the Same Coin. *Proceedings of the 7th International Symposium on Aviation Psychology*, 906-911. Colombus, OH, USA.

Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors*, *37*, 32-64.

Endsley, M. R. (1996). Situation awareness measurement in test and evaluation. In T. G. O'Brien & S. G. Charlton (Éds.), *Handbook of human factors testing and evaluation.* (p. 159-180). Lawrence Erlbaum Mahwah, NJ.

Endsley, M. R. (2000). Theoretical Underpinnings of Situation Awareness: A Critical Review. In M. R. Endsley & D. J. Garland (Éds.), *Situation Awareness Analysis and Measurement* (p. 3-21). Mahwah, NJ: Lawrence Erlbaum Associates.

Endsley, M. R. (2016). Designing for Situation Awareness: An approach to user-centered design. In *Designing for Situation Awareness: An approach to user-centered design*. CRC Press.

Endsley, M. R., Bolstad, C. A., Jones, D. G., & Riley, J. M. (2003). Situation Awareness Oriented Design: From User's Cognitive Requirements to Creating Effective Supporting Technologies. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *47*, 268-272. SAGE Publications Sage CA: Los Angeles, CA.

Endsley, M. R., & Garland, D. J. (2000). *Situation awareness analysis and measurement* (M. R. Endsley & Garlan, Éds.). Lawrence Erlbaum Associates, Inc.

Endsley, M. R., Hoffman, R. R., Kaber, D., & Roth, E. (2007). Cognitive engineering and decision making: An overview and future course. *Journal of Cognitive Engineering and Decision Making*, *1*, 1-21.

Endsley, M. R., & Jones, W. M. (2001). A model of inter and intra team situation awareness: Implications for design, training and measurement. *New trends in cooperative activities: understanding system dyanmics in complex environments*, 46-67.

Endsley, M. R., Selcon, S. J., Hardiman, T. D., & Croft, D. G. (1998). Comparative analysis of SAGAT and SART for evaluations of situation awareness. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *42*, 82-86. SAGE Publications Sage CA: Los Angeles, CA.

Erer, K. S. (2007). Adaptive usage of the Butterworth digital filter. *Journal of Biomechanics*, *40*, 2934-2943.

Fabre-Thorpe, M., Delorme, A., Marlot, C., & Thorpe, S. (2001). A limit to the speed of processing in ultra-rapid visual categorization of novel natural scenes. *Journal of Cognitive Neuroscience*, *13*, 171-180.

Fastrez, P., Campion, B., & Collard, A.-S. (2009). Le tri de cartes. *Document numérique*, *12*, 23-45.

Fleiss, J. L., Levin, B., & Myunghee, C. P. (1982). Statistical Methods for Rates and Proportions. In D. J. Balding, N. A. C. Cressie, N. I. Fisher, I. M. Johnstone, J. B. Kadane, L. M. Ryan, … J. L. Teugels (Éds.), *Statistical Methods for Rates and Proportions* (p. 510). John Wiley & Sons, Inc.

Foltz, P. W. (2018). Automating the assessment of team collaboration through communication analysis. *Design recommendations for intelligent tutoring systems*, *6*, 185-197.

Foushee, H. C., & Helmreich, R. L. (1988). Group interaction and flight crew performance. In *Academic Press series in cognition and perception. Human factors in aviation.* (p. 189-227). San Diego, CA, US: Academic Press.

Frank, M. C., Vul, E., & Johnson, S. P. (2009). Development of infants' attention to faces during the first year. *Cognition*, *110*, 160-170.

Freeman, J. B., & Ambady, N. (2010). MouseTracker: software for studying real-time mental processing using a computer mouse-tracking method. *Behavior research methods*, *42*, 226-241.

Frisch, S., Dshemuchadse, M., Görner, M., Goschke, T., & Scherbaum, S. (2015). Unraveling the sub-processes of selective attention: Insights from dynamic modeling and continuous behavior. *Cognitive processing*, *16*, 377-388.

Fuchs, S., Hale, K., Stanney, K. M., Berka, C., Levendowski, D., & Juhnke, J. (2006). Physiological Sensors Cannot Effectively Drive System Mitigation Alone. In *Foundations of Augmented Cognition* (2e éd., p. 193-200).

Fuchs, S., & Schwarz, J. (2017). Towards a dynamic selection and configuration of adaptation strategies in Augmented Cognition. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *10285*, 101-115.

Funke, G. J., Knott, B. A., Salas, E., Pavlas, D., & Strang, A. J. (2012). Conceptualization and Measurement of Team Workload: A Critical Need. *Human Factors*, *54*, 36-51.

Gamer, M., Lemon, J., Fellows, I., & Singh, P. (2012). Various coefficients of interrater reliability and agreement. *R package version 0.84*, *1*.

Gllckman, A., Zimmer, S., Montero, C., Montero, R., Montero, C., Guerette, P. J., & Campbell, W. J. (1987). *The evolution of teamwork skills: An empirical assessment with implications for training - Technical Report*. Norfolk, VA: Old Dominion University Center for Applied Psychological Studies.

Goldberg, J. H., & Kotval, X. P. (1999). Computer interface evaluation using eye movements: Methods and constructs. *International Journal of Industrial Ergonomics*, *24*, 631-645.

Goldberger, A. L., Amaral, L., Hausdorff, J. M., Ivanov, P., Peng, C.-K., & Stanley, H. (2002). Fractal dynamics in physiology: Alterations with disease and aging. *Proceedings of the National Academy of Sciences of the United States of America*, *99 Suppl 1*, 2466-2472.

Gordon, S. E., & Gill, R. T. (1997). Cognitive task analysis. In C. E. Zsambok & G. Klein (Éds.), *Naturalistic decision making* (p. 131-140). Lawrence Erlbaum Associates, Inc.

Gorman, J. C., Cooke, N. J., Winner, J. L., Duran, J. L., Pedersen, H. K., & Taylor, A. R. (2007). Knowledge training versus process training: The effects of training protocol on team coordination and performance. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *51*, 382-386. SAGE Publications Sage CA: Los Angeles, CA.

Gorman, J., Cooke, N. J., & Winner, J. L. (2006). Measuring team situation awareness in decentralized command and control environments. *Ergonomics*, *49*, 1312-1325.

Gorman, J., Grimm, D. A., Stevens, R. H., Galloway, T., Willemsen-Dunlap, A. M., & Halpin, D. J. (2020). Measuring real-time team cognition during team training. *Human factors*, *62*, 825-860.

Grinsted, A., Moore, J. C., & Jevrejeva, S. (2004). Application of the cross wavelet transform and wavelet coherence to geophysical time series. *Nonlinear Processes in Geophysics*, *11*, 561-566.

Guastello, S. J., Marr, D., Perna, C., Castro, J., Gomez, M., & Peressini, A. F. (2016). Physiological synchronization in emergency response teams: Subjective workload, drivers and empaths. *Nonlinear Dynamics, Psychology, and Life Sciences*, *20*, 223-270.

Guhe, M., Gray, W. D., Schoelles, M. J., Liao, W., Zhu, Z., & Ji, Q. (2005). Non-intrusive measurement of workload in real-time. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *49*, 1157-1161. SAGE Publications Sage CA: Los Angeles, CA.

Hackman, J. R. (1987). The design of work teams. In J. Lorsch (Éd.), *Handbook of organizational behavior* (p. 315-342). Englewood Cliffs, NJ: Prentice-Hall.

Hafed, Z. M., & Clark, J. J. (2002). Microsaccades as an overt measure of covert attention shifts. *Vision Research*, *42*, 2533-2545.

Hancock, P. A., & Meshkati, N. (1988). Human mental workload. In P. A. Hancock & N. Meshkati (Éds.), *Human mental workload.* Oxford, England: Elsevier Science Publishers B.V.

Hauland, G. (2019). Measuring team situation awareness by means of eye movement data. *Proceedings of HCI International 2003*, *3*, 230-234.

Hauss, Y., Gauss, B., & Eyferth, K. (2001). SALSA-a new approach to measure situational awareness in air traffic control. Focusing attention on aviation safety. *Proceedings of the 11th International Symposium on Aviation Psychology*, 1-20. Colombus, OH, USA.

Hayashi, N., Someya, N., & Fukuba, Y. (2010). Effect of intensity of dynamic exercise on pupil diameter in humans. *Journal of Physiological Anthropology*, *29*, 119-122.

Henning, R. A., Armstead, A. G., & Ferris, J. K. (2009). Social psychophysiological compliance in a four-person research team. *Applied Ergonomics*, *40*, 1004-1010.

Hessels, R. S., Niehorster, D. C., Kemner, C., & Hooge, I. T. C. (2017). Noise-robust fixation detection in eye movement data: Identification by two-means clustering (I2MC). *Behavior Research Methods*, *49*, 1802-1823.

Hjelmfelt, A. T., & Pokrant, M. A. (1998). Coherent tactical picture. *CNARM*, 97-129.

Ho, N., Johnson, W., Panesar, K., Wakeland, K., Sadler, G., Wilson, N., … Brandt, S. (2017). Application of human-autonomy teaming to an advanced ground station for reduced crew operations. *2017 IEEE/AIAA 36th Digital Avionics Systems Conference (DASC)*, 1-4. IEEE.

Hoffman, R. R., & McNeese, M. D. (2009). A History for Macrocognition. *Journal of Cognitive Engineering and Decision Making*, *3*, 97-110.

Hogervorst, M. A., Brouwer, A. M., & van Erp, J. B. F. (2014). Combining and comparing EEG, peripheral physiology and eye-related measures for the assessment of mental workload. *Frontiers in Neuroscience*, *8*, 14.

