



Développement d'une plate-forme d'évaluation personnalisable et adaptable pour l'étude du comportement émotionnel en situation de multisollicitation

Régis Mollard, Marion Wolff, Nadine Couture, Alexis Clay

DANS **LE TRAVAIL HUMAIN** 2012/3 (VOL. 75), PAGES 253 À 277

ÉDITIONS **PRESSES UNIVERSITAIRES DE FRANCE**

ISSN 0041-1868

ISBN 9782130594116

DOI 10.3917/th.753.0253

Article disponible en ligne à l'adresse

<https://www.cairn.info/revue-le-travail-humain-2012-3-page-253.htm>



Découvrir le sommaire de ce numéro, suivre la revue par email, s'abonner...

Flashez ce QR Code pour accéder à la page de ce numéro sur Cairn.info.



Distribution électronique Cairn.info pour Presses Universitaires de France.

La reproduction ou représentation de cet article, notamment par photocopie, n'est autorisée que dans les limites des conditions générales d'utilisation du site ou, le cas échéant, des conditions générales de la licence souscrite par votre établissement. Toute autre reproduction ou représentation, en tout ou partie, sous quelque forme et de quelque manière que ce soit, est interdite sauf accord préalable et écrit de l'éditeur, en dehors des cas prévus par la législation en vigueur en France. Il est précisé que son stockage dans une base de données est également interdit.

DÉVELOPPEMENT D'UNE PLATE-FORME
D'ÉVALUATION PERSONNALISABLE ET ADAPTABLE
POUR L'ÉTUDE DU COMPORTEMENT ÉMOTIONNEL
EN SITUATION DE MULTISOLLICITATION

par R. MOLLARD*, M. WOLFF*, N. COUTURE** et A. CLAY***

SUMMARY

DEVELOPMENT OF AN ADAPTABLE AND CUSTOMIZABLE EVALUATION PLATFORM FOR THE STUDY OF EMOTIONAL BEHAVIOR IN SITUATIONS OF MULTIPLE SOLICITATIONS

This paper presents a platform for the evaluation of user experience pertaining to innovative concepts in evolutionary socio-technical systems. Developing a system involves performing formative evaluations (during the design phase) and summative evaluations (at the end of the design phase).

Designing or redesigning systems is creating new human-human interactions and human-system interactions through technological innovations, but also creating new uses. Anticipating these uses and evaluating these interactions requires setting up experiments and tests using methods and tools to capture, analyze, and interpret the activities and behaviors of individuals while they are interacting. Beyond the use (utility, usability, sense), the emotion of users may be a factor in the performance of a system. The platform combines prototype methods and tools. In this paper, we present a method for assessing the emotional processes of subjects faced with multiple solicitations during the performance of a monitoring task. Through a usage scenario, this paper also presents the evaluation protocol, and the methods chosen for their relevance to the situation. The protocol used is based on an attentional test based on the MultiAttribute Task (MAT), where subjects are being diverted from their main task by agents/stressors/distracters. With the exception of heart rate, the parameters used to characterize the subject's behavior, both in terms of performance and in terms of emotional reactions, are sensitive to the effect of task complexity. This was also the case for subjective assessments (difficulty of the task, self-assessment of performance, emotions, etc.). Using multivariate analysis (Principal Component Analysis) it has been possible through the joint analysis of performance and verbalizations to identify individual strategies depending on the type of distractor that the subject should manage. The emotional response is evaluated from different angles with a gradient of difficulty. It can be interpreted according to the usage scenario, and it can be related to the level of performance.

* Université Paris-Descartes, Lati/Axe Ergonomie – 45, rue des Saints-Pères, F-75270 Paris Cedex 06 & ESTIA, F-64210 Bidart – email : regis.mollard ou marion.wolff@parisdescartes.fr

** ESTIA & LaBRI, UMR 5800, F-33400 Talence, France. – email : n.couture@estia.fr

*** ESTIA, F-64210 Bidart, France – email : a.clay@estia.fr

The platform and the associated method described in this paper highlight the importance of timing of a maximum means in real time, while making choices to limit the intrusiveness of the tools used, and to select appropriate methods for performing statistical analysis.

Keywords: *Use-Test, Multiple solicitations, Evaluation, Interactive behavior, Multivariate analysis.*

I. INTRODUCTION

Pendant longtemps les évaluations de l'interaction Humain-Système (Humain-Humain et Humain-Machine) ont été réalisées selon des critères fonctionnels et des facteurs de performance. Cette interaction ne se limite pas à ces éléments, elle est également conditionnée par le ressenti et l'état émotionnel des personnes. Depuis maintenant deux décennies, l'importance des émotions dans l'interaction a été mise en évidence par différentes publications (Jordan, 2004 ; Norman, 2004) issues de domaines divers : psychologie, ergonomie, psychophysiology, ingénierie, etc. Mais rares sont les études qui prennent en compte l'ensemble de ces paramètres, car les difficultés à surmonter sont multiples : il s'agit dans un premier temps de synchroniser des recueils de données quantitatives de performance en relation avec des mesures physiologiques et des indicateurs comportementaux (expressions faciales, directions du regard). Dans un second temps, il faut pouvoir analyser l'ensemble de ces données multivariées quantitatives, tout en gérant les artefacts, et également pouvoir prendre en compte des données qualitatives (observations, verbalisations) utiles pour la compréhension de certains comportements ou réactions.

L'objectif principal de cette étude est de valider les modalités de recueil synchronisé de données multiples, de vérifier que les procédures de rejet des artefacts fonctionnent et surtout, pouvoir tester la sensibilité des indicateurs retenus pour caractériser les comportements des sujets confrontés à de multiples sollicitations. Cette étude a été menée avec une nouvelle plateforme, la Plate-forme d'Évaluation de Prototypage et de tests d'usageS (PEPSS) mise en place à l'ESTIA par une équipe pluridisciplinaire (Couture, Wolff, Mollard & Todeschini, 2010). Le protocole de test-utilisateur présenté dans cet article constitue un exemple d'utilisation de la PEPSS, dont la particularité est d'être reconfigurable afin de pouvoir simuler de multiples évaluations d'interactions Humain-Machine/Humain-Humain.

Cette plateforme est définie comme un centre de tests pour produits innovants. À cet effet, des situations expérimentales transposables à différents domaines sont mises en place afin de pouvoir analyser le comportement des utilisateurs placés face à des concepts innovants et étudier les aspects liés, non seulement à l'utilisabilité du produit, mais aussi à l'organisation du travail et au bien-être au travail, en proposant différents tests situationnels où les sujets auront à exécuter des tâches nécessitant de l'attention et durant lesquelles des éléments perturbateurs interviendront (alarmes, sollicitations humaines...). Ces comportements sont ainsi abordés sous plusieurs aspects tant du point de vue de la performance que

du point de vue émotionnel : vécu de la situation, gestion des émotions positives et négatives.

Le domaine de la reconnaissance d'émotions par ordinateur est un domaine très actif, marqué par un intérêt croissant depuis sa création. La première conférence internationale rassemblant les acteurs du domaine à la fois en informatique et en psychologie, ACHI, s'est déroulée en 2005 et a été reconduite en 2007 et 2009. Depuis les travaux de Picard en 1997, de nombreux systèmes ont vu le jour, permettant une reconnaissance sur les divers canaux de communication émotionnelle (visage, gestuelle, voix, réactions physiologiques et nerveuses), d'abord indépendamment puis simultanément (Pantic, Sebe, Cohn & Huang, 2005 ; Jaimes & Sebe, 2007 ; Zeng, Pantic, Roisman & Huang, 2009). Les systèmes de reconnaissance ont donc au fil des années considéré les différents vecteurs de la communication émotionnelle, en commençant par le visage et la voix, pour ensuite s'intéresser à la gestuelle et enfin aux réactions physiologiques de l'émotion.

Il existe de nombreuses définitions de l'émotion. Les émotions primaires telles que définies par la littérature (Jaimes & Sebe, 2007) doivent être déclenchées par des stimuli universels, apparaître spontanément, déclencher des pensées ou sensations spécifiques, être présentes chez d'autres primates que chez l'humain. Par ailleurs, la réaction doit être rapide à apparaître et à disparaître, et le traitement cognitif du stimulus doit être automatique. Seules six émotions répondent à ces critères : la joie, la peur, la colère, la surprise, la tristesse et le dégoût.

Les émotions secondaires (combinaisons d'émotions primaires comme la honte, résultant de la combinaison de la peur et de la colère) ainsi que des états (humeurs) peuvent déclencher des réactions comparables aux émotions primaires. Nous nous basons également sur le modèle de Scherer (2004) issu des théories de l'évaluation cognitive et nous considérons dans ce cadre les émotions utilitaires qui sont des épisodes relativement brefs de réponses synchronisées de tous ou de la plupart des systèmes organiques en réponse à l'évaluation d'un événement interne ou externe étant d'une importance majeure pour des besoins ou pour des buts personnels.

