

AN HSI EXAMPLE : MOHICAN (MONITORING DE LA PERFORMANCE HUMAINS-MACHINES PAR L'ANALYSE DE LA CONFIANCE ET DE LA COOPÉRATION)

Human Systems Integration Course
Chloé MOREL

Présentation du projet

Projet MMT Man Machine Teaming

Qu'est-ce que le projet MMT ?

- **Le projet « Man Machine Teaming » (MMT) a pour but de préparer les technologies d'intelligence artificielle nécessaires à l'aviation de combat du futur.** Il a été lancé officiellement le 16 mars 2018 en présence de Florence Parly, ministre des Armées et de Joël Barre, délégué général pour l'armement.
- MMT est une initiative lancée et financée par la **Direction générale de l'armement (DGA)**.
- Confié à **Dassault Aviation et Thales** qui disposent déjà d'une expertise reconnue en matière d'interfaces homme/machine et de systèmes aéronautiques de combat.
- Un quart des études ont été confiées aux laboratoires, ETI-PME innovantes et startups spécialisés dans **l'intelligence artificielle, la robotique et les nouvelles interfaces homme/machine.** L'objectif est de faire émerger un écosystème industriel permettant aux innovations d'être détectées, évaluées et, à terme, mûries et intégrées dans **les développements des futurs avions de combat** dès la rénovation à mi-vie du Rafale à l'horizon 2030.

Objectifs du projet MOHICAN

→ **Proposer et tester une méthode d'évaluation de la performance du teaming pilote-assistant virtuel au sein du cockpit d'un avion de combat simulé.**

- Modèles de confiance et de collaboration
- Proposition de métriques de confiance et de collaboration
 - Tenir compte du contexte et de l'environnement du pilote pour l'analyse de la confiance et de la collaboration
 - Indices basés sur l'expérience opérationnelle
 - Construction de métriques basée sur des prototypes virtuels tangibles
 - Expérimentations et développements de systèmes (AV)

Le partenaire : Synapse Défense



Expertise des opérations aériennes militaires



ENTRAINEMENT COLLABORATIF

Nous construisons, planifions, animons et débriefons l'exécution des exercices internationaux ou des entraînements quotidiens.



ANALYSE DE LA PERFORMANCE DES FORCES

Nous collectons, traitons et analysons les données de l'entraînement pour évaluer objectivement la performance des forces aériennes



MODÉLISATION DES OPÉRATIONS AÉRIENNES

Nous créons et animons un environnement "universel" de simulation des opérations aériennes, creuset pour l'apprentissage des IA tactiques.