Hogg, D. N., Follesø, K., Volden, F. S., & Torralba, B. (1994). *SACRI: A measure of situation awareness for use in the evaluation of nuclear power plant control room systems providing information about the current process state*.

Höglund, F., Berggren, P., & Nählinder, S. (2010). Using shared priorities to measure shared situation

awareness. *Proceedings of the 7th International ISCRAM Conference, 1*. Seattle,WA, USA.

Hollnagel, E. (1993). Models of cognition: procedural prototypes and contextual control. *Le Travail Humain, 56*, 27-51.

Hollnagel, E., & Woods, D. D. (1999). Cognitive systems engineering: New wine in new bottles. *International Journal of Human Computer Studies, 51*, 339-356.

Hooey, B. L., Gore, B. F., Wickens, C. D., Scott-nash, S., Socash, C., Salud, E., … Foyle, D. C. (2011). Modeling pilot situation awareness. In P. C. Cacciabue (Éd.), *Human modelling in assisted transportation* (p. 207-213). Springer.

Ifrah, G. (1994). *Histoire universelle des chiffres. Tome 1 : L'intelligence des hommes racontée par les nombres et le calcul.* (2eme éd.). Paris: Robert Lafont.

Jackson, I., & Sirois, S. (2009). Infant cognition: Going full factorial with pupil dilation. *Developmental Science, 12*, 670-679.

Jacob, R. J. K., & Karn, K. S. (2004). Eye tracking in human-computer interaction and usability research: Ready to deliver the promises. In Hyona, Radach, & Deubel (Éds.), *The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research* (p. 573-605). Oxford, England: Elsevier Science Publishers B.V.

Janisse, M. P. (1977). *Pupillometry: The psychology of the pupillary response.* (Hemisphere, Éd.). Washington: Hemisphere.

Jin, P. (1992). Toward a reconceptualization of the law of initial value. *Psychological Bulletin, 111*, 176-184.

Johannsdottir, K. R., & Herdman, C. M. (2010). The Role of Working Memory in Supporting Drivers' Situation Awareness for Surrounding Traffic. *Human Factors, 52*, 663-673.

Johansson, B., Hellgren, C., Oskarsson, P.-A., & Svensson, J. (2013). Supporting situation awareness on the move - The role of technology for spatial orientation in the field. *ISCRAM 2013 Conference Proceedings - 10th International Conference on Information Systems for Crisis Response and Management*.

Jorna, P. (1993). Heart rate and workload variations in actual and simulated flight. *Ergonomics, 36*, 1043-1054.

Kaber, D. B., Prinzel, L. J., Wright, M. C., & Clamann, M. P. (2002). *Workload-Matched Adaptive Automation Support of Air Traffic Controller Information Processing Stages*. National Aeronautics and Space Administration.

Kanas, N. (2005). Interpersonal issues in space: Shuttle/Mir and beyond. *Aviation, Space, and Environmental Medicine, 76*, B126-34.

Kang, O., & Wheatley, T. P. (2017). Pupil dilation patterns spontaneously synchronize across individuals during shared attention. *Journal of Experimental Psychology: General, 146*, 569.

Kaster, A., Tappert, E., Ruckert, C., & Becker, R. (2010). *Design of ergonomic user interfaces for Asymmetric Warfare (Gestaltung ergonomischer Benutzungsschnittstellen für Asymmetric Warfare). Final Report*. Wachtberg, Allemagne: Fraunhofer-Institute for Communication, Information processing and ergonomics – FKIE.

Kiekel, P. A., Cooke, N. J., Foltz, P. W., & Shope, S. M. (2001). Automating Measurement of Team Cognition through Analysis of Communication Data. *Usability Evaluation and Interface Design*, 1382-1386.

Kieslich, P. J., Wulff, D. U., Henninger, F., Haslbeck, J. M. B., & Schulte-Mecklenbeck, M. (2018). Mouse-and hand-tracking as a window to cognition: A tutorial on implementation, analysis, and visualization. *Manuscript in preparation.*

Kilingaru, K., Tweedale, J. W., Thatcher, S., & Jain, L. C. (2013). Monitoring pilot "situation awareness". *Journal of Intelligent & Fuzzy Systems*, *24*, 457-466.

Klein, G. (2000). Cognitive task analysis of teams. In J. M. Schraagen, S. F. Chipman, & V. L. Shalin (Éds.), *Cognitive Task Analysis* (p. 417-429). Lawrence Erlbaum Associates, Inc.

Klein, G. (2008). Naturalistic decision making. *Human Factors*, *50*, 456-460.

Klein, G., Ross, K. G., Moon, B. M., Klein, D. E., Hoffman, R. R., & Hollnagel, E. (2003). Macrocognition. *IEEE Intelligent Systems*, *18*, 81–85.

Klingner, J., Kumar, R., & Hanrahan, P. (2008). Measuring the task-evoked pupillary response with a remote eye tracker. *Eye Tracking Research and Applications Symposium (ETRA)*, 69-72.

Klippel, A., Hirtle, S., & Davies, C. (2010). You-are-here maps: Creating spatial awareness through map-like representations. *Spatial Cognition and Computation*, *10*, 83-93.

Knapen, T., De Gee, J. W., Brascamp, J., Nuiten, S., Hoppenbrouwers, S., & Theeuwes, J. (2016). Cognitive and ocular factors jointly determine pupil responses under equiluminance. *PLoS ONE*, *11*, 13.

Kocejko, T., & Wtorek, J. (2013). Gaze tracking in multi-display environment. *2013 6th International Conference on Human System Interactions (HSI)*, 626-631.

Koelewijn, T., Zekveld, A. A., Festen, J. M., & Kramer, S. E. (2012). Pupil dilation uncovers extra listening effort in the presence of a single-talker masker. *Ear and Hearing*, *33*, 291-300.

Kott, A., & Milks, W. (2009). Approaches to validation of information fusion systems. *12th International Conference on Information Fusion*, 882-889. Seattle,WA, USA: IEEE.

Lacey, J. I. (1956). The evaluation of autonomic responses: Toward a general solution. *Annals of the New York Academy of Sciences*, *67*, 125-163.

Lemercier, A., Guillot, G., Courcoux, P., Garrel, C., Baccino, T., Schlich, P., … Baccino, T. (2015). Pupillometry of taste : Methodological guide – from acquisition to data processing-and toolbox for MATLAB. *The Quantitative Methods for Psychology*, *10*, 179-195.

Letsky, M. P., & Warner, N. W. (2017). Macrocognition in teams. In M. P. Letsky, N. W. Warner, S. M. Fiore, E. Salas, & C. A. P. Smith (Éds.), *Macrocognition in Teams: Theories and Methodologies* (2nd éd., p. 1-13). CRC Press.

Lini, S. (2013). *L'anticipation et sa représentation dans les interfaces homme-système en aéronautique : une approche d'ingénierie cognitique*. Université de Bordeaux.

Linkov, I., Galaitsi, S., Trump, B., Keisler, J., Cybertrust, & Kott, A. (2020). *From Explainable to Actionable and Interpretable Artificial Intelligence*. 91-96.

Marion, D. (2018). *Sharing big display : développement des technologies et métaphores d ' interactions nouvelles pour le partage collaboratif d ' affichage en groupe ouvert.* Université de Bordeaux.

Mathôt, S., Aarts, E., Verhage, M., Veenvliet, J. V., Dolan, C. V., & van der Sluis, S. (2013). *A simple way to reconstruct pupil size during eye blinks*.

Mathôt, S., Fabius, J., Van Heusden, E., & Van der Stigchel, S. (2018). Safe and sensible preprocessing and baseline correction of pupil-size data. *Behavior Research Methods*, *50*, 94-106.

Mathôt, S., Grainger, J., & Strijkers, K. (2017). Pupillary Responses to Words That Convey a Sense of Brightness or Darkness. *Psychological Science*, *28*, 1116-1124.

Matthews, M. D., & Beal, S. A. (2002). Assessing situation awareness in field training exercises. *ARI DTIC report 1795*.

Mays, L. E., & Gamlin, P. D. (1995). Neuronal circuitry controlling the near response. *Current opinion in neurobiology*, *5*, 763-768.

McDougal, D. H., & Gamlin, P. D. (2010). The influence of intrinsically-photosensitive retinal ganglion cells on the spectral sensitivity and response dynamics of the human pupillary light reflex. *Vision research*, *50*, 72-87.

McGuigan, F. J., & Andreassi, J. L. (1980). *Psychophysiology: Human behavior and physiological response*. New York: Oxford University Press.

McLaren, J. W., Erie, J. C., & Brubaker, R. F. (1992). Computerized analysis of pupillograms in studies of alertness. *Investigative Ophthalmology & Visual Science*, *33*, 671-676.

Megaw, T., & Sharples, S. (2015). The definition and measurement of mental workload. In *Evaluation of Human Work, 4th Edition* (p. 515-548). CRC Press.

Mehta, R. K., & Parasuraman, R. (2013). Neuroergonomics: a review of applications to physical and cognitive work. *Frontiers in human neuroscience*, *7*, 889.

Meireles, L., Alves, L., & Cruz, J. (2018). Conceptualization and Measurement of Individual Situation Awareness (SA) in Expert Populations across Operational Domains: a systematic review of the literature with a practical purpose on our minds. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *62*, 1093-1097. SAGE Publications Sage CA: Los Angeles, CA.

Missa, J.-N., & Debru, C. (1993). *L'Esprit-cerveau : la philosophie de l'esprit à la lumière des neurosciences*. Paris: J. Vrin.

Mønster, D., Håkonsson, D. D., Eskildsen, J. K., & Wallot, S. (2016). Physiological evidence of interpersonal dynamics in a cooperative production task. *Physiology and Behavior*, *156*, 24-34.