II. DE L'INTERACTION HUMAIN-MACHINE À L'EXPÉRIENCE UTILISATEUR

II.1. ERGONOMIE ET ÉMOTIONS

L'ergonomie s'est attachée à améliorer la performance des systèmes (humains et machines) au travers de l'utilité, de l'utilisabilité (facile à utiliser, facile à apprendre, facile à retenir, peu d'erreurs, satisfaction) et de l'accessibilité (utilisable par le plus grand nombre) des machines lors de leurs conceptions et de leurs évaluations. Si cette démarche a longtemps été adaptée au travail, le champ étymologique de l'ergonomie, pour lequel les choix de l'opérateur en matière d'organisation du travail et d'outils sont restreints, il n'en va pas de même lorsque le choix est possible. D'une part,

il y a un glissement de l'ergonomie, du travail vers les produits et services de la vie quotidienne ; dans ce contexte, l'utilisateur peut pleinement exercer ses choix et là, les critères pertinents ne sont plus uniquement l'utilité ou l'utilisabilité (Overbeeke, Djajadiningrat, Hummels & Wensveen, 2000), mais plutôt l'émotion procurée et l'image de soi renvoyée. D'autre part, les chercheurs ont montré que même dans l'activité de travail, l'émotion peut avoir une influence sur la cognition. Certaines études ont montré qu'une émotion positive peut entraîner l'augmentation de la prise de risque et de la rapidité de la prise de décision complexe. La cognition peut influencer l'état émotionnel, et le partage de ressources attentionnelles face à des objectifs contradictoires est susceptible d'induire de la frustration lors de la prise de décisions non satisfaisantes. De même, lorsque la prescription des émotions est forte, contradictoire et pressante, elle peut compromettre la relation au client (Jeantet, 2003). Il est donc nécessaire de changer de paradigme vers une approche prenant en compte l'être dans sa globalité (travail et vie privée). L'expérience utilisateur est fondée sur : la dimension affective et les émotions, la beauté perçue du produit, la qualité hédonique du produit, le plaisir de l'utilisateur (amusement, bien-être), la richesse personnelle et l'image de soi. Et l'on peut dire sur l'esthétisme que « les belles choses » donnent envie d'être utilisées (Jordan, 2004) et donnent l'impression de pouvoir travailler mieux. L'esthétique de l'interface influencerait ainsi la perception de l'utilisabilité du système (Tractinsky, Katz & Ikar, 2000). D'autres auteurs soulignent la forte relation entre la beauté du produit et l'identité sociale de celui à qui il appartient, mais il est également important de mesurer la relation qu'il y a entre l'attrait et l'utilisabilité des objets (Hassenzahl, 2004 ; McDougall, Reppa, Smith & Playfoot, 2009).

Ainsi, l'attractivité des nouveaux outils interactifs peut devenir aussi importante que la fonctionnalité et la fiabilité. L'esthétique a valeur symbolique et métaphorique, surtout pour les objets tangibles qui, par certaines de leurs formes, transmettent une signification à l'utilisateur. Enfin avec l'esthétique et le design de l'interface, on peut également aborder l'esthétique d'interaction – la beauté d'interaction et d'utilisation que certains auteurs n'hésitent pas à souligner (Djajadiningrat, Overbeeke & Wensveen, 2000), en insistant sur la richesse possible des interactions ou sur l'esthétique pragmatique (Petersen, Iversen, Krogh & Ludvigsen, 2004).

De nombreuses pistes sont donc à explorer visant à utiliser le côté ludique des objets issus des nouvelles technologies, leur esthétique et le plaisir qu'ils peuvent apporter aux utilisateurs, tout en étudiant les aspects ergonomiques de ces futures applications, ce qui va dans le sens des nouvelles conceptions gravitant autour de la notion d'expérience utilisateur (voir à ce sujet le cadre intégrateur, inclusif et holistique, proposé par Barcenilla et Bastien, 2009).

II.2. LES RÉPONSES ÉMOTIONNELLES

Les réponses à une émotion sont décrites dans l'état de l'art de Mauss et Robinson (2009) : l'expérience subjective, la réponse du système nerveux périphérique et autonome, la réponse du système nerveux central, le comportement.

L'émotion peut être évaluée de manière discrète (oui ou non) ou en termes de valence et d'intensité. Le modèle de représentation choisi dans les travaux présentés ici est un modèle discret où chaque type d'émotion est désigné par un label spécifique (c'est-à-dire un mot), par exemple « joie », « peur », etc. Cette approche catégorielle est un avantage certain en informatique. En effet, reconnaître une émotion d'un ensemble discret revient à choisir une catégorie (c'est-à-dire une émotion) parmi celles proposées dans l'ensemble. La catégorisation de signaux est un problème largement étudié en informatique, qui dispose d'outils adéquats à sa résolution (comme les réseaux de neurones). Les émotions basiques d'Ekman (1992) que nous avons retenues sont issues de la théorie évolutionniste. Pour Darwin, les émotions sont des réactions à des stimuli, préparant le corps à agir d'une certaine façon. Issues de l'évolution, elles se sont développées pour offrir une réponse extrêmement rapide et adaptée à la perception de la situation. Il est ainsi possible de voir les émotions comme des réactions préprogrammées à certains événements. Par exemple, la peur prépare le corps à la fuite et la colère, à l'attaque. Les émotions peuvent être appréhendées de différentes manières, aussi bien d'un point de vue subjectif que physiologique (Lottridge & Moore, 2009).

L'expérience subjective peut être évaluée par l'analyse du discours (indices lexicaux, morphologiques, syntaxiques, sémantiques, pragmatiques) ou encore l'autoévaluation à partir d'échelles ou d'outils construits et validés.

La réponse du système nerveux périphérique peut être analysée notamment par les mesures de conductance électrodermale, de fréquence cardiaque ou de pression sanguine ; celle du système nerveux central par électroencéphalographie ou imagerie par résonance magnétique.

Enfin, la réponse comportementale peut être recueillie par la mesure de la fréquence du signal acoustique pour les modifications vocales (ton, timbre, fréquence fondamentale et intensité), électromyographie des muscles corrugateurs *supercilli*, *orbiculis oculi* ou des zygomatiques pour le comportement facial, la mesure des déplacements de points particuliers du visage pour l'expression faciale, la mesure du diamètre de la pupille, la mesure des mouvements des segments corporels de la totalité du corps, ou encore une observation par un tiers.

Selon Lottridge et Moore (2009), ces différents types d'évaluations ont été classés en quatre approches disciplinaires fondées sur :

- le postpositivisme (recueil de données objectives : mesures physiologiques, réponse électrodermale ou fréquence cardiaque) ;
- le constructivisme (verbalisation du ressenti et la production de données qualitatives) ;
- le pragmatisme (autoévaluation à partir d'outils validés expérimentalement sur la base de données statistiques telles que PrEMO (Lottridge & Moore, 2009) ou sur l'échelle SAM *Self-Assessment Manikin scale* [Gil, 2009]) ;
- le plaidoyer participatif (évaluation sociale au travers, par exemple, de la reconception critique d'un produit ou l'évaluation de la frustration ou de la satisfaction en situation de travail).

Ces quatre approches ayant chacune leur intérêt font l'objet de champs de recherche particuliers.

Les innovations informatiques proposent également aujourd'hui de nouveaux outils pour évaluer les émotions comme en témoigne l'étude récente de Becker, Aucouturier, Mougenot et Yamanaka (2010) pour étudier l'effet de la présentation d'images de couleur sur l'émotion provoquée par un son (toux ou éternuement). En ingénierie informatique, de nouveaux outils se sont également développés pour tenter de reconnaître les émotions par le mouvement en temps réel *via* un logiciel spécifique qui a été récemment mis au point : *eMotion* (Clay, Couture & Nigay, 2009 *a* ; 2009 *b* ; 2010). Le mouvement du sujet est capturé par une combinaison de capture du mouvement *Moven* de chez *XSens*¹ et suit les trois niveaux classiques de capture (des informations du monde réel, ici par la combinaison), d'analyse (des données de capture en caractéristiques pertinentes pour la reconnaissance d'émotions, ici des caractéristiques du mouvement) et d'interprétation (des caractéristiques pour fournir une émotion). Ce nouvel outil, initialement lié au domaine artistique, et plus particulièrement à la danse, a prouvé récemment son efficacité (Clay *et al.*, 2009 *b* ; Clay, Delord, Couture & Domenger, 2010) ; il pourrait être tout à fait complémentaire aux autres évaluations précitées mettant en œuvre des recueils de données quantitatives et qualitatives, si dans le domaine étudié les sujets peuvent être évalués en situation dynamique et non statique.

Pour provoquer les réponses du sujet, on peut utiliser comme tâches : la recherche d'icônes sur le bureau d'un ordinateur (McDougall *et al.*, 2009), des batteries standardisées d'images, comme l'International Affective Picture System ou des séquences de films vidéo (Buisine *et al.*, 2006), de musique (Gil, 2009). Lors de l'utilisation d'images, on peut faire également varier le contraste pour étudier les interdépendances perception/émotion (Gombert, 2009). Il est également possible d'utiliser la réussite ou l'échec à une tâche donnée, la lecture de mots ou de phrases décrivant des émotions (Gil, 2009).

L'étude que nous proposons dans cet article met en œuvre un test utilisateur où le sujet, en position debout, est évalué lors d'une tâche de surveillance en situation statique, les outils d'étude des émotions seront par conséquent d'ordre classique : recueil de performances lors d'une tâche de surveillance d'écran, de données physiologiques et de verbalisations. Lors de cette première expérience PEPSS, l'utilisation de l'outil *eMotion* et de la combinaison *Moven* (Clay *et al.*, 2009 *a*) a été également tentée. Bien que cette tentative se soit révélée infructueuse pour ce test spécifique, cela a permis d'évaluer très précisément les limites d'utilisation de tels outils et de réfléchir à des protocoles plus adaptés qui seront mis en place ultérieurement.

III. TEST DE MULTISOLLICITATION : PROTOCOLE

Le protocole de test utilisateur mis en œuvre par la PEPSS pour cette étude simule les tâches qu'un opérateur aurait à effectuer en parallèle (multisolllicitations) : situation d'accueil de clientèle/visiteurs et/ou de

1. Distribuée par la société Xsens : <http://www.xsens.com/en/general/mvn>

surveillance d'événements (situation type du guichet d'accueil où l'opérateur doit fournir des renseignements multiples tout en effectuant des tâches de surveillance).