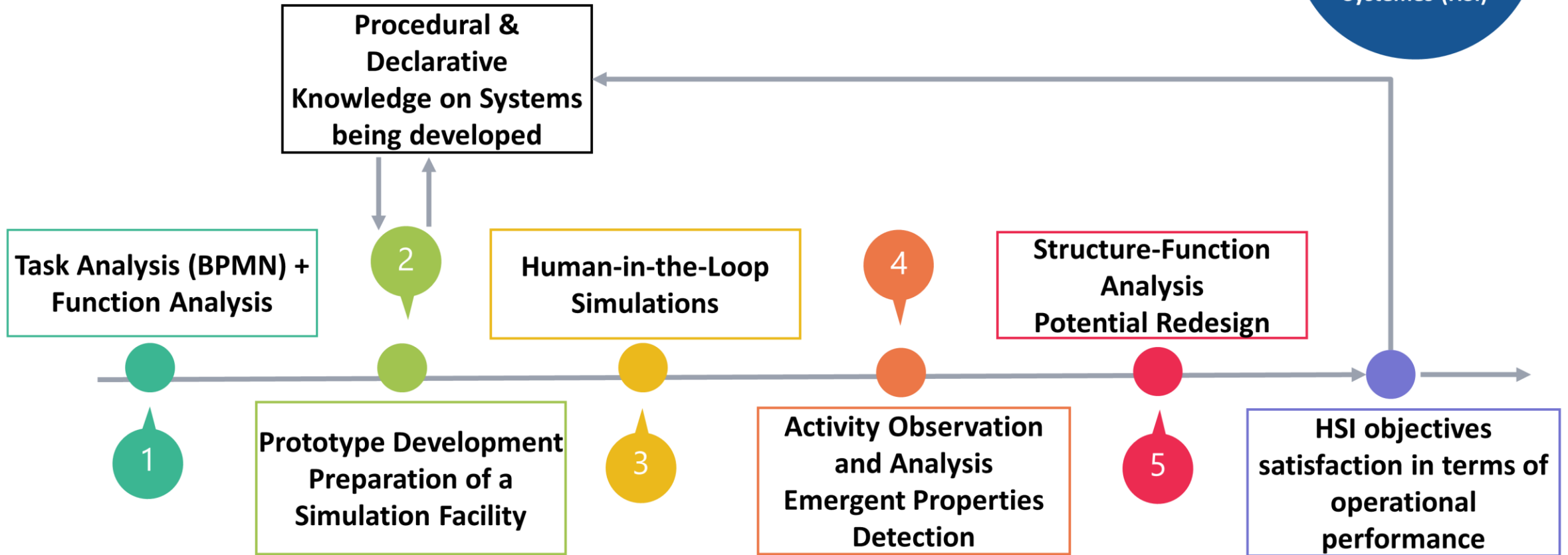


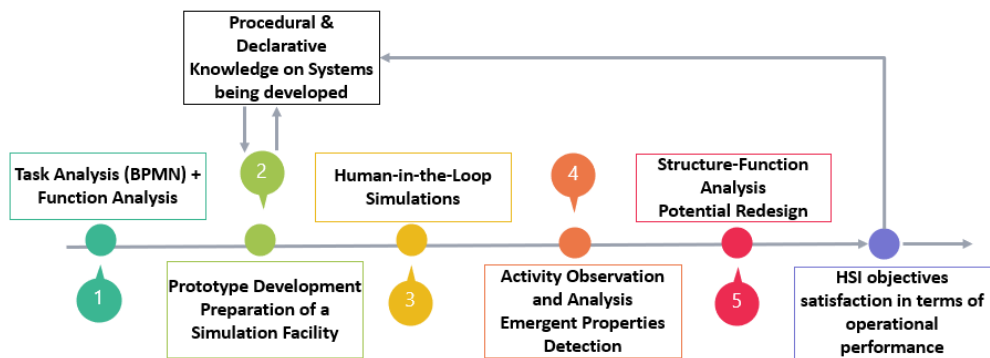
ETUDES TACTIQUES ET CONSEILS

Nous apportons un accompagnement opérationnel aux industriels : analyse du besoin, soutien au développement de nouveaux produits, expérimentation et analyses.

Méthodologie proposée

Solution méthodologique proposée pour l'analyse de la confiance et de la collaboration





1

- a. Analyse des phases de vol
- b. Analyse des tâches et analyse fonctionnelles (méthode CFA)
- c. Réalisation des BPMN

2

- a. Préparation des simulations
- b. Développement des outils d'évaluation du contexte/situation disponible (évaluation objective physiologique et évaluation subjective de la situation)
- c. Mise en place des expérimentations avec le prototype développé

3

- a. Réalisation des simulations (méthode SHDLB : simulation humain-dans-la-boucle) basées sur des scénarios denses et réalistes contextualisant les modèles BPMN précédemment décrits
- b. Analyse de la situation disponible/contexte
- c. Extraction des données brutes

4

- a. Analyse de l'activité réelle
- b. Réalisation des BPMN
- c. Recueil et analyse des critères de confiance/collaboration via la grille d'évaluation (prise en compte agrégateur dur/mou, confiance constatée/cumulée, situation disponible)
- d. Identification des fonctions émergentes

5

- a. Modification réalisée sur le système à partir des fonctions émergentes identifiées (démarche HCD)
- b. Modification de l'analyse de tâche à partir des fonctions émergentes identifiées
- c. Itération

Analyse de tâches

Tâche versus activité

En ergonomie, on distingue deux concepts clés (Montmollin, 1995) :

- **La tâche**
- **L'activité**

Tâche prescrite/Tâche réelle = Activité

La tâche prescrite : définit de façon théorique le travail, ce qui est donné à faire à l'opérateur.