Moore, K., & Gugerty, L. (2010). Development of a novel measure of situation awareness: The case for eye movement analysis. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *54*, 1650-1654. SAGE Publications Sage CA: Los Angeles, CA.

Morton, M., & Roberts, T. (2011, septembre 1). *Joint Space Operations Center (JSpOC) Mission System (JMS)*.

Nagel, T. (1989). *What Does It All Mean? A Very Short Introduction to Philosophy*. New York: Oxford University Press.

Nakayama, M., & Shimizu, Y. (2004). Frequency analysis of task evoked pupillary response and eye-movement. *Proceedings of the 2004 symposium on Eye tracking research & applications*, 71-76.

NASA. (2013). *International Space Station (ISS) EVA Suit Water Intrusion - Report*.

National Research Council. (1990). Distributed Decision Making. In *Distributed Decision Making*. National Academies Press.

NATO Innovation Hub. (2018). *Cyber Psychology and NATO : Sutdy Report*.

NATO SACT Transformation Network Branch. (2018). *Chiefs of Transformation Conference Analysis Report*.

Newell, A., & Simon, H. A. (1972). *Human problem solving* (Vol. 104). Englewood Cliffs, NJ: Prentice-Hall.

Nofi, A. (2000). Defining and Measuring Shared Situational Awareness. In *Final Report* (Vol. 5). Center for Naval Analyses.

O'Neill, W. D., & Zimmerman, S. (2000). Neurologial interpretations and the information in the cognitive pupillary response. *Methods of information in medicine*, *39*, 122-124.

Orasanu, J., & Fischer, U. (1991). *Information transfer and shared mental models for decision making*.

Ososky, S., Schuster, D., Jentsch, F., Fiore, S., Shumaker, R., Lebiere, C., … Stentz, A. (2012). The importance of shared mental models and shared situation awareness for transforming robots from tools to teammates. *Proceedings of the SPIE, Volume 8387, id. 838710 (2012).*, *8387*, 838710.

Ownby, M., & Kott, A. (2006). Reading the mind of the enemy: Predictive analysis and command effectiveness. *CCRTS*, 28.

Parasuraman, R., & Galster, S. M. (2013). Sensing, assessing, and augmenting threat detection: behavioral, neuroimaging, and brain stimulation evidence for the critical role of attention. *Frontiers in human neuroscience*, *7*, 273.

Parasuraman, R., Hockey, G. R. J., Gaillard, A. W. K., & Burov, O. (2003). Adaptive automation matched to human mental workload. *NATO Science Series Sub Series I Life and Behavioural Sciences*, *355*, 177-193.

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics Part A:Systems and Humans.*, *30*, 286-297.

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2008). Situation Awareness,Mental Workload,and Trust in Automation:Viable,Empirically Supported Cognitive Engineering Constructs. *Journal of Cognitive Engineering and Decision Making*, *2*, 140-160.

Perla, P., Markowitz, M., Nofi, A., Weuve, C., & Loughran, J. (2000). *Gaming and shared situation awareness*. Center for Naval Analyses.

Peysakhovich, V., Causse, M., Scannella, S., & Dehais, F. (2015). Frequency analysis of a task-evoked pupillary response: Luminance-independent measure of mental effort. *International Journal of Psychophysiology*, *97*, 30-37.

Peysakhovich, V., & Hurter, C. (2018). Scanpath visualization and comparison using visual aggregation techniques. *Journal of Eye Movement Research*, *10*, 1-14.

Peysakhovich, V., Vachon, F., & Dehais, F. (2017). The impact of luminance on tonic and phasic pupillary responses to sustained cognitive load. *International Journal of Psychophysiology*, *112*, 40-45.

Posner, M. I., Snyder, C. R., & Davidson, B. J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology: General*, *109*, 160-174.

Prébot, B., Cavel, C., Calice, L., Mahaut, M., Leduque, A., & Salotti, J. M. (2019). Team performance analysis of a collaborative spatial orientation mission in Mars analogue environment. *Proceedings of the 70th International Astronautical Congress, IAC*, 21-25. Washington DC, DC, USA: IAF.

Prebot, B., Salotti, J. M., Vennin, C., & Claverie, B. (2020). Shared spatial situation awareness as a team performance indicator in collaborative spatial orientation task. *Advances in Intelligent Systems and Computing*, *956*, 106-115.

Prébot, B., Schwarz, J., Fuchs, S., & Claverie, B. (2020). From “knowing what” to “knowing when”: Exploring a concept of situation awareness synchrony for evaluating sa dynamics in teams. In R. A. Sottilare & J. Schwarz (Éds.), *Adaptive Instructional Systems. HCII 2020. Lecture Notes in Computer Science* (Vol. 12214, p. 505-518). Springer, Cham.

Privitera, C. M., Renninger, L. W., Carney, T., Klein, S., & Aguilar, M. (2010). Pupil dilation during visual target detection. *Journal of Vision, 10*, 3-3.

Privitera, Claudio M., & Stark, L. W. (2006). A binocular pupil model for simulation of relative afferent pupil defects and the swinging flashlight test. *Biological Cybernetics*. https://doi.org/10.1007/s00422-005-0042-8

Prytz, E., Rybing, J., Jonson, C.-O., Petterson, A., Berggren, P., & Johansson, B. (2015). An exploratory study of a low-level shared awareness measure using mission-critical locations during an emergency exercise. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *59*, 1152-1156. SAGE Publications Sage CA: Los Angeles, CA.

Rasmussen, J. (1983). Skills, rules, and knowledge; signals, signs, and symbols, and other distinctions in human performance models. *IEEE transactions on systems, man, and cybernetics*, 257-266.

Rasmussen, J., Brehmer, B., & Leplat, J. (1991). Distributed decision making: Cognitive models for cooperative work. In *Distributed decision making: Cognitive models for cooperative work.*

Raymond, C., Prun, D., & Cegarra, J. (2016). *Associer l’ingénierie système et le Cognitive Work Analysis pour soutenir l’intégration Homme-Système*.

Reilly, J., Kelly, A., Kim, S. H., Jett, S., & Zuckerman, B. (2019). The human task-evoked pupillary response function is linear: Implications for baseline response scaling in pupillometry. *Behavior Research Methods*, *51*, 865-878.

Roscoe, A. H. (1993). Heart rate as a psychophysiological measure for in-flight workload assessment. *Ergonomics*, *36*, 1055-1062.

Rossi, J.-P. (1999). *Méthodes de recherche en psychologie* (Dunod).

Salas, E., Cannon-Bowers, J., & Johnston, J. H. (1997). How can you turn a team of experts into an expert team?: Emerging training strategies. In *Naturalistic decision making.* (p. 359-370).

Salas, E., Cooke, N. J., & Rosen, M. A. (2008). On Teams, Teamwork, and Team Performance: Discoveries and Developments. *Human Factors*, *50*, 540-547.

Salas, E., Dickinson, T. L., Converse, S., & Tannenbaum, S. I. (1992). Toward an understanding of team performance and training. *Teams: Their Training and Performance;*, *Volume 12*, 3–29.

Salas, E., Prince, C., Baker, D. P., & Shrestha, L. (1995). Situation awareness in team performance: Implications for measurement and training. *Human factors*, *37*, 123-136.

Salas, E., Reyes, D. L., & Woods, A. L. (2017). The Assessment of Team Performance: Observations and Needs. In *Methodology of Educational Measurement and Assessment* (p. 21-37). Springer International Publishing.

Sallaberry, J. (2020). Cognition individuelle et cognition collective. *Ingénierie cognitique*, *4*, 1-16.

Sallaberry, J., & Claverie, B. (2018). *Introduction aux sciences humaines et sociétales*. L’Harmattan.

Salmon, P. M., Stanton, N. A., Walker, G. H., & Green, D. (2006). Situation awareness measurement: A review of applicability for C4i environments. *Applied Ergonomics*, *37*, 225-238.

Salmon, P. M., Stanton, N. A., Walker, G. H., & Jenkins, D. P. (2008). What really is going on? Review

of situation awareness models for individuals and teams. *Theoretical Issues in Ergonomics Science*, *9*, 297-323.

Salmon, P. M., Stanton, N. A., Walker, G. H., Jenkins, D. P., Baber, C., Jenkins, D. P., … Young, M. S. (2008). What really is going on? Review of situation awareness models for individuals and teams. *Theoretical Issues in Ergonomics Science*, *9*, 297-323.

Salmon, P. M., Stanton, N. A., Walker, G. H., Jenkins, D. P., & Rafferty, L. (2010). Is it really better to share? Distributed situation awareness and its implications for collaborative system design. *Theoretical Issues in Ergonomics Science*, *11*, 58-83.

Salotti, J. M., Bahmani, C., Bégout, P., Dubresson, A., Salvan, L., Chalard, P., & Melin, G. (2016). Space camp project in bordeaux. *Proceedings of the International Astronautical Congress, IAC*.

Salotti, J. M., Prébot, B., & Cinelli, I. (2019). Situation awareness issues during outdoor activity. *Geophysical Research Abstracts*, *21*, 2484-2491.

Saner, L. D., Bolstad, C. A., Gonzalez, C., & Cuevas, H. M. (2009). Measuring and Predicting Shared Situation Awareness in Teams. *Journal of Cognitive Engineering and Decision Making*, *3*, 280-308.

Saner, L. D., Bolstad, C. A., Gonzalez, C., & Cuevas, H. M. (2010). Predicting Shared Situation Awareness in Teams: A Case of Differential SA Requirements. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *54*, 314-318. SAGE Publications Sage CA: Los Angeles, CA.