Dans ce protocole, la tâche principale du sujet est d'effectuer une surveillance d'événements informatisés (dérives de curseurs, contrôle de voyants) *via* un test attentionnel. Le test utilisé pour la PEPSS est inspiré de la *MultiAttribute Task Battery* (MATB) fréquemment employée dans les recherches en aéronautique pour mesurer la charge mentale des opérateurs (Comstock & Arnegard, 1992). Durant ce test, en plus de cette tâche attentionnelle, un certain nombre de distracteurs-stresseurs interviennent régulièrement, venant perturber l'opérateur dans sa tâche. Le protocole mis en œuvre pour ce test de multisollicitations se déroule en trois phases, chacune d'une durée de six minutes. Durant chacune de ces phases, trois types de distracteurs vont intervenir pendant l'exécution de la tâche principale : sollicitations de personnes venant demander la localisation d'un individu, alarmes sonores et visuelles à stopper, demande de codes d'accès *via* un talkie-walkie. Chaque phase débute de la même manière : la première minute, le sujet n'effectue que la tâche principale, puis lors de chaque minute suivante les distracteurs interviennent (Couture *et al.*, 2010).

Ces trois phases sont de difficulté croissante : lors de la phase 1, les sollicitations interviennent l'une après l'autre, elles sont ensuite couplées en phase 2, pour apparaître enfin en simultané et en continu lors de la phase 3. Après chaque phase sont données aux sujets des échelles d'évaluation du niveau de leur performance, de confiance dans cette évaluation, de difficulté de la tâche, de l'effort fourni, ainsi que des échelles d'autoévaluation des émotions « *Self-Assessment Manikin scale* (SAM) » (Gil, 2009) se déclinant en trois échelles illustrées par des pictogrammes que le sujet doit cocher pour indiquer son état émotionnel : plaisir, éveil et dominance de la situation (ces trois échelles seront illustrées dans les pages suivantes, Figure 2).

Le sujet évalué est en position debout, situation d'accueil appliquée de plus en plus fréquemment en secteur public, tertiaire ou encore de surveillance dans l'industrie (Mollard, Gibert & Wolff, 2010), face à un pupitre sur lequel est posé un ordinateur. Il est équipé de différents capteurs liés au matériel utilisé, et ses réactions aux distracteurs peuvent être enregistrées en temps réel ainsi que ses performances.

Deux caméras vidéo synchrones sont utilisées pour enregistrer d'une part les mouvements de la tête et les expressions faciales (vue de face), et d'autre part les changements d'attitudes posturales du sujet (caméra de scène).

À la fin du test, un débriefing est proposé au sujet lors duquel les verbalisations sont enregistrées.

III.1. TÂCHE PRINCIPALE :

TEST ATTENTIONNEL DE TYPE *MULTIATTRIBUTE TASK* (MAT)

Le « MAT », test attentionnel de suivi de processus, est la tâche principale que le sujet doit effectuer. Le sujet doit surveiller quatre curseurs et deux voyants. Les curseurs se déplacent verticalement. Lorsque l'un des

curseurs vient en butée, haute ou basse, le sujet doit cliquer sur la touche de fonction correspondante (F1, F2, F3, F4) du clavier pour ramener le processus dans un état stable. (voir Figure 1 ci-après). Un des voyants est allumé, l'autre est éteint, lorsqu'un des boutons change d'état, le sujet doit cliquer sur la touche de fonction correspondante (F5 ou F6) du clavier pour ramener le processus dans son état initial. Si le sujet met plus de quinze secondes à réagir quand le curseur atteint la butée, le processus revient dans un état normal.

L'ordre d'apparition des événements est fixe et se reproduit 24 fois sur une période de 6 minutes (4 par minute) selon une procédure prédéfinie. Compte tenu du nombre de sollicitations, il est considéré que le sujet ne peut pas les mémoriser pour éviter un effet d'apprentissage en phases 2 et 3, ce qui a été vérifié lors des expérimentations.

Ce test permet d'évaluer les performances du sujet en termes de temps de réponse, nombre de non-réponses (lorsque les quinze secondes sont écoulées sans réponse du sujet).

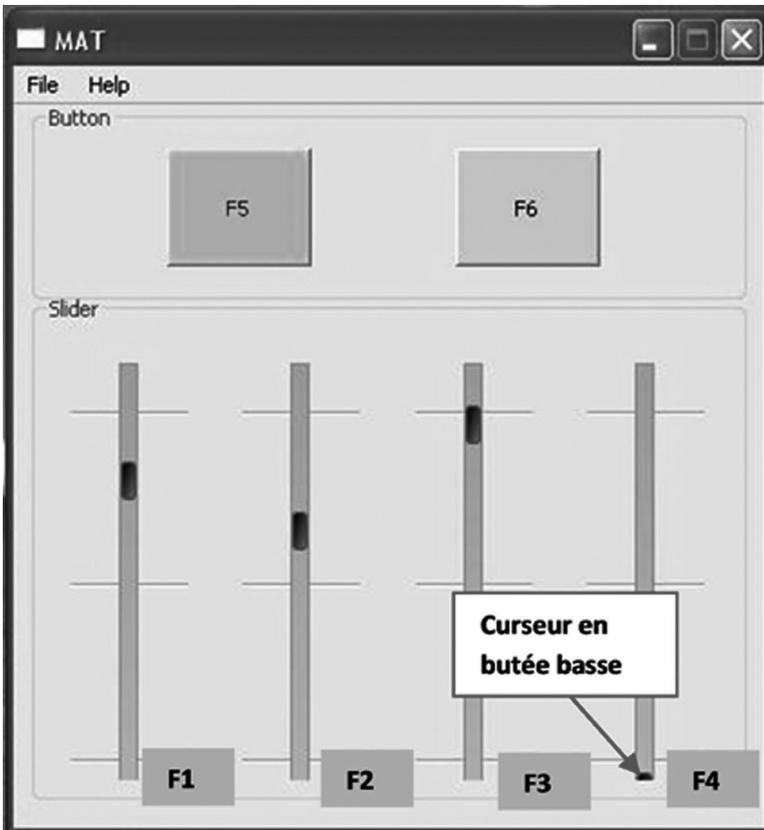


Figure 1 : Test MAT
Figure 1: The MAT test

III.2. DISTRACTEURS

Les distracteurs sont au nombre de trois et interviennent régulièrement toutes les minutes à partir de la deuxième minute de chaque phase (chacune d'une durée de six minutes) et dans un ordre aléatoire : l'un après l'autre pour la phase 1, couplés pour la phase 2, puis simultanément et en continu pour la phase 3 afin de complexifier progressivement la tâche du sujet :

1. Des compères viennent demander au sujet le numéro de salle correspondant à la localisation d'une personne. Ce numéro est indiqué dans une liste fournie au préalable au sujet, qui doit chercher la bonne correspondance nom/n° de salle parmi une vingtaine de choix possibles, puis l'indiquer verbalement à la personne concernée.
2. Des compères appellent le sujet (*via* un talkie-walkie) pour lui demander le code de sécurité d'un site répertorié dans une liste fournie au sujet, qui doit chercher la bonne correspondance site demandé/code parmi six choix possibles, puis l'indiquer verbalement *via* le talkie-walkie.
3. Une alarme sonore retentit accompagnée d'une alarme visuelle. Le sujet doit alors saisir au clavier un mot de passe dans une fenêtre présente à l'écran. En « magicien d'Oz », un compère éteint les alarmes dès que le mot de passe est saisi.

III.3. DONNÉES PHYSIOLOGIQUES ET COMPORTEMENTALES

La direction du regard et les expressions faciales sont enregistrées au moyen du système d'*eye tracking* « FaceLAB5 ». Ainsi, selon le système d'encodage des actions faciales « FACS » d'Ekman (1992), il devient possible à partir de points caractéristiques du visage de déterminer des expressions faciales et de les associer à des émotions.

Les réponses physiologiques (fréquence cardiaque FC et réponse électrodermale notées SC pour *Skin Conductance*), le codage des activités et des séquences vidéo, sont acquises de manière synchrone à l'aide du dispositif « ProComp » et du logiciel « Captiv », couplés avec le matériel « FaceLAB5 »¹.

III.4. DONNÉES PHYSIOLOGIQUES ET COMPORTEMENTALES

Après chaque phase, chacune d'une durée de six minutes, le sujet doit évaluer : son niveau de performance, son niveau de confiance dans l'évaluation précédente, la difficulté de la tâche et l'effort fourni. Ces quatre évaluations sont présentées sous forme de fiches, comportant des échelles analogiques graduées (ou non) de 1 à 10. Les trois échelles SAM (représentées par différents pictogrammes à cocher) concernant le plaisir, l'excitation et la dominance sont également présentées au sujet (voir Figure 2 ci-après).

1. Matériel distribué par l'équipe TEA (Technologie, Ergonomie, Application) : <http://www.teaergo.com>

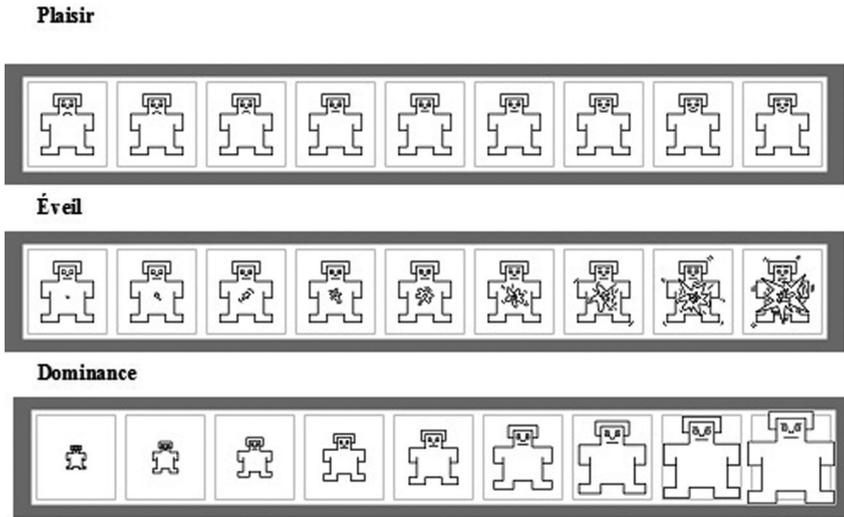


Figure 2 : Échelles « Self-Assessment Manikin » (SAM) proposées par Gil (2009)

Figure 2: Self-Assessment Manikin scale (SAM) suggested by Gil (2009)

À la fin de l'expérimentation, afin de pouvoir relier les différentes performances du sujet à son ressenti (Cahour, 2006), un « débriefing » a été organisé selon un guide d'entretien ciblant l'évaluation générale du test, le ressenti, les difficultés, les stratégies mises en place pour assurer la tâche principale, pour neutraliser les distracteurs, etc. Les verbalisations sont enregistrées numériquement.

