



**HAL**  
open science

# Le rôle de l'utilisateur dans les systèmes de traitements automatiques

Sylvain Fleury

► **To cite this version:**

Sylvain Fleury. Le rôle de l'utilisateur dans les systèmes de traitements automatiques. Psychologie. Université Rennes 2, 2014. Français. NNT : 2014REN20001 . tel-00935100

**HAL Id: tel-00935100**

**<https://theses.hal.science/tel-00935100>**

Submitted on 23 Jan 2014

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

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



**THESE / Université Rennes 2**  
*sous le sceau de l'Université Européenne de Bretagne*  
pour obtenir le titre de  
**DOCTEUR de l'Université de Rennes 2**  
*Mention : Psychologie cognitive*  
**Ecole doctorale Sciences Humaines et Sociales**

présentée par  
**Sylvain Fleury**  
Centre de Recherches en Psychologie, Cognition  
et Communication (EA-1285)

# Le rôle de l'utilisateur dans les systèmes de traitements automatiques

**Thèse soutenue le 09/01/2014**  
devant le jury composé de :

**Mireille BÉTRANCOURT**  
Professeure des Universités, Université de Genève / *rapporteur*

**Christine CHAUVIN**  
Professeure des Universités, Université de Bretagne Sud / *rapporteur*

**Eric ANQUETIL**  
Professeur des Universités, INSA Rennes / *examineur*

**Eric JAMET**  
Professeur des Universités, Université Rennes 2 / *Directeur de thèse*

**SOUS LE SCEAU DE L'UNIVERSITÉ EUROPÉENNE DE BRETAGNE**

UNIVERSITÉ RENNES 2

École Doctorale Sciences Humaines et Sociales

Unité de Recherche – CRPCC (EA 1285)

Centre de Recherche en Psychologie, Cognition et Communication

**LE RÔLE DE L'UTILISATEUR DANS LES SYSTÈMES DE TRAITEMENTS  
AUTOMATIQUES**

Thèse de Doctorat

Discipline : Psychologie Cognitive

Présentée par Sylvain FLEURY

Directeur de thèse : Éric JAMET

Soutenue le 9 janvier 2014

Jury :

Mireille BÉTRANCOURT, Professeure des Universités, Université de Genève (*rapporteur*)

Christine CHAUVIN, Professeure des Universités, Université de Bretagne Sud (*rapporteur*)

Eric ANQUETIL, Professeur des Universités, INSA Rennes (*examineur*)

Éric JAMET, Professeur des Universités, Université Rennes 2 (*Directeur de thèse*)

# Avant-propos

Ce travail de thèse a fait l'objet de quatre soumissions d'articles dans des revues internationales dont deux acceptations. Le contenu de ces articles a été entièrement réécrit afin de s'insérer plus harmonieusement dans ce manuscrit.

**Étude 2.** Fleury, S., & Jamet, E. (in press). Facilitating the comparison of multiple visual items on screen : the example of electronic architectural plan correction. *Applied Ergonomics*.

**Étude 3.** Fleury, S., & Jamet, E. (soumis). Putting Users Back in the Interaction Loop for the Automatic Supervised Interpretation of Technical Documents.

**Étude 5.** Fleury, S., Jamet, E., Ghorbel, A., Lemaitre, A., & Anquetil, E. (soumis). Application of the Resources Model to the Supervision of an Automated Process.

**Conclusion générale.** Fleury, S., Jamet, E., Loup-Escande, E., Ghorbel, A., Lemaitre, A., & Anquetil, E. (in press). Towards Specifications for Automatic Recognition Software : An Example of a User-Centred Design. *Journal of Software Engineering and Applications*.

# Remerciements

Après plus de 100000 km effectués en train pour aller de mon domicile au laboratoire, soit environ 2,5 fois le tour de la Terre ou plus de 15 fois la distance parcourue pour se rendre à Washington et y réaliser une conférence, après l'équivalent de 60 jours entiers passés dans ce transport, ce travail de thèse est arrivé à son terme.

Je tiens à remercier tout particulièrement Eric pour sa confiance, sa disponibilité et pour m'avoir patiemment appris le métier de chercheur.

Mes remerciements vont également aux membres du jury pour avoir accepté de consacrer du temps à examiner cette thèse et de participer à ma soutenance.

Je remercie les collègues de l'IRISA et de Script&Go pour m'avoir fait bénéficier de leurs compétences en informatique et de leur regard technique sur mon travail, mais aussi toute l'équipe du CRPCC, les enseignants, étudiants et ingénieurs pour leur sympathie, leur aide et pour les échanges souvent fructueux.

J'adresse un immense merci à Samantha, Nicolas, Jean, Natacha ainsi que mes parents, pour la tâche de « détection d'erreurs » à laquelle ils se sont livrés sur ce manuscrit.

Enfin, merci aux participants pour leur rôle essentiel dans ce travail.

# Table des matières

<b>Introduction générale</b>	<b>9</b>
<b>1.1. MobiSketch : une démarche d'ergonomie de conception</b>	<b>10</b>
<b>1.2. L'interprétation automatique de plans</b>	<b>12</b>
<b>1.3. L'automatisation et le rôle de l'opérateur</b>	<b>13</b>
1.3.1. Les 3 niveaux de contrôle dans les interactions homme-machine	13
1.3.2. La complémentarité homme-machine	16
1.3.3. Les ironies de l'automatisation	19
1.3.4. Le syndrome de l'humain hors de la boucle	20
1.3.5. Les niveaux d'automatisation	22
<b>1.4. Problématique de la thèse</b>	<b>24</b>
<b>Étude 1 : Analyse des activités de complétion, recopie et correction de plans d'architecture</b>	<b>26</b>
<b>1. INTRODUCTION</b>	<b>26</b>
1.1. Traitement des plans	26
1.2. Production écrite et traitement des plans	28
1.3. Traitement des informations visuo-spatiales	30
<b>2. METHODE</b>	<b>32</b>
2.1. Participants	32
2.2. Matériel	33
2.3. Procédure	34
2.4. Mesures	35
<b>3. RESULTATS</b>	<b>36</b>
3.1. Epreuve de complétion de plan	37
3.2. Epreuve de recopie de plan	37
3.3. Epreuve de correction de plan	38
3.4. Tests d'empans	39
<b>4. DISCUSSION</b>	<b>40</b>
<b>Étude 2 : Comment présenter l'interprétation aux utilisateurs ? Effet de l'intégration spatiale et de la séquentialité</b>	<b>42</b>
<b>1. INTRODUCTION</b>	<b>42</b>
1.1. Traitement de deux sources visuelles	43
1.1.1. Le principe de proximité compatibilité	44
1.1.2. Effet du partage de l'attention	45
1.1.3. Le principe de contiguïté spatiale	47
1.2. Le guidage de l'exploration visuelle d'un document	50
<b>2. EXPERIENCE 1</b>	<b>54</b>
2.1. METHODE	55
2.1.1. Participants	55
2.1.2. Matériel	55
2.1.3. Procédure	57
2.2. Résultats	57

2.2.1. Efficience	58
2.2.2. Efficacité	58
2.3. Discussion	59
<b>3. EXPERIENCE 2</b>	<b>61</b>
3.1. Matériel et méthode	62
3.1.1. Participants	62
3.1.2. Matériel	62
3.1.3. Procédure	62
3.2. Résultats	63
3.2.1. Efficience	63
3.2.2. Efficacité	63
3.2.3. Temps de fixation	64
3.2.4. Contrôle de l'ensemble des erreurs	65
3.3. Discussion	66
<b>4. CONCLUSION</b>	<b>67</b>
<b>Étude 3 : Effets de l'interruption utilisateur sur la correction des erreurs</b>	<b>69</b>
<b>1. INTRODUCTION</b>	<b>69</b>
1.1. L'oubli en mémoire à court terme	70
1.2. Le modèle du processeur humain	71
1.3. L'interférence en mémoire à court terme	72
1.4. Hypothèses	74
<b>2. Méthode</b>	<b>75</b>
2.1. Participants	75
2.2. Matériel et procédure	75
<b>3. Résultats</b>	<b>76</b>
3.1. Avec Interruption vs Sans Interruption	76
3.2. Description des comportements d'interruptions	77
3.3. Interruptions effectives	78
<b>4. Discussion</b>	<b>79</b>
<b>Étude 4 : Test utilisateur du prototype fonctionnel avec interruptions système</b>	<b>83</b>
<b>1. INTRODUCTION</b>	<b>83</b>
<b>2. METHODE</b>	<b>84</b>
2.1. Participants	84
2.2. Matériel et procédure	85
<b>3. RESULTATS</b>	<b>87</b>
3.1. Données comportementales	87
3.1.1. Interruptions	87
3.1.2. Durée de la tâche	89
3.1.3. Repérage des erreurs	91
3.2. Données de l'entretien	92
3.2.1. Interruptions du système	92
3.2.2. Sollicitations de l'utilisateur	93
3.2.3. Remarques générales	94
<b>4. DISCUSSION</b>	<b>94</b>
<b>Étude 5 : Impact des sollicitations système sur la correction des erreurs</b>	<b>101</b>
<b>1. INTRODUCTION</b>	<b>101</b>

1.1. Détection de cible avec assistance automatique	103
1.2. Le modèle des ressources	105
1.2.1. Les ressources informationnelles	105
1.2.2. Les stratégies d'interaction	107
1.3. Application du modèle des ressources à la reconnaissance de plans	108
1.4. La confiance dans le système	110
1.5. Hypothèses	116
<b>2. METHODE</b>	<b>119</b>
2.1. Participants	119
2.2. Matériel	119
2.3. Procédure	120
<b>3. RESULTATS</b>	<b>122</b>
3.1. Durée de la tâche	122
3.2. Erreurs	123
3.2.1. Erreurs signalées par le système	123
3.2.2. Erreurs non signalées par le système	124
3.3. Echelle de confiance dans les systèmes automatiques	125
<b>4. DISCUSSION</b>	<b>126</b>
<b>5. CONCLUSION</b>	<b>132</b>
<b><i>Étude 6 : Évaluation du prototype auprès des utilisateurs finaux</i></b>	<b>133</b>
<b>1. Introduction</b>	<b>133</b>
1.1. Évaluation de l'intention d'usage : le TAM	134
<b>2. Evaluation 1</b>	<b>136</b>
2.1. Méthode	137
2.1.1. Participants	137
2.1.2. Matériel et procédure	137
2.2. Résultats	138
2.2.1. Utilité et utilisabilité perçues des fonctions du système	138
2.2.2. Questions d'usage	139
2.3. Discussion	141
<b>3. Evaluation 2</b>	<b>142</b>
3.1. Méthode	142
3.1.1. Participants	142
3.1.2. Matériel et procédure	143
3.2. Résultats	143
3.2.1. Continuum papier/DAO	143
3.2.2. Esquisse de projet	144
3.2.3. Édition de plans en mobilité	145
3.2.4. Relation client	146
3.3. Discussion	146
<b><i>Conclusion générale</i></b>	<b>149</b>
<b>1. Les niveaux d'automatisation au cours du projet</b>	<b>149</b>
<b>2. Préconisations</b>	<b>150</b>
2.1. Comment présenter les informations à l'écran ?	151
2.2. Faut-il permettre aux utilisateurs d'interrompre le système ?	153
2.3. Quelle forme donner aux fonctions interactives ?	154
2.4. Comment faciliter la vérification de l'interprétation ?	156
2.5. Le système doit-il communiquer ses incertitudes aux utilisateurs ?	158

<b>3. Apports théoriques</b>	<b>160</b>
<b>4. Limites et perspectives</b>	<b>161</b>
<b><i>Bibliographie</i></b>	<b>164</b>
<b><i>Annexes</i></b>	<b>182</b>
<b><i>Index des Illustrations</i></b>	<b>200</b>
<b><i>Index des Tableaux</i></b>	<b>202</b>

# Introduction générale

L'automatisation peut être définie comme « l'exécution par une machine (généralement un ordinateur) d'une fonction qui, avant, était assumée par un humain » (Parasuraman & Riley, 1997, p. 231, notre traduction). Pourtant, des opérateurs se trouvent encore impliqués dans des tâches largement automatisées pour être en mesure de réagir en cas de problème. La raison essentielle du maintien des humains dans des systèmes fortement automatisés est leur indispensable capacité à faire face aux situations imprévues (Reason, 1993). Cette problématique de la place des hommes par rapport aux systèmes automatiques pose question depuis plusieurs décennies, y compris au sein du grand public, au point par exemple de devenir l'objet d'une caricature dans le magazine « The New Yorker » (voir figure 1).

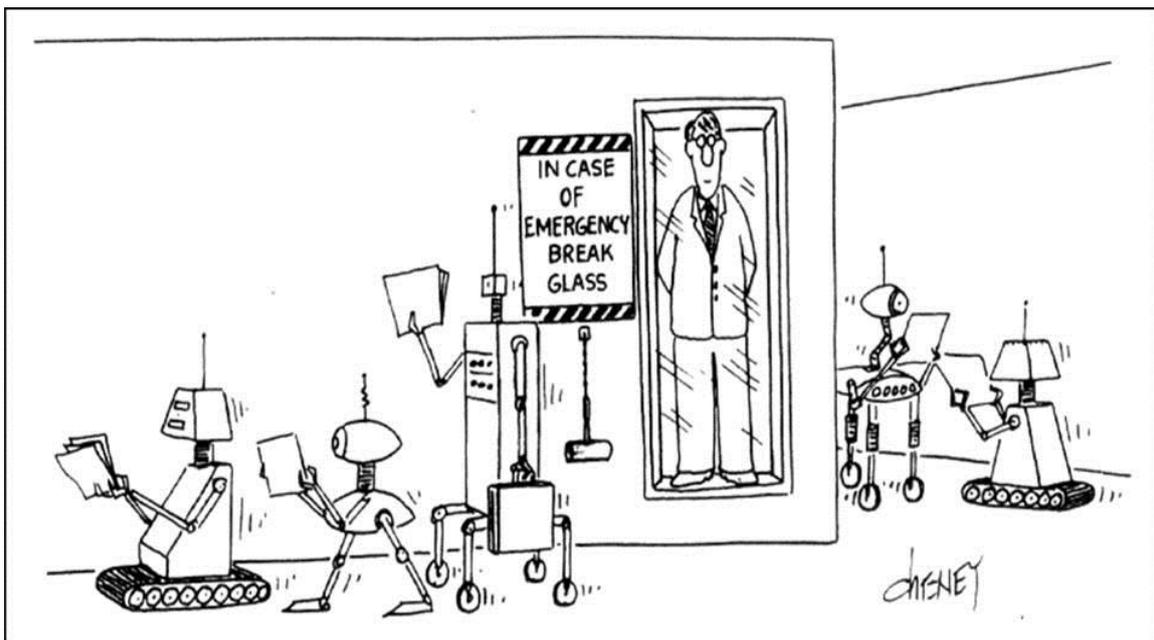


Figure 1. "Robots milling around the office halls, pass man in glass booth with sign that reads: In Case of Emergency Break Glass" (Cheney, 1989)

Les métiers dans lesquels l'optimisation de l'interaction homme-machine en vue de diminuer les erreurs a été la plus étudiée sont ceux qui présentent un fort enjeu en termes de sécurité comme les opérateurs de centrales nucléaires, les contrôleurs

aériens ou même les pilotes d'avions. Pourtant, dans les secteurs professionnels au sein desquels les erreurs ne génèrent pas de danger pour les personnes, celles-ci peuvent tout de même être à l'origine d'importantes pertes de temps ou d'argent.

## **1.1. MobiSketch : une démarche d'ergonomie de conception**

Le travail décrit dans le présent manuscrit s'inscrit dans cette thématique de l'automatisation et des interactions homme-machine. Il a été réalisé dans le cadre du projet ANR MobiSketch<sup>1</sup>. L'objectif de celui-ci était l'élaboration d'un logiciel d'interprétation automatique de documents techniques structurés. Autrement appelée rétro-conversion, conversion rétrospective ou plus simplement reconnaissance automatique, l'interprétation permet d'extraire d'un document numérisé une représentation interprétée éditable. Cette fonctionnalité vise à constituer un continuum entre le document sous sa forme papier et ce même document sous sa forme interprétée. Il s'agit de rendre exploitable par un ordinateur des documents présents sur papier en les transformant dans des formats standards.

Les trois partenaires pour ce projet étaient l'Institut de Recherche en Informatique et Systèmes Aléatoires (IRISA) pour la recherche sur la reconnaissance de documents, le Centre de Recherche en Psychologie, Cognition et Communication (CRPCC) qui travaillait sur les aspects ergonomiques et Script&Go dont le rôle était d'apporter une connaissance métier et un travail sur les interfaces.

Deux orientations complémentaires ont été mise en œuvre lors de la conception du logiciel : une conception technocentrée d'une part et une conception anthropocentrée d'autre part. L'orientation technocentrée conduite par l'IRISA visait à concevoir et optimiser un logiciel innovant en améliorant ses possibilités techniques et en franchissant des verrous technologiques. L'orientation anthropocentrée opérée par le

---

<sup>1</sup> ANR-09-CORD-015

CRPCC avait quant à elle pour objectif la conception d'un logiciel adapté à l'utilisateur final (Wilson, Bekker, Johnson, & Johnson, 1996).

Dans cette seconde orientation, l'ergonomie suivait une démarche de conception centrée-utilisateur. La norme ISO 9241-210 (2010) identifie 6 principes qui caractérisent la conception centrée-utilisateur : 1) la conception est basée sur une compréhension explicite des utilisateurs, des tâches et des environnements, 2) les utilisateurs sont impliqués tout au long de la conception et du développement, 3) la conception est dirigée et affinée par les évaluations centrées utilisateurs, 4) le processus est itératif, 5) la conception prend en compte la totalité de l'expérience utilisateur, 6) l'équipe de conception regroupe des compétences et des perspectives pluridisciplinaires.