Scerbo, M. W. (1996). Theoretical perspectives on adaptive automation. In *Human factors in transportation. Automation and human performance: Theory and applications.* (p. 37-63). Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc.

Scheidt, D., & Schultz, K. (2011). On Optimizing Command and Control Structures. *16th International Command and Control Research and Technology Symposium*.

Schmidt, K. (1991). Cooperative work: A conceptual framework. *Distributed Decision Making Cognitive Models for Cooperative Work*, *21*, 75-110.

Schreiber, B. T., Bell, H. H., & Raspotnik, W. B. (1998). Investigating communication and situation awareness in air combat. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *42*, 21-25. SAGE Publications Sage CA: Los Angeles, CA.

Schwarz, J., & Fuchs, S. (2017). Multidimensional real-time assessment of user state and performance to trigger dynamic system adaptation. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *10284*, 383-398.

Schwarz, J., Fuchs, S., & Flemisch, F. (2014). Towards a more holistic view on user state assessment in adaptive human-computer interaction. *Conference Proceedings - IEEE International Conference on Systems, Man and Cybernetics*, *2014-Janua*, 1228-1234.

Sellers, J., Helton, W. S., Näswall, K., Funke, G. J., & Knott, B. A. (2014). Development of the Team Workload Questionnaire (TWLQ). *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *58*, 989-993. SAGE Publications Sage CA: Los Angeles, CA.

Sexton, J. B., & Helmreich, R. L. (2000). Analyzing cockpit communications: the links between language, performance, error, and workload. *Human Performance in Extreme Environments*, *5*, 63-68.

Shomstein, S., & Behrmann, M. (2006). Cortical systems mediating visual attention to both objects and spatial locations. *Proceedings of the National Academy of Sciences of the United States of America*,

103, 11387-11392.

Sirois, S., & Brisson, J. (2014). Pupillometry. *Wiley Interdisciplinary Reviews: Cognitive Science*, *5*, 679-692.

Sokolov, Y. N. (1963). *Perception and the conditioned reflex*.

Sonnenwald, D. H., Maglaughlin, K. L., & Whitton, M. C. (2004). Designing to support situation awareness across distances: an example from a scientific collaboratory. *Information processing & management*, *40*, 989-1011.

Spain, R., Geden, M., Min, W., Mott, B., & Lester, J. (2019). Toward computational models of team effectiveness with natural language processing. *CEUR Workshop Proceedings*, *2501*, 30-39.

St. John, M., & Kobus, D. a. (2003). *DARPA Augmented Cognition Technical Intagration Experiment (TIE)*.

St. John, M., Kobus, D. a., Morrison, J. G., & Schmorrow, D. (2004). Overview of the DARPA Augmented Cognition Technical Integration Experiment. *International Journal of Human-Computer Interaction*, *17*, 131-149.

Stanton, N. A., Salmon, P. M., Walker, G. H., Salas, E., & Hancock, P. A. (2017). State-of-science: situation awareness in individuals, teams and systems. *Ergonomics*, *60*, 449-466.

Steinhauer, S. R., & Hakerem, G. (1992). The pupillary response in cognitive psychophysiology and schizophrenia. In D. Friedman & G. Bruder (Éds.), *Annals of the New York Academy of Sciences* (Vol. 658, p. 182-204). New York: NY Academy of Sciences.

Steinhauser, N. B., Pavlas, D., & Hancock, P. A. (2009). Design principles for adaptive automation and aiding. *Ergonomics in Design*, *17*, 6-10.

Stevens, R. H., Galloway, T., Berka, C., Behneman, A., Wohlgemuth, T., Lamb, J., & Buckles, R. (2011). Linking models of team neurophysiologic synchronies for engagement and workload with measures of team communication. *20th Annual Conference on Behavior Representation in Modeling and Simulation 2011, BRiMS 2011*, 122-129.

Stevens, R. H., Galloway, T., & Lamb, C. (2014). Submarine Navigation team resilience: Linking EEG and behavioral models. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. SAGE Publications Sage CA: Los Angeles, CA.

Stevens, R. H., Gorman, J., Amazeen, P., Likens, A., & Galloway, T. (2013). The organizational neurodynamics of teams. *Nonlinear dynamics, psychology, and life sciences*, *17*, 67-86.

Strang, A. J., Funke, G. J., Knott, B. A., Galster, S. M., & Russell, S. M. (2012). Effects of cross-training on team performance, communication, and workload in simulated air battle management. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *56*, 1581-1585. SAGE PublicationsSage CA: Los Angeles, CA.

Strang, A. J., Funke, G. J., Russell, S. M., & Thomas, R. D. (2013). Physio-behavioral coupling as an index of team processes and performance: Overview, measurement, and empirical application. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *8027 LNAI*, 219-228. Springer, Berlin, Heidelberg.

Strauch, B. (2017). Investigating Human Error: Incidents, Accidents, and Complex Systems. In *Investigating Human Error: Incidents, Accidents, and Complex Systems* (Second). CRC Press.

Sulistyawati, K., Wickens, C. D., & Chui, Y. P. (2009). Exploring the Concept of Team Situation Awareness in a Simulated Air Combat Environment. *Journal of Cognitive Engineering and*

Decision Making, 3, 309-330.

Szafir, D., & Mutlu, B. (2012). Pay Attention! Designing Adaptive Agents That Monitor and Improve User Engagement. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 11–20. New York, NY, USA: Association for Computing Machinery.

Taylor, R. M. (1990). Situational Awareness Rating Technique (SART): The development of a tool for aircrew systems design. *Situational Awareness in Aerospace Operations, 17*, 25-53.

The MathWorks Inc. (2012). *MATLAB and Statistics Toolbox Release 2012b*. Natick, Massachusetts, United States: The MathWorks.

Tognoli, E. (2008). EEG coordination dynamics: Neuromarkers of social coordination. In A. Fuchs & V. K. Jirsa (Éds.), *Coordination: Neural, Behavioral and Social Dynamics. Understanding Complex Systems* (p. 309-323). Springer, Berlin, Heidelberg.

Tognoli, E., Lagarde, J., DeGuzman, G. C., & Kelso, J. A. S. (2007). The phi complex as a neuromarker of human social coordination. *Proceedings of the National Academy of Sciences, 104*, 8190-8195.

Tomarken, A. J. (1995). A psychometric perspective on psychophysiological measures. *Psychological Assessment, 7*, 387-395.

Torrence, C., & Compo, G. P. (1998). A Practical Guide to Wavelet Analysis. *Bulletin of the American Meteorological Society, 79*, 61-78.

Tsang, P. S., & Vidulich, M. A. (2006). Mental Workload and Situation Awareness. In G. Salvendy (Éd.), *Handbook of Human Factors and Ergonomics* (p. 243-268). John Wiley & Sons, Inc.

van de Merwe, K., van Dijk, H., & Zon, R. (2012). Eye movements as an indicator of situation awareness in a flight simulator experiment. *The International Journal of Aviation Psychology, 22*, 78-95.

Varela, F. J., & Lavoie, P. (1996). *Invitation aux sciences cognitives*. Editions du Seuil.

Veltman, J. A., & Gaillard, A. W. K. (1998). Physiological workload reactions to increasing levels of task difficulty. *Ergonomics, 41*, 656-669.

Vidulich, M. A., & Hughes, E. R. (1991). Testing a subjective metric of situation awareness. *Proceedings of the Human Factors Society Annual Meeting, 35*, 1307-1311. SAGE Publications Sage CA: Los Angeles, CA.

Vidulich, M. A., & Tsang, P. S. (2015). The Confluence of Situation Awareness and Mental Workload for Adaptable Human-Machine Systems. *Journal of Cognitive Engineering and Decision Making, 9*, 95-97.

Vukovic, M., Sethu, V., Parker, J., Cavedon, L., Lech, M., & Thangarajah, J. (2019). Estimating cognitive load from speech gathered in a complex real-life training exercise. *International Journal of Human-Computer Studies, 124*, 116-133.

Waag, W. L., & Houck, M. R. (1994). Tools for assessing situational awareness in an operational fighter environment. *Aviation, Space, and Environmental Medicine, 65*, A13-A19.

Wass, S. V., Forssman, L., & Leppänen, J. (2014). Robustness and precision: How data quality may influence key dependent variables in infant eye-tracker analyses. *Infancy, 19*, 427-460.

Weiss, M. W., Trehub, S. E., Schellenberg, E. G., & Habashi, P. (2016). Pupils dilate for vocal or familiar music. *Journal of Experimental Psychology: Human Perception and Performance, 42*, 1061-1065.

Wickens, C. D. (2001). Workload and situation awareness. *Stress, Workload and Fatigue: Theory,*

research and practise., 443-450.

Wickens, C. D., Hollands, J. G., Banbury, S., & Parasuraman, R. (1992). *Engineering psychology and human performance* (C. D. Wickens, J. G. Hollands, S. Banbury, & R. Parasuraman, Éds.). New York: Psychology Press.

Wickens, C. D., & McCarley, J. S. (2007). *Applied attention theory*. CRC press.

Wilder, J. (1957). The law of initial value in neurology and psychiatry: Facts and problems. *Journal of Nervous and Mental Disease*, *125*, 73-86.

Winn, M. B., Wendt, D., Koelewijn, T., & Kuchinsky, S. E. (2018). Best Practices and Advice for Using Pupillometry to Measure Listening Effort: An Introduction for Those Who Want to Get Started. *Trends in Hearing*, Vol. 22, p. 1-32.

Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology*, *18*, 459-482.

Zagermann, J., Pfeil, U., & Reiterer, H. (2016). Measuring Cognitive Load Using Eye Tracking Technology in Visual Computing. *Proceedings of the Sixth Workshop on Beyond Time and Errors on Novel Evaluation Methods for Visualization*, 78-85.