Elle comprend:

- Des objectifs
- Des procédures
- Des moyens mis à disposition (outils, durée)
- Les caractéristiques de l'environnement

Exemple de tâche prescrite :

une recette de cuisine !



Préparation :

Préchauffez le four à 180°C (Ch 6).

Faites fondre le chocolat.

Ajoutez le beurre.

Mélangez les œufs, le sucre, la levure puis la farine.

Versez le chocolat fondu, puis mélangez.

Versez la préparation dans un moule.

Faites cuire environ 25 min.

Tâche prescrite/tâche réelle = Activité

La tâche réelle : Comportement réel de l'opérateur sur son lieu de travail, qui comprends :

- Les logiques d'action
- Les gestes et postures
- Les raisonnements
- Les réactions physiologiques
- Les communications
- Les verbalisations

Activité = l'écart entre la tâche prescrite et la tâche réelle

Tâche versus activité

- L'intervention ergonomique repose sur une **analyse de l'activité réelle** des êtres humains dans le système.
- L'ergonomie permet de concevoir des modèles plus fidèles de l'activité réelle, de manière à **orienter la conception des systèmes**.
 - ✓ **Systèmes de travail**
 - ✓ **Systèmes informatiques/ Interfaces Hommes-machines**
 - ✓ **Systèmes architecturaux (accessibilité)**
 - ✓ **Systèmes complexes**

BPMN

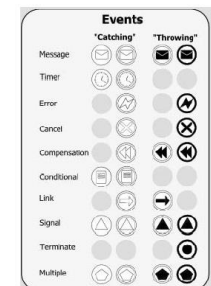
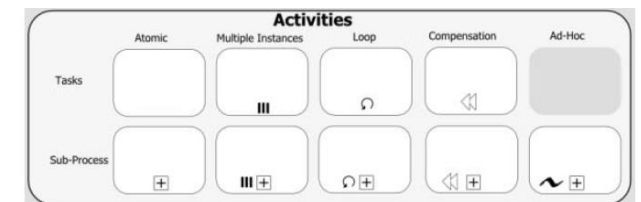
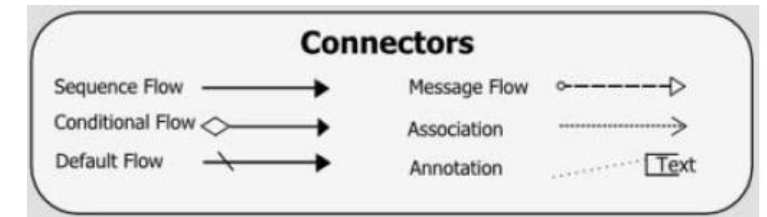
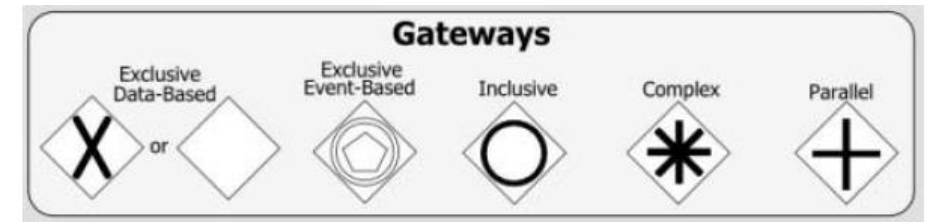
- Le formalisme BPMN (***Business Process Modelling Notation***) est développé par la BPMI (*Business Process Management Initiative*) depuis 2005. Il a été conçu pour être facilement utilisable et compréhensible par tous les acteurs de l'entreprise.
- Plateforme open source **permettant de modéliser des processus métiers complexes** : Bonitastudio

<https://fr.bonitasoft.com/>

BPMN

Il s'organise autour de six types de sigles différents symbolisant les événements, les activités, les jonctions ou branchement, les connexions et les « bassins » :

- Les évènements
- Les activités
- Les connexions
- Les branchements
- Les pistes (Pool)
- Des annotations et des associations entre éléments liés



Analyse fonctionnelle

Fonctions cognitives en suivant les uses cases associées aux ressources cognitives