III.5. SUJETS, CONSIGNES ET ÉQUIPEMENTS

Lors d'un prétest, trois sujets (étudiants en 2^e année d'école d'ingénieurs à l'ESTIA) ont participé dans le but d'ajuster les différents paramètres du test-utilisateur, de vérifier la capture de la fréquence cardiaque, de la réponse électrodermale et de la position des regards, en liaison avec la synchronisation des outils informatiques.

Lors du test-utilisateur, quatre sujets (deux élèves ingénieurs à l'ESTIA de sexe masculin, âgés de 24 ans et deux secrétaires de sexe féminin, appartenant au personnel administratif de l'ESTIA, moyenne d'âge de 36 ans) ont pu être évalués sur l'ensemble de la durée de l'expérimentation (1 heure environ). Compte tenu du faible nombre de sujets, cette étude se place dans un champ exploratoire, bien que l'analyse des données, qui combine à la fois du multidimensionnel et de l'inférence, lui donne déjà une certaine validité (cf. *infra*).

Avant la procédure de test, le sujet est accueilli et un questionnaire concernant son âge, genre, ainsi que ses connaissances quant à sa réactivité émotionnelle lui est proposé et l'anonymat est garanti. L'autorisation de filmer et de diffuser est également demandée.

Il a été également vérifié que les sujets n'étaient pas atteints d'*alexithymia* (difficulté à exprimer verbalement ses émotions) et qu'ils ne présentaient pas de phobies susceptibles d'être provoquées par le matériel de test.

Le test est alors présenté au sujet ainsi que le matériel de recueil et les équipements qu'il aura à utiliser (pose du cardiofréquencemètre, du capteur de réponse électrodermale), et il est également averti des différents distracteurs qui viendront le perturber. Cependant, on ne lui précise pas ni quand ni comment ils vont survenir lors de l'exécution de sa tâche. Une fois le sujet équipé, le matériel lui est présenté en détail et les différents réglages sont effectués (ajustement du matériel *d'eye tracking*, synchronisation informatique avec les différents capteurs de fréquence cardiaque, de réponse électrodermale et avec les caméras vidéo). Ensuite, le sujet est invité à tester la tâche « MAT » et peut demander le cas échéant des précisions. Quand il est prêt, l'expérimentation débute.

III.6. PLAN EXPÉRIMENTAL

Le plan expérimental ne comporte que des données appariées. Les sujets passent les trois phases, chacune d'une durée de six minutes, de complexité croissante (facteur Phase noté P3 avec les trois modalités p1, p2 et p3) auxquelles sont associées les réponses aux quatre tests proposés : (t1) le test MAT et les trois distracteurs : (t2) sollicitations humaines, (t3) alarmes, (t4) demande de code d'accès *via* un talkie-walkie (facteur Test noté T4 avec les quatre modalités t1, t2, t3 et t4 susmentionnées).

L'ordre des phases est constant et l'ordre de présentation des tests t2, t3 et t4 à l'intérieur de chaque phase est aléatoire (la tâche t1 étant maintenue constante lors de la première minute de chaque phase).

Lors de la phase 1, chacun des distracteurs apparaît l'un après l'autre, ce qui laisse le temps au sujet de se familiariser avec la tâche.

Lors de la phase 2, un peu plus difficile à réaliser, les distracteurs sont présentés au sujet simultanément deux par deux (par exemple : sollicitations humaines avec alarmes qui se déclenchent en même temps ou demande de code d'accès au même moment). Aucune instruction n'est donnée au sujet quant à la gestion en simultané des différents distracteurs. Chacun est donc libre de gérer « sa stratégie » de réponse comme il l'entend et de prioriser ainsi une réponse par rapport à une autre selon sa personnalité. La priorité de réponse à un distracteur plutôt qu'à un autre est encore plus accentuée en phase 3, lors de laquelle les trois distracteurs apparaissent en simultané. Comme le sujet ne peut pas gérer les trois distracteurs en même temps, il se trouve alors obligé de prioriser ses réponses. Une des questions que l'on peut se poser à ce propos est de savoir si les sujets vont gérer les apparitions des distracteurs de la même façon et quelles sollicitations seront traitées en priorité.

IV. ANALYSE DES DONNÉES

Comme l'indique Park (2008), si les mesures psychophysiologiques peuvent être de bons indicateurs des émotions, elles doivent être analysées avec prudence, en tenant compte entre autres des nombreux artefacts dus à la sensibilité des appareils utilisés. Ceci nous a conduits d'une part à vérifier régulièrement la fiabilité d'enregistrement des appareils, et d'autre part à effectuer également des « lissages » de données concernant notamment la fréquence cardiaque (FC) et la réponse électrodermale (*Skin Conductance sc*). En ce qui concerne les analyses des expressions faciales, bien que « FaceLAB » permette de traduire mathématiquement un bon nombre d'expressions (positionnement de points en x, y, z dans un repère orthonormé) et de pondérer chacun des critères des différentes expressions faciales, certains points cruciaux pour la détermination de l'émotion ne sont pas pris en compte actuellement. Cette variable « expression faciale » est par conséquent encore en cours d'analyse et ne sera pas présentée dans les résultats ci-après.

Néanmoins, l'analyse des enregistrements vidéo a permis de caractériser les réponses comportementales des sujets lors de l'accomplissement de la tâche. Ces relevés ont été étudiés au niveau descriptif et donnent un aperçu des réactions des sujets. À titre d'exemple, on a pu comptabiliser certains ajustements posturaux au niveau des membres inférieurs (oscillations au niveau du genou) non liés à la réalisation des tâches, qui augmentaient avec la complexité de la tâche. Pour les quatre sujets testés, le nombre de ces ajustements est multiplié par trois entre la phase 1 et la phase 3. Cet indicateur fera donc l'objet d'investigations dans des expériences ultérieures.

IV.1. VARIABLES DÉPENDANTES RETENUES POUR LES ANALYSES

Les variables dépendantes (VD) quantitatives retenues pour les analyses sont les suivantes :

- sc: Skin Conductance, réponse électrodermale, mesurée en microsiemens (μS)
- FC : Fréquence cardiaque
- Omissions (MAT/Alarmes)
- Temps de réponse (MAT/Alarmes)
- Échelles analogiques pour les évaluations subjectives (Performance, Confiance, Difficulté, Effort, SAM Plaisir, SAM Éveil, SAM Dominance).
- Pourcentage de scrutation de l'écran affichant le test attentionnel MAT (tâche principale à effectuer).

Nous avons également considéré des VD qualitatives issues de l'analyse des verbalisations et des observations, qui ont permis d'illustrer la stratégie favorisée par le sujet (priorisation de la tâche MAT, des sollicitations humaines, de la gestion des alarmes ou des appels *via* le talkie-walkie).

IV.2. MÉTHODE D'ANALYSE

Afin de pouvoir déterminer des profils exploratoires de comportements à partir de l'ensemble des variables analysées et de pouvoir également valider ces premiers résultats malgré le faible nombre de participants, nous avons effectué une analyse par minute et non par sujet. Les minutes sont ainsi considérées du point de vue de l'analyse des données comme les individus statistiques du tableau de résultats.

Ce tableau est ainsi constitué de 72 lignes (4 sujets \times 3 phases \times 6 minutes) auxquelles on a adjoint 13 colonnes : les 13 variables dépendantes (VD) présentées au paragraphe précédent (cf. *supra*). Des facteurs structurants (ou variables qualitatives), au sens de l'analyse des données (Le Roux & Rouanet, 2004), tels qu'ils sont décrits dans les études méthodologiques ergonomiques (Wolff, 2003 ; Wolff, Burkhardt & De La Garza, 2005 ; Couture *et al.*, 2010) ou psychologiques (Maurer, Delfour, Wolff & Adrien, 2010) ont complété les colonnes de ce tableau de données. Il s'agit des facteurs *SUJET* (repère des réponses de chacun des quatre sujets), *PHASE* (repère de chaque phase concernée) et *STRATÉGIE* utilisée par le sujet (combinaison des observations et des verbalisations).

Des analyses univariées ont été effectuées à partir de chacun de ces facteurs avant de mettre en œuvre une analyse multivariée (une Analyse en Composantes Principales standard : ACP). Quand les données l'ont permis (homogénéité des variances, conditions de normalité respectées), une analyse inférentielle de type ANOVA a pu être effectuée (Wolff, 2003).

IV.3. ANALYSES UNIVARIÉES

À l'exception de la fréquence cardiaque, les paramètres retenus pour caractériser le comportement du sujet, tant du point de vue de la performance que de celui des réactions émotionnelles, se révèlent sensibles à l'effet de la complexité de la tâche. Il en est de même pour les évaluations subjectives (difficulté de la réalisation de la tâche, autoévaluation de la performance, émotions, etc.). Le fait que la fréquence cardiaque ne varie pas sensiblement lors de la situation expérimentale n'est en fait pas tant surprenant car s'il est admis, suite notamment aux nombreux travaux menés dans le domaine aéronautique, qu'elle peut varier en fonction de l'effort mental du sujet et de son niveau de stress (Roscoe, 1978, 1993), il a été déjà démontré que ces variations sont observées essentiellement pour des niveaux très élevés de charge mentale (Cabon & Mollard, 2003 ; Hoogeboom & Mulder, 2004). En effet, lors d'un effort mental non surchargeant, le rythme cardiaque peut devenir plus régulier, ce qui se traduit généralement par un intervalle interbattement plus faible que lors du repos.