Ziemke, T., Schaefer, K. E., & Endsley, M. (2017). Situation awareness in human-machine interactive systems. *Cognitive Systems Research, 46*, 1-2.

Zsambok, C. E., & Klein, G. (2014). *Naturalistic decision making*. Psychology Press.

Annexes

Annexe 1

QUESTIONNAIRE POST-SITUATIONNEL
Expérience Explore MARS

Zone à remplir par l'expérimentateur

- Modalité (M1 ou M2) : _____
- Indiv ID (I_1; I_2…): _____
- Date et heure de début : _____
- Temps de complétion : _____ mins
- Scénario: _____

Zone à remplir par l'individu

- Rôle : _____
- Age: _____
- Genre • F • M

- A quel point avez-vous l'habitude de travailler avec votre coéquipier ?

Jamais 1 2 3 4 5 Très souvent

- Estimez la difficulté du placement de l'objet à retrouver

Très facile 1 2 3 4 5 Extrêmement dur

- Estimez votre sens de l'orientation

Très mauvais 1 2 3 4 5 Excellent

- Avez-vous l'habitude d'utiliser une carte ou un GPS ?

Jamais 1 2 3 4 5 Très régulièrement

Performance

- Comment jugeriez-vous la performance de l'équipe ?

Très faible	1	2	3	4	5	Parfaite

- Comment jugeriez-vous la performance de votre coéquipier ?

Très faible	1	2	3	4	5	Parfaite

- Comment jugeriez-vous votre performance ?

Très faible	1	2	3	4	5	Parfaite

- Diriez-vous que votre partage de conscience de situation était :

Très faible	1	2	3	4	5	Très bonne

- Avez-vous mis en place une stratégie de communication ? Si oui, décrivez.

- *(Laisser vide si c'est la première fois que vous passez l'expérience)*
  Comparé à l'expérience précédente, vous avez trouvé la tâche :
  • plus difficile • pareille • plus facile

  Pourquoi ?

- Qui parlait le plus ? • Moi • Mon Coéquipier

- Qui posait le plus de questions ? • Moi • Mon Coéquipier

- Qu'est-ce qui vous aurait aidé à être plus performant ?

Annexe 2 : FKIE SA Analysis and Macros

Work Report	Wachtberg, GE
	20/12/2018

Ce document présente une partie des travaux réalisés dans le cadre de la collaboration avec le laboratoire allemande FKIE. Il comporte 3 sections distinctes.
La première détaille le développement d'un score de conscience de situation (SA Ratio) visant à obtenir plus de nuances dans l'évaluation de la SA, en considérant que le fait de manquer un élément de faible importance ne devrait pas avoir le même poid dans l'évaluation que le fait de manquer un évènement à risque. Pour cela les réponses aux question ont été analysées au regard de l'intention et du comportement de l'opérateur (observation video) et une formule a été proposée.
La deuxième section détaille le processus de détection de première perception d'un élément par position du regard, méthode dont nous nous sommes servis dans le Chapitre 4
La troisième section présente la documentation relative aux programmes MACRO développés pour l'analyse des données sous Excel, basée sur la structure des données recueillies au FKIE.
Cette documentation à été fournie au laboratoire FKIE à mon départ.
L'ensemble des Macro Excel sont disponibles en me contactant : baptiste.prebot@ensc.fr

Second SA evaluation

The goal of this second evaluation is to compare the list of tasks the user thinks he has to do (responses to the questionnaire) to what he actually had to do at this moment, thus evaluating his perception of the tasks at hand (Level 1 SA). This second evaluation is also done to get more nuances in the SA evaluation, considering that missing a low importance task (as ID out) is not the same as missing a Warning and, thus should not be given the same negative evaluation.

For every Pause, tasks at hand have been listed.
This list is then compared to the one obtained from the user questionnaire.

In order to take into account the importance of the task perceived (or not), two tasks hierarchy systems have been used.

The task priority already used in the system, attributing values to tasks.

- Engagement (Bekämpfen) – **500 pts**
- Warning (Warnen) – **400 pts**
- Identification within ISR (Identifizieren innerhalb der ISR)– **300 pts**
- NRTT (NRTT-Kontakt anlegen) – **200 pts**
- Identification outside ISR (Identifizieren außerhalb der ISR) – **100 pts**

(a need for re-identification because a track changed its behaviour, is considered as an ID In ISR)

Then, answered are pondered depending on their status and importance. The perception of the task that caused the performance decrement is more important than perceiving or missing any other task. In the same way, thinking you have a non-present task to do is a less important error than missing an actual task. In this way the following hierarchy of results have been made

Perception of Decrement Task (Listing or Missing - T_{Perf}) > Missing of regular task (T_{miss}) > Listing regular task (T_{ok}) > Listing a non-present task (T_{Add})

Coefficients have been chosen empirically to keep coherence between SA evaluations.

- Task causing decrement perceived: **+ 2,2** * T_{Perf}
- Task causing decrement missed: **- 2,2** * T_{Perf}
- Normal Task missing: -**1,2** * T_{miss}
- Normal Task perceived: T_{ok}
- Non-present Task perceived: **-0,4** * T_{Add}

Where

- T_{Perf} is the value of a task that caused the performance decrement
- T_{miss} is the value of a task that is to do but not present in the user's answers
- T_{ok} is the value of a task to do correctly present in the user's answers
- T_{Add} is the value of a task present in the user's answers that is not to do
-

Based on this, the user performance ratio is calculated as follow:

$$\frac{\pm 2{,}2 \sum T_{Perf} + \sum T_{ok} - 1{,}2 \sum T_{miss} - 0{,}4 \sum T_{Add}}{2{,}2 \sum T_{Perf} + \sum T_{ok} + \sum T_{miss}}$$

Where:

$\pm 2{,}2 \sum T_{Perf} + \sum T_{ok} - 1{,}2 \sum T_{miss} - 0{,}4 \sum T_{Add}$ is the pondered sum of its correct answers and errors.

$2{,}2 \sum T_{Perf} + \sum T_{ok} + \sum T_{miss}$ is the theoretical maximum score of the user in this situation.

This score is between -1 and 1. In order to obtain a SA score between 0 and 1 (or percentage) we considered and assessment where 0.5 (or 50%) would be considered as "equally partially right, partially wrong SA". The following final formula take also into account the negative impact of negative previous SA score.

$$SA\ ratio = 0{,}5 + 1/2 \left(\frac{\pm 2{,}2 \sum T_{Perf} + \sum T_{ok} - 1{,}2 \sum T_{miss} - 0{,}4 \sum T_{Add}}{2{,}2 \sum T_{Perf} + \sum T_{ok} + \sum T_{miss}} \right)$$

In some special cases, score is still negative (when no right answer and lot of T_{Add}). In these cases, SA being completely wrong, the score is brought back to 0.

Some differences between the lists can be due to the interpretation of the question by the user and does not necessary mean he is wrong. In some cases users may have been doing a task seconds before the Pause, not really knowing he it counts as a "task to do", and not adding it to his answers when he

was actually conscious of it.
In order to compensate this questionnaire bias and distinguish between errors and misunderstanding or anticipations, videos have been analysed and special cases treated as so.

The following cases were taken into account as follow:

- If a user answered Warn or Engage that was going to be listed a few seconds later, we did not counted it as an error and did not take this answer into account, considering that he was actually anticipating on high level of importance tasks.
- Cases where Decrement is caused by IDkorrektur were taken out, considering that no answer were proposed for this case in the user questionnaire and that the ID correction task is more referring to a Level 2 situation awareness (integration – wrong mental model of the situation) evaluation than just the perception of elements of the situation.

These updates are shown in the Auswertung_alleVP_LE.csv file, “list Tasks at hand” and “SA Score” columns

Task Perception Detection & Analysis

This part of the study focuses on assessing the possibility of using Eye Tracking data to determine when a subject first looks at a new event occurring in the simulation.

To do so, a list of task have been established by selecting 23 tasks from the 100 occurring in a full simulation scenario of the AMIGO experiment.

In order to more clearly link the eye tracking activity to the task, the selected task has to be the only task at hand when occurring, meaning that:

- If a task occurs less than 1 minute after another task, it is not selected.
- If a task occurs when the subject is dealing with another task, it is not selected.

Only tasks of Identification OOR and NRTT have been selected, because being the less time sensitive ones and the most easy to spot.

To further limit the eventual parasitizing of ET activity due to other tasks, only task during BASELINE and MONOTONIE Periods have been selected.

A first selection of 23 tasks – 4 NRTT and 19 ID OOR- have been done.

Thanks to the log file, Eye tracking data for each task has been extracted, from 1 second before the task apparition (Timestamp from the log file) to 1s after the Action on the task (Track “Hooked” for ID ,” Open” for NRTT – Timestamps from log file)

This period between the occurrence of the task and the first action on it, necessary contains the PERCEPTION period.

Based on a first plot of the ET data for each task and subject, the task list has been refined by taking out 5 tasks (Tasks 4, 6, 9, 16 and 22) that were to sensible to other event happening in the simulation.

The list of participant used has also been refined to 7 participants based on the quality of the ET data available (VP 2, 4, 7, 10 and 11 have been taken out)

This new selection is composed of 18 tasks, all Identification OOR, and 7 subjects, creating a total of 126 sets of Eye Tracking Data.

If more than 15% of ET data was missing, especially around the estimated time of perception, the Data for this Task and VP was considered unusable.

If the time between the event and the action on it is superior to 15s, the data was considered unusable due to the number of parasitizing activities happening in the meantime.

This used the Perception_Analyzer_Macro.