Fonctions	Ressources cognitives
Actionner	Motricité, proprioception/validation par pilote
Communiquer	Perception/validation par pilote
	Attention/Mesure temporelle (réaction au stimulus)
	Mémoire ; apprentissage/penser à voix haute
Comprendre	Mémoire/énumération + cause sur rejeu
	Raisonnement/explicitation
Décider	Mémoire/énumération sur rejeu
	Prise de décision/arbre de décision
Écouter	Perception/validation par pilote
	Attention/Mesure temporelle (réaction au stimulus)
	Mémoire énumération sur rejeu
Ordonner	Perception/validation par pilote
	Attention/Mesure temporelle (réaction au stimulus)
	Mémoire ; apprentiss/penser à voix haute
Percevoir	Perception/validation par pilote
	Mémoire/énumération+cause sur rejeu
	Raisonnement/Mesure temporelle (rapidité d'inférence)
Piloter	Mémoire ; apprentiss/penser à voix haute

Dispositif expérimental

Simulations réalisées

	SIMU1	SIMU2	SIMU3	SIMU4	SIMU5
Date	23 et 24 Mars	9 Juillet	9 Juillet	22 septembre	22 septembre
Pilotes	3	2	2	1	1
Assistant virtuel	Jester +	Jester BASIC	JESTER BASIC	Jester ADVANCED	Jester ADVANCED
Niveau de difficulté de l'expérimentation	+	+	++	++	+++

Logiciels utilisés

- Module F14 de DCSWorld (Digital Combat Simulator) : Ce simulateur de vol de combat et de vol civil développé par Eagle Dynamics en association avec The Fighter Collection, paru en 2012. Le jeu est constitué de modules permettant de piloter plus d'une trentaine d'aéronefs différents et notamment le F14, en solo ou multijoueur, en vol libre ou dans le cadre d'opérations militaires de grande ampleur, dans différents contextes historiques. Le cockpit F14-TOMCAT.DCS world propose un mode de jeu [monoplace + assistant virtuel (appelé JESTER)] :offrant le choix de distribuer les tâches au sein du cockpit, et déléguant des fonctions tactiques à JESTER, Le pilote sollicite également JESTER par un menu contextuel et/ou par Voice Recognition.
- TAC view : Outil d'analyse de vol universel qui vous permet de facilement enregistrer, analyser et comprendre tout vol afin d'améliorer vos compétences bien plus rapidement qu'à l'aide de débriefings conventionnels.

Montage de simulation



+



+

**ADD – ON
MÉTIERS D'AIDE
À LA DÉCISION**



Assistant virtuel

VIRTUAL ASSISTANT IMPROVEMENTS TO ENHANCE TRUST & COLLABORATION

JESTER +

(audio + text)

JESTER + BASIC
VIRTUAL ASSISTANT

(V0 -15th may)

JESTER + ADVANCED
VIRTUAL ASSISTANT

(V0 – 1st september)

DATA COLLECTING TO INCREASE BASIC MEASURES AVAILABILITY FOR HMT KPIs

DCS DATA STREAMING

AUDIO & VIDEO

TACVIEW DATA EXPORTS

EYE TRACKERS FULL DATA PACKAGE

Outils d'évaluation

- Bonita studio (Bonisoft) : Réalisation des BPMN
- Lunettes d'eye tracking (Tobii): Analyse de l'activité
- The Observer XT (Noldus): Encodage des vidéos
- Cardiofréquencemètre : mesure de la FC
- Pilot Load Index (Algorithme Synapse): mesures du contexte/situation disponible
- Grille d'évaluation : analyse des critères de confiance et collaboration
- Echelles standardisées : SUS, NASA-TLX, SART, Scale of trust
- Questionnaires : analyse de la confiance a priori et a posteriori
- GEM : généralisation des analyses par un groupe d'expert

Outils de capture FH

Cardiofréquencemètre : Montre GARMIN

UTILISATION DE LA CEINTURE GARMIN CARDIOFRÉQUENCEMÈTRE (MESURE LA FRÉQUENCE CARDIAQUE INSTANTANÉE EN BPM) :

- plus fiable que la mesure au poignet
- moins intrusif



Eye tracking : Lunettes Tobii

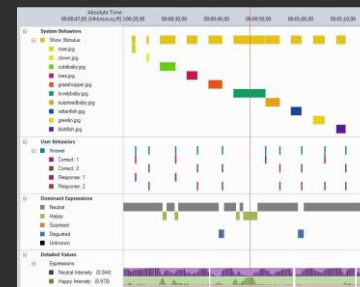
consiste à suivre et enregistrer en continu et en temps réel, la position du regard, de l'utilisateur de l'écran.