Le rythme cardiaque peut aussi dépendre du niveau d'expérience du sujet (Billings, Gerke & Chase, 1973) ou de l'association du risque et de la responsabilité (Roman, 1975). L'utilisation d'un autre indice aurait certainement été plus adaptée à notre expérience : le calcul de la *variabilité cardiaque*. Cette mesure se révèle particulièrement plus sensible à l'effort

mental que la fréquence cardiaque dans des situations relativement bien contrôlées du point de vue des communications verbales et de l'activité physique de l'opérateur (Cabon, Farbos & Mollard, 2000 ; Cabon & Mollard, 2003).

De plus, dans notre expérience, les personnes n'ont vraisemblablement jamais été en situation de charge mentale très élevée, car elles ont priorisé certaines tâches plutôt que réaliser l'ensemble des tâches demandées.

D'autres difficultés d'analyses ont été aussi constatées : les différences globales de comportements/performances selon le type de distracteur utilisé (sollicitation humaine, alarme ou appel *via* un talkie) n'ont pas été simples à évaluer en analyses statistiques classiques, tant descriptives qu'inférentielles, compte tenu des couplages effectués de manière aléatoire lors des phases 2 et 3.

En revanche, lors de l'analyse multivariée, il a été possible *via* l'analyse conjointe des performances et des verbalisations de mettre à jour des stratégies individuelles de réponses selon le type de distracteur que le sujet devait gérer. Il a été possible également d'extraire quelques situations et d'effectuer des analyses individuelles de comportements (voir paragraphe IV.4.2).

Au niveau des analyses globales, à titre d'illustration, seront présentés quelques résultats concernant notamment le pourcentage de scrutation de l'écran pour la tâche MAT, la réponse électrodermale (SC), et certaines échelles analogiques relatives aux évaluations subjectives.

IV.3.1. Pourcentage de scrutation de l'écran (tâche MAT)

Cette variable peut être considérée comme un indicateur de la concentration du sujet sur l'exécution de la tâche principale. Lors de la phase 1, le pourcentage moyen de scrutation de l'écran s'élève à 74,3 % (écart-type : 2,0) et diminue progressivement en phase 2 (moy = 66,3 % ; écart-type : 5,7) pour chuter totalement lors de la phase 3 avec un score égal à 28,4 % (écart-type : 5,1).

On remarque que les sujets ont une baisse de concentration qui diminue de plus de la moitié entre la phase 1 et la phase 3, pour l'ensemble des relevés sur les six minutes enregistrées. Les écarts-types sont plus élevés pour les phases 2 et 3, car les sujets n'ont pas eu les mêmes réactions : les deux sujets masculins ont tenté de poursuivre la tâche malgré les distracteurs, alors que les deux sujets féminins ont quasiment abandonné la tâche, privilégiant les réponses aux sollicitations extérieures.

Lors des expériences ultérieures, il serait donc intéressant de poursuivre cette étude en approfondissant l'effet du genre sur les performances en situation de multisollicitation.

IV.3.2. Réponse électrodermale (variable SC – Skin Conductance)

La réponse électrodermale (mesure de la conductance) augmente significativement entre la phase 1 et la phase 3 (résultats de l'ANOVA : $F[2,69] = 7,65$; $p = 0,001$), et indique que les sujets réagissent à la complexité du test. En

conformité avec les études antérieures, la sc ici peut être considérée comme un bon indicateur de la réaction émotionnelle (voir Figure 3 avec le tracé des moyennes et l'illustration des erreurs types pour chaque phase concernée).

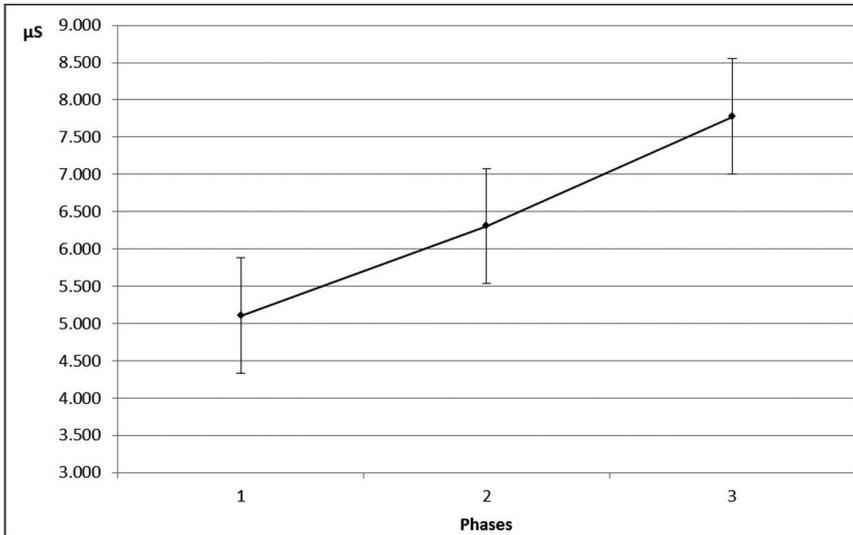


Figure 3 : Évolution de la sc par phase
Figure 3: The sc evolution phase by phase

IV.3.3. Étude des omissions et des temps de réponse (TR) pour les tâches informatisées

Des analyses ont été effectuées pour les omissions de réponse concernant le test de surveillance des échelles avec curseurs MAT ou la tâche de sollicitation par les alarmes (saisir un mot de passe pour stopper les alarmes). Les résultats sont similaires dans les deux cas.

L'augmentation du nombre d'omissions et des temps de réponse, que ce soit pour les échelles du test MAT ou la gestion des alarmes, montre qu'en phase 1 et 2 le sujet parvient à maintenir son attention sur « MAT » (augmentation modérée des omissions et des TR), mais qu'il n'y parvient plus en phase 3. Quelle que soit l'analyse, les résultats sont très significatifs (tests pour MAT : $F[2, 69] = 18,20$; $p = 0,0000$, $F[2, 69] = 18,505$; $p = 0,000$, tests pour les alarmes : $F[2, 69] = 7,731$; $p = 0,001$, $F[2, 69] = 12,51$; $p = 0,000$). Ces résultats sont conformes aux constatations effectuées sur la variable « pourcentage de temps de scrutation de l'écran ».

IV.3.4. Étude des évaluations subjectives de la performance, de la confiance, de la difficulté et de l'effort fourni

Quelle que soit l'échelle analogique d'évaluation subjective proposée, les différences entre les phases sont très significatives (performance : $F[2, 69] = 56,946$; $p = 0,000$, confiance : $F[2, 69] = 9,945$; $p = 0,000$, difficulté : $F[2, 69] = 96,900$; $p = 0,000$, effort : $F[2, 69] = 26,664$; $p = 0,000$).

La performance décroît avec la difficulté d'exécution de la tâche alors que la perception de l'effort augmente. Conformément aux résultats antérieurs, la plate-forme paraît très bien adaptée pour mesurer les performances des sujets en situation de surcharge de travail élevée.

IV.3.5. Étude des évaluations subjectives SAM plaisir, SAM éveil et SAM dominance

Hormis l'échelle SAM plaisir, pour laquelle les résultats sont mitigés (l'expérience plaît beaucoup aux sujets, mais ils n'apprécient pas trop la surcharge de travail imposée et qui augmente de phase en phase !), les résultats obtenus aux deux autres échelles sont très significatifs (éveil : $F [2, 69] = 38,132$; $p = 0,000$, dominance : $F [2, 69] = 9,163$; $p = 0,000$). L'autoévaluation de l'éveil est de plus en plus élevée, probablement en relation avec l'intensité des sollicitations qui augmente de phase en phase et qui requiert de plus en plus d'attention de la part du sujet, lequel a la sensation de ne plus pouvoir dominer ses émotions.

IV.4. ANALYSE EN COMPOSANTES PRINCIPALES STANDARD (ACP)

L'objectif d'une ACP, appelée analyse des corrélations ou encore analyse géométrique des données, est de rechercher à partir de données multidimensionnelles des proximités et des différences entre groupements d'individus afin d'en faire ressortir des profils de comportements. Le principe de l'ACP est de représenter les données sous forme de nuages de points dans des espaces géométriques et de fonder l'interprétation sur les proximités et les oppositions entre les points à partir d'axes factoriels (appelés également variables factorielles) qui résument ces informations (pour un exposé théorique, voir Le Roux & Rouanet, 2004 et pour des exemples d'application : Wolff, 2003 ; Wolff *et al.*, 2005 ; Maurer *et al.*, 2010).

Pour les interprétations, deux types de variables seront analysés : les variables dites « actives » (dans cette étude les dix premières variables numériques exposées au § IV.1) qui vont participer à la construction de chacun des axes et les variables dites « supplémentaires » (les trois variables « SAM ») qui ne participeront pas à la construction des axes mais qui donneront un complément d'informations. Les interprétations se font également à partir de la matrice des corrélations entre toutes les variables. Ci-après les résultats pour le « nuage des variables ». Sera ensuite présenté le « nuage des individus-stratégies » dont la construction particulière est expliquée au paragraphe IV.4.2.

IV.4.1. Analyse et interprétation du « nuage des variables »

Nous retiendrons pour l'analyse des résultats trois axes factoriels qui représentent plus de 86 % de la variance totale (pour un exposé détaillé de la méthode et des analyses des calculs, voir Le Roux & Rouanet, 2004 et Wolff, 2003).

Chacun de ces axes, représentant une dimension spécifique de l'expérimentation, sera interprété ci-après avec l'analyse des nuages des variables.

Le plan factoriel des axes 1 et 2 est illustré Figure 4. Pour les interprétations décrites ci-après, les contributions relatives à la variance de chaque variable aux différents axes retenus ont été également prises en compte.