L'application de ces principes conduit à 4 tâches principales. La première tâche consiste à comprendre et à spécifier le contexte d'utilisation du système. La seconde tâche vise à spécifier les besoins des utilisateurs et les exigences des différents acteurs. L'objectif de la troisième tâche est de produire des solutions de conception (*e.g.* scénario, maquette, prototype). Enfin, la quatrième tâche concerne l'évaluation des solutions à tous les niveaux du projet depuis les premières étapes de la conception jusqu'à l'utilisation sur le long-terme afin de préciser les choix de conception. Pour réaliser cette quatrième tâche, deux principales méthodes d'évaluations centrées utilisateurs existent : les tests auprès d'utilisateurs et les évaluations expertes basées sur les travaux concernant l'utilisabilité.

Dans ce projet, nous avons effectué des allers-et-retours entre la littérature et le recueil de données empiriques. Ce recueil a été réalisé à travers l'utilisation de la méthode expérimentale et des évaluations de prototypes par le biais d'entretiens ainsi que d'enquêtes. Ce manuscrit décrit le travail d'ergonomie réalisé dans le cadre du projet MobiSketch, mais ne traite pas des travaux menés en parallèle sur les aspects techniques du système (Ghorbel, 2012).

## 1.2. L'interprétation automatique de plans

Pour certaines activités comme la conception de plans d'architecture, les professionnels passent généralement par une étape de dessin manuscrit sur papier (Manolopoulou, 2005; Bilda & Gero, 2006) avant de reconstruire le plan dans un logiciel de Dessin Assisté par Ordinateur (DAO). Bien que les résultats du projet soient applicables à d'autres types de documents techniques structurés, l'exemple qui a été choisi pour les tests est celui des plans d'architecture manuscrits en 2 dimensions.

Plus généralement, au-delà de cette application spécifique, de nombreux documents techniques sont interprétés dans un but d'archivage, mais aussi avec un objectif d'édition au sein de logiciels de DAO. Ce besoin conduit à la mise en œuvre de nombreuses recherches pour l'amélioration de logiciels de reconnaissance automatique de symboles pour les plans cadastraux (Ogier, Mullot, Labiche, & Lecourtier, 1995), les plans électriques (Ouyang & Davis, 2009), les cartes routières (Chiang & Knoblock, 2011; Chiang, Leyk, & Knoblock, 2013), les logos (Lladós, Valveny, Sanchez, & Marti, 2002) ou encore les partitions musicales (Bainbridge & Bell, 2001; Pugin, 2006; Rebelo *et al.*, 2012; Raphael & Wang, 2011).

Les études sur la reconnaissance automatique de plans d'architecture sont elles aussi particulièrement nombreuses (*e.g.* Ah-Soon & Tombre, 1997; Ahmed, Weber, Langenhan, Dengel, & Petzold, 2013; Lladós *et al.*, 2002; Valveny & Marti, 1999; Koutamanis & Mitossi, 1993). Ces travaux portent généralement sur un aspect spécifique de l'interprétation de plan, comme l'apprentissage incrémental de nouveaux symboles (Ah-Soon & Tombre, 1998; Ah-Soon & Tombre, 2001), le partitionnement des pièces d'un bâtiment (Macé, Locteau, Valveny, & Tabbone, 2010) ou encore l'interprétation en 3D d'un plan (Elsen, Demaret, Yang, & Leclercq, 2012).

L'optimisation de logiciels d'interprétation automatique de plans pourrait permettre un gain de temps pour des professionnels. Par exemple, une recherche de Lu, Tai, Su et Cai (2005) montrait que leur système était capable d'interpréter en moins d'une heure

des plans d'une complexité telle qu'ils estimaient qu'un humain mettrait au moins une semaine et demi à les traiter. Cependant, le traitement réalisé par ce type de logiciel comporte forcément un risque d'erreurs. Lu *et al.* (2005) ont estimé à partir de 200 dessins que leur système d'interprétation automatique atteignait un taux de reconnaissance de 90%. Les évaluations d'Ahmed, Liwicki, Weber et Dengel (2011) sur leur propre méthode d'analyse montraient quant à elles un taux de reconnaissance de 79% (les corpus de plans utilisés ne sont pas identiques dans les différentes évaluations). Enfin, des évaluations de notre propre prototype en cours de développement montraient un taux de reconnaissance de 91% sur des plans relativement simples (Ghorbel, Lemaitre, & Anquetil, 2012). Cela implique qu'un opérateur doit être présent et en mesure de corriger ces erreurs. D'ailleurs, cette contrainte de vérification n'est pas spécifique à l'interprétation automatique. En effet, la recopie manuelle de schémas papiers sur des logiciels de DAO impose elle aussi de contrôler les erreurs. De la même manière, les systèmes de beautification, *i.e.* de modification des contenus graphiques pour les rendre visuellement plus attrayants sans en changer le sens (Sezgin, Stahovich, & Davis, 2007), sont aussi sujets à un risque d'erreurs (Orbay & Kara, 2011) et doivent donc faire l'objet de vérifications.

## **1.3. L'automatisation et le rôle de l'opérateur**

Le logiciel développé dans le cadre du projet MobiSketch nécessite l'implication d'un opérateur pour corriger les erreurs. Or, son rôle exact dans lors de l'interprétation de plan doit être défini.

### **1.3.1. Les 3 niveaux de contrôle dans les interactions homme-machine**

Les interactions des opérateurs avec des systèmes techniques complexes ont été décrites dans un modèle particulièrement influent proposé par Rasmussen

(1979, 1983, 1986). Le modèle de Rasmussen (1983) décrit trois niveaux de contrôle qu'un opérateur peut exercer dans une tâche donnée. Ces niveaux se distinguent selon le type de ressources cognitives qu'ils conduisent à mobiliser (voir figure 2). Ainsi, les comportements peuvent être basés sur les habiletés, les règles ou les connaissances.

Le contrôle basé sur les habiletés renvoie à des configurations de comportements sensorimoteurs automatisés, ne nécessitant donc pas d'effort conscient. Ces comportements peuvent être mis en œuvre lors d'un simple examen des feedbacks durant lequel, dès la détection d'un signal d'erreur donné, une réaction motrice est effectuée par l'opérateur (Rasmussen, 1983). Ces actions sont réalisées de manière quasi automatique suite à un apprentissage relativement long. Du fait de cette automatisation, ces actions nécessitent peu de ressources pour être effectuées. À titre d'exemple de comportement basé sur les habiletés, on peut citer le maintien de la trajectoire d'une voiture sur la route par le biais de légers ajustements continus avec le volant. Cette tâche requière de l'entraînement, mais elle n'est que très peu coûteuse en ressources cognitives et repose sur le traitement en continu de feedbacks analogiques récupérés dans l'environnement. Les feedbacks utilisés dans les comportements basés sur les habiletés sont ce que Rasmussen (1983) a appelés des « signaux ». Les signaux sont des données sensorielles traitées comme des variables continues et qui n'ont pas de sens en dehors du contexte spatio-temporel précis dans lequel se trouve l'opérateur (Rasmussen, 1983).

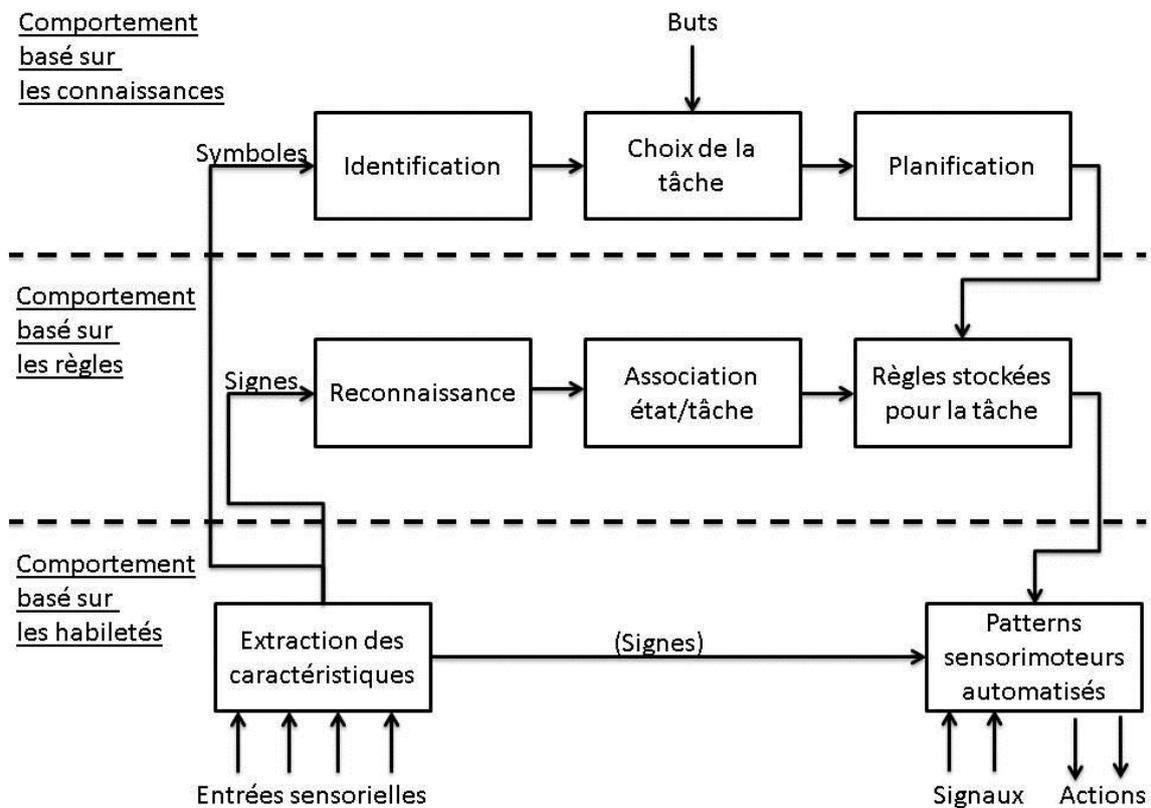


Figure 2. Version simplifiée du modèle des 3 niveaux de performances des opérateurs (d'après Rasmussen, 1983)

Le second niveau de comportement décrit par Rasmussen (1983) est basé sur les règles. Les règles évoquées dans le modèle sont des procédures, soit des séquences d'actions stockées en mémoire. Elles sont issues d'un apprentissage, supervisé ou non, de déductions empiriques ou d'une planification consciente de la tâche. L'opérateur identifie une situation dans laquelle, pour atteindre ses objectifs, il doit appliquer une procédure qui lui est connue. Ce type de comportement renvoie par exemple à des tâches de surveillance et de contrôle. L'apparition d'un signe spécifique dans l'environnement débloque une règle stockée en mémoire et déclenche sa mise en œuvre. Rasmussen (1983) définit les signes comme des percepts physiques indiquant un état de l'environnement et servant à activer des séquences de comportements mémorisés.

Enfin, le troisième niveau de comportement est basé sur les connaissances. Ce type de comportement est mis en œuvre de manière forcément consciente et repose sur un

raisonnement. Ces comportements apparaissent lorsqu'il n'y a pas de procédure adaptée stockée en mémoire par l'opérateur. C'est ce qui permet de faire face aux situations nouvelles ou imprévues. Ce processus nécessite plus de temps et de ressources cognitives que les deux autres niveaux. Les actions basées sur les connaissances reposent sur la manipulation de symboles. Rasmussen (1983) décrit les symboles comme étant le sens qui peut être véhiculé par les signes. Un signe fait partie du monde physique alors qu'un symbole fait partie du monde sémantique (Rasmussen, 1983).

Dans ce modèle, l'opérateur effectue des boucles successives pour passer des stimuli à l'action. Pour ce faire, il peut emprunter des chemins plus ou moins longs. La boucle complète consiste à réaliser des raisonnements pour décider des actions à mettre en œuvre. Il est aussi possible pour l'opérateur de prendre des raccourcis en utilisant des procédures mémorisées (les règles) ou ses habiletés motrices. Si un opérateur apprend de nouvelles règles pour répondre à une situation donnée, cela lui permettra, lorsqu'il rencontrera cette situation, de ne plus avoir à parcourir l'ensemble de la boucle jusqu'aux comportements basés sur les connaissances. En effet, il pourra plus simplement mobiliser les procédures apprises, *i.e.* mettre en œuvre des comportements basés sur les règles, ce qui lui permettra de réduire le coût cognitif de la réalisation de la tâche.

### **1.3.2. La complémentarité homme-machine**

Dans une perspective de coopération homme-machine, le logiciel développé dans le projet MobiSketch renvoie à ce que Ferber (1999) appelait un agent cognitif téléonomique ou plus simplement un agent rationnel ou intentionnel. En effet, il est capable de tester « mentalement » plusieurs hypothèses concurrentes avant de prendre la décision d'afficher une interprétation jugée pertinente (Ghorbel, Almaksour, Lemaitre, & Anquetil, 2011; Ghorbel, Macé, Lemaitre, & Anquetil, 2011). Le type de fonctionnement du prototype permet de considérer que le système final sera en mesure d'apprendre durant le travail d'interprétation si l'utilisateur corrige les erreurs pendant le processus de reconnaissance (Ghorbel, Anquetil, & Lemaitre, 2012).

Cela implique que les erreurs, lorsqu'elles sont identifiées rapidement, permettraient l'amélioration des performances lors des reconnaissances suivantes, en évitant la propagation en cascade des erreurs. Il y a donc un intérêt à ce que l'utilisateur interagisse avec le système pendant l'analyse.

Le questionnement ergonomique du projet MobiSketch s'inscrit dans une problématique très générale liée à l'automatisation. Selon Fitts (1951), il y aurait une complémentarité fondamentale entre les humains et les machines, du fait que ces deux entités seraient toutes deux performantes dans des activités respectivement différentes. Les machines seraient notamment meilleures que les humains pour répondre rapidement à un signal, stocker de l'information brièvement et l'effacer complètement, raisonner de manière déductive ou encore effectuer plusieurs opérations complexes à la fois. À l'inverse, les humains seraient généralement meilleurs que les machines pour percevoir des configurations de lumière ou de son, improviser et utiliser des procédures flexibles, stocker des informations pour de longues périodes et les rappeler au moment approprié, raisonner de manière inductive ou encore exercer un jugement (Fitts, 1951).

La figure 3 présente différents types de tâches en fonction de leur prévisibilité et de leur degré d'automatisation. Ce graphique illustre bien la complémentarité entre l'homme et la machine décrite par Fitts (1951). Plus la tâche à réaliser est répétitive et prévisible, plus la machine est capable de la résoudre facilement. L'exemple le plus extrême d'automatisation totale donné par Sheridan et Verplank (1978) est le mécanisme d'une horloge. Un tel mécanisme est entièrement automatique et prévisible. Les « robots actuels » quant à eux traiteraient des tâches comportant une part modérée d'imprévisibilité avec l'aide d'opérateurs alors que les tâches dotées d'une très forte composante imprévisible seraient traitées par des humains. Le contrôle supervisé, noté au centre du graphique, renvoie à un niveau intermédiaire d'automatisation pour un niveau tout aussi moyen d'imprévisibilité.

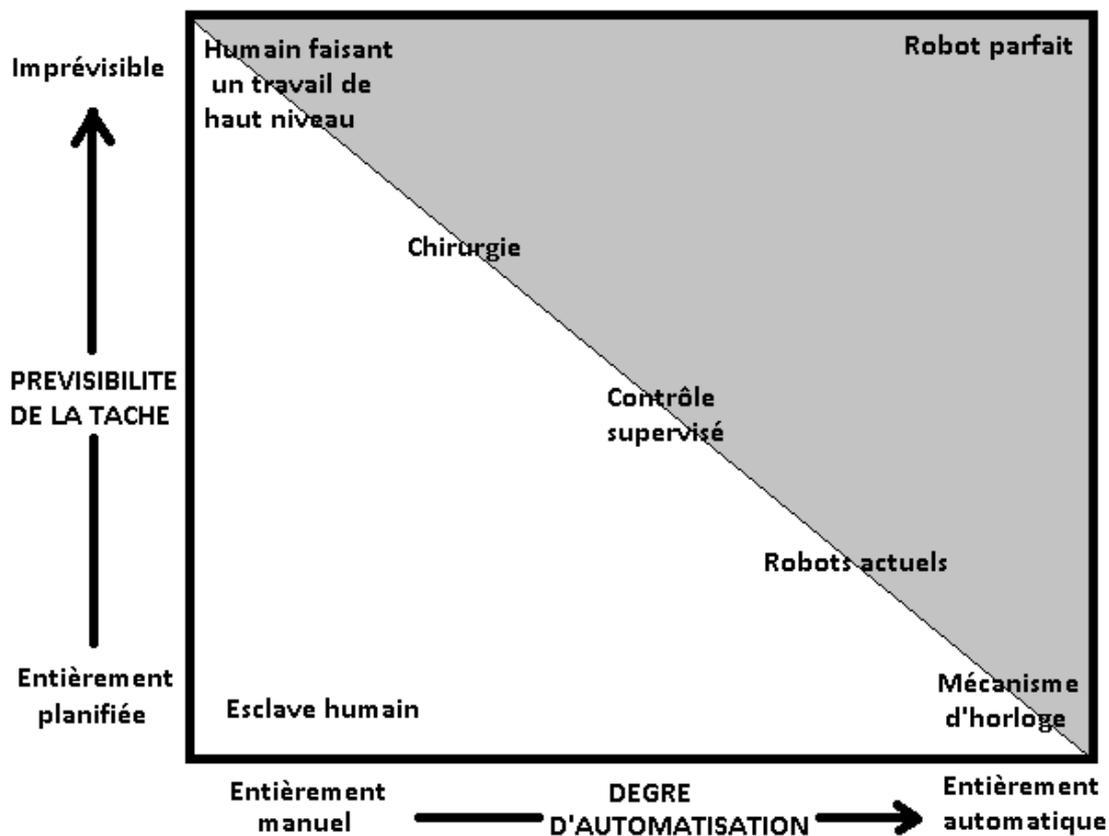


Figure 3. Type de tâche en fonction de la prévisibilité et du degré d'automatisation (d'après Sheridan & Verplank, 1978)

En termes d'imprévisibilité, la reconnaissance automatique de plan ne constitue pas une situation extrême. Cette imprévisibilité n'est en effet jamais nulle, puisqu'il n'est pas possible d'anticiper la totalité des situations que l'on peut rencontrer dans ce domaine à cause de la variété et de l'évolution des symboles utilisés d'une part et de la qualité graphique variable des plans à traiter d'autre part. À l'inverse, l'imprévisibilité n'est pas totale puisqu'il y a bien un ensemble de symboles et de configurations de symboles susceptibles d'être présents dans la majorité des cas. En se fiant à Sheridan et Verplank (1978), puisque le niveau d'imprévisibilité de la tâche est moyen, le système optimal pour la résoudre devrait être caractérisé par un degré intermédiaire d'automatisation. Un système trop automatisé risquerait d'avoir des difficultés à gérer les imprévus alors qu'un système trop manuel risquerait de rendre la tâche très

répétitive pour l'opérateur, au risque d'affecter sa vigilance et donc de dégrader ses performances (Fitts, 1951).