Capture des mouvements oculaires par enregistrement des points de fixation et des saccades des regards de l'utilisateur.



The Observer XT Noldus

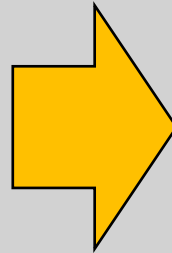
- Permet d'instrumenter l'observation des usages
- Représente les comportements de manière précise et quantitative
- Intègre les données comportementales et physiologiques
- Crée des clips vidéo des données les plus intéressantes
- Importation de données provenant de différentes sources



Besoin de contextualiser : PILOT LOAD INDEX

- Identifier le contexte pour ajuster les principes d'évaluation de la performance du teaming
 - Les macros-fonctions => 1^{ère} étape
 - La gestion du risque
 - La charge mentale du pilote
- Un algorithme qui modélise la charge cognitive du pilote

Critères d'évaluation de la confiance/collaboration



Trust metrics' criteria

Criteria	Measures	Measurement tools
Attentional management	Time	Chronometer
Competence (mission)	Subjective scale	Questionnaire
Complacency, over-trust	Expert judgement	Video recording
Knowledgeability (with virtual assistant)	Subjective scale	Questionnaire
Training level (skills)		
Feedback (interaction, processing, action effects) Virtual assistant limitations	Binary	Counting
Mission complexity consts (WL)	Subj. scale or t_R/t_A	Questionnaire
Predictability	Subjective scale	Questionnaire
Redundancy (cross-checking/verification)	Subjective scale	Questionnaire
Risk awareness & Safety	Subjective scale	Questionnaire
Situation awareness	Subjective scale	Questionnaire
Transparency (Uncertainty, Vulnerability)	Subjective scale	Questionnaire

Collaboration metrics' criteria

Criteria	Measures	Measurement tools
Adaptability (HSI in context)	Subjective scale	Questionnaire
Intersubjectivity (Common ground, shared cognition)	Subjective scale (verbal protocols)	Questionnaire
Control vs. management (Directability)	Binary, verbal protocols	Counting
Delegation of authority	Binary	Counting
Efficiency in collaboration, interaction	Scale	Questionnaire
Egocentric versus collaborative	Subjective scale	Questionnaire

SÉLECTION PRENANT EN COMPTE DIFFÉRENTS PARAMÈTRES :

- Signification, explicitation et terminologie (ontologie)
- Pertinence vis-à-vis du projet MOHICAN
- Mesurabilité

Critères d'évaluation de la confiance/collaboration

2^{ÈME} SÉLECTION PRENANT EN COMPTE DIFFÉRENTS PARAMÈTRES :

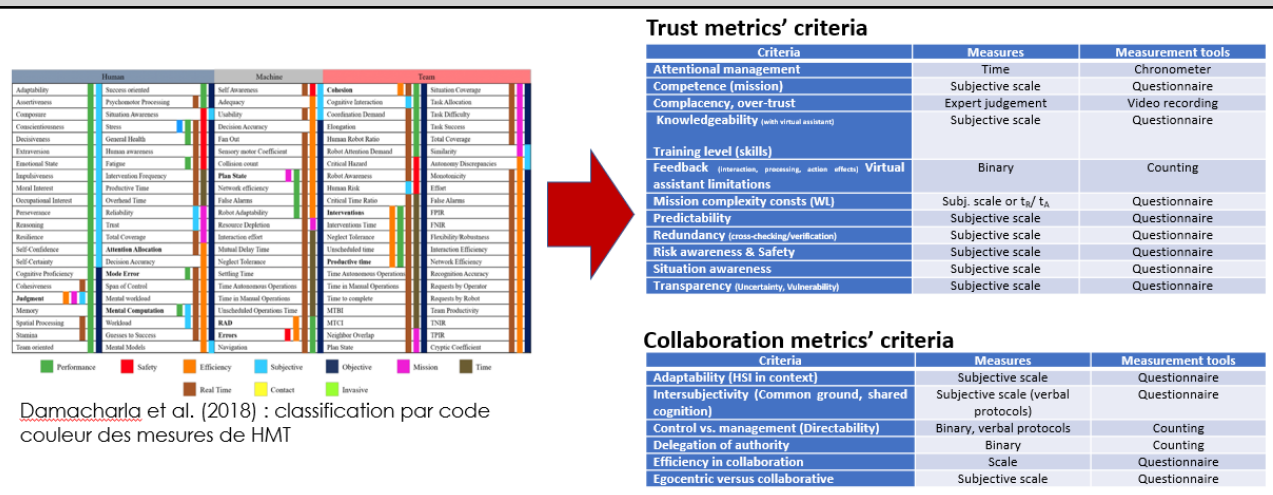
- Mesurabilité (mesures objectives versus subjectives)
- Pertinence vis-à-vis de la mission
- Influences au cours du temps (critères évolutifs versus stables)