Manual estimation of perception period

Gaze position data, layered over the screen recording, allowed us to manually estimate when the subject perceives the task, by annotating the time (in mm:ss after start) when the subject gaze towards the new event.

This has been done manually by watching the videos after each task for each of the 12 subjects. Understandably this only gives a rough estimation (only precise to the second), subject to numerous influence factors – subjectivity of the researcher, responsivity of him…

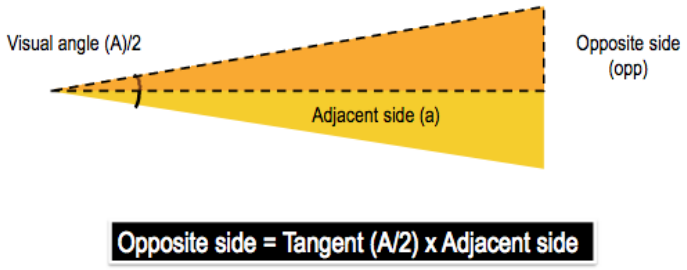
Using the gaze position data to determine a more precise perception time

By using the gaze position data, we can determine more accurately when the subject perceives the event. To do so, we first need to create a dynamic Area of Interest (AOI) around the moving track.

The size of the AOI, centred on the position of the Track, has to take into account the extent of the central visual (foveal) field with which people look and the accuracy and resolution of the eye tracking hardware.

- Foveal vision is commonly around 2°
- The accuracy of the SMI iView X eye tracker was 0.5 to 1 degree of visual angle

The AOI has been calculated the following way:



The radius of the AOI circle, is based on the measured distance (given in centimetres) between the eye and the target.

$$r = \text{Tan}((2+1)/2) * \text{Distance to screen}$$

Then converted in pixels (resolution of the HP ZR2440w screen 1920*1200 px, size 520*320 mms), 37,2pixels/cms

$$R = r * 37{,}2$$

We will be looking for the Time of first fixation in the AOI surrounding the Task

Then, for each ET data, the distance of the gaze from the Task is calculated and compared to the Rayon of the AOI, to determine if the gaze is within the AOI or not.

The timestamp of when the gaze enter the AOI is then noted in the Sum-Up sheet.

Some more details about the algorithm

- For the AOI calculation:
  - In the log file, two types of data were used to get the position of the track. Unfortunately these two position data were not measured from the same origin point

    #Track gives position in pixels, 0 being the top left corner of the screen
    #Kontakt, also gives position in pixel, 0 being the top left corner OF THE MAP area.

An offset has to be added to #Kontakt Data to correspond to #Track Data:

Xtrack = Xkontakt + 484 pixels

Ytrack = Ykontakt + 29 pixels

Excel Macro for Eye-Tracking Data extraction and analysis

This program Calculate Fixation, Saccades and Pupil Width Data and evaluate them in regard of the chosen Standard Deviation (High or Low) for every Pause recorded.

Two versions are available, one for the SMI and one for the Tobii.
SMI uses the Fixation and Saccade classification made by the Eye-tracker, Tobii does not.

Installing the macros

Paste the "ET_Analyzer_Macro" file to the XLSTART folder on the computer

(Where ? http://wheatblog.com/2011/08/where-is-the-excel-personal-macro-workbook-located/)

The macros are then available through the VIEW ("ANSICHT") Tab in Excel

What is in there?

Main program to run:

- **ET _Analyzer()** : Only program to run
  - Open Pop Up window of settings, to define :
    - Baseline wanted Start time (seconds after Start event)
    - Baseline wanted Length
    - Duration of the period before Pause to analyse
    - % of Standard Deviation to use for evaluation
  - Create a sheet with only EyeTracking data
  - Create 1 sheet/Pause with ET data of the chosen timespan before the Pause
  - Create a SUM UP sheet, regrouping the ET analysis for each Pause detected:
    - Number of Fixations, Average Fixation Length, Fixation Length Evaluation based on Baseline SD
    - Number of Saccades , Average Saccade Length, Saccade Length Evaluation based on Baseline SD
    - Average Pupil width , Pupil Width Evaluation based on Baseline SD

Secondary programs:

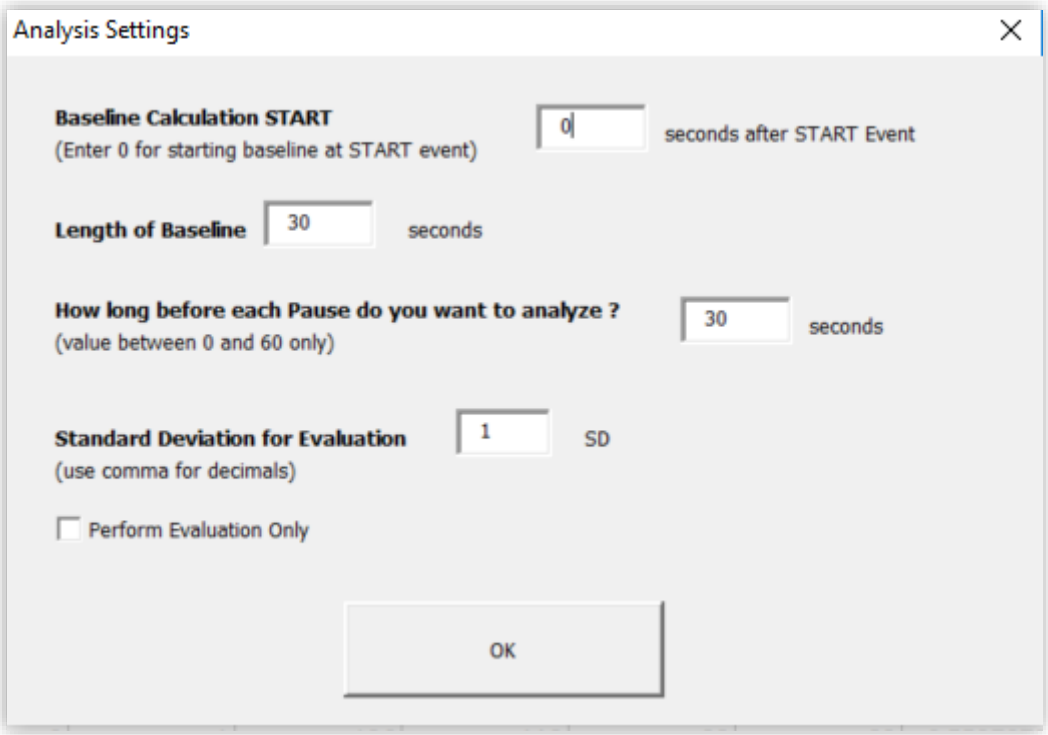
- BaselinesPupilCalcul(): For Baseline only, Calculate Average Pupil width and Standard deviation
- IsolateFix() : For Baseline only, Detect Fixations and create sheet with only them, Calculate Number of Fixations, Average Fixation Length and Standard Deviation
- IsolateScdes():For Baseline only, Detect Saccades and create sheet with only them, Calculate Number of Saccades, Average Saccade Length and Standard Deviation
- BaselineID(): Identify the ET baseline period based on the chosen length of it, and create new sheet with it.
- SDAssess(): Gives Evaluation – High or Low – depending on chosen Standard deviation %
- IsolateETData() : Extract the Eyetracking data from the raw data file and create new sheet with them named ET_DATA_ONLY

- Calcul() : Calculate Everything \o/
- Createplot() : 'NOT USED , Can automatically generate plots based on given ranges on SUMUP File
- PupilSize_FormatCheck() : NOT USED, To use if Data are imported into Excel without telling that the "." is a used as "," for numbers

How to use it?

Import the iMotions .txt raw data set into Excel. (On import, define “.” As Decimal separator and “/” as Thousands separator).
Go to the VIEW tab, in the MACRO section, choose VIEW MACROS
Select **ET _Analyzer** and click on RUN

A Settings Pop Up windows appears



Enter the chosen parameters.
Since Pauses have to be at least separated by 1min, entering a timespan superior to 60s for the third parameter, would lead to incoherent analysis.

If a first evaluation has already been done (so the SUM UP sheet already exists), the program allows to re-run only the evaluation part with a different SD to use.
To do so, check the “Perform Evaluation Only” box, and choose a SD to use.

Click OK.
Let the magic happen

WARNING:
In cases where the Eye Tracker stopped during the experiment there are Pauses happening after the last ET data recorded. In order to avoid creating sheets for these Pauses without available ET data, the program checks if there is more ET data after the detected Pause and, if not, deletes the sheet and ends the program (meaning that was the last Pause to work with). In this case, a pop up window will

ask you to confirm the deletion of the sheet. **So do not freak out, just confirm.**

Excel Macro for BioHarness Data extraction and analysis

Works similarly to the Eye-Tracking Analysis Macro, this macro does the following:

- Create one sheet per Pause containing the BioHarness Data for the chosen duration before Pause
- Calculate :
    - Average **Heart Rate** (HR) on the chosen period
    - Average **Heart Rate Variability** (HRV) on the chosen period
    - Average **Respiration Rate** (RR) on the chosen period
- Assess these average based on the chosen Standard Deviation percentage.
- Sum up all these calculated data and evaluations in a SUM UP sheet

Refer to the EyeTracking Analysis Macro section for more details about the used functions.

Excel Macro for Task Perception Analysis

The Macro program take in input the list of tasks and their time of happening, in mm:ss from the start.

Overall the program extract range of eye tracking data from 1s (1000ms) before the real apparition of Task, to 1s after the user acted on it (opening the NRTT or Selecting the Track), for each task.