### **1.3.3. Les ironies de l'automatisation**

Dans un article emblématique intitulé « *Ironies of automation* », Bainbridge (1983) décrivait l'incompréhension régnant entre les concepteurs de systèmes automatiques et les opérateurs. Selon cet auteur, les concepteurs auraient tendance à percevoir les opérateurs comme des agents peu fiables et peu efficaces. Ils chercheraient donc à les remplacer par des machines dès lors qu'il est possible de le faire. Le premier constat, qualifié d'ironique par Bainbridge (1983), renvoie précisément au fait que les erreurs de conception des systèmes automatiques sont souvent des sources majeures de difficultés pour les opérateurs. Le second constat est que les concepteurs, tout en cherchant à écarter les opérateurs dès que possible, les laissent prendre en charge les tâches qu'ils ne parviennent pas à automatiser. Souvent, les tâches que les opérateurs ont à réaliser seraient donc arbitraires au regard de leurs propres capacités, car dépendantes de ce qui a pu être automatisé ou non par les concepteurs (Bainbridge, 1983).

Norman (1989) a décrit ce que l'on pourrait considérer comme une troisième ironie de l'automatisation. Selon lui, l'automatisation des activités physiques ou cognitives tend à diminuer les compétences des opérateurs. Le problème ne provient pas directement de la présence d'un automate, mais de certaines de ses caractéristiques qui, souvent, ne sont pas appropriées (Norman, 1989). Les opérations automatiques sont généralement effectuées de manière pertinente dans les conditions normales. En revanche, en situation perturbée, les feedbacks tendent à devenir inadéquats et l'humain doit alors contrôler la totalité de la conduite de la tâche. Lorsque la situation dépasse les capacités de l'automate, les feedbacks inadéquats sont à l'origine de difficultés pour l'opérateur. Ce mauvais fonctionnement des systèmes automatiques serait causé par leur niveau intermédiaire d'intelligence : suffisamment élevé pour faire ce qui devrait être fait par des humains, mais trop faible pour gérer correctement les anomalies. Il faudrait donc soit des automates plus intelligents capables de

répondre à toutes les anomalies, soit des automates moins intelligents ne travaillant que sur des tâches faciles pour des machines. Pour améliorer les performances des systèmes homme-machine, la solution consisterait donc à ajuster le niveau d'intelligence des systèmes automatiques et surtout à prendre en compte le fait que ces derniers devraient être prêts à réagir correctement en cas d'erreur en fournissant des feedbacks pertinents à l'opérateur. L'amélioration passerait ainsi par une technologie plus souple, plus flexible, capable de mieux communiquer dans les situations perturbées.

### **1.3.4. Le syndrome de l'humain hors de la boucle**

Pour faire référence aux mauvais fonctionnements de systèmes homme-machine, Endsley (1995) parlait de « syndrome de l'humain hors de la boucle ». Cette expression fait référence aux situations dans lesquelles, une activité étant automatisée, l'humain ne se trouve plus dans la boucle de contrôle au sens où il n'exerce plus de contrôle direct sur la tâche. Le syndrome de l'humain hors de la boucle a plusieurs conséquences négatives sur le fonctionnement des systèmes homme-machine (Kaber & Endsley, 1997) :

- L'opérateur ne parvient pas à percevoir les changements des paramètres du système et à intervenir lorsque c'est nécessaire (diminution de la vigilance)
- Il accorde une sur-confiance au système (complaisance)
- Il perd la conscience de la situation ou de l'état du système (Endsley, 1996)
- Il voit ses compétences de contrôle direct diminuer

Ces conséquences ont un impact sur les performances des opérateurs, à la fois en situation normale et en situation perturbée. Lorsque l'opérateur sort de la boucle de contrôle à cause de l'allocation de ses fonctions à un automate, les interactions entre l'humain et la machine sont limitées et la conscience qu'a l'opérateur de l'état du système est réduite. En situation normale, la faible conscience de l'état du système a tendance à diminuer la rapidité de réponse des opérateurs, voire à diminuer la

vigilance vis-à-vis des erreurs potentielles (Kaber & Endsley, 1997). En situation perturbée, l'opérateur devrait effectuer le travail que la machine ne parvient pas à réaliser correctement, mais il pourrait avoir perdu ses compétences du fait de la sortie de la boucle de contrôle qui le prive d'entraînement.

Selon Kaber et Endsley (1997), lorsque l'humain est extrait de la boucle de contrôle, il doit généralement effectuer une vérification des feedbacks envoyés par l'ordinateur pour s'assurer que le système ne commet pas d'erreur. Si les feedbacks indiquent la présence d'erreurs, l'opérateur va intervenir pour prendre directement le contrôle de la tâche et réaliser lui-même les traitements jusqu'à revenir à une situation normale. Il peut être amené à valider les choix de l'automate ou au contraire à les contester. En situation normale, les opérateurs ne sont donc pas impliqués dans les processus de contrôle et ce manque d'implication contribue aux erreurs humaines par la perte de compétence et de conscience de la situation qui génère des effets de surprise du fait que l'opérateur ne peut anticiper (Sarter, Woods, & Billings, 1997).

Lorsque l'humain travaille avec un système automatique, l'opérateur peut jouer trois rôles principaux (Kaber & Endsley, 1997) :

- La supervision, qui consiste à vérifier les indicateurs afin de déterminer si une intervention est nécessaire
- La surveillance, qui vise à détecter un éventuel événement critique
- Le traitement passif, qui renvoie au fait de devoir accepter ou rejeter les actions proposées par le système

Tout comme les systèmes totalement automatiques, la supervision, la surveillance et le traitement passif du système peuvent mettre les utilisateurs partiellement ou totalement en dehors de la boucle de contrôle. Le syndrome de l'humain hors de la boucle, dans ce cas, peut conduire à l'échec de l'opérateur à surveiller les changements de paramètres du système et donc à intervenir au moment opportun (Kaber & Endsley, 1997). Les autres risques évoqués sont la sur-confiance dans le système, la perte de conscience de la situation et la diminution des compétences de contrôle direct. Dans le cas spécifique de l'interprétation automatique de plans d'architecture, la perte de compétence n'est pas véritablement un risque. En effet, ce

traitement automatique ne remplace les traitements normalement réalisés par l'humain que pour certaines étapes des projets de conceptions. De ce fait, un professionnel de l'architecture maintient son entraînement dans le traitement des plans grâce à son travail lors de toutes les autres étapes. En revanche, l'interprétation automatique de plans d'architecture pourrait déclencher une perte de conscience de la situation. Un des objectifs du travail d'ergonomie de ce système est donc d'améliorer la conscience de la situation des utilisateurs.

Selon Wickens (2002), il y aurait une covariation entre la conscience de la situation et la charge mentale durant la tâche. La conscience de la situation (mesurée par le « *Situation Awareness Global Assessment Technique* » ; Endsley, 1995) et la charge mentale (mesurée par le NASA-TLX ; Hart & Staveland, 1988) seraient toutes deux impactées par les variations du niveau d'automatisation de la tâche (Kaber & Endsley, 2003). Une augmentation du niveau d'automatisation tendrait à réduire la charge mentale requise, mais à diminuer en contrepartie le niveau de conscience de la situation (Wickens, 2002). Une bonne conception du système pourrait cependant diminuer la charge mentale sans dégrader la conscience de la situation (Wickens, 2002). Cela impliquerait d'envoyer une grande quantité de feedbacks pertinents à l'utilisateur afin de maintenir, malgré l'automatisation, un niveau élevé de conscience de la situation.

### **1.3.5. Les niveaux d'automatisation**

Entre l'automatisation totale d'une tâche et sa réalisation manuelle, il existe plusieurs niveaux intermédiaires possibles. Endsley et Kaber (1999) ont proposé une typologie des niveaux possibles d'automatisation allant de l'automatisation totale d'une tâche à sa réalisation manuelle en passant par 8 niveaux intermédiaires. Les tâches que cette graduation permet de décrire auraient en commun 3 principales caractéristiques (Endsley & Kaber, 1999) :

- Plusieurs objectifs à la fois

- Plusieurs tâches à la fois, chacune avec un degré de pertinence différent pour les objectifs à atteindre
- Une forte demande dans un temps limité

Ces tâches plus ou moins automatisées reposent sur 4 sous-tâches (Endsley, 1997) : la surveillance, la génération, la sélection et l'implémentation. La surveillance (*monitoring*) consiste à vérifier les feedbacks afin de connaître l'état du système. La génération (*generating*) est une activité de formulation des stratégies possibles pour atteindre les buts. La sélection (*selecting*) renvoie à la décision d'utiliser une stratégie donnée. Enfin, l'implémentation (*implementing*) consiste à mettre en œuvre la stratégie choisie. Les niveaux d'automatisation décrits par Endsley et Kaber (1999) renvoient aux différentes configurations possibles d'allocation de ces 4 sous-tâches à l'humain ou à la machine (voir tableau 1).

Tableau 1. Les niveaux d'automatisations (Kaber & Endsley, 1997; Endsley & Kaber, 1999)

Niveau d'automatisation	Fonctions			
	Surveillance	Génération	Sélection	Implémentation
1. Contrôle manuel	Humain	Humain	Humain	Humain
2. Aide à l'action	Humain/ Ordinateur	Humain	Humain	Humain/ Ordinateur
3. Traitement par lots	Humain/ Ordinateur	Humain	Humain	Ordinateur
4. Contrôle partagé	Humain/ Ordinateur	Humain/ Ordinateur	Humain	Humain/ Ordinateur
5. Aide à la décision	Humain/ Ordinateur	Humain/ Ordinateur	Humain	Ordinateur
6. Prise de décision mixte	Humain/ Ordinateur	Humain/ Ordinateur	Humain/ Ordinateur	Ordinateur
7. Système rigide	Humain/ Ordinateur	Ordinateur	Humain	Ordinateur
8. Prise de décision automatique	Humain/ Ordinateur	Humain/ Ordinateur	Ordinateur	Ordinateur
9. Contrôle supervisé	Humain/ Ordinateur	Ordinateur	Ordinateur	Ordinateur
10. Automatisation totale	Ordinateur	Ordinateur	Ordinateur	Ordinateur

En phase de conception d'un système automatisé, la question de l'allocation des fonctions à l'humain et à la machine nécessite une approche basée à la fois sur les sciences humaines et les sciences de l'ingénieur (Morel, Chauvin, Rossi, & Berruet, 2008). En effet, non seulement les caractéristiques cognitives des opérateurs doivent être prises en compte ainsi que leur conscience de la situation et leur éventuelle perte de compétence, mais il convient aussi de s'intéresser aux caractéristiques de la machine. Par exemple, le système développé dans le cadre du projet MobiSketch est capable d'apprendre de ses erreurs au fur et à mesure de l'analyse (Ghorbel, Anquetil, & Lemaitre, 2012). Cette caractéristique a un impact sur le rôle qui pourrait être donné aux opérateurs pendant la tâche (Fleury, Ghorbel, Lemaitre, Anquetil, & Jamet, 2013).

## **1.4. Problématique de la thèse**

Dans le projet MobiSketch, le développement d'un système de reconnaissance automatique de plans d'architecture risque d'extraire l'humain de la boucle de contrôle. Le risque de perte de compétence est faible du fait que les utilisateurs continueront à traiter des plans dans les autres étapes des projets architecturaux. Cependant, il y a un risque de diminution de la conscience de la situation et de perte de vigilance. Or, du fait du risque d'erreurs d'interprétation, un utilisateur doit vérifier les interprétations pour pouvoir intervenir en cas d'erreur. Cette vérification risque donc d'être particulièrement difficile.

La problématique générale de cette thèse est la suivante : comment améliorer la performance du système homme-machine ? Cela passe par un questionnement sur une diminution des symptômes liés à la sortie de la boucle de contrôle des opérateurs. D'un point de vue plus opérationnel, on se demande comment faciliter le repérage et la correction par l'humain des erreurs éventuelles d'interprétation effectuées par la machine. Ce travail s'inscrit dans une démarche d'ergonomie de conception. Ainsi, les thématiques traitées ont été choisies en fonction des problèmes rencontrés au cours du développement. Pour répondre aux différentes questions ergonomiques qui ont émergé durant le projet, des modèles théoriques issus de la psychologie cognitive et

de l'ergonomie ont été mobilisés. Pour cette raison, ce manuscrit est constitué selon une structure particulière. Chaque chapitre correspond à une étude, *i.e.* à un questionnement en termes d'application, une revue de la littérature et la description des expériences, tests ou enquêtes menés pour répondre aux interrogations ne trouvant pas de réponse dans les théories existantes. Ce choix de structure a été jugé préférable à un plan plus classique afin de mieux refléter le fait que ce travail a d'abord été guidé par les questions appliquées sous-tendant le développement du système d'interprétation automatique de plans. Les modèles théoriques sont présentés au fur et à mesure des descriptions des expériences car ils ont effectivement été mobilisés progressivement au cours de l'avancée du projet.

# **Étude 1 : Analyse des activités de complétion, recopie et correction de plans d'architecture**

## **1. INTRODUCTION**

L'objectif de cette première expérience était de délimiter le terrain d'étude qui allait être celui de ce travail de thèse. MobiSketch consiste à concevoir un système innovant doté de nombreuses fonctionnalités qui peuvent soulever un grand nombre de questionnements ergonomiques. Par exemple, ces questions pourraient concerner le transfert des compétences en dessin du papier à la tablette (pour la composition à la volée), l'édition manuscrite d'un plan numérique en situation de mobilité (pour compléter ou corriger un plan) ou encore la correction des erreurs d'interprétation automatique. L'étude présentée ici avait pour but de conduire des participants à réaliser différentes tâches sur des plans d'architecture afin d'observer et analyser leurs comportements, mais aussi leurs éventuelles difficultés, ainsi que l'intervention éventuelle de caractéristiques cognitives spécifiques au traitement des informations visuo-spatiales. Cela devait permettre d'orienter les travaux suivants sur les types de tâches qui posaient problème et de prendre en compte les caractéristiques cognitives impliquées.

### **1.1. Traitement des plans**

Les schémas et les plans en psychologie cognitive ont été étudiés de diverses manières. Ils peuvent être des représentations sémantiques (Mani & Johnson-Laird,

1982), des produits d'activités de copie (Budd, Houtz, & Lambert, 2008; Maycock, Liu, & Klein, 2009), une aide à l'apprentissage de cours (Butcher, 2006) ou plus généralement des supports externes allégeant les traitements cognitifs (Larkin & Simon, 1987; Suwa & Tversky, 2002). Certains travaux dans le domaine de l'architecturologie tentent d'apporter des éléments de compréhension des mécanismes cognitifs impliqués dans la conception de projets architecturaux (e.g. Raynaud, 1999 ; Lecourtois, 2005). Cette thématique est généralement abordée sous l'angle de l'usage des plans papier par les professionnels. Les architectes sont capables de concevoir un projet sans passer par la réalisation d'esquisses préalables (Bilda, Gero, & Purcell, 2006; Bilda & Gero, 2006) mais, malgré cela, ils passent presque systématiquement par cette étape car le papier demeure un des instruments de la conception architecturale (Unwin, 2007; Manolopoulou, 2005). Ces esquisses réalisées au début, puis au fur et à mesure de la conception du projet, constituent des représentations externes cristallisant les idées et permettant ainsi d'alléger les traitements cognitifs visuo-spatiaux (Bafna, 2008; Suwa & Tversky, 2002; Tversky & Suwa, 2009). Les architectes utilisent le dessin dans la conception pour externaliser leurs idées dans une optique de communication (avec les collègues, les clients...), mais aussi pour les sauvegarder et les utiliser comme support au raisonnement (Do & Gross, 1996, 2001; Do, 2002).

En plus de son utilité pratique, la réalisation d'esquisses manuscrites sur papier est une pratique valorisée. Le dessin manuscrit de l'architecte lui permet une certaine expression de son style personnel là où l'outil informatique tend à uniformiser les détails (Senyapili & Basa, 2006). Réalisé à la main, le dessin sur papier a un caractère unique et même une dimension artistique qu'un logiciel de DAO gomme pour ne laisser apparaître que les aspects techniques. De plus, dans une enquête réalisée dans une école d'architecture américaine, 85% des étudiants et une écrasante majorité des enseignants interrogés estimaient que les compétences de conception sont directement liées aux compétences de dessin (Lyn & Dulaney, 2009). Si l'usage du dessin manuscrit reste très commun dans la conception de projets architecturaux, les pratiques peuvent tout de même évoluer du fait des nouvelles possibilités techniques.