LISTE DE 9 CRITÈRES POUR ANALYSER LA QUALITÉ DE CHAQUE INTERACTION : EFFICACITÉ - EFFICIENCE - FIABILITÉ/ROBUSTESSE - PERTINENCE - TRANSPARENCE - FLEXIBILITÉ/ADAPTABILITÉ - QUALITÉ DU FEEDBACK - ALLÈGEMENT PERÇU DE LA TÂCHE - ABSENCE DE GÊNE.

4 ÉCHELLES STANDARDISÉES :

- UTILISABILITÉ (SUS)
- CHARGE MENTALE (NASA-TLX)
- SITUATION AWARENESS (SART)
- SCALE OF TRUST

QUESTIONNAIRES : QUESTIONNAIRE SUR LA CONFIANCE A PRIORI, QUESTIONNAIRE SUR LA CONFIANCE APRÈS INTERACTION : CONNAISSANCE ET COMPRÉHENSION DE L'AV, EFFORT, MAÎTRISE DES RISQUES, PRÉDICTIBILITÉ, BÉNÉFICE, PERFORMANCE DANS LES SITUATIONS CRITIQUES, QUANTITÉ D'ATTENTION REQUISE, GÊNE, NIVEAU DE STRESS, CONTRÔLE DE LA SITUATION, NIVEAU DE VIGILANCE, PRISE DE RISQUE, NÉCESSITÉ ETC...



Mesures de bas niveau

Elaboration d'une grille d'analyse :

- permet d'étudier l'ensemble des interactions qui ont lieu pendant les expérimentations ;
- permet de recueillir l'ensemble des **mesures objectives et subjectives** lié aux différents critères de confiance et de collaboration ;

=> élaborée de manière itérative.

Objectifs :

- qualifier le niveau de confiance et de collaboration pour chaque interaction ;
- prendre en compte la situation disponible à l'aide des indices ;
- différencier la confiance constatée de la confiance cumulée ;
- évaluer la sur-confiance ;
- Améliorer l'AV via une démarche HCD.

Echelles standardisées

⇒ SUS (System Usability Scale) : Évaluation de l'utilisabilité;

- Cette échelle permet d'évaluer l'utilisabilité d'un système. Pour rappel, l'utilisabilité est le degré selon lequel un système peut être utilisé par des utilisateurs identifiés pour atteindre des buts définis avec efficacité, efficacité et satisfaction, dans un contexte d'utilisation spécifié. Ce questionnaire est composé de 10 questions avec une échelle sur 5.

⇒ NASA-TLX (TASK Load Index) : Indice de charge de travail mentale;

- Echelle permettant de mesurer la charge de tâche subjective. Elle se compte de 6 items sur une échelle de 20 graduations : Exigence mentale, exigence physique, pression temporelle, performance, effort, frustration.

Echelles standardisées

⇒ *SART (Situation Awareness Rating Technique) : Évaluation de la Situation Awareness;*

- Cette méthode repose sur l'évaluation de dix dimensions liées à la conscience de la situation : Connaissance de la situation, complexité de la situation, niveau d'attention, variabilité de la situation, quantité d'information traitée, excitation, instabilité de la situation, qualité de l'information, de la concentration de l'attention et de la familiarité avec la situation. Les dix dimensions sont à évaluer avec une échelle de 1 (faible) à 7 (élevé).