L'axe 1 (horizontal, Figure 4) prend en compte 51 % de la variance totale. Il représente essentiellement les performances du sujet. À droite sur le graphique, on trouve les variables relatives aux temps, aux omissions et aux évaluations de la difficulté, de l'effort, et de l'éveil qui sont toutes en assez forte corrélation positive (corrélations significatives avec des coefficients variant de $r = 0,87$ à $r = 0,51$ avec l'échelle SAM éveil). Elles sont opposées aux variables d'évaluation de la performance et de la confiance (à gauche sur le graphique), qui sont corrélées négativement et significativement avec le groupement précédent (de $r = -0,88$ à $r = -0,50$ avec les omissions) : plus les temps de réponse et le nombre d'omissions sont élevés, plus les autoévaluations de la difficulté, de l'effort et de l'éveil tendent à augmenter, alors que celles concernant la performance et la confiance en soi diminuent.

L'axe 2 (vertical, Figure 4), avec 21,5 % de la variance prise en compte, représente les variables physiologiques, celles des évaluations de la difficulté et de l'effort (en haut du graphique). La SC et la FC sont corrélées positivement et significativement entre elles ($r = 0,51$) ainsi qu'avec les évaluations concernant la difficulté et l'effort. Elles sont opposées à SAM dominance (en bas sur le graphique) avec laquelle elles sont en corrélation négative. Ainsi, lorsque les indices physiologiques augmentent avec la charge de travail, le sentiment de dominer la situation devient moins important.

L'axe 3 (non représenté sur la Figure 4) ne prend en compte que 13,5 % de la variance totale du nuage, l'essentiel (72,5 %) ayant déjà été expliqué par les deux axes précédents. Il ne sera donc pas interprété à l'aide des corrélations entre les variables, mais selon la corrélation des variables avec cet axe ou selon la contribution élevée de ces variables, ces indicateurs pouvant donner des informations supplémentaires pertinentes en complément des interprétations déjà fournies.

Cet axe peut ainsi être interprété comme un axe du ressenti du sujet (les variables SC, SAM plaisir, les autoévaluations de la performance indiquant une confiance dans le jugement [CONF] et les omissions au test MAT [OMEC] y sont bien représentées : corrélations avec l'axe non négligeables et contributions à la variance assez élevées). Lorsque la SC augmente, les sujets tendent à effectuer plus d'omissions (les deux variables SC et Omissions – OMEC – pour le test MAT sont toutes deux en corrélation négative avec cet axe ; donc à interpréter ensemble). Toutefois, les sujets ne semblent pas évaluer cette baisse de performance correctement, car la variable d'évaluation « Confiance dans le jugement » est en corrélation positive avec cet axe et se situe ainsi en opposition avec SC et OMEC.

On peut aussi en conclure qu'ils prennent certainement plaisir à effectuer le test même si celui-ci est complexe car la variable SAM plaisir est également en assez forte corrélation négative avec cet axe tout comme les variables SC et OMEC (et donc en opposition avec la variable CONF). Ce qui a été confirmé par la suite lors du débriefing effectué.

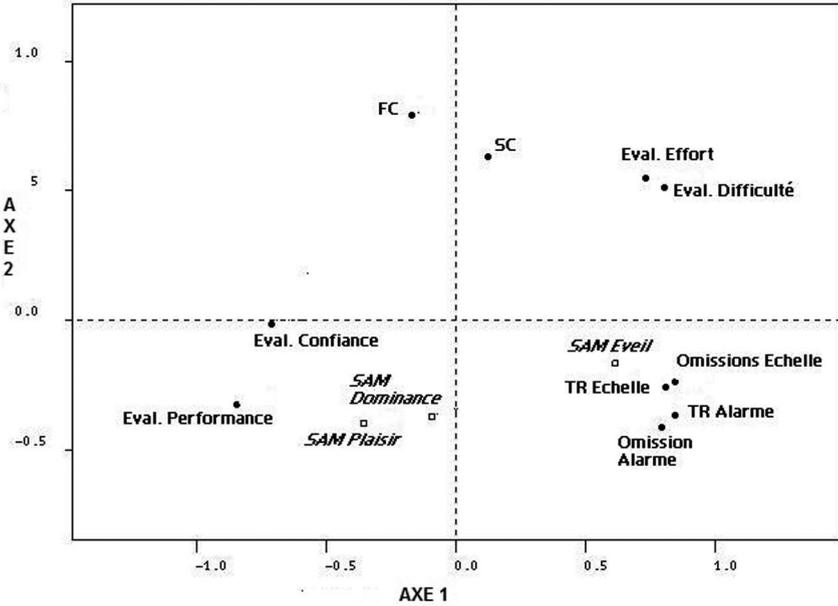


Figure 4 : Nuage des variables actives (ronds pleins) et des variables supplémentaires (carré vide) ; plan 1-2

Figure 4: Cloud of active variables (full circle) and supplementary variables (empty squares); plane 1-2

IV.4.2. Analyse et interprétation du « nuage des stratégies »

À partir des facteurs structurants constitués (Sujet, Phase et Stratégie, voir § IV.2 Méthode d'analyse), nous avons constitué des nuages dérivés de points moyens élaborés à partir des résultats individuels. Celui concernant les stratégies des sujets est présenté ci-après (Figure 5). Ce facteur Stratégie a été élaboré à partir de la combinaison verbalisations recueillies lors des débriefings/performances des sujets et la concordance a été vérifiée. Chaque individu statistique-ligne a été ainsi indexé en fonction de la stratégie utilisée (priorisation de la tâche MAT ou des réponses aux alarmes ou des demandes de codes *via* le talkie-walkie). Les interprétations se font à partir de celles élaborées pour le nuage des variables et également à la lumière des résultats individuels, des observations (*via* la vision des vidéos enregistrées) et des débriefings. Les quatre sujets concernés ont ainsi établi des stratégies différentes.

La stratégie qui consiste à privilégier les réponses au test MAT est représentée sur l'axe 1 et directement liée à la recherche de la performance. Les autoévaluations vont dans ce sens également. Les autres sollicitations deviennent alors secondaires pour eux. Dans ce cas, le sujet (un homme) ne pense qu'à sa propre réussite, ne veut pas être interrompu et se concentre

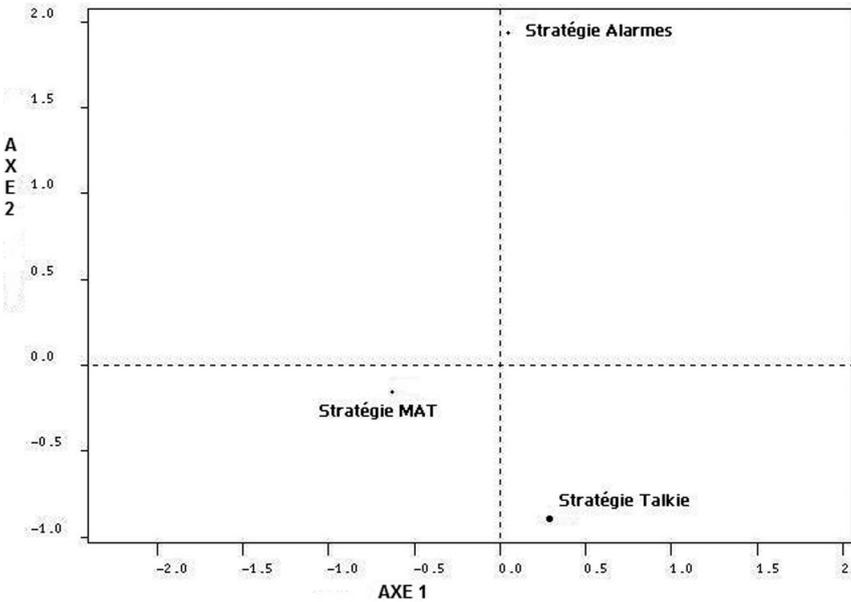


Figure 5 : Nuage dérivé des points moyens selon les stratégies de réponses du sujet ; plan 1-2

Figure 5: Derived cloud of the subject responses strategies' mean points; plane 1-2

énormément sur sa tâche. L'analyse des résultats individuels indique peu de variations de la SC lors de la survenue des distracteurs à partir de la minute 2 de chaque phase, sauf en phase 3 où le sujet commence à être un peu plus tendu (légère augmentation de la FC et de la SC). Lors du débriefing, le sujet a lui-même confirmé cette concentration extrême, qui est d'ailleurs également visible sur la vidéo enregistrée.

La stratégie consistant à privilégier l'arrêt des alarmes au détriment de la tâche MAT a une forte contribution sur l'axe 2, qui est représenté essentiellement par les variables physiologiques. Cela peut s'expliquer par le fait que l'alarme sonore (son strident) et visuelle a été représentative d'une véritable situation d'urgence, et le sujet concerné (un homme) a donc délaissé la tâche MAT pour pouvoir stopper au plus vite ces signaux désagréables provoquant chez lui une réaction physiologique importante, indiquée très précisément par l'analyse des résultats individuels (augmentation de la SC lors des différentes apparitions de l'alarme). Le sujet était tellement perturbé par l'alarme qu'il a négligé non seulement la tâche MAT mais aussi les autres sollicitations. Ceci est confirmé par l'analyse des vidéos et également par le sujet lui-même lors du débriefing.

À titre illustratif, on trouvera dans la Figure 6 ci-après, les réactions physiologiques de ce sujet (Sujet 4) lors de la deuxième minute de la phase 2 (impliquant l'apparition conjointe des deux stimuli : appel au talkie et déclenchement de l'alarme).