Ainsi, les professionnels pourraient être amenés à dessiner directement sur des tablettes tactiles, à rétro-convertir des plans (manuscrits ou non) ou encore à éditer des plans existants dans un environnement numérique. La diffusion de ces nouveaux outils passe notamment par une meilleure connaissance des composantes cognitives impliquées dans le traitement des plans d'architecture. Une recherche de Gobert (1999) indique que les composantes impliquées dépendent du type de tâche effectuée avec les plans. Cette expérience apporte des informations concernant la compréhension des plans d'architecture et leur visualisation en 3D. En revanche, il n'existe pas à notre connaissance d'étude décrivant les traitements cognitifs impliqués dans la complétion de plans, leur dessin ou leur correction.

## **1.2. Production écrite et traitement des plans**

En l'absence de modèle de la production de plan, une première étape de réflexion a consisté à découper cette activité en plusieurs types de sous-tâches en s'inspirant de la structure des modèles de la production de texte (voir Chanquoy & Alamargot, 2002 pour une revue). Le premier modèle de la production de texte a été élaboré par Hayes et Flower (1980; Flower & Hayes, 1981). Il tentait de rendre compte des processus impliqués dans l'écriture en lien avec d'une part, la mémoire à long terme et d'autre part, l'environnement de la tâche. La mémoire à long terme constitue la base de connaissances dans laquelle l'individu va récupérer les informations utiles en fonction de ses objectifs. Les connaissances ainsi rapatriées sont ensuite organisées et des buts sont déterminés. L'étape suivante est la mise en texte. Il s'agit de l'écriture à proprement parler et de ce qu'elle implique en termes de construction syntaxique et rhétorique, ainsi que de choix lexicaux. Vient enfin l'étape de révision qui consiste à relire le texte déjà écrit, pour éventuellement effectuer des modifications ou des corrections sur celui-ci.

Après la publication de ce modèle, les recherches portant sur la compréhension et la rédaction de textes se sont scindées en deux approches principales (Olive & Piolat, 2005; Torrance & Galbraith, 2006). La première approche est capacitaire. Elle est basée sur une vision de la mémoire de travail comme un réservoir unique de ressources (*e.g.*

Just & Carpenter, 1992). Selon cette approche, un enfant qui maîtriserait peu les processus d'écriture mobiliserait plus de ressources pour les mettre en œuvre. Ainsi, les performances de rédaction de texte dépendraient de l'automatisation des processus permettant d'alléger le coût des traitements (McCutchen, 1996). La seconde approche, initiée par Kellogg (1996), prend appui sur le modèle de la mémoire de travail élaboré par Baddeley (1986). Dans une perspective modulaire, Kellogg (1996) ou encore Swanson et Berninger (1996) ont proposé un modèle mettant en lien les processus de rédaction de textes avec les différentes instances de la mémoire de travail (Baddeley, 1986). Selon ce modèle, l'activité de production de texte est composée de 3 processus, eux-mêmes composés chacun de deux sous-processus : la formulation, l'exécution et le contrôle. La formulation consiste à planifier les idées à formuler et ensuite à les traduire en phrases. L'exécution correspond à la programmation de l'écriture et à sa réalisation motrice. Enfin, le contrôle renvoie à la relecture du texte rédigé et à son édition éventuelle. Selon ce modèle, les composantes de la mémoire de travail interviendraient de manière différenciée selon les étapes de traitement (voir figure 4). L'étape de formulation solliciterait à la fois le calepin visuo-spatial, l'administrateur central et la boucle phonologique ; l'exécution ne mettrait en jeu que l'administrateur central ; le contrôle ferait appel à l'administrateur central et à la boucle phonologique (Kellogg, 1996).

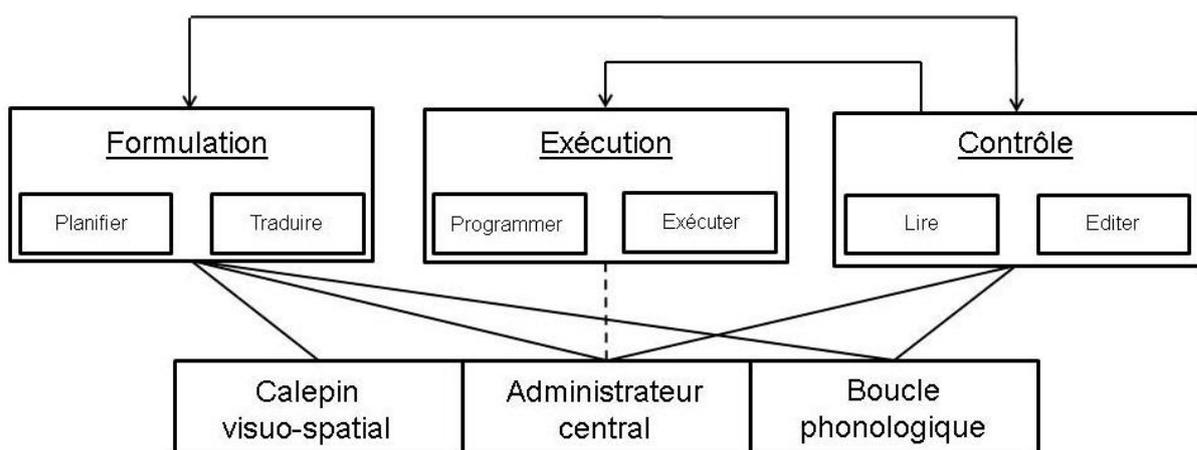


Figure 4. Production de textes et mémoire de travail (Kellogg, 1996)

Les structures en trois étapes des modèles de Hayes et Flower (1980) et de Kellogg (1996) sont relativement proches et il est possible d'imaginer des étapes similaires dans la production de plans. En s'inspirant de ce modèle, on peut penser que le traitement des plans d'architecture pourrait, pour en faciliter l'étude, être découpé en trois activités différentes : la traduction d'un bâtiment réel en plan en deux dimensions (étape de formulation constituée d'une sous étape de transformation des formats des informations), la recopie à l'identique d'un plan en deux dimensions (étape d'exécution) et la correction d'un plan selon un modèle (étape de contrôle). Le matériel traité dans ces différentes tâches est de type visuo-spatial. Les 3 activités susmentionnées devraient donc principalement solliciter ce que Baddeley (1986) a appelé le calepin visuo-spatial.

### **1.3. Traitement des informations visuo-spatiales**

Le modèle de la mémoire de travail de Baddeley (1986, 1992) permet de faire des hypothèses sur la nature des processus mnésiques sollicités dans les traitements cognitifs effectués sur des plans d'architecture. Selon ce modèle, le traitement des informations verbales et des informations visuo-spatiales est effectué par deux structures distinctes : la boucle phonologique et le calepin visuo-spatial. Pour autant, cela ne signifie pas que la boucle phonologique ne puisse pas, dans certains cas, être partiellement recrutée pour une tâche visuo-spatiale ou inversement (Dunlap, 1997). Le calepin visuo-spatial est lui-même subdivisé en plusieurs sous-composantes. Cependant, la caractérisation de ces composantes n'est pas établie de manière claire. Le premier découpage proposé reposait sur la nature des stimuli : une composante traiterait les stimuli visuels et l'autre les stimuli spatiaux (Baddeley, 1992; Logie, 1995; Hecker & Mapperson, 1997). Comme la boucle phonologique, le calepin visuo-spatial serait constitué d'un cache passif distinct d'un système actif de traitement (Logie, 1986; Frick & De Rose, 1986). C'est la composante spatiale qui est décrite comme active, *i.e.* utilisée consciemment et volontairement. La composante visuelle serait

quant à elle passive, fonctionnant de manière automatique (Logie, 1995). Une étude de Bruyer et Scailquin (1998) a contribué à appuyer ce découpage. Dans cette expérience, une tâche concurrente « auditivo-spatiale » (déterminer si un son provient de la gauche ou de la droite) effectuée par les participants interférait avec une tâche principale de génération ou de rotation d'images, mais pas avec une tâche de maintien d'images en mémoire. Cette configuration de résultats corrobore, selon les auteurs, l'équivalence de la distinction entre visuel et spatial avec la distinction entre passif et actif. Selon Bruyer et Scailquin (1998), la génération ou la rotation d'images nécessite la mobilisation de ressources attentionnelles et donc la composante active du calepin visuo-spatial. En revanche, le maintien en mémoire d'une image peut être effectué par sa composante passive.

Plus récemment, Vergauwe, Barrouiller et Camos (2009) ont mis en évidence une interférence entre une tâche de traitement et une tâche de stockage, y compris lorsque ces tâches renvoient l'une à un contenu spatial et l'autre à un contenu visuel. Les auteurs en concluent que les composantes visuelle et spatiale sollicitent une part de ressources communes tout comme les processus de traitement et de stockage.

Plutôt que cette division du calepin visuo-spatial en composantes visuelle et spatiale, Pickering, Gathercole, Melanie et Lloyd (2001) proposaient de le scinder en fonction du caractère dynamique ou statique de l'information traitée. De nombreuses recherches ont d'ailleurs montré que le traitement de stimuli visuels présentés de manière séquentielle interfère avec le traitement de stimuli spatiaux mais pas avec des stimuli visuels statiques (Pearson & Sahraie, 2003; Quinn, 1994; Smyth, 1996; Quinn & Ralston, 1986; Smyth, Pearson, & Pendleton, 1988). Malgré l'importante proximité fonctionnelle entre la composante spatiale et les mécanismes de traitements de séquences, ils ne se confondent pas totalement (Smyth & Pendleton, 1989). D'autres auteurs ont montré une association entre la simultanéité de la présentation des stimuli visuels et une sollicitation exclusive de la composante visuelle (Della Sala, Gray, Baddeley, Allamano, & Wilson, 1999; Darling, Della Sala, Logie, & Cantagallo, 2006). Ils parlent alors de composante visuelle (*visual component*) que Logie (1995) définit comme un système passif permettant l'encodage et le stockage d'informations

visuelles et de composante spatio-séquentielle (*spatio-sequential components*) décrite comme un système actif d'autorépétition permettant de stocker des séquences de mouvements et de transformer des images mentales (Della Sala *et al.*, 1999; Darling *et al.*, 2006; Frenkel & Bourdin, 2009). Baddeley (2002) précise de plus que le calepin visuo-spatial sert non seulement au traitement et au maintien des informations visuo-spatiales, mais il permet également d'intégrer les informations visuelles avec les informations spatiales.

Par conséquent, le calepin visuo-spatial est vraisemblablement mobilisé dans les tâches de traitement de plans d'architecture. Il reste à déterminer de quelle manière il intervient dans différents types de tâches s'appuyant sur des plans et dans quelle mesure ces interventions sont cruciales pour la réussite de ces tâches. L'expérience 1 décrite ici présentait deux objectifs complémentaires : le premier était d'apporter un éclairage général sur les difficultés éventuelles rencontrées par des personnes manipulant des plans en fonction du type de tâche réalisée. Cette expérience avait donc un caractère exploratoire. Le second objectif était d'évaluer le rôle des deux sous-composantes du calepin visuo-spatial dans ces activités de manipulation de plans afin de pouvoir en tenir compte de manière pertinente lors des études suivantes.

## **2. METHODE**

### **2.1. Participants**

Les participants à cette expérience étaient 40 volontaires recrutés à partir de la base de données des volontaires du Laboratoire d'Observation des USages des Technologies de l'Information et de la Communication (LOUSTIC), novices dans la manipulation de plans d'architecture. Le participant le plus jeune avait 18 ans et 5 mois, le plus âgé avait 30 ans et 4 mois, pour une moyenne d'âge de 24 ans ( $SD = 2,71$ ). Cet échantillon était composé de 14 jeunes actifs et de 26 étudiants Rennais provenant de disciplines variées.

## 2.2. Matériel

L'ensemble des tâches était effectué à l'aide d'un stylet sur un ordinateur portable hybride HP TouchSmart PC en position tablet PC. Les plans ont été réalisés par les participants sur une interface de dessin volontairement sommaire en termes de fonctionnalités permettant de tracer des traits noirs au crayon et d'effacer en cas d'erreur. Elle permettait aussi, lorsque les participants validaient une production, de faire apparaître automatiquement une nouvelle image préalablement sélectionnée par l'expérimentateur. Cette fonctionnalité spécifique permettait de simuler une fonction que ne possède pas réellement cette interface : la rétro-conversion d'un plan manuscrit. Cette fonctionnalité est simulée par l'apparition à l'écran, à la place du plan dessiné par le participant, d'un plan finalisé (voir Figure 5).

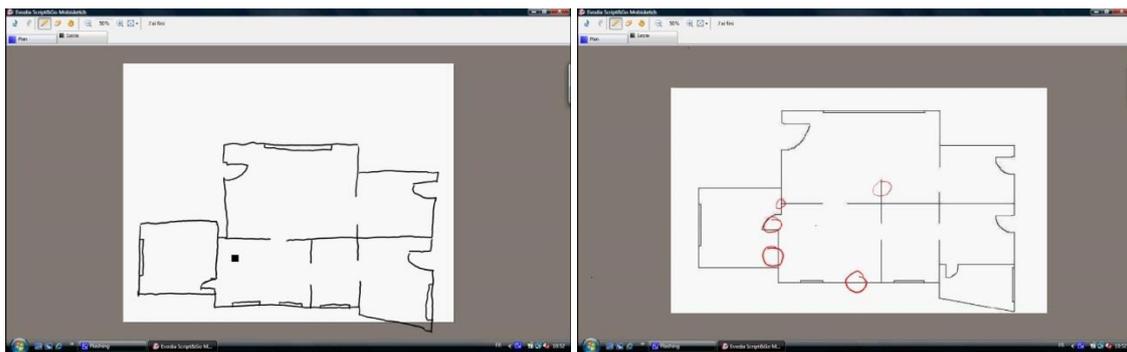


Figure 5. Captures d'écran du plan dessiné par un participant (à gauche) et du plan « interprété » (à droite) sur lequel le participant a entouré les erreurs

Dans cette expérience, une version numérique de l'épreuve des blocs de Corsi a été utilisée afin de mesurer l'empan spatial des participants (Jones, Farrand, Stuart, & Morris, 1995). Il s'agit d'une planche sur laquelle sont disposés neuf carrés. Ces carrés sont désignés successivement dans un ordre donné et le participant doit mémoriser cet ordre pour le reproduire. Le nombre de localisations qu'il faut mémoriser augmente au fur et à mesure du déroulement du test et ce dernier est terminé lorsque le participant échoue deux fois de suite à reproduire une séquence d'une longueur donnée (Berch, Krikorian, & Huba, 1998). L'empan spatial est alors le nombre maximum de carrés retrouvés en un seul essai. Dans cette expérience, l'épreuve des

blocs de Corsi était administrée sur le Tablet PC. Les séquences étaient donc présentées sous la forme de changements de couleur successifs des cases à l'écran, au rythme d'une case par seconde. La désignation de la séquence par le participant se faisait de manière tactile.

Le test des motifs visuels, autrement appelé « épreuve des grilles de Wilson », était aussi utilisé dans cette expérience dans le but de fournir une mesure de l'empan visuel des participants (Della Sala *et al.*, 1999; Darling *et al.*, 2006). Ce test est constitué de grilles dans lesquelles une case sur deux est noircie. Les participants doivent mémoriser les emplacements des cases noircies et doivent ensuite les reproduire sur une grille vierge (voir figure 6). À chaque essai réussi, le nombre de cases noircies augmente de un. L'empan visuel correspond au nombre maximum de cases que le participant parvient à replacer correctement sur la grille.

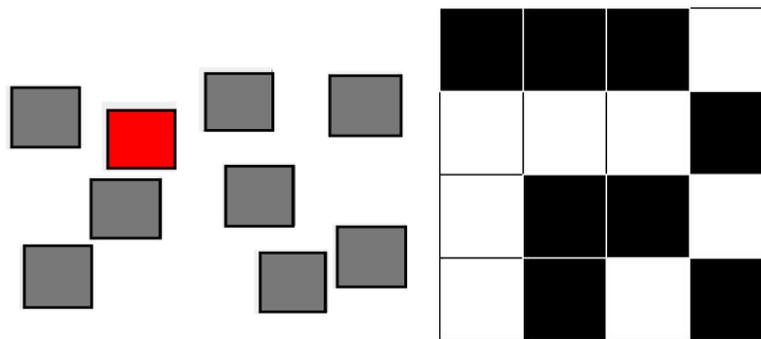


Figure 6. Captures d'écran des épreuves électroniques des blocs de Corsi (à gauche) et des motifs visuels (à droite)

## 2.3. Procédure

Les participants commençaient par passer l'épreuve des blocs de Corsi ainsi que le test de Wilson. Dans un second temps, ils devaient travailler sur des plans et ce, dans trois exercices successifs. Dans le premier exercice, l'expérimentateur présentait aux participants le Tablet PC sur lequel était affiché le plan incomplet du LOUSTIC dans l'interface de dessin. La pièce dans laquelle les participants se trouvaient pendant l'expérience avait été effacée du plan. Leur objectif consistait à la redessiner à la main. Pour le deuxième exercice, l'expérimentateur montrait aux participants un modèle de

plan imprimé sur une feuille de papier. Les participants avaient alors pour objectif de reproduire ce plan dans l'interface de dessin. L'expérimentateur expliquait aux participants que le plan qu'ils réalisaient à main levée allait ensuite être interprété puis converti par le logiciel pour aboutir à un plan finalisé. Par exemple, les lignes tracées à main levée allaient être transformées en ligne droite. Cela signifiait que dans le cas idéal, le résultat de la conversion allait être exactement équivalent au modèle. En réalité, cette explication était fautive. Le plan finalisé présenté aux participants était préparé à l'avance et demeurait identique pour tout le monde. Il comportait quatre erreurs que les participants devaient entourer (voir figure 7, les erreurs sont ici entourées en rouge). Pour ce faire, ils comparaient le plan « converti » sur Tablet PC au modèle sur papier (conservé entre temps).