⇒ *Scale of Trust in Automated Systems: Évaluation de la confiance ;*

- Cette échelle constituée de 12 items avec une échelle de Lickert en 7 points évalue la confiance entre un humain et une machine.

Questionnaires pré-test

- Questionnaire sur le profil des pilotes ;
- Questionnaire pré-test sur la confiance a priori :

Ce questionnaire a été élaboré à partir de la thèse de Rajaonah, B. (2006). Il comprend 4 parties (Chaque item est associé à une échelle analogique visuelle linéaire de 10 cm (valeur usuelle, selon Rude & Retel, 2000) :

- **Partie 1** : propension à faire confiance (11 questions) ;
- **Partie 2** : L'attitude face aux technologies et face à l'utilisation de l'AV (6 questions) ;
- **Partie 3** : la prise de risque associée à l'utilisation d'un AV (4 questions) ;
- **Partie 4** : Les bénéfices potentiels (4 questions) ;
- **Partie 5** : Les qualités que doit posséder l'AV (5 questions).

Questionnaires post-test

- **Questionnaire post-test sur la confiance pour chaque type d'interaction en pré et post-test.** Ce questionnaire détaille sur une échelle analogique visuelle linéaire de 10 cm (valeur usuelle, selon Rude & Retel, 2000), la confiance accordée pour chaque fonctionnalité de Jester.
- **Entretiens semi-directif en 4 parties :**
 - **Partie 1 :** Questions générales
 - **Partie 2 :** questions sur la confiance
 - **Partie 3 :** questions sur l'utilisabilité
 - **Partie 4 :** questions sur la collaboration

Questionnaires post-test

- Questionnaire post-test sur la confiance après interaction avec l'AV

Ce questionnaire a été élaboré à partir de la thèse de Rajaonah, B. (2006). Il comprend 26 questions : Connaissance de l'AV, Plaisant, Qualité de l'interaction, Facilité d'utilisation, Effort, Compréhension de l'AV, Maîtrise des risques et des ressources, Prédicibilité de l'AV, Bénéfice, Performance dans les situations critiques, Quantité d'attention requise, gêne, Allègement de la tâche, Niveau de stress, Dépendance, Facilite la tâche, Contrôle de la situation, Utilité, Nécessité, Prise de risque, Niveau de vigilance, Niveau de confiance, Confiance dans l'AV, Confiance dans la relation avec l'AV, Confiance en moi.