In details, for every task and every subject:

- Create a Sum-up sheet for every task
- Find the real timestamp (in ms) of apparition of the Task
- Find the first action made on the task and its timestamp
- Select EyeTracking data from 1s before the Task to 1s after the Action timestamp
- Extract this data in the sum up sheet
- Detect when Gaze Data enter a defined Area Of Interest around the Task (=Perception Time of the Task)

The program can plot and export the resulting diagram for the evolution of the pupil width for each task for each subject with the Task occurrence, Action and Perception Timestamps Marked.

Kang, Z., Mandal, S., Crutchfield, J., Millan, A., & McClung, S. N. (2016). Designs and algorithms to map eye tracking data with dynamic multielement moving objects. *Computational Intelligence and Neuroscience*, *2016*. https://doi.org/10.1155/2016/9354760

Mandal, S., & Kang, Z. (2015). Eye tracking analysis using differently shaped areas of interest to represent multi- element moving objects. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *59*(1), 1515–1519. https://doi.org/10.1177/1541931215591328

Supporting the eye tracking analysis of multiple moving targets: Design concept and algorithm - IEEE Conference Publication. (o. J.). Abgerufen 19. Dezember 2018, von https://ieeexplore.ieee.org/abstract/document/6974418

Annexe 3

Analyse comparative des solutions logicielles de simulation de tâche à disposition

Nous présentons ici les critères de comparaison des solutions logicielles de simulation de tâche envisagés pour la première expérience et présentons ces solutions. A la suite de cette analyse, il a été décidé, dans un premier temps, d'utiliser la solution Wargame Red Dragon, clé en main, ne nécessitant pas de paramétrage particulier, et dans un second temps de lancer la conception et le développement d'un outil spécifique répondant mieux à nos besoins, sur la base de la solution Warship Commander Task.

Contexte et critères de comparaison

Pour cette première expérience, les contraintes relativement importantes ont façonné les choix du logiciel.
Cette première phase ayant pour but de tester des hypothèses et dégrossir les métriques, les sujets de l'expérience sont des étudiants et personnes non expertes. L'absence d'outil issu d'un véritable programme C2 n'est donc pas un souci puisque la tâche (et le logiciel) doit être suffisamment simple pour être compréhensible et maîtrisable par des personnes non-initiées au domaine militaire ou au C2, en un temps raisonnable, tout en restant une tâche relevant du control d'opération militaire.
-Ce fut donc la première contrainte imposée à l'outil à choisir ; proposer une tâche d'opération militaire tout en restant suffisamment simple et maîtrisable en un temps court.
-Il a également semblé primordial, afin d'évaluer correctement la cognition d'équipe et notamment les workloads individuels, d'avoir une maîtrise complète sur les scénarios. Ainsi une seconde contrainte est d'avoir la main sur le paramétrage des scénarios, que ce soit leur déroulement, les actions demandées, la répartition des tâches entre les opérateurs ou encore la difficulté et son évolution.
-Nous plaçant dans une situation de travail d'équipe, une troisième contrainte est bien sûr que la tâche proposée demande une collaboration entre au minimum deux opérateurs possédant chacun des rôles distincts mais complémentaires (CF DEF TRAVAIL COLLABORATIF).
Au vu du matériel de mesure à disposition au moment de ce choix, le nombre d'opérateurs impliqués dans le travail collaboratif, et donc le scénario en découlant, a été fixé à 2.
Pour cela trois outils ont été évalués au regard de ces différents critères et besoins

Solutions logicielles évaluées

Wargame Red Dragon

Ce logiciel, déjà présent au laboratoire car utilisé dans de précédents projets étudiants le développement de stratégies de groupe ; à servi de base de réflexion pour établir les besoins en terme de fonctionnalités et caractéristiques attendues de la part du logiciel de simulation de tâche de C2 à utiliser.
Wargame Red Dragon est un jeu vidéo de stratégie dans lequel des joueurs sont amenés à coopérer afin de vaincre une équipe adverse contrôlée par l'ordinateur. Les joueurs contrôlent des armées contemporaines, composées d'infanteries, de véhicules terrestres et aériens. L'objectif est alors de conquérir et défendre le maximum de zones géographiques par rapport à l'adversaire.
Si sa difficulté demande effectivement aux joueurs de mettre en place des stratégies de groupe, celle-ci n'est pas maîtrisable. L'équipe adverse étant générée par une Intelligence Artificielle, nous n'avons pas la main sur le déroulé du scénario ni sur la difficulté de celui-ci.

Certes la tâche est fidèle et le réalisme poussé, mais dans le cadre d'une étude scientifique comme celle-ci, il a semblé préférable de privilégier une tâche plus simple, peut-être moins réaliste, mais dans un environnement où l'ensemble des paramètres sont parfaitement maîtrisables et maîtrisés.
La complexité des situations et des nombreuses tâches à réaliser semblent pas si facilement abordable pour des non-initiés au domaine du jeu vidéo de stratégie et demanderait alors un temps de prise en main du jeu et de l'interface trop long pour notre utilisation.

C3 Fire

Développé par C3learningLabs, à l'université de Linkoping en Suède, C3 fire se base sur un scénario de lutte contre un incendie de forêt. Utilisé pour l'entraînement, la recherche et l'évaluation du travail d'équipe et de la prise de décision distribuée, il apparaît dans de nombreuses études notamment par la Swedish Defense Research Agency (FOI). Ce logiciel permet à une équipe d'opérateurs expérimentés de prendre différents rôles d'un centre de commandement et de contrôle de combat de feu de forêt. Ils doivent alors interagir avec la simulation et entre eux pour combattre l'incendie simulé. La tâche consiste à diriger un ensemble de brigades d'intervention (pompiers, ambulances, secours…) de manière à éteindre le feu et gérer la crise le plus efficacement possible.
Une version C3 Conflict (donc axée militaire) est en cours de développement. Cependant l'avancement du projet n'est pas encore suffisant pour nous permettre d'avoir accès à une version béta utilisable.

Warship Commander Task software (DARPA)

Développé dans le cadre d'une expérience intitulée DARPA Augmented Cognition en 2004 visant à évaluer un ensemble d'outils de mesure psychophysiologique de la cognition individuelle ; ce logiciel propose une tâche de commandement et contrôle "quasi réaliste".
Les participants monitorent les niveaux de risques d'aéronefs présents autour d'un convoi militaire et doivent y répondre en fonction de règles d'engagement précises.

Le principal défaut de cette solution est qu'elle n'est pas prévue pour étudier la collaboration. La tâche proposée est conçue pour un seul opérateur.
Cependant, de par la similarité des buts des sujets d'étude il nous a semblé intéressant de nous pencher un peu plus sur cette solution et de baser la conception d'une tâche de commandement et control collaborative sur le travail déjà réalisé pour le WCT. Les scénarios ayant déjà été étudiés de manière à posséder des niveaux de difficultés bien précis, les variables étant claires et maîtrisable...

Le tableau ci-dessous résume les besoins et les différentes solutions explorées.

TABLEAU COMPARATIF DES DIFFERENTES SOLUTIONS ET BESOINS

Critères	Wargame Red Dragon	WCT	C3 fire	Pondération du Critère (1-3)
Tâche collaborative/multi utilisateurs	oui	non	oui	3
Rôles différenciés	oui	non	oui	3
Tâche militaire	oui	oui	non	2
Tâche suffisamment simple pour non-experts	non	oui	peu	2
Main sur le paramétrage des scénarios	non	oui	peu	3
Score	oui	oui	non	1

Annexe 4

Post-synchronisation des Eye-trackers : fonctionnement du programme Matlab de correction des timestamps.

Lors de l'expérience collaborative conduite en 2017, 3 eye-trackers ont été utilisés afin d'enregistrer l'activité oculaire de 3 coéquipiers durant une tâche collaborative.
Il s'est avéré qu'une défaillance de l'horloge interne des eye-trackers a faussé les relevés. Les timestamps, marqueurs temporels de chaque prise d'échantillon, ne sont alors pas basés sur l'heure réelle d'enregistrement. Ainsi à chaque mise sous tension de l'appareil, l'horloge interne est réinitialisée au 01/01/2007. Les eye trackers n'ayant pas, et ne pouvant pas, être tous mis sous-tensions exactement en même temps, une synchronisation à posteriori de leurs relevés est donc nécessaire. Cela passe par une correction des timestamps pour les faire correspondre à l'heure et la date réelle de l'enregistrement.
La suite du document détaille le procédé utilisé afin d'effectuer cette correction et post-synchroniser les données issues des différents eye-trackers pour une même passation.

A l'issue de la session d'expérience le programme d'acquisition enregistre les données mesurées par l'eye tracker dans 3 fichiers distincts. 2 fichiers concernent respectivement les données mesurées pour l'oeil droit et l'oeil gauche, le troisième fichier est la liste des timestamps associés à ces relevés. Le nom de fichier automatiquement généré contient la date et l'heure de son enregistrement. Voici un exemple de nom de fichier enregistré : "A2_session1_partie1_Tobii_timeStampAll_2018_1_16_16_44_32.438"
Cette heure d'enregistrement doit correspondre au dernier timestamp enregistré par l'eye tracker. Or ce n'est pas le cas. L'objectif du programme est donc de traduire en timestamp correct cette heure récupérée dans le titre et de reconstituer la liste des timestamps réels à partir de celui-ci.

Le programme MATLAB développé prend en entrée le nom du fichier et récupère la date et l'heure de l'enregistrement dans celui-ci. Pour cela la chaîne de caractères est parsée. En utilisant les "_" comme séparateurs, l'année, le mois, le jour, l'heure, les minutes, les secondes et les millisecondes, sont récupérées.
La concaténation de l'année, du mois, du jour, heures, minutes, secondes et millisecondes est ensuite transformée en date (utilisation de la fonction matlab *datetime*) Dans le cas précédent, cela donne : "16/01/2018 16:44:32.348"
Cette date est ensuite traduite en timestamp UNIX (temps écoulé depuis le 01/01/1970, en ms), format initial des timestamp enregistrés par l'eye-tracker : 1516121072409.
Ce timestamp représente l'instant auquel le fichier à été enregistré. Il correspond donc également à la version corrigée du dernier timestamp enregistré par l'eye tracker.