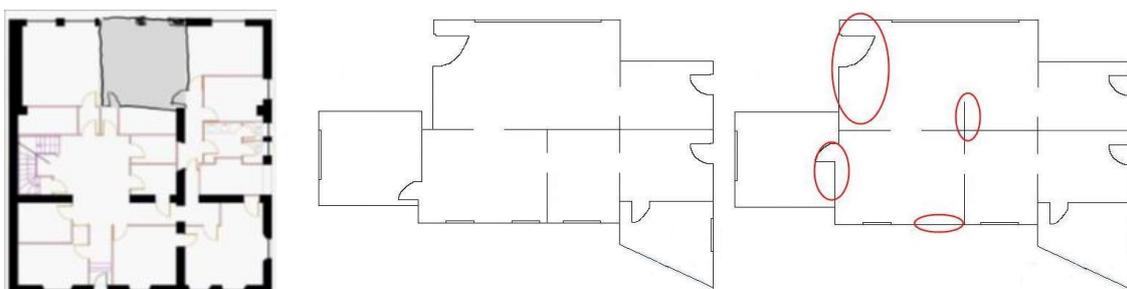


Figure 7. Plan du LOUSTIC complété par un participant (gauche), modèle sur papier (centre) et plan « converti » (droite)

## 2.4. Mesures

L'épreuve des blocs de Corsi et le test de Wilson permettaient d'attribuer à chaque participant un score d'empan spatial et un score d'empan visuel. Pendant les trois tâches portant sur les plans (complétion, copie et comparaison), plusieurs mesures étaient effectuées. Chacune des trois tâches donnait lieu à des productions manuscrites (voir figure 8). Des critères de qualité ont été définis dans le but d'obtenir un score par participant et par tâche. Ainsi, la complétion du plan de LOUSTIC est notée sur 4 points, un point étant attribué pour chaque élément (porte ou fenêtre) placé sur le bon mur. La tâche de repérage des erreurs était notée sur quatre points aussi, un par erreur repérée. Enfin, l'épreuve de recopie de plan était quant à elle

notée sur 22 points car cette notation prend en compte non seulement l'emplacement des différents ouvrants (5 portes, 6 fenêtres et 4 ouvertures intérieures), mais aussi l'agencement des pièces et les longueurs relatives des fenêtres.

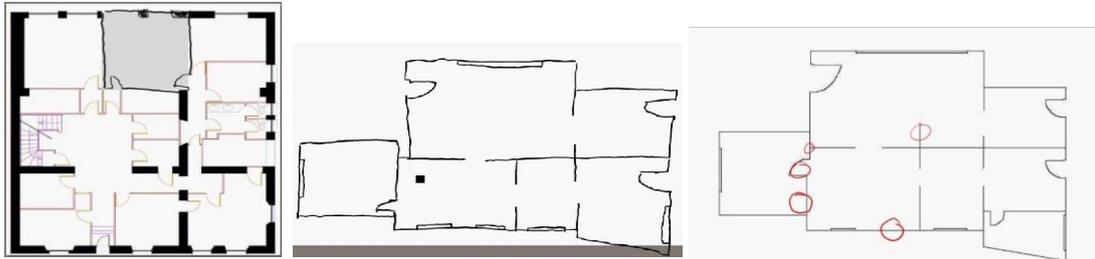


Figure 8. Exemples de productions des participants pour les épreuves de complétion (gauche), de recopie (centre) et de correction (droite)

Durant la réalisation des trois tâches liées aux plans, les participants portaient des lunettes caméra. La particularité de ces lunettes est de pouvoir filmer grâce à une caméra miniature dissimulée dans la monture. Cet outil permettait de mesurer le nombre d'allers et retours effectués entre le modèle et la production en cours. Afin de ne pas influencer les mouvements de tête des participants, l'existence de cette caméra leur a été cachée en leur disant que ces lunettes servaient à mesurer en temps réel leurs diamètres pupillaires pendant la tâche. Par ailleurs, un enregistrement vidéo de l'écran, grâce au logiciel de capture CamStudio, était effectué afin de connaître les temps de réalisation des différentes tâches.

### 3. RESULTATS

Concernant chacune des trois épreuves (complétion, recopie et correction), les données présentées ici sont la performance évaluée par un score, la durée de la tâche et le nombre d'allers et retours visuels. Les liens éventuels entre ces variables ont été évalués par le biais de corrélations.

### 3.1. Epreuve de complétion de plan

La performance à la tâche de complétion de plan était mesurée par un score allant de 0 à 4 correspondant au nombre d'ouvrants correctement placés sur le plan. La distribution des scores montre que cette tâche était plutôt facile pour l'échantillon testé puisque sur 40 participants, 32 ont le score maximal ( $M = 3,675$ ,  $SD = 0,52$  ; voir figure 9). L'analyse corrélacionnelle n'a pas permis d'établir de lien significatif entre le score de complétion et la durée de la tâche ( $r(38) = 0,029$ ,  $p = 0,872$ ) ni entre le score de complétion et le nombre d'allers et retours visuels entre l'environnement et le plan à compléter ( $r(38) = 0,055$ ,  $p = 0,763$ ).

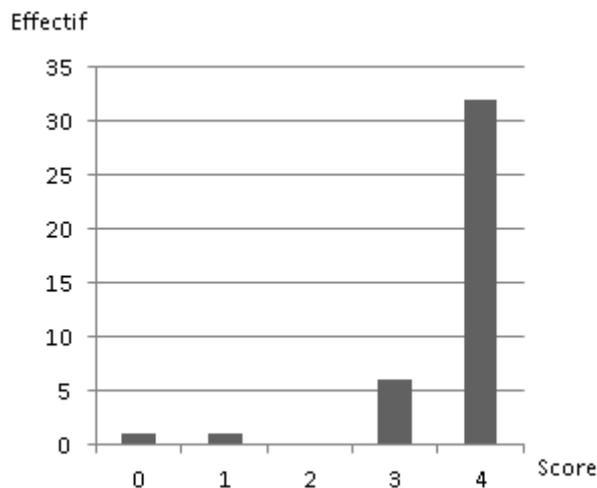


Figure 9. Effectif (en ordonnée) en fonction du score à la tâche de complétion (en abscisse)

### 3.2. Epreuve de recopie de plan

La performance à la tâche de recopie de plan était mesurée par un score allant de 0 à 22 correspondant au score de qualité du plan recopié. La distribution de ces scores indique que cette tâche, comme la tâche de complétion, était plutôt facile pour l'échantillon testé ( $M = 21,15$ ,  $SD = 0,638$  ; voir figure 10). L'analyse des données par corrélacions n'a pas permis d'établir de lien significatif entre le score de recopie et la durée de la tâche ( $r(38) = 0,066$ ,  $p = 0,695$ ) ni entre le score de recopie et le nombre

d'allers et retours visuels entre l'environnement et le plan à compléter ( $r(38) = 0,124, p = 0,458$ ).

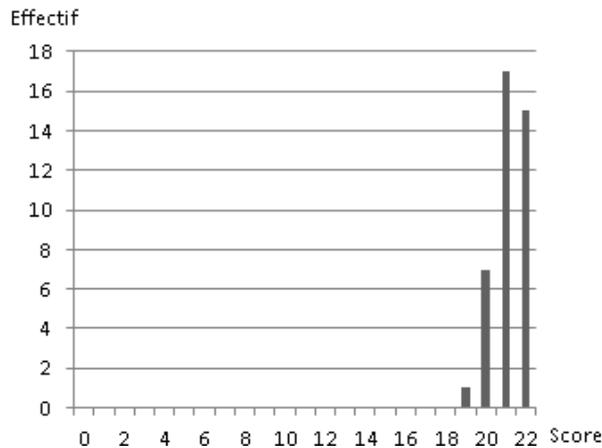


Figure 10. Effectif (en ordonnée) en fonction du score à l'épreuve de recopie (en abscisse)

### 3.3. Epreuve de correction de plan

La performance à la tâche de correction de plan était mesurée par un score allant de 0 à 4 renvoyant au nombre d'erreurs entourées. Contrairement aux résultats des deux épreuves précédentes, la distribution des scores de correction (voir figure 11) semble indiquer que cette tâche en apparence très simple a tout de même constitué une difficulté pour un nombre important de participants ( $M = 2,825, SD = 0,895$ ).

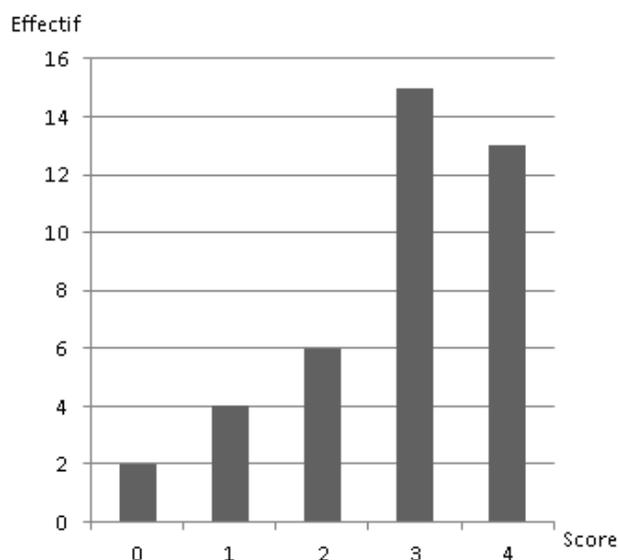


Figure 11. Effectif (en ordonnée) en fonction du score à l'épreuve de correction (en abscisse)

L'analyse corrélacionnelle n'a pas permis d'établir de lien significatif entre le score de correction et la durée de la tâche ( $r(38) = 0,253, p = 0,115$ ), mais a révélé en revanche une corrélation positive significative entre le score de correction et le nombre d'allers et retours visuels entre les deux plans ( $r(38) = 0,351, p = 0,026$  ; voir figure 12). Plus les participants faisaient d'allers et retours et plus ils repéraient d'erreurs.

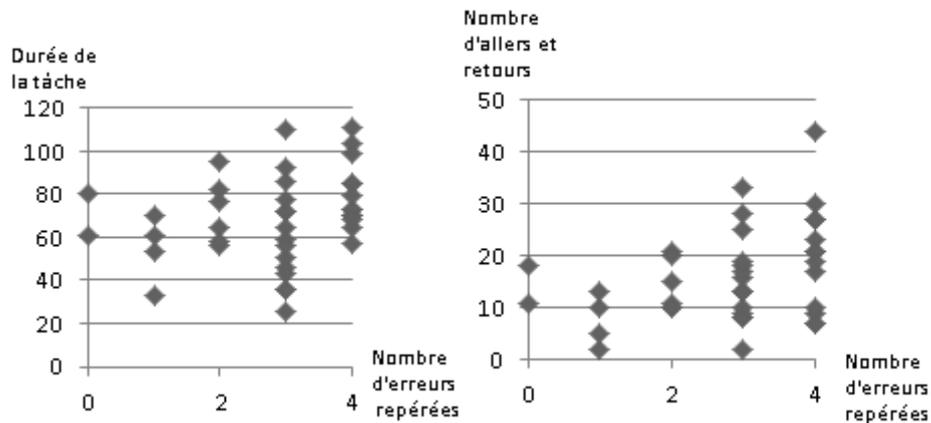


Figure 12. Durée de la tâche (à gauche) et nombre d'allers et retours modèle/cible (à droite) en fonction du nombre d'erreurs repérées (en abscisse)

### 3.4. Tests d'empan

Pour déterminer l'intervention éventuelle des composantes visuelle et spatiale du calepin visuo-spatial, des corrélations ont été calculées entre les scores aux tests d'empan (blocs de Corsi et grilles de Wilson) et les scores aux trois tâches, mais aussi entre les nombres d'allers et retours visuels et les durées des trois épreuves de manipulation de plans. Sur les 18 corrélations testées (voir annexes 1, 2 et 3), seule la corrélation entre le score à l'épreuve de recopie et l'empan visuel est statistiquement significative ( $r(38) = 0,393, p = 0,012$ ). Les participants dont l'empan visuel était élevé produisaient un plan plus ressemblant au modèle que ceux dont l'empan visuel était plus faible.

## 4. DISCUSSION

Cette première étude avait deux objectifs principaux. Le premier était d'observer les activités de complétion, de recopie et de correction de plans afin de repérer d'éventuelles difficultés spécifiques. Le second objectif de cette étude était de vérifier si les composantes visuelle et spatiale du calepin visuo-spatial avaient un rôle décisif dans le niveau de performance pour ces trois tâches. De manière étonnante, la tâche de correction, qui pouvait sembler la plus facile au premier abord, s'avère être au contraire celle qui génère le plus de difficultés. Le repérage des différences sur un plan pourtant très simple n'a été réussi que par 13 participants sur les 40 testés, soit 32,5% d'entre eux. Par ailleurs, l'analyse statistique révèle une corrélation positive et significative entre le nombre d'allers et retours visuels effectués entre le modèle et le plan à vérifier et le nombre d'erreurs repérées. Ainsi, faire de nombreux allers et retours visuels entre les deux sources à comparer permettrait donc, dans une certaine mesure, d'améliorer les performances pour cette tâche. Ce résultat laisse penser que la difficulté de cette tâche de repérage des erreurs pourrait être liée à la distance entre les deux sources visuelles. Cette distance conduit les participants à devoir maintenir en mémoire une ou plusieurs informations d'un plan et à déplacer leur attention vers le second plan pour comparer les informations de la zone correspondante sur celui-ci. Le résultat indiquant que les participants faisant plus d'allers et retours ont de meilleures performances suggère que ceux-ci pourraient mémoriser une plus faible quantité d'informations. Par exemple, un participant donné peut contrôler une porte et une fenêtre en une seule fois alors qu'un autre peut faire un aller et retour pour la porte et un second pour la fenêtre. Or l'oubli des informations en mémoires à court terme tend à apparaître d'autant plus rapidement que le nombre de ces informations est important (Melton, 1963).

Concernant les composantes visuelles et spatiales, la corrélation positive observée entre les empan visuels et les scores de qualité des plans recopiés suggère l'implication de la composante visuelle du calepin visuo-spatial dans la tâche de recopie de plan. Ce résultat semble cohérent puisque la tâche de recopie consiste à

stocker une image mentale du modèle dans le cache visuel pour le reproduire sur la feuille de dessin. Ce résultat rejoint également les interprétations de Bruyer et Scailquin (1998) selon lesquelles seul le maintien d'images en mémoire, sans transformation, sollicite la composante visuelle sans solliciter la composante spatiale. La recopie n'implique pas d'effectuer des manipulations mentales sur l'image car cette dernière est reproduite à l'identique, c'est-à-dire en conservant sa taille et son orientation. Ainsi, la composante spatiale du calepin n'est pas ou peu sollicitée. L'absence de corrélation des emports mesurés avec les performances aux tâches de complétion et de correction ne prouve pas que ces tâches ne sollicitent pas le calepin visuo-spatial des participants. Cela indique néanmoins que si cette composante intervient dans ces tâches, elle n'a pas un rôle particulièrement crucial et n'impacte pas de manière sensible les performances.

La difficulté dans l'activité de correction étant détectée, il convient de la caractériser afin de trouver des moyens de la maîtriser. Des tests sur le prototype de logiciel de rétro-conversion ont montré un taux d'erreur d'interprétation de 9% sur des plans simples (Ghorbel, Lemaitre, & Anquetil, 2012). Bien que les performances du système soient susceptibles d'évoluer, il existe des plans nettement plus complexes et une variété non définie des symboles. Le fait que les plans à rétro-converter soient plutôt issus du format papier implique des dégradations partielles éventuelles ou des dessins manuscrits de qualité variable qui peuvent augmenter la difficulté de la reconnaissance. Dans ces conditions, le risque d'erreur d'interprétation est donc toujours présent. Il est donc indispensable qu'un humain vérifie l'interprétation, repère les erreurs et les corrige. La tâche que doit remplir sans faute cet opérateur consiste alors à comparer deux types d'informations visuelles : le document initial et son interprétation. Il est ainsi nécessaire de répondre à la problématique suivante : comment présenter l'information rétro-convertie à l'opérateur de manière à ce qu'il soit le plus à-même de corriger toutes les erreurs ?

# **Étude 2 : Comment présenter l'interprétation aux utilisateurs ? Effet de l'intégration spatiale et de la séquentialité<sup>2</sup>**

## **1. INTRODUCTION**

L'expérience décrite dans la partie précédente a montré que la correction de plans constituait une difficulté importante puisque dans cette tâche en apparence très simple, seuls 32,5% des participants sont parvenus à repérer toutes les erreurs. En termes d'application, ce taux de réussite observé est loin d'être satisfaisant.

Les deux expériences présentées dans cette seconde étude avaient pour objectif de répondre à deux principales interrogations. La première consistait à se demander comment présenter l'information interprétée aux utilisateurs de manière à ce qu'ils soient les plus à-même de repérer toutes les erreurs. Dans les premiers prototypes, l'interprétation était présentée à côté du plan. Un format alternatif pourrait consister à présenter l'interprétation superposée au plan. La réponse à cette question passe par une analyse de la littérature concernant le traitement de deux sources visuelles et par l'expérimentation. La seconde question était la suivante : est-il préférable de montrer la reconnaissance automatique en temps réel plutôt que de la cacher ? L'affichage de la reconnaissance en temps réel irait dans le sens d'une amélioration de la conscience de la situation pour les utilisateurs (Endsley, 1995) et s'inscrit donc dans une logique

---

<sup>2</sup> Fleury, S., & Jamet, E. (in press). Facilitating the comparison of multiple visual items on screen : the example of electronic architectural plan correction. *Applied Ergonomics*.

d'insertion des utilisateurs dans la boucle de contrôle. Ce choix de conception pourrait avoir un impact sur l'exploration visuelle de l'interprétation présentée et donc sur la performance des utilisateurs dans la détection des erreurs.