26 QUESTIONS :

-CONNAISSANCE DE L'AV

-PLAISANT

-QUALITÉ DE L'INTERACTION

-FACILITÉ D'UTILISATION

-EFFORT

-COMPRÉHENSION DE L'AV

-MAÎTRISE DES RISQUES ET DES RESSOURCES

-PRÉDICTIBILITÉ DE L'AV

-BÉNÉFICE

-PERFORMANCE DANS LES SITUATIONS CRITIQUES

-QUANTITÉ D'ATTENTION REQUISE

- GÊNE

-ALLÈGEMENT DE LA TÂCHE

-NIVEAU DE STRESS

-DÉPENDANCE

-FACILITE LA TÂCHE

-CONTRÔLE DE LA SITUATION

-UTILITÉ

-NÉCESSITÉ

-PRISE DE RISQUE

-NIVEAU DE VIGILANCE

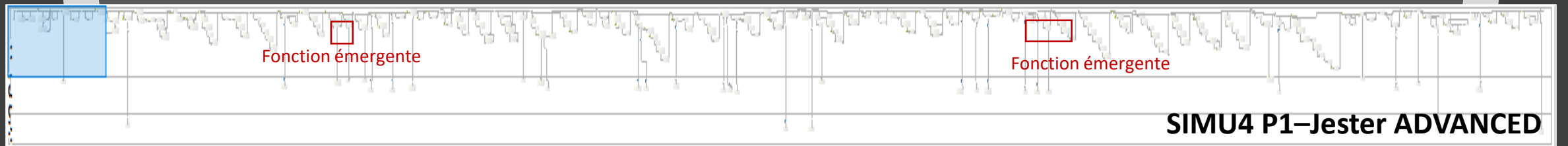
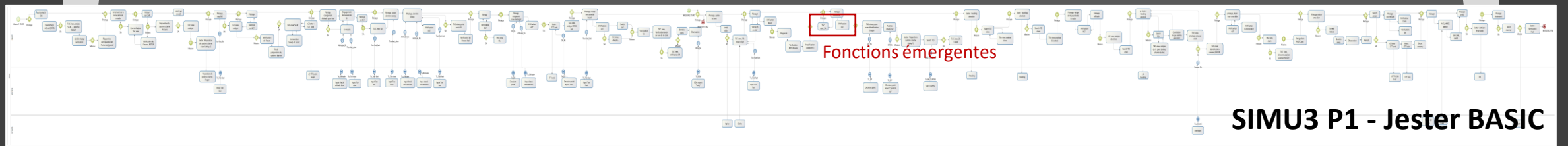
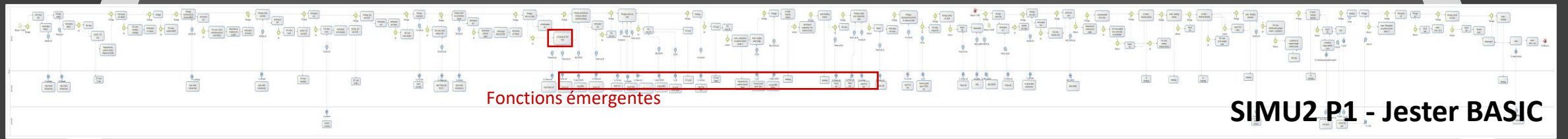
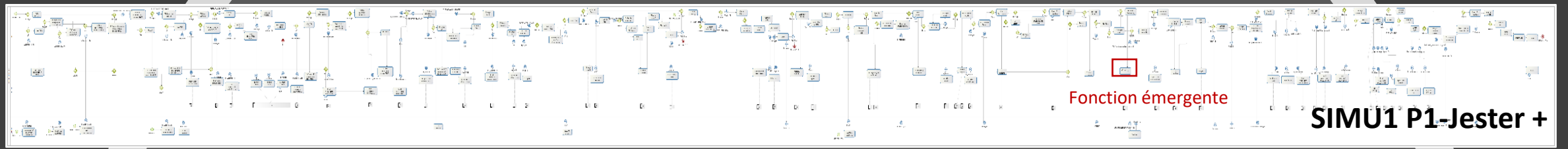
-NIVEAU DE CONFIANCE

-CONFIANCE DANS L'AV

-CONFIANCE DANS LA RELATION AVEC L'AV

-CONFIANCE EN MOI

Analyse



Analyse confiance et collaboration

- **Exemple de perte de confiance cumulée : « Bloc Altitude »**
 - **Exemple de sur-confiance : « Speed to fast »**
- => Comparaison entre critère d'efficacité et fiabilité du système.
- **Exemple de variation de la confiance et de la collaboration en fonction du contexte, identification d'une fonction émergente**
- => critère d'efficacité
-

Bilan

Travail réalisé et limites

Prise en compte du contexte pour l'analyse de la confiance et de la collaboration

- Indice théorique de charge cognitive du pilote;
- Utilisable en temps réel et permet de remplacer l'utilisation de la FC (Pertinence hors simu) ;
- Adapte les principes de mesure de confiance et de collaboration dans le temps (Confiance instantanée versus cumulée : Acquisition et perte de la confiance).

LIMITES

- Algorithme d'indice de tâches se base sur les macro-fonctions mais pas sur l'activité (Théorie Vs activité réelle)
- Algorithme doit être étendu à d'autres contextes & amélioré
- Algorithme doit être testé avec un panel de pilotes plus diversifié

Modèles et métriques de confiance et de collaboration

- Mise en évidence de critères de performance de confiance et collaboration analysés dans une grille ;
- Analyse de la confiance a priori ;
- Analyse globale de la confiance et de la collaboration avec l'utilisation de l'échelle standardisée Échelle de confiance dans les systèmes automatisés + questionnaires

LIMITES

- Le questionnaire n'est pas standardisé
- mesures subjectives prépondérantes dans le modèle

Validation de la démarche HCD : permet d'optimiser la construction de l'AV

Merci pour votre attention !