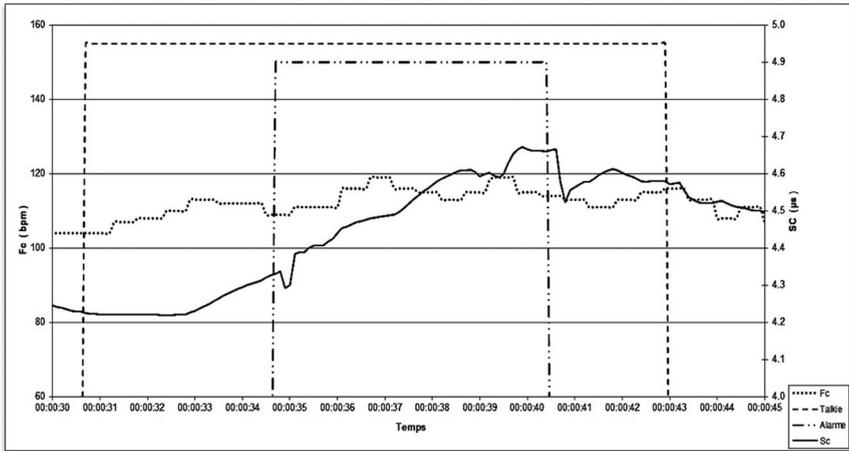


Figure 6 : Illustration des réactions physiologiques du sujet 4 lors la minute 2 de la phase 2 (apparition simultanée de 2 stimuli)

Figure 6: Illustration of the physiological responses of subject 4 at the minute 2 of phase 2 (simultaneous occurrence of two stimuli)

Sur ce graphique, on constate d'une part que l'appel au talkie (dont la durée est symbolisée par le rectangle en traitillés, démarrant juste avant le temps 00 :31 et se terminant au temps 00 :43) et d'autre part le déclenchement de l'alarme au temps 00 :35. La FC exprimée en battements par minute (bpm) est indiquée sur l'axe des ordonnées de gauche (en pointillés sur le graphique) et la SC exprimée en microsiemens (μ s) est positionnée sur l'axe des ordonnées de droite (en trait plein sur le graphique).

On peut constater, conformément à ce qui a été énoncé précédemment, que la FC n'est pas un bon indicateur ici : elle n'indique aucune variation notable, ni à l'apparition du talkie ni à celle de l'alarme sonore et visuelle. *A contrario*, la SC augmente de 50 % lors du déclenchement de l'alarme.

Au contraire, les deux sujets féminins qui ont choisi de donner la priorité aux réponses demandées pour les codes d'accès *via* le talkie-walkie ont des réactions physiologiques moindres, mais délaissent également les autres sollicitations, pensant répondre à ce qui est – selon elles – une priorité. Les analyses individuelles indiquent des FC et SC moins importantes que pour les deux sujets masculins, mais aussi des performances au test MAT de moins en moins élevées au fur et à mesure que la tâche se complexifie. Par ailleurs, les sollicitations humaines ne les intéressent plus, pas plus que l'alarme qui ne semble pas les gêner outre mesure. L'analyse des vidéos montre bien d'ailleurs cette focalisation sur les appels liés au talkie. Lors du debriefing, les sujets ont indiqué que pour elles, la réponse à donner au talkie était primordiale, car la situation revêtait alors un caractère des plus critiques. S'agit-il d'une influence des séries télévisées ? L'une d'entre elles a en effet indiqué que la séquence talkie ressemblait à une séquence d'une célèbre série télévisée américaine !

Ces résultats ne sont qu'une exploration des différents comportements possibles des personnes mises en situation de multisollicitation. Il serait intéressant d'approfondir ces premiers résultats avec un nombre plus important de sujets et de pouvoir vérifier si ces stratégies sont plutôt liées à la réaction émotionnelle, au genre ou à la situation. Par ailleurs, la gestion des priorités joue ainsi un rôle crucial dans les situations où l'attention doit être divisée ; ceci n'est pas sans rappeler les études classiques menées sur l'attention sélective ou partagée que bon nombre d'auteurs ont pu mettre en évidence depuis les travaux de Cherry (1953) sur l'effet *cocktail party* ou de Neisser et Becklin (1975).

De même, chacun d'entre nous sait que l'efficacité de nos réactions à un événement dépend également de nos ressources physiologiques. Ces pistes seraient certainement à approfondir lors d'expériences ultérieures.

V. CONCLUSION

Le test-utilisateur qui vient d'être présenté se situe, rappelons-le, dans un champ exploratoire compte tenu du faible nombre de sujets testés. Toutefois, l'analyse des données mise en œuvre permet, par une approche méthodologique originale (inversion de matrice) de pouvoir utiliser des tests inférentiels et des analyses multidimensionnelles, avec de faibles échantillons. Ce choix méthodologique donne ainsi la possibilité de valider les résultats de cette première expérience et de dégager des pistes de réflexion très encourageantes qu'il conviendra d'explorer au cours des prochaines évaluations.

L'objectif initial de cette première approche est atteint. Nous avons pu, au cours des différentes expérimentations (prétest et test proprement dit), valider les modalités de recueil synchronisé de multiples données, vérifier que les procédures de rejet des artefacts fonctionnaient et surtout, tester la sensibilité des indicateurs retenus pour caractériser les comportements des sujets confrontés à des multiples sollicitations. La réponse émotionnelle est évaluée sous différents angles avec un gradient de difficulté contrôlé ; elle peut être interprétée en fonction du contexte et en relation avec la qualité de la performance.

Au travers de la PEPSS, nous espérons offrir, grâce aux capacités de reconfiguration du site, à la versatilité des procédures de test et aux méthodes maîtrisées d'analyses des données, un outil qui permette à la fois de tester les usages et les réactions émotionnelles face à des concepts et/ou des dispositifs innovants. L'expression émotionnelle d'un individu est influencée par sa personnalité bien sûr, mais également par sa position (assis ou debout), les contraintes qui s'exercent sur lui (protections par exemple) ou l'agencement de son environnement de travail. L'expression émotionnelle par le mouvement est particulièrement sensible à ces divers facteurs et il est donc pertinent d'observer et de relever ce ressenti (bien-être, mal-être, émotions) en vue d'améliorer l'environnement de travail.

Les futurs travaux vont porter notamment sur l'évaluation et les tests d'usages de systèmes innovants basés sur une interaction tangible qui sont encore très peu formalisés. Nous comptons explorer cette dimension usage de l'Interaction Tangible, en nous appuyant sur la plateforme PEPSS présentée dans cet article, en mettant en place des bancs d'expérimentations pour des tests d'usage et des évaluations de nos systèmes centrés sur l'Interaction Tangible pour déterminer des protocoles généralisables. Enfin, dans l'avenir, nous prévoyons également de nous concentrer sur la caractérisation du plaisir lié à l'utilisation de surfaces interactives dans des domaines variés tels que la domotique, l'aéronautique et les services.

REMERCIEMENTS

Les auteurs remercient vivement tous les contributeurs de la PEPSS et tout particulièrement E. Delord, L. Nigay et L. Todeschini pour leur précieuse aide, ainsi que tous les étudiants et personnels de l'ESTIA qui ont accepté de participer aux prétests ainsi qu'aux évaluations.

BIBLIOGRAPHIE

- Barcenilla J. & Bastien J.-M. C. (2009), « L'acceptabilité des nouvelles technologies : quelles relations avec l'ergonomie, l'utilisabilité et l'expérience utilisateur ? », *Le Travail humain*, 72 (4), 311-331.
- Becker A., Aucouturier J.-J., Mougenot C. & Yamanaka T. A. (2010), « Situated Experimental Protocol to Study Emotional Responses to an Interactive Object », *Paper Presented at the Third International Workshop on Kansei*, Fukuoka, Japan, August. En ligne <http://www.jj-aucouturier.info/papers/KANSEI-2010-1.pdf>
- Billings C. E., Gerke R. J. & Chase R. C. (1973), « Stress and Strain in Student Helicopter Pilots », *Aerospace Medicine*, 44, 1031-1035.
- Buisine S., Abrilian S., Niewiadomski R., Martin J.-C., Devillers L. & Pelachaud C. (2006), « Perception of Blended Emotions: from Video Corpus to Expressive Agent », in J. Gratch *et al.* (eds.), *IVA'06 International Conference on Intelligent Virtual Agents*, *LNAI 4133* (p. 93-106), Berlin Heidelberg, Springer-Verlag. En ligne <http://stephanie.buisine.free.fr/publis/IVA06.pdf>
- Cabon P., Farbos B. & Mollard R. (2000), « Gaze Analysis and Psychophysiological Parameters: a Tool for the Design and the Evaluation of Man-Machine Interfaces », *EEC Report n° 350*, Bretigny, France, Eurocontrol Experimental Center.
- Cabon P. & Mollard R. (2003), « Prise en compte des aspects physiologiques dans la conception et l'évaluation des IHM », in G. Boy (éd.), *Ingénierie cognitive*, Paris, Hermès Sciences Publications, p. 99-138.
- Cahour B. (2006), « Les affects en situation d'interaction coopérative : proposition », *Le Travail humain*, 69 (4), 379-400.
- Cherry C. (1953), « Some Experiments on the Recognition of Speech with One and Two Ears », *Journal of the Acoustical Society of America*, 25, 975-979.
- Clay A., Couture N. & Nigay L. (2009 a), « Engineering Affective Computing: a Unifying Software Architecture », in *Proceedings of the 3rd IEEE International Conference on Affective Computing and Intelligent Interaction and Workshops, 2009 (ACII'09)*, Amsterdam, The Netherlands: IEEE, p. 1-6. En ligne <http://iihm.imag.fr/publis/2009/ACII-2009ClayCoutureNigay.pdf>