## **1.1. Traitement de deux sources visuelles**

Les premières recherches durant lesquelles des participants furent amenés à comparer des images pour en extraire les différences et les similitudes portaient sur le repérage de concept grâce au traitement d'exemples visuels simples (Cahill & Hovland, 1960 ; Hovland & Weiss, 1953). Dans l'étude de Cahill et Hovland (1960), un concept renvoie à un ensemble d'attributs géométriques spécifiques. Les participants devaient déduire le concept à partir d'exemples négatifs. Les exemples négatifs sont des stimuli visuels pour lesquels au moins un attribut du concept est absent. Cette tâche nécessitait donc de traiter et de mettre en lien plusieurs informations visuelles. Les résultats de l'expérience de Cahill et Hovland (1960) suggèrent que l'extraction du concept à partir d'un ensemble d'images est plus facile lorsque ces images sont présentées simultanément que lorsqu'elles sont présentées successivement. En effet, la présentation successive induit une nécessité de maintien plus long en mémoire des images précédemment consultées qu'en situation de présentation simultanée. Plus récemment, Kim et Astion (2000) ont reproduit le même type d'expérience en testant trois formats de présentation des images : dans le format successif, les images étaient présentées une à une ; dans le format simultané, la présentation des images s'effectuait par paires, chaque paire étant différente de la précédente ; dans le format ancré, les images étaient aussi présentées par paires, mais une des deux images restait fixe alors que la seconde changeait. Les auteurs n'ont pas seulement confirmé la supériorité de la présentation simultanée par rapport à la présentation successive des images, mais ont également observé que le maintien à l'écran d'une même image, plutôt que le changement systématique des deux, facilitait aussi la tâche puisque l'image déjà connue n'avait pas à être ré-explorée pour être comparée à la seconde image. Les auteurs expliquent les résultats en invoquant des différences de niveau de

sollicitation des capacités mnésiques sans entrer plus dans les détails (Cahill & Hovland, 1960; Kim & Astion, 2000).

Dans le domaine de l'ergonomie, le traitement de deux sources visuelles mutuellement référencées a été traité par de nombreuses études et selon des approches variées. Les trois principales approches vont être présentées : le principe ergonomique de proximité-compatibilité, l'effet de la dissociation de l'attention et l'effet de la contiguïté spatiale.

### 1.1.1. Le principe de proximité compatibilité

Le principe de proximité compatibilité est une recommandation ergonomique selon laquelle deux sources d'informations visuelles qui nécessitent d'être traitées de manière co-référente devraient être disposées spatialement proches l'une de l'autre. À l'inverse, si deux sources d'informations visuelles doivent être traitées de manière séparée, alors il est préférable de les séparer physiquement pour éviter un encombrement trop important (Wickens & Carswell, 1995). Plusieurs travaux de Wickens et ses collaborateurs ont porté sur le traitement d'informations visuelles en situation de pilotage. Dans ce contexte, les contraintes sont nombreuses puisqu'il faut à la fois que le traitement cognitif soit rapide et efficace, mais également que les informations affichées prennent le moins de place possible puisque la taille du cockpit est limitée. Kroft et Wickens (2001) ont confronté les élèves pilotes à deux cartes. L'une présentait les caractéristiques du terrain (relief, lignes électriques, ponts...) et l'autre affichait à la fois le trafic aérien et la météo. Les participants devaient répondre à des questions à choix multiples. La moitié d'entre elles ne nécessitaient pas de traiter ensemble les deux canaux d'informations. Par exemple, la question « l'avion X va-t-il traverser une zone de conditions météorologiques difficiles ? » (notre traduction) peut être résolue en ne traitant que la carte comprenant à la fois les lignes aériennes et les informations météorologiques. Pour répondre correctement à l'autre moitié des questions, les informations des deux cartes devaient obligatoirement être intégrées. Par exemple, pour répondre à la question « l'avion X va-t-il survoler la rivière Y ? »

(notre traduction, Kroft & Wickens, 2001), il faut pouvoir comparer l'emplacement de la rivière sur une carte avec la ligne aérienne de l'autre carte.

Dans cette étude, la rapidité de réponse aux questions nécessitant l'intégration mentale des deux cartes est plus élevée lorsque celles-ci sont superposées que lorsqu'elles sont côte à côte. Kroft et Wickens (2001) expliquent ce résultat en termes d'économie de ressources. La réponse aux questions nécessitant le traitement des deux cartes mettrait en jeu trois types de traitements cognitifs : l'attention sélective pour rechercher les informations utiles, l'attention focalisée pour lire ces informations et l'attention divisée pour intégrer mentalement les informations d'une carte avec celles de l'autre carte. Les étapes de recherche et de lecture sont forcément présentes pour répondre à toutes les questions. En revanche, l'étape d'intégration mentale n'est nécessaire que dans les cas où la question implique le traitement de deux cartes et que celles-ci sont présentées séparément. Lorsque les cartes sont superposées, l'intégration des informations est réalisée de manière physique et n'a plus à être effectuée mentalement par les participants (Kroft & Wickens, 2001). Cette étape d'intégration mentale des informations prend du temps et sa suppression permet donc de réaliser la tâche plus rapidement. À l'inverse, cette étude montre aussi que la superposition des cartes a un impact négatif sur les performances pour les questions ne nécessitant pas l'intégration des informations des deux cartes. Pour ces questions, la séparation des deux cartes permettrait un désencombrement à l'écran qui faciliterait la recherche et la lecture des informations (Kroft & Wickens, 2001). Les résultats de cette étude sont cohérents avec le principe de proximité-compatibilité puisqu'ils soutiennent l'idée selon laquelle des informations à traiter ensemble devraient être proches et que, lorsqu'il n'est pas nécessaire qu'elles soient traitées ensemble, il est préférable de les éloigner.

### 1.1.2. Effet du partage de l'attention

L'effet du partage de l'attention fait référence à un partage entre des informations spatialement ou temporellement distribuées. Dans cette partie, seul le partage de l'attention entre les informations spatialement distribuées est traité. Selon Tricot

(1998), l'effet du partage de l'attention « se produit quand les sujets doivent traiter des sources d'information multiples dont l'intégration doit être opérée mentalement afin que le sens puisse être inféré du matériel présenté » (p. 44). La première étude mettant en évidence cet effet fut celle de Tarmizi et Sweller (1988) portant sur la réalisation d'exercices de géométrie. Sur un autre exercice de mathématique, Sweller, Chandler, Tierney et Cooper (1990) montraient une diminution de la durée nécessaire pour réaliser la tâche lorsque les informations textuelles étaient intégrées aux informations graphiques. Selon ces auteurs, l'intégration mentale des multiples sources d'information requière des ressources cognitives qui peuvent être économisées en intégrant physiquement ces informations.

Dans une série d'expériences, Chandler et Sweller (1991) ont montré un effet positif sur l'apprentissage de l'intégration du texte dans l'image lorsque l'intégration mentale de ces deux sources est indispensable à la compréhension du document. En revanche, cet effet n'apparaît pas lorsque l'intégration mentale des deux sources n'est pas indispensable à la compréhension (Chandler & Sweller, 1991). L'effet du partage de l'attention reposerait sur le fait que l'apprentissage du document ne peut débuter que lorsque les éléments issus des deux sources ont été mis en lien, mais ne concerne pas les cas dans lesquels les deux sources visuelles disparates peuvent être traitées indépendamment l'une de l'autre.

Cet effet de l'intégration apparaît aussi lorsque les deux sources sont textuelles, à condition que la compréhension d'un texte soit dépendante de la compréhension de l'autre (Chandler & Sweller, 1992). Dans le cas des articles scientifiques, la compréhension de la partie « résultats » dépend souvent de la lecture préalable de la partie « méthode ». Ainsi, une étude de Chandler et Sweller (1992) montrait une amélioration de la compréhension d'un article lorsque les informations à traiter ensemble pour être comprises étaient regroupées, en comparaison à une version originale de l'article, séparant ces informations. L'effet du partage de l'attention sur les performances a par la suite été observé sur des manuels d'utilisation de logiciel (Sweller & Chandler, 1994). Le contenu de ce type de document a, selon les auteurs de l'étude, un fort niveau d'interactivité. Ce niveau d'interactivité élevé serait

indispensable à l'apparition de l'effet du partage de l'attention (Sweller & Chandler, 1994) car cela implique que les éléments doivent être simultanément maintenus en mémoire de travail pour être traités ensembles (Chandler & Sweller, 1996). Un haut niveau d'interactivité dans un document génèrerait une charge cognitive élevée pour traiter ce document. Cela expliquerait que l'effet du partage de l'attention n'apparaisse que lorsque l'interactivité est élevée. Cela appuierait également l'idée selon laquelle l'effet du partage de l'attention découlerait bien d'un phénomène de surcharge cognitive (Chandler & Sweller, 1996). De nombreuses études ont ainsi montré un effet du partage de l'attention entre deux sources spatialement disparates avec des documents à forte interactivité, comme des circuits électriques (*e.g.* Tindall-Ford, Chandler, & Sweller, 1997; Kalyuga, Chandler, & Sweller, 1999), un document pédagogique décrivant le fonctionnement du cœur (Erhel & Jamet, 2006, 2011) ou encore des problèmes mathématiques (*e.g.* Mwangi & Sweller, 1998; Bobis, Sweller, & Cooper, 1993; Bétrancourt & Caro Dambreville, 1998; voir Ginns, 2006 pour une revue).

En outre, l'effet de la dissociation de l'attention peut être interprété à travers la théorie de la charge cognitive (Ayres & Sweller, 2005). Cierniak, Schieiter et Gerjets (2009) vont dans ce sens en mettant non seulement en évidence l'effet du partage de l'attention, mais en démontrant en plus que la séparation des sources à traiter génère une augmentation de la charge cognitive grâce à des mesures de cette dernière. En améliorant les caractéristiques ergonomiques d'un document pédagogique et notamment en évitant l'effet de dissociation de l'attention, il est possible de réduire la charge cognitive extrinsèque pour favoriser l'allocation de ressources en charge utile et donc améliorer la compréhension du document (Cooper, 1998; Ayres & Sweller, 2005).

### 1.1.3. Le principe de contiguïté spatiale

Moreno et Mayer (1999) ont proposé de remplacer l'expression trop générale d'attention partagée (*split attention effect*), qui renvoie à plusieurs phénomènes très différents, par l'effet de la contiguïté spatiale, temporelle ou effet de la modalité selon

les cas. Dans le domaine des documents pédagogiques, le principe de contiguïté spatiale s'énonce ainsi : « les étudiants apprennent mieux lorsque les images et les mots correspondants sont proches que lorsqu'ils sont éloignés sur la page ou sur l'écran » (Mayer, 2001, p. 81, notre traduction).

Ce principe est appuyé par un grand nombre d'études. Dans la plus ancienne d'entre elles, Mayer (1989) demandait à des étudiants d'apprendre un document portant sur le mode de fonctionnement des systèmes de direction mécanique, à air comprimé et hydraulique. Les résultats de cette étude indiquaient que l'apprentissage est plus efficace lorsque le texte est accompagné d'une illustration légendée que lorsqu'il est présenté avec une illustration non légendée. Chaque passage du texte faisait référence à une zone spécifique de l'image, ce qui obligeait les participants à effectuer des recherches visuelles sur l'image. Cette recherche serait diminuée quand les légendes sont insérées dans l'image. Dans une étude analogue, Mayer et Gallini (1990) montraient que l'intégration du texte dans les différentes zones correspondantes de l'image améliorait les performances, plus encore que l'insertion de légendes seules. Là encore, ce résultat peut être expliqué en termes de diminution de la recherche visuelle. Avec un document portant sur la formation des éclairs, Mayer, Steinhoff, Bower et Mars (1995) ont appuyé ces conclusions en mettant en évidence un effet positif sur l'apprentissage de l'intégration du texte dans les zones correspondantes de l'image. Avec le même contenu sur la formation des éclairs, Moreno et Mayer (1999) distinguaient dans une même étude les effets de modalité, de contiguïté temporelle et de contiguïté spatiale. Les participants devaient apprendre le document soit avec les paragraphes de texte intégrés dans les zones correspondantes de l'animation, soit avec le texte dans une zone adjacente à l'animation, soit encore avec le texte diffusé de manière sonore (diffusion synchrone ou asynchrone par rapport à l'animation). Les auteurs ont ainsi mis en évidence deux effets différents. Le premier est l'effet de la multimodalité caractérisé par de meilleures performances des participants bénéficiant du texte sonore par rapport aux autres participants. Le second effet mis en évidence est celui de la contiguïté spatiale puisque le groupe de participants ayant appris le document avec le texte intégré à l'animation avait de meilleures performances que

celui qui disposait du document avec le texte adjacent à l'animation. En revanche, les auteurs ne montraient pas d'effet significatif lié à la contiguïté temporelle puisque aucune différence n'apparaissait comme liée au caractère synchronisé des informations sonores avec l'animation. Moreno et Mayer (1999) n'ont pas apporté de détail pour expliquer l'effet de la contiguïté spatiale et se sont contentés de préciser que les participants avec le texte séparé de l'animation réussissaient moins bien à cause de leur charge cognitive plus élevée, en comparaison à ceux pour qui le texte est intégré à l'animation. Il faut préciser qu'une réplique récente de cette étude par Schmidt-Weigand, Kohnert et Glowalla (2010) avec le même matériel expérimental ne parvient pas à mettre en évidence d'effet statistiquement significatif de la contiguïté spatiale, mais retrouve néanmoins l'effet de la modalité. Cette absence de reproduction des résultats de Moreno et Mayer (1999) pourrait être due à une taille d'échantillon trop faible (Schmidt-Weigand *et al.*, 2010) puisqu'ils étaient 13 à 14 par condition contre 40 à 41 dans l'expérience initiale.

Holsanova, Holmberg et Holmqvist (2009) ont eux aussi testé l'effet de la contiguïté spatiale en incorporant des enregistrements des mouvements oculaires. Comme attendu, leurs participants ont obtenu des performances plus élevées lorsque les informations textuelles étaient intégrées à proximité des informations graphiques correspondantes que lorsque le texte et l'image étaient séparés. Le parcours oculaire typique dans cette expérience consistait, dans le format Séparé, à d'abord explorer les images avant de lire le texte alors que dans le format Intégré, les participants avaient tendance à traiter successivement les paires d'informations mutuellement référencées (un paragraphe, puis la zone de l'image correspondante, puis un autre paragraphe, etc). Selon les auteurs, l'intégration du texte et de l'image facilite, pour les participants, la mise en lien des informations textuelles avec les informations imagées correspondantes. Le comportement oculaire induit serait donc plus propice à l'intégration mentale des informations textuelles et imagées dans le format Intégré que dans le format Séparé. Dans deux expériences, Johnson et Mayer (2012), ont montré un effet positif de l'intégration spatiale du texte et de l'image sur l'apprentissage du fonctionnement d'un système de freinage de voiture.

L'enregistrement des mouvements oculaires montrait que les participants, avec le format Intégré, effectuaient plus de transitions visuelles entre le texte et les images qu'avec le format Séparé, favorisant ainsi l'intégration en mémoire des informations. L'intégration tendrait à encourager les apprenants à s'engager dans des stratégies cognitives appropriées durant l'apprentissage. Cependant, dans la troisième expérience de Johnson et Mayer (2012), aucune différence significative n'était observée entre les performances des participants du groupe Intégré et celles des participants pour qui le texte était adjacent à l'image, mais avec des numéros intégrés dans l'image et correspondant aux numéros attribués aux paragraphes du texte. De la même manière, Florax et Ploetzner (2010) ne sont pas parvenu à mettre en évidence de différence significative de performance d'apprentissage entre un format Intégré et un format dans lequel le texte est segmenté, numéroté et l'image légendée. Les auteurs suggèrent que souvent, dans les études sur l'effet de la contiguïté spatiale, l'intégration du texte dans l'image implique, en plus de la variation de distance, une segmentation du texte et une mise en évidence des liens entre chaque paragraphe de texte et la zone correspondante de l'image. Néanmoins, trois études sur des matériels différents ont mis en évidence une supériorité significative du format Intégré par rapport au format « numéroté » (Kester, Kirschner, & Van Merriënboer, 2005; Bodemer, Ploetzner, Feuerlein, & Spada, 2004; Purnell, Solman, & Sweller, 1991). L'effet de l'intégration du texte dans l'image irait donc au-delà de la simple mise en correspondance des informations entre elles. Il y aurait en plus une facilitation de la recherche visuelle des informations correspondantes liée à la proximité elle-même.

## **1.2. Le guidage de l'exploration visuelle d'un document**

La recherche d'erreurs d'interprétation de plan implique une comparaison de sources visuelles multiples, mais aussi une exploration de ces dernières. Or, l'exploration d'une source visuelle ne se fait pas forcément par balayages réguliers et dépend des caractéristiques de l'image. Par exemple, dans une exploration libre d'une image, les

individus ont tendances à regarder d'abord le centre de l'écran (Bindemann, 2010), peut-être, comme le prétendent certains auteurs, parce que les photographes ont tendance à viser l'objet « intéressant » et donc à le placer plutôt au centre de l'image (Dorr, Martinez, Gegenfurtner, & Barth, 2010). Un grand nombre de caractéristiques d'une image ou d'une animation influence la façon dont celle-ci est explorée (Schütz, Braun, & Gegenfurtner, 2011). Sur un contenu dynamique, il est d'ailleurs possible de réaliser des captures attentionnelles de différentes manières (voir Ludwig, Ranson, & Gilchrist, 2008 pour une revue). Par exemple, le déplacement d'une figure à l'écran a tendance à déclencher des comportements de poursuite visuelle (Schütz *et al.*, 2011; Dorr *et al.*, 2010). L'attention peut aussi être capturée par des changements soudains de contraste (Franconeri, Simons, & Junge, 2004), des mouvements (Abrams & Christ, 2003) ou encore l'apparition soudaine d'une figure à l'écran (Godjin & Theeuwes, 2002; Theeuwes, Kramer, Hahn, & Irwin, 1998).