Le programme calcule ensuite l'écart, en microsecondes, entre chaque timestamp enregistré et le précédent. La liste des timestamps corrigés est ensuite reconstituée à partir du nouveau timestamp et le report 1 à 1 des écarts. Cette nouvelle liste est alors enregistrée dans un nouveau fichier.

Ce programme est appliqué à l'ensemble des 72 fichiers de timestamps récoltés durant l'expérience

Le programme est disponible sur demande, à l'adresse baptiste.prebot@ensc.fr

Annexe 5 : Grilles d'observation inter-juges

Grille de l' approche N°1 : de la collaboration à la similarité des dilatations pupillaires (pp. 135)

Phase	Partie	Debut	Fin	Joueurs	Collaborent	Ne collaborent pas	Commentaires
1	2	11	10	2-3			
2	4	6.30	5.30	1-3			
3	20	6	5.30	2-3			
4	14	7.30	7	1-2-3			
5	20	6.40	6.10	1-2-3			
6	9	11	10.30	1-2			
7	14	3.30	3	1-3			
8	3	8.30	7.30	1-2			
9	9	5.30	5	1-2			
10	7	9	8.30	1-2			
11	11	5.15	4.45	1-2-3			
12	16	11.45	11.15	2-3			
13	4	6.30	5.30	1-2			
14	16	11.45	11.15	1-2-3			
15	2	11	10	1-3			
16	4	13	12.30	1-2-3			
17	12	6.30	6	1-2			
18	3	7	6.30	1-3			
19	11	2.30	2	1-2			
20	22	6.30	6	1-1			
21	12	2	1.30	1-2			

Grille de l' approche N°2 : de la similarité des dilatations pupillaires à la collaboration (pp. 137)

Phase	Par-tie	Dé-but (min)	Fin (min)	Type de collaboration					Commentaires
				1 avec 2	2 avec 3	1 avec 3	Ø	Les trois	
1	1	12	11						
2	1	7	6						
3	2	10	9						
4	3	8	7						
5	4	8	7						
6	4	2.30	1.30						
7	8	13	12						
8	9	3.30	2.30						
9	10	0.30	0						
10	11	9.30	8.30						
11	13	7.30	6.30						
12	13	9	8						
13	14	6.20	5.20						
14	14	13	12						
15	15	12	11.30						
16	16	5.30	5						
17	18	5	4						
18	20	9.30	9						
19	22	3.30	3						
20	23	5	4.30						
21	24	9.30	8.30						

Annexe 6

Note on Data centricity in Command and Control and Decision Making

Par Baptiste PREBOT, Ingénieur Cogniticien et Doctorant DGA – ENSC
baptiste.prebot@ensc.fr

A l'intention de l'État-major des Armées dans le cadre de la participation française à la Chiefs of Transformation Conference 2018 (COTC18) – "Disruptive Advances Shaping Warfare"

Les processus de collaboration et décision mis en œuvre dans le C2 reposent sur la mobilisation de nombreux acteurs et l'analyse d'une grande quantité d'information. La donnée à toujours eu un rôle central car d'elle découlent l'information et l'intelligence, éléments primordiaux dans le contexte actuel de Network Centric Warfares où le partage d'informations et la fusion de données sont la clef vers la supériorité informationnelle recherchée.

Ces dernières décennies ont vu la multiplication des sources de données. Drones, Open data, internet des objets, réseaux sociaux…sont venus enrichir le déjà impressionnant volume de données produit par le C4ISR. De par la diversité de ses formats et sources, cette masse de données représente une ressource stratégique pour mieux comprendre et anticiper la situation et le contexte. Peindre une représentation de la situation plus précise et dynamique est un besoin fondamental dans les conflits actuels de plus en plus complexe, au nombre et à la diversité des acteurs grandissants.

Toute prise de décision repose sur la compréhension et représentation de la situation actuelle. Il va de soi que plus cette compréhension est fine, poussée et juste, plus la décision prise pourra être optimale. Ces données, en les croisant, les combinant, les exploitant, représentent donc la solution pour fournir au C2 l'agilité, la réactivité et la capacité de prédiction nécessaires pour mener les opérations d'aujourd'hui et de demain, reposant sur une représentation partagée de la dynamique de crise.

Cependant la quantité de données disponible est telle que non seulement elle menace de surcharger les réseaux, mais surpasse également la capacité de traitement des opérateurs Humains, et n'est donc pas exploitée à son plein potentiel.
Nous possédons aujourd'hui les moyens technologiques suffisants pour nous fournir une matière première que nous ne sommes pas en mesure d'exploiter comme il se doit.

L'intégration au C2 d'outils d'Intelligence Artificielle permettant l'obtention d'une collaboration symbiotique entre les Hommes et la machine, est la seule façon de gérer une telle complexité. Big Data, Data Analytics, Cognitive Computing et autres Cognitives Tools, comme nous les appelons, doivent permettre de fusionner et interpréter en quasi temps-réel, les sources de renseignements habituelles, les données de terrain et les informations recueillies via l'analyse des réseaux sociaux, de « l'internet des objets » et des Open Data.

Les outils de « Cognitive Computing », doivent permettre d'augmenter la compréhension de la dynamique du conflit et de la stratégie ennemie à tout niveau de granularité du cycle de la décision. Les Big Data et Analytics peuvent aider les hommes à surveiller les activités, automatiser les questions de routine, mais également monitorer et prédire les CCIRs, développer un war-gaming plus réaliste et automatique, suggérer des plans d'actions plus appropriés…

La perspective de l'analyse en temps réel permet jusqu'à entrevoir le « reshaping » de la OODA loop où planification, observation, évaluation et décision seraient effectués en continue (Voir Desclaux 2016).

Eloignons-nous un instant du domaine militaire et regardons ce qui se fait pour le commerce en ligne, où le Big Data et les Analytics sont le centre de la stratégie des acteurs de ce domaine. Le monitoring et tracking de l'activité sur internet permet aux sites ou aux agences publicitaires de profiler l'utilisateur, d'anticiper ses attentes, ses hésitations, ses gouts, cela afin de lui suggérer aussi bien une nouvelle paire de chaussures à acheter qu'un politicien à élire.
S'il est possible de profiler ainsi le visiteur lambda d'un site de e-commerce, alors pourquoi ne serait-il pas possible de faire de même pour une ville ou une région durant un conflit ?

L'apport de ces outils est nécessaire, obligatoire, mais avec eux doit venir une refonte de la structure même, des rôles de chacun, de la répartition des tâches et l'organisation des processus. Ils ont le potentiel de profondément bouleverser l'organisation du C2, à tous les niveaux hiérarchiques. Redéfinir qui prend quelle décision, quand, pourquoi et comment.

De toutes parts nous entendons parler d'automatisation, d'armes et de véhicules autonomes… si bien que l'autonomie est devenue une orientation reflexe dans le développement de tout nouveau système. L'automatisation accrue est communément vue comme la « next-step » évidente. Or cette révolution doit passer, non pas par de l'automatisation à tout va, mais par un réel « teaming » homme-machine. Une hybridation de l'homme par la machine et de la machine par l'homme. Afin de tirer avantage de la complémentarité de chacun pour former un système le plus performant possible. Il ne s'agit plus de réfléchir à la limitation de l'un ou de l'autre mais à l'optimisation de la performance du couple, de l'équipe Homme-Machine.

Le man-machine teaming est un changement de paradigme clef qui doit être adopté pour penser les systèmes et organisations de demain dans le sens de l'hybridation de l'un et l'autre et non plus du remplacement de l'un par l'autre.

Le futur est indéniablement data-centric, la donnée étant la ressource la plus précieuse que nous ayons et que nous possédons en si grande quantité. Bien sûr des risques déjà existants en seront exacerbés. La sensibilité à la manipulation de l'information est de fait plus critique et les menaces cyber sont grandissantes, rendant la cyber-sécurité un enjeu clef. Dans le même temps, la collaboration entre les pays et les armées autour de ces données et de leurs systèmes d'exploitation est primordiale pour décupler la puissance de leur analyse.

Il nous faut maintenant travailler au développement des systèmes et modèles organisationnels nous permettant la pleine exploitation du potentiel de ces données.

Quelques références :

Claverie, B., Desclaux, G. (2016). Command and control : un système de systèmes pour accompagner la complexité.. Communication et Organisation, Presses Universitaires de Bordeaux, 2016, Engagement entrepreneurial et territoires, 50, pp.255-276

Desclaux, G., Claverie, B., Chaudron, L., Le Guyader, H., Prebot, B., Salotti, J.M. (2016). Cognitive computing and augmented intelligence in man-system integration: Impact on C2 HQ key processes. COTC'16 : 2016 Ch iefs of

Transformation Conference « Innovate, Adapt, Tra nsform : Prepare Now for 2030 ». NATO Allied Command Transformation. Norfolk (Virginia - USA) . 13-15 December 2016. Proceedings.

Desclaux, G., Prebot, B. (2018). Command and Control at the Autonomy and Cognitive Era : For a decision cycle augmented by the symbiosis between human and systems. in 23rd ICCRTS International Command & Control Research & Technology Symposium, Orlando, FL (November 2018). Proceedings

Endsley, M. R. (2017). From here to autonomy: lessons learned from human–automation research. Human factors, 59(1), 5-27.