- Clay A., Couture N. & Nigay L. (2009 b), « Towards Emotion Recognition in Interactive Systems: Application to a Ballet Dance Show », in *WinVR'09, Proceeding of the World Conference on Innovative Virtual Reality*, Châlon-sur-Saône, France, ASPE, p. 19-24. En ligne <http://iihm.imag.fr/publis/2009/WINVR2009.pdf>
- Clay A., Couture N. & Nigay L. (2010), « eMotion : un outil pour personnaliser la reconnaissance d'émotions », in *Proceedings of the Ergo'IA 2010 Conference*, New York: ACM, p. 59-66. En ligne <http://iihm.imag.fr/publis/2010/ERGOIA-2010-ClayCoutureNigay.pdf>
- Clay A., Delor E., Couture N. & Domenger G. (2010), « Augmenting a Ballet Dance Show Using the Dancer's Emotion: Conducting Joint Research in Dance and Computer Science », *Arts and Technology of Lecture Notes of the Institute of Computer Sciences, Social Informatic and Telecommunications Engineering*, 30, 148-156, Berlin Heidelberg, Springer-Verlag. En ligne http://www.estia.fr/~ncouture/NRC_Publis/artsit09.pdf
- Comstock J. L. & Arnegard R. J. (1992), « The Multiattribute Task Battery for Human Operator Workload and Strategic Behavior Research », *Technical Report 104174*, Hampton, VA, NASA Langley Research Center.
- Couture N., Wolff M., Mollard R. & Todeschini L. (2010), « Étude du comportement et des émotions dans un contexte de multisollicitations : approche exploratoire avec la Plate-forme d'évaluation, de prototypage et de tests d'usages (PEPSS) », in *Proceedings of the Ergo'IA 2010 Conference*. New York : ACM, p. 67-74. doi : 10.1145/1868650.1868661
- Djajadiningrat J. P., Overbeeke C. J. & Wensveen S. A. G. (2000), « Augmenting Fun and Beauty: a Pamphlet », in W.E. Mackay (ed.), *Proceedings of DARE 2000 Designing Augmented Reality Environment*, New York, ACM, p. 131-134. doi :10.1145/354666.354680
- Ekman P. (1992), « Are There Basic Emotions? », *Psychological Review*, 99 (3), 550-553. En ligne <http://www.paulekman.com/wp-content/uploads/2009/02/are-there-basic-emotions1.pdf>
- Gil S. (2009), « Comment étudier les émotions en laboratoire ? », *Revue électronique de psychologie sociale*, 4, 15-24. En ligne <http://reps.psychologie-sociale.org/>
- Gombert F. (2009), *Acquisition et optimisation du logiciel Neuroscan d'électroencéphalographie dans un protocole expérimental précis* (Rapport de stage), Compiègne, France, Université technologique de Compiègne.
- Hoogeboom P. J. & Mulder L. J. M. (2004), « Physiological Indices for the Estimation of Momentary Changes in Cognitive Workload and Mental State », in D. de Ward, K. A. Brookhuis & C. M. Weikert (eds.), *Human Factors in Design*, Maastricht, the Netherlands, Shaker Publishing, p. 147-159. En ligne http://www.hfes-europe.org/books/firstpage_2003/14.pdf
- Jaimes A. & Sebe N. (2007), « Multimodal Human-Computer Interaction: a Survey », *Computer Vision and Image Understanding*, 108 (1-2), 116-134.
- Jeantet A. (2003), « L'émotion prescrite au travail », *Travailler* 1 (9), 99-112. doi : 10.3917/trav.009.0099.
- Jordan P. W. (2004), *Designing Pleasurable Products*, NY, USA, CRC press, Taylor & Francis Group.
- Le Roux B. & Rouanet H. (2004), *Geometric Data Analysis*, Dordrecht, Kluwer Academic Publishers.
- Lottridge D. & Moore G. (2009), « Designing for Human Emotion: Ways of Knowing », *New Review of Hypermedia and Multimedia*, 15 (2), 147-172. doi : 10.1080/13614560903165136.
- Mauss I. B. & Robinson M. D. (2009), « Measures of Emotion: a Review », *Cognition & Emotion*, 23 (2), 209-237. doi: 10.1080/02699930802204677.
- McDougall S., Reppa I., Smith G. & Playfoot D. (2009), « Beyond Emoticons: Combining Affect and Cognition in Icon Design », in *Proceeding of Engineering*

- Psychology and Cognitive Ergonomics, HCI 2009*, Berlin, Germany, Springer-Verlag, p. 71-80. doi: 10.1007/978-3-642-02728-4_8.
- Mollard R., Gibert A. & Wolff M. (2010), « Transformation du travail et prévention des risques : aménagement des espaces d'activités et conception des postes de travail dans les bureaux de poste », in *Proceedings of the Ergo'IA 2010 Conference*, New York, ACM, p. 193-199. doi : 10.1145/1868650.1868682.
- Maurer M., Delfour F., Wolff M. & Adrien J.-L. (2010), « Dogs, Cats and Horses: Their Different Representations in the Minds of Typical and Clinical Populations of Children », *Anthrozoos: a Multidisciplinary Journal of the Interactions of People & Animals*, 23, 4, 365-382.
- Neisser U & Becklin R. (1975), « Selective Looking: Attending to Visually Specified Events », *Cog. Psychol.*, 7, 480-494. En ligne [http://dx.doi.org/10.1016/0010-0285\(75\)90019-5](http://dx.doi.org/10.1016/0010-0285(75)90019-5)
- Norman D. (2004), *Emotional Design: why we Love (or Hate) Everyday Things*, NY, Basic Book.
- Overbeeke K., Djajadiningrat T., Hummels C. & Wensveen S. (2000), « Beauty in Usability: Forget about Ease of Use! », *Pleasure with Products, Beyond Usability*, NY, USA, CRC Press, Taylor & Francis Group.
- Park B. (2008), « Psychophysiology as a Tool for HCI Research: Promises and Pitfalls », in *Proceedings of the 13th International Conference on Human-Computer*, Berlin-Heidelberg, Germany, Springer-Verlag, p. 141-148. doi : 10.1007/978-3-642-02574-7_16
- Pantic M., Sebe N., Cohn J. F. & Huang T. (2005), « Affective Multimodal Human-Computer Interaction », in *MULTIMEDIA '05: Proceedings of the 13th Annual ACM International Conference on Multimedia*, NY, USA, ACM, p. 669-676. doi : 10.1145/1101149.1101299.
- Petersen M. G., Iversen O., Krogh P. & Ludvigsen M. (2004), « Aesthetic Interaction – A Pragmatic Aesthetics of Interactive Systems », in *Proceedings of ACM DIS'04*, Cambridge, MA, USA, ACM, p. 269-276. En ligne <http://www.cse.chalmers.se/research/group/idc/ituniv/kurser/06/idproj/p269-petersen.pdf>
- Roman J. (1975), « Risk and Responsibility as Factors Affecting Heart Rate in Test Pilots. The Flight Research Program-2 », *Aerospace Medicine*, 36, 518-523.
- Roscoe A. H. (1978), « Stress and Workload in Pilots », *Aviation Space and Environment Medicine*, 49, 630-636.
- Roscoe A. H. (1993), « Heart Rate as Physiological Measure for in-Flight Workload Assessment », *Ergonomics*, 36, 9, 1055-1062.
- Scherer K. R. (2004), « Which Emotions Can Be Induced by Music? What Are the Underlying Mechanisms? And how Can we Measure them? », *Journal of New Music Research*, 33 (3), 239-251. En ligne http://www.affective-sciences.org/system/files/2004_Scherer_JNMR.pdf
- Tractinsky N., Katz A. S. & Ikar D. (2000), « What Is Beautiful Is Usable », *Interacting with Computers*, 13, 127-145. doi : 10.1016/S0953-5438(00)00031-X
- Wolff M. (2003), « Apports de l'analyse géométrique des données pour l'analyse de l'activité », in J.-C Sperandio & M. Wolff (éds.), *Formalismes de modélisation pour l'analyse du travail et l'ergonomie*, Paris, PUF, p. 195-227.
- Wolff M., Burkhardt J. M. & De la Garza C. (2005), « Analyse exploratoire de "points de vue" : une contribution pour outiller les processus de conception », *Le Travail humain*, 68 (3), 253-284.
- Zeng Z., Pantic M., Roisman G. I & Huang T. S. (2009), « A Survey of Affect Recognition Methods: Audio, Visual, and Spontaneous Expressions », *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(1), 39-58. En ligne <http://doc.utwente.nl/64555/1/ICMI07-ZengEtAl-FINAL.pdf>

RÉSUMÉ

Cet article présente une plate-forme qui permet d'évaluer l'expérience utilisateur face à des concepts innovants dans des systèmes sociotechniques évolutifs. Développer un système nécessite d'effectuer des évaluations qui soient formatives (tout au long de la phase de conception) et sommatives (en fin de conception). Concevoir ou reconcevoir des systèmes, c'est créer de nouvelles interactions humain-humain et humain-système. Au travers d'innovations technologiques c'est aussi créer de nouveaux usages. Prévoir ces usages et évaluer ces interactions nécessite de mettre en place des expérimentations et des essais faisant appel à des méthodes et des outils permettant d'enregistrer, d'analyser et d'interpréter les activités mais aussi les comportements des personnes en interaction. Au-delà de l'usage (utilité, utilisabilité, sens), l'émotion ressentie au cours de l'action peut être un facteur déterminant de la conduite et de la performance d'un système. La plate-forme PEPSS réunit à la fois des prototypes, des méthodes et des outils associés. Dans cet article, nous présentons une méthode permettant d'évaluer les processus émotionnels d'opérateurs confrontés à des multisollicitations pendant la réalisation d'une tâche de surveillance. Le protocole utilisé se base sur un test attentionnel multisollicitation (dérivé de la MultiAttribute Task ou MAT), où les sujets se voient détournés de leur tâche principale par des agents distracteurs/stresseurs. La plate-forme PEPSS et la méthode statistique associée que nous décrivons dans cet article soulignent l'importance de la synchronisation d'un maximum de moyens en temps réel, tout en effectuant des choix afin de limiter l'intrusivité des outils mis en œuvre.

Mots-clés : *Test d'usage, Multisollicitation, Évaluation, Comportement interactif, Analyse multivariée.*

Manuscrit reçu : janvier 2011

Accepté par F. Darses après révision : mai 2011