Un certain nombre d'études ont montré que les différents moyens de réaliser des captures attentionnelles peuvent constituer des outils permettant de guider l'attention des participants pour faciliter l'exploration d'un document multimédia. Les indices visuels peuvent aider à sélectionner l'information pertinente, à expliciter la structure d'un document ou encore à faciliter la coréférenciation des informations (de Koning, Tabbers, Rikers, & Paas, 2009). Ainsi, Craig, Gholson et Driscoll (2002) ont montré que dans un document multimédia comportant à la fois des informations textuelles, imagées et une explication sonore, le changement soudain de couleur des éléments visuels au moment où l'information sonore y fait référence améliore les performances d'apprentissage. Guider au moment adéquat l'attention des participants sur l'information visuelle dont parle l'explication sonore faciliterait la coréférenciation des informations. Ce résultat est cohérent avec le principe de la contiguïté temporelle décrit par Moreno et Mayer (1999). Cette amélioration des performances d'apprentissage grâce au changement de couleur de l'information pertinente au moment approprié par rapport à un contenu auditif a été confirmée par une recherche de Ozelik, Arsalan-Ari et Cagiltay (2010). Les enregistrements des mouvements oculaires sur ce type de document ont montré que le changement de couleur (passage

du noir au rouge) de l'information au moment adéquat permet de guider l'attention des participants sur cette information, *i.e.* qu'ils effectuent un plus grand nombre de fixations et de plus longue durée sur cette zone au bon moment que lorsqu'ils n'ont pas accès à ces indices visuels. Le changement de couleur peut aussi permettre l'amélioration de la coréférentiation de la légende intégrée dans un schéma et des informations textuelles situées à côté de ce schéma. C'est ce qu'ont démontré l'expérience de Kalyuga *et al.* (1999) et celle de Ozcelik, Karakus, Kursun et Cagiltay (2009) dans lesquelles l'utilisation d'un code couleur permettait d'améliorer les performances d'apprentissage d'un document constitué d'une image et d'un texte lui faisant référence. Lin et Atkinson (2011) ont mis en évidence le même type de résultats avec des indices visuels d'une autre nature. Dans cette recherche, le guidage favorisant la coréférenciation des informations visuelles et sonores était amélioré par l'apparition d'une flèche pointant l'information visuelle pertinente. Pour une tâche de détection de cible à l'écran, un guidage basé sur la poursuite visuelle a été testé par Nickles, Melloy et Gramopadhye (2003). Cette étude montrait que les performances de recherche d'une cible parmi un ensemble de figures était facilité par le déplacement d'un curseur induisant un balayage visuel régulier. Le fait de poursuivre visuellement le curseur conduirait à effectuer une exploration plus exhaustive de l'image et augmenterait donc ainsi les chances de trouver la figure cible.

Dans le domaine des documents multimédias, plusieurs études ont montré que la présentation séquentielle des informations visuelles pouvait être un moyen de guider l'exploration d'un document et que cela pouvait améliorer les performances d'apprentissage. Ainsi, Bétrancourt et Tversky (2000) ont montré qu'une modulation de l'ordre d'apparition des éléments spatiaux à l'écran pouvait influencer la compréhension du document. Un format Séquentiel de présentation des informations a donc un impact sur la manière dont le document est appris, notamment parce qu'il impose un ordre dans le traitement des informations. Cet impact de l'ordre de l'apparition séquentielle des éléments sur la construction du modèle mental a été confirmé par une seconde étude de Bétrancourt, Bisseret et Faure (2001) dans laquelle des participants devaient apprendre des cartes de village et de bibliothèques

présentées de manière statique ou séquentielle. Dans le format Séquentiel, les éléments pouvaient apparaître un à un ou par groupes fonctionnels (les éléments qui renvoient à la même fonction ou à la même thématique) ou encore par groupes de proximité (les éléments proches les uns des autres apparaissent en même temps). Lors de la phase de restitution, les participants avaient tendance à rappeler les éléments dans un ordre proche de celui qui leur a été présenté. Cela indique que l'ordre d'apparition des éléments impacte la structuration de leur représentation en mémoire.

Dans deux expériences de Jamet (2006), les participants devaient apprendre un document constitué d'un schéma, d'informations textuelles intégrées et d'informations sonores. Les participants pour lesquels les éléments visuels apparaissaient au fur et à mesure de la diffusion des informations sonores ont obtenu des scores plus élevés aux épreuves de rappel et d'inférence qui suivaient l'apprentissage. Selon l'auteur, cette différence proviendrait d'une facilitation du traitement synchrone des sources visuelles et auditives, ce qui est là encore cohérent avec le principe de la contiguïté temporelle (Moreno & Mayer, 1999). Jamet et Arguel (2008) ont obtenu le même type de résultats. Dans cette étude, la présentation séquentielle des informations visuelles et sonores de manière synchronisée améliorerait la performance d'apprentissage du document. Cet effet apparaît si le document décrit une procédure, mais aussi lorsque le contenu n'a pas de composante temporelle. Jamet, Gavota et Quaireau (2008) ont eux aussi mis en évidence un effet positif de la présentation séquentielle des informations visuelles sur l'apprentissage d'un document multimodal. Ils ont d'ailleurs montré dans la même étude un effet positif de la mise en saillance des informations pertinentes par changement de couleur, les deux processus de guidage (séquentialité et mise en saillance) ayant des effets additifs sur la qualité de l'apprentissage. La présentation séquentielle peut donc être un choix de conception qui pourrait permettre de guider l'exploration d'un document. Ainsi, cela donne la possibilité au concepteur d'imposer à la fois un ordre et un rythme d'exploration des éléments visuels.

## 2. EXPERIENCE 1

Le principe de la contiguïté spatiale décrit par Mayer (2001) a été mis en évidence dans des expériences portant sur l'apprentissage de documents multimédias constitués d'un texte et d'une image (ou d'une animation) se faisant mutuellement référence. L'application de ce principe à la conception d'une interface permettant à des utilisateurs de rechercher et corriger les erreurs d'interprétation de plans pose un problème puisque cette recherche d'erreur n'est pas une tâche d'apprentissage et qu'elle porte sur deux sources imagées. Cependant, certains travaux ont démontré que l'effet de la contiguïté spatiale apparaissait de manière similaire sur d'autres types de documents et sur d'autres tâches. Ainsi, Chandler et Sweller (1992) montraient que l'intégration des sources d'informations mutuellement référencées améliore la performance de compréhension, même lorsque les deux sources sont textuelles. Kester *et al.* (2005) obtenaient le même type de résultats pour une tâche de résolution de problème sur un circuit électrique. Enfin, Bauhoff, Huff et Schawn (2012) ont étendu le champ d'application du principe de la contiguïté spatiale à la comparaison d'images pour identifier les différences. Ils considèrent d'ailleurs que l'explication de leurs résultats est la même que celle proposée par Mayer (2001) pour la contiguïté spatiale, *i.e.* que l'éloignement des sources mutuellement référencées obligerait les participants à effectuer des recherches visuelles coûteuses en ressources cognitives et diminuerait ainsi leurs chances d'avoir en mémoire de travail au même moment deux informations des deux sources à mettre en lien. La première hypothèse de l'expérience décrite ici était que le champ d'application du principe de la contiguïté spatiale peut être encore plus étendu que le proposaient Bauhoff *et al.* (2012) en s'appliquant aussi à la recherche d'erreurs d'interprétation de plan par comparaison visuelle de l'interprétation avec le plan initial. L'explication en termes de diminution des recherches visuelles qui alourdissent la charge cognitive devrait s'appliquer de la même manière sur la recherche d'erreur. Concrètement, l'intégration de l'interprétation et du plan initial devrait fortement diminuer ces recherches visuelles et donc améliorer les performances des participants.

Par ailleurs, la réussite d'une recherche d'erreurs dans des plans passe par une certaine exhaustivité de l'exploration. Or, d'un point de vue logiciel, la reconnaissance des plans se fait progressivement, élément par élément. Il est donc possible de profiter de ce caractère progressif en affichant à l'écran la reconnaissance en temps réel afin de guider l'attention des participants. La seconde hypothèse de l'étude présentée était donc que l'affichage à l'écran en temps réel de la reconnaissance automatique génère un guidage de l'attention des participants qui les conduit à effectuer un contrôle plus exhaustif des interprétations. L'affichage de la reconnaissance devrait ainsi améliorer la performance de détection des erreurs.

## **2.1. METHODE**

### **2.1.1. Participants**

Cette étude a porté sur 54 étudiants et jeunes actifs (19 hommes et 35 femmes) recrutés à partir de la base de données du LOUSTIC. Le plus jeune avait 18 ans et 9 mois et le plus âgé avait 31 ans et 3 mois. L'âge moyen des participants était de 23 ans et 2 mois pour un écart-type de 38 mois.

### **2.1.2. Matériel**

L'ensemble des tâches des participants dans cette expérience était réalisée sur une tablette Asus Eee Slate de 12.1 pouces de diagonale. Les épreuves de comparaison de plans ont été effectuées avec un prototype de logiciel de reconnaissance automatique sur trois plans différents. À partir d'un plan manuscrit, ce prototype est en mesure de reconnaître les différents symboles utilisés pour créer une interprétation du plan. Dans le cadre de l'expérience, il se limitait à reconnaître les murs qu'il représentait par des segments noirs, les portes qu'il signalait par des cases rouges et les fenêtres qu'il représentait par des cases bleues. Lorsque l'épreuve démarrait, le plan initial apparaissait immédiatement à l'écran. Pendant toute la durée de la reconnaissance (de 30 secondes à 1 minute selon la complexité du plan), les participants n'avaient pas la possibilité d'entourer les erreurs. Chaque participant devait ainsi comparer trois

paires de plans successivement et dans une des trois conditions expérimentales. Dans la condition séparée, le plan manuscrit apparaissait à l'écran et aucune interaction n'était possible tant que la reconnaissance n'était pas terminée. Ensuite, l'interprétation apparaissait à côté du plan manuscrit et les participants avaient la possibilité d'entourer les erreurs (voir figure 13). La condition Intégré était très proche de la précédente puisque seul le plan manuscrit apparaissait à l'écran pendant toute la durée de la reconnaissance. En revanche, au lieu d'apparaître à côté du plan manuscrit, l'interprétation apparaissait par-dessus celui-ci lorsque l'analyse se terminait. Enfin, dans le format Séquentiel, le plan manuscrit apparaissait, comme dans les deux formats précédents, immédiatement. En revanche, le processus de reconnaissance automatique était visible en temps réel par les participants puisque l'interprétation se construisait progressivement par-dessus le plan manuscrit. Les trois plans utilisés étaient les véritables plans des trois niveaux des locaux du LOUSTIC. Ils étaient composés respectivement de 11, 11 et 33 symboles. La proportion d'erreurs sur chaque plan a été maintenue constante et volontairement élevée afin d'aboutir à un niveau de difficulté adéquat. Ils contenaient respectivement 5, 5 et 15 erreurs.

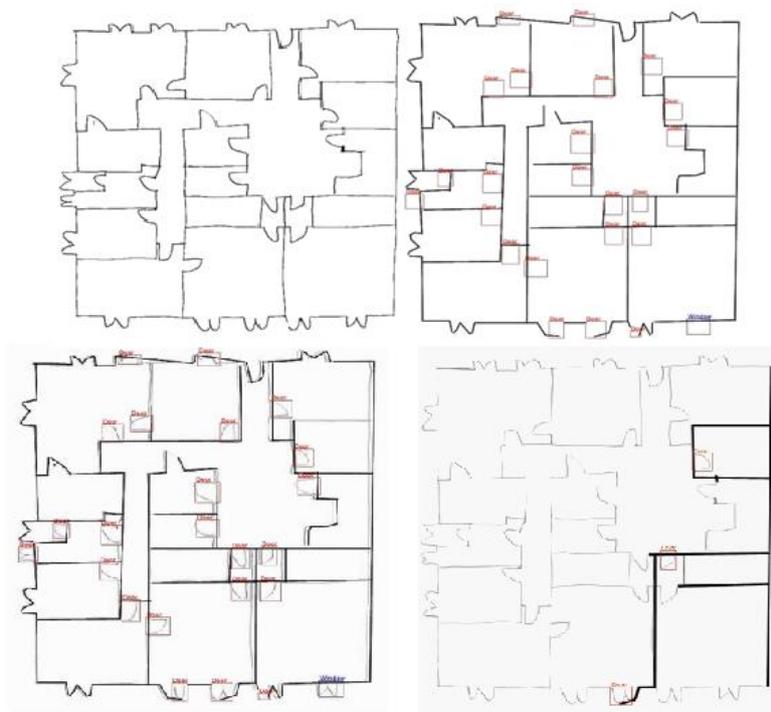


Figure 13. Formats Séparé (en haut), Intégré (en bas à gauche) et Séquentiel (en bas à droite)

### 2.1.3. Procédure

Tous les participants commençaient par passer une version informatisée de l'épreuve des blocs de Corsi et des grilles de Wilson afin d'évaluer leurs empan spatio-séquentiel et visuel. Un plan d'architecture était ensuite présenté à l'écran. Ce plan était interprété par le prototype et les participants devaient entourer les erreurs sur le plan résultant de cette interprétation. L'expérimentateur précisait aux participants que le système symbolisait les murs par des traits noirs, l'identification d'une porte par un carré rouge et l'identification d'une fenêtre par un carré bleu. Les différents types d'erreurs réalisées par le prototype consistaient à reconnaître une porte au lieu d'une fenêtre, une fenêtre au lieu d'une porte ou encore de reconnaître un mur au lieu d'une porte ou d'une fenêtre. Chaque participant passait ainsi 3 plans différents dans la même condition expérimentale. L'ordre de présentation des 3 plans était contrebalancé à l'intérieur de chaque groupe expérimental. L'activité de comparaison de plans a été évaluée sur deux aspects : l'efficacité et l'efficience.

## 2.2. Résultats

Afin de vérifier que les empan visuel et spatial n'ont eu pas d'impact sensible sur les performances, une analyse corrélacionnelle a été effectuée. Celle-ci ne révèle pas de corrélation statistiquement significative entre l'empan visuel et la durée de la tâche ( $r(52) = -0,091, p = 0,523$ ) ou le nombre d'erreurs repérées ( $r(52) = 0,182, p = 0,197$ ) ni entre l'empan spatial et la durée de la tâche ( $r(52) = -0,154, p = 0,276$ ) ou le nombre d'erreurs repérées ( $r(52) = -0,065, p = 0,647$ ). Des comparaisons de scores d'empan visuel et spatial en fonction des groupes expérimentaux ont été réalisées afin de vérifier qu'ils étaient équivalents. Concernant l'empan visuel, l'homoscédasticité étant acceptable ( $F(2, 49) = 0,284, p = 0,754$ ), cette comparaison a été effectuée par le biais d'une analyse de variances. Celle-ci nous permet d'admettre l'hypothèse d'égalité des groupes en termes de scores d'empan visuels ( $F(2, 49) = 1,099, p = 0,537$ ). En revanche, l'homoscédasticité des scores d'empan spatial n'étant pas acceptable ( $F(2, 49) = 3,101, p = 0,054$ ), la comparaison des scores a été effectuée à l'aide d'un test de

Kruskal-Wallis permettant de qualifier l'hypothèse d'égalité comme acceptable ( $p = 0,520$ ).

### 2.2.1. Efficience

La comparaison des durées de la tâche n'était effectuée qu'entre les groupes Séparé et Intégré. En effet, bien que les participants du groupe Séquentiel ne puissent pas entourer les erreurs avant la fin de la reconnaissance (comme dans les deux autres groupes), ils avaient tout de même la possibilité de commencer mentalement le repérage d'erreurs plus tôt que les autres. Ainsi, comparer la durée de la tâche de la condition Séquentiel avec celle des autres conditions n'aurait pas de sens. Afin de vérifier l'homogénéité des distributions des temps de réalisation de la tâche de comparaison de plan, un test de Levene a été effectué. Celui-ci révèle une homoscedasticité acceptable ( $F(1, 32) = 0,440, p = 0,512$ ). Une analyse de variances montre que la durée de la tâche est significativement plus élevée dans la condition Séparé ( $M = 228,12, SD = 48,182$ ) que dans la condition Intégré ( $M = 170,65, SD = 42,560$ ),  $F(1, 32) = 13,586, MSE = 2066,426, p = 0,001$  (voir tableau 2). Ce résultat va dans le sens de notre hypothèse selon laquelle l'intégration des sources diminue les recherches visuelles et permet donc d'améliorer les performances.

Tableau 2. Moyennes et écart-types des durées de la tâche dans les groupes Séparé et Intégré

	Durée de la tâche	
	<u>M</u>	<u>SD</u>
Séparé	228,12	48,182
Intégré	170,65	42,560
Séquentiel	103,72	39,716

### 2.2.2. Efficacité

L'efficacité des différents formats est comparée à travers une table de contingence regroupant les réussites (repérage de toutes les erreurs) et les échecs à la tâche par les participants. Cet indicateur binaire a été choisi étant donné que, dans ce type de

tâche, le nombre d'erreurs non identifiées est presque toujours très faible. Ainsi, la variance du nombre d'erreurs identifiées est aussi très peu élevée. Le calcul du rapport de vraisemblance montre qu'il existe une relation statistiquement significative entre la fréquence de la réussite à la tâche et le format de présentation ( $LR(2, N = 52) = 7,127, p = 0,028$ ). Pour localiser la différence, un rapport de vraisemblance est calculé pour comparer ces fréquences entre les groupes Séparé et Intégré. Aucune différence significative n'est révélée par ce test ( $LR(1, N = 34) = 0,485, p = 0,485$ ). Enfin, un troisième rapport de vraisemblance est calculé pour comparer les groupes Intégré et Séquentiel. Celui-ci révèle que la réussite complète à la tâche est significativement plus fréquente dans le format Séquentiel ( $LR(1, N = 35) = 6,660, p = 0,01$ ) (voir tableau 3). Ce résultat va dans le sens de notre hypothèse selon laquelle la séquentialité améliore l'efficacité de la recherche d'erreur en guidant l'exploration du plan de manière plus exhaustive.

Tableau 3. Proportion de participants ayant repéré toutes les erreurs dans les groupes Séparé, Intégré et Séquentiel

	Réussite complète de la tâche		
	Oui	Non	Total
Séparé	8	9	17
Intégré	6	11	17
Séquentiel	14	4	18
Total	28	24	52

## 2.3. Discussion

La première hypothèse de cette étude était que la superposition des plans allait constituer une aide pour les participants. Les résultats valident cette hypothèse puisque les participants ont réalisé la tâche dans une durée significativement plus

courte avec des plans superposés qu'avec des plans séparés. Pour rechercher les erreurs, les participants devaient stocker en mémoire de travail simultanément le sens et la localisation d'un symbole d'une part et l'interprétation proposée par le système pour ce symbole d'autre part, afin de comparer ces informations entre elles. Pour faire cela, ils devaient dans un premier temps récupérer en mémoire de travail une information (symbole ou interprétation), effectuer une recherche visuelle de l'information correspondante sur l'autre source, puis comparer mentalement ces deux informations pour déterminer s'il s'agit ou non d'une erreur. La supériorité du format Intégré par rapport au format Séparé laisse penser que l'interprétation de l'effet de la contiguïté spatiale proposée par Mayer (2001) est aussi applicable à la recherche d'erreurs. Cela signifierait que l'intégration diminue la recherche visuelle, activité coûteuse en termes de ressources cognitives, facilitant ainsi le regroupement en mémoire de travail des deux informations à comparer. Ce résultat confirme que le principe de la contiguïté spatiale, déjà étendu à la comparaison d'images par Bauhoff *et al.* (2012) s'applique aussi à la recherche d'erreurs sur des plans. Néanmoins, il est surprenant que l'intégration améliore l'efficacité mais pas l'efficience des participants. Ce résultat peut être expliqué par le fait que dans notre étude, le temps que prennent les participants pour effectuer les corrections est libre. Ainsi, ils avaient la possibilité de compenser leurs difficultés en allouant plus de temps à la tâche. À l'inverse, le temps de consultation des documents était généralement contraint dans les expériences portant sur l'effet de la contiguïté spatiale comme celles d'Erhel et Jamet (2011) ou celle de Mayer et Gallini (1990).

La seconde hypothèse pour cette expérience prévoyait que l'apparition séquentielle du plan interprété allait améliorer l'efficacité de l'interaction. Cette hypothèse est validée par les résultats puisque le repérage des erreurs est meilleur avec le format Séquentiel qu'avec le format Intégré. Ce gain pourrait provenir d'un effet de guidage attentionnel proposant aux participants un ordre d'exploration du plan qui permet, s'il est suivi, de s'assurer que l'ensemble du plan est contrôlé, ce qui n'est pas forcément le cas dans les formats statiques.

L'absence de corrélation entre les performances et les scores d'empans visuel et spatial confirme les résultats obtenus lors de la première étude qui indiquait que ces deux composantes n'ont pas d'impact décisif sur la manière dont une tâche de correction de plan se déroule. Les mesures d'empans n'ont ainsi plus été effectuées lors des expériences suivantes.

## **3. EXPERIENCE 2**

L'expérience précédente a permis de montrer que la superposition des plans et l'apparition séquentielle constituent des aides au repérage d'erreurs. La seconde expérience consistait à répliquer la première en y ajoutant une mesure d'oculométrie afin de fournir des éléments interprétatifs plus précis. Dans l'expérience 1, la diminution de la durée de la tâche dans la condition Intégré par rapport à la condition Séparé a été interprétée comme étant la conséquence de la recherche visuelle effectuée par les participants du groupe Séparé qui, ayant mémorisé une information sur une source, recherchent l'information correspondante sur l'autre source. Ainsi, la première hypothèse (hypothèse 1) de l'expérience 2 était que les participants du groupe Séparé passent plus de temps à fixer en dehors des zones d'erreurs que les participants du groupe Intégré.

L'amélioration de la précision du repérage des erreurs en condition Séquentiel a été interprétée dans l'expérience précédente comme résultante d'un guidage attentionnel par l'apparition soudaine d'éléments à l'écran imposant aux participants un ordre d'exploration du document, qui les conduisait à contrôler l'ensemble des erreurs. Afin de vérifier cette interprétation, l'hypothèse selon laquelle le nombre d'erreurs non contrôlées par les participants est plus élevé dans la condition Intégré que dans la condition Séquentiel a été testée (hypothèse 2).

## **3.1. Matériel et méthode**

### **3.1.1. Participants**

Cette seconde étude a porté sur 54 étudiants et jeunes actifs (23 hommes et 31 femmes) recrutés à partir de la base de volontaires du LOUSTIC. Le plus jeune avait 18 ans et 6 mois et le plus âgé avait 29 ans et 8 mois. L'âge moyen des participants était de 23 ans et 5 mois pour un écart-type de 31 mois.

### **3.1.2. Matériel**

Dans cette expérience, les conditions expérimentales étaient identiques à celles de la précédente (Séparé, Intégré et Séquentiel). La seule différence résidait dans l'enregistrement des mouvements oculaires pendant la tâche avec un « eye-tracker » Tobii T60 intégré dans un écran TFT 17 pouces d'une résolution maximum de 1280 x1024 pixels. Ce matériel non tactile impliquait que le repérage des erreurs soit fait non sur une tablette mais par l'utilisation d'une souris.

### **3.1.3. Procédure**

Comme dans l'expérience précédente, il était demandé aux participants d'entourer des erreurs d'interprétation de plans d'architecture. Chaque participant passait ainsi 3 plans dans la même condition expérimentale. L'ordre de présentation des plans était contrebalancé à l'intérieur de chaque groupe expérimental. L'activité de comparaison de plans a été évaluée sur deux aspects, la durée de l'activité et la réussite complète à la tâche. Ces mesures ont été complétées par l'enregistrement des mouvements oculaires des participants pendant l'épreuve de comparaison de plans.

## 3.2. Résultats

### 3.2.1. Efficience

Un test de Levene a été effectué pour vérifier l'homogénéité des distributions des temps de réalisation de la comparaison de plan. Celui-ci révèle une homoscedasticité acceptable ( $F(1, 32) = 0,040, p = 0,842$ ). Comme dans l'expérience précédente, l'analyse de variances des durées de la tâche en fonction du format de présentation révèle une différence significative entre les formats Séparé ( $M = 237,81, SD = 52,453$ ) et Intégré ( $M = 189,56, SD = 54,149$ ),  $F(2, 48) = 11,680, MSE = 2847,40, p = 0,013$  (voir tableau 4).

Tableau 4. Moyennes et écart-types pour la durée de la tâche dans les groupes Séparé, Intégré et Séquentiel

	Durée de la tâche	
	<u>M</u>	<u>SD</u>
Séparé	237,81	52,453
Intégré	189,6	54,149
Séquentiel	147,12	46,595

### 3.2.2. Efficacité

Le calcul du rapport de vraisemblance montre qu'il existe une relation statistiquement significative entre la fréquence de la réussite à la tâche (voir tableau 5) et le format de présentation ( $LR(2, N = 51) = 10,552, p = 0,005$ ). Afin d'identifier plus précisément la différence, un rapport de vraisemblance est calculé pour comparer ces fréquences entre les groupes Séparé et Intégré. Aucune différence significative n'est révélée par ce test ( $LR(1, N = 34) = 0,425, p = 0,515$ ).

Enfin, un troisième rapport de vraisemblance est calculé pour comparer les groupes Intégré et Séquentiel. Celui-ci révèle que la réussite complète à la tâche est significativement plus fréquente dans le format Séquentiel ( $LR(1, N = 35) = 9,808, p = 0,002$ ). Les résultats concernant la durée et la réussite à la tâche sont conformes à

ceux obtenus dans l'expérience précédente. Les résultats en termes d'efficience et d'efficacité de la correction obtenus dans la première expérience présentée dans cette partie sont répliqués.

Tableau 5. Proportion de participants ayant repéré toutes les erreurs

	Réussite complète de la tâche		
	Oui	Non	Total
Séparé	8	8	16
Intégré	7	11	18
Séquentiel	15	2	17
Total	29	23	51

### 3.2.3. Temps de fixation

Un test de Levene sur les temps de fixation oculaire sur les erreurs révèle une homoscedasticité insuffisante ( $F(1, 32) = 3,358, p = 0,076$ ). L'effet de l'intégration sur le temps cumulé de fixations sur les erreurs est donc testé par le biais d'un test non paramétrique de Mann-Whitney. Celui-ci ne révèle aucune différence significative de temps cumulé de fixation sur les erreurs entre les conditions Séparé ( $M = 74,19, SD = 24,17$ ) et Intégré ( $M = 73,28, SD = 36,44$ ),  $U = 141,5, p = 0,931$  (voir tableau 6).

Le test de Levene sur les temps cumulés de fixation en dehors des zones d'erreurs révèle une homoscedasticité acceptable ( $F(1, 32) = 0,038, p = 0,847$ ). L'effet de l'intégration sur le temps cumulé des fixations en dehors des zones d'erreurs est testé par le biais d'une analyse de variances. Celle-ci révèle que le temps de fixation en dehors des zones d'erreurs est significativement plus élevé dans le groupe Séparé ( $M = 162,62, SD = 40,19$ ) que dans le groupe Intégré ( $M = 115,33, SD = 41,89$ ),  $F(1, 32) = 11,456, MSE = 1653,680, p = 0,002$ .

Ce résultat va dans le sens de l'hypothèse sur la recherche visuelle. À titre indicatif, le nombre moyen d'allers et retours visuels entre les deux plans à comparer dans le format Séparé est de 63,9 ( $SD = 23,97$ ) soit 1 aller et retour toute les 3,75 secondes en moyenne.

Tableau 6. Moyennes et écart-type des temps de fixations cumulés sur les erreurs et en dehors des zones d'erreur

	Temps de fixation cumulé en dehors des zones d'erreur		Temps de fixation cumulé sur les erreurs	
	<u>M</u>	<u>SD</u>	<u>M</u>	<u>SD</u>
Séparé	162,62	40,19	74,19	24,17
Intégré	115,33	41,89	73,28	36,44
Séquentiel	69,21	44,12	77,9	26,5

### 3.2.4. Contrôle de l'ensemble des erreurs

L'enregistrement des mouvements oculaires a permis de différencier les participants ayant visuellement contrôlé toutes les erreurs de ceux en ayant oublié. Afin de comparer le nombre de participants ayant contrôlé l'ensemble des erreurs dans les différents groupes, un rapport de vraisemblance a été calculé, révélant une différence statistiquement significative entre les groupes,  $LR(2, N = 51) = 6,377, p = 0,041$  (voir tableau 7).

Pour identifier plus précisément la différence, un second rapport de vraisemblance a été calculé pour comparer ces fréquences entre les groupes Séparé et Intégré. Aucune différence significative n'a été révélée par ce test,  $LR(1, N = 34) = 0,133, p = 0,715$ . Enfin, un troisième rapport de vraisemblance a été calculé pour comparer les groupes Intégré et Séquentiel. Celui-ci révèle que le contrôle de l'ensemble des erreurs est significativement plus fréquent dans le format Séquentiel,  $LR(1, N = 35) = 4,207, p = 0,040$ . Ce résultat accrédite l'interprétation selon laquelle la séquentialité génère un contrôle plus exhaustif des plans.

Tableau 7. Proportion de participants ayant contrôlé l'ensemble des erreurs

Contrôle visuel de l'ensemble des erreurs.			
	Oui	Non	Total
Séparé	9	7	16
Intégré	9	9	18
Séquentiel	14	3	17
Total	32	19	51

### 3.3. Discussion

L'expérience 2 était une réplique de la première présentée dans ce chapitre et à laquelle ont été ajoutés des enregistrements des mouvements oculaires pendant la tâche de comparaison de plans. Les résultats obtenus confirment les précédents puisque là encore, l'intégration des plans améliore l'efficacité de la comparaison alors que la séquentialité en améliore l'efficacité. La première hypothèse était que les participants du groupe Séparé ne trouveraient pas instantanément les informations référentes, mais devraient effectuer des recherches visuelles. Cette hypothèse est validée par les résultats de l'expérience puisque le temps cumulé moyen de fixation en dehors des zones d'erreurs est plus important dans la condition Séparé que dans la condition Intégré. En revanche, le temps de fixation cumulé sur les zones d'erreurs n'est pas significativement différent dans les groupes Séparé et Intégré. Ce résultat indique qu'en moyenne, le gain de temps dû à l'intégration des plans ne correspond pas à une diminution du temps passé à traiter les erreurs mais à l'économie du temps de recherche visuelle afin de corréler les informations des deux sources. Le fait de constater dans cette étude la présence d'une recherche visuelle dans le format Séparé renforce l'idée selon laquelle le principe de la contiguïté spatiale pourrait être étendu à la correction de plans. En effet, non seulement l'amélioration des performances avec

l'intégration spatiale des sources est cohérente avec le principe de contiguïté spatiale, mais en plus, la recherche visuelle observée grâce aux enregistrements des mouvements oculaires indique que l'interprétation de ce principe par Mayer (2001) semble aussi s'appliquer à la correction de plans. Cela s'explique par l'existence d'un point commun central entre les deux tâches : les participants doivent coréférer des informations disparates en effectuant des allers et retours visuels. Ces allers et retours visuels impliquent des fixations involontairement en dehors des zones à contrôler qui constituent une perte de temps. Cette difficulté ne semble pas provenir d'une spécificité des plans d'architecture, mais plutôt de la situation de comparaison zone par zone de deux sources visuelles disparates. Ce choix de disposer côte à côte deux éléments dont il est nécessaire de comparer les sous-parties impliquerait donc qu'une part du temps consacré à la tâche de comparaison soit mobilisée dans des ajustements visuels permettant la coréfération.

La seconde hypothèse de l'expérience était que l'amélioration de l'efficacité dans la condition Séquentiel était due au guidage attentionnel conduisant les participants à effectuer un contrôle plus exhaustif du plan qu'en l'absence de guidage. Cette hypothèse est validée car les participants du groupe Séquentiel sont significativement plus nombreux à avoir contrôlé visuellement l'ensemble des erreurs. Le caractère statique du format Intégré implique que les participants qui ne mettent pas en place de stratégie d'exploration organisée s'exposent au risque d'oublier de contrôler certaines erreurs. Dans la condition Séquentiel, les erreurs apparaissent une par une à l'écran. Chacune des erreurs en apparaissant peut générer une capture attentionnelle (Godjin & Theeuwes, 2002; Theeuwes *et al.*, 1998) et se trouve donc plus certainement contrôlée qu'en condition Intégré.

## **4. CONCLUSION**

La principale limite de cette étude est qu'elle repose sur une tâche de repérage d'erreurs alors que dans le logiciel final, l'objectif n'est évidemment pas d'entourer les erreurs mais aussi de les corriger, ce qui implique une étape supplémentaire de

sélection de la bonne réponse. Cette simplification de la tâche découle d'une raison technique : à ce stade du projet, il n'était pas envisageable d'utiliser un prototype fonctionnel incluant cette fonction de correction des erreurs.

Les deux expériences présentées dans cette partie montrent qu'une présentation séquentielle des informations devrait favoriser leur coréférenciation ainsi qu'une exploration exhaustive des documents. Pendant l'étape de contrôle par l'utilisateur, il est donc préférable de superposer l'interprétation et le document initial afin d'améliorer l'efficacité de la correction. De plus, afficher à l'écran l'interprétation qui se construit en temps réel peut permettre une amélioration de l'efficacité du contrôle. Cette séquentialité n'impose d'ailleurs pas de contrainte particulière à des utilisateurs qui pour des raisons d'organisation préféreraient n'effectuer qu'un contrôle a posteriori. Le format Séquentiel est donc un choix de conception avantageux parce qu'il peut améliorer les performances sans contrepartie, mais en plus, il permet d'envisager des interactions entre l'utilisateur et le logiciel pendant le travail de reconnaissance. Cette approche plus coopérative de la correction est présentée dans la suite de ce manuscrit.

# Étude 3 : Effets de l'interruption utilisateur sur la correction des erreurs<sup>3</sup>

## 1. INTRODUCTION

Les expériences présentées dans la partie précédente de ce manuscrit montrent d'une part que la comparaison de plans est réalisée plus rapidement lorsqu'il s'agit de comparer des plans superposés que lorsqu'ils sont placés côte à côte et d'autre part que la présentation séquentielle de l'interprétation augmente l'efficacité de la recherche d'erreur. Les effets de la superposition ont été interprétés comme étant dus à une suppression des recherches visuelles ayant pour conséquence une facilitation de la coréférenciation des informations à comparer. L'effet positif de la séquentialité serait quant à lui lié au guidage attentionnel généré par l'apparition progressive des interprétations des éléments du plan. Le guidage imposerait un ordre d'exploration donné du plan conduisant les participants à en effectuer un contrôle plus exhaustif.

Cependant, avec le format Séquentiel, les participants n'avaient pas la possibilité d'entourer les erreurs durant la reconnaissance. Ils devaient attendre que l'interprétation soit terminée pour pouvoir commencer à les entourer. Comme évoqué en introduction de ce manuscrit, le fonctionnement du prototype permet de considérer que le système final sera capable d'apprendre durant le travail d'interprétation à condition que l'utilisateur corrige des erreurs pendant le processus de reconnaissance (Ghorbel, Anquetil, & Lemaitre, 2012). Cela implique que les corrections rapidement réalisées pourraient améliorer les performances du logiciel

---

<sup>3</sup> Fleury, S., & Jamet, E. (soumis). Putting Users Back in the Interaction Loop for the Automatic Supervised Interpretation of Technical Documents.

pour les interprétations suivantes en évitant la propagation en cascade d'erreurs de reconnaissances.

Dans la classification des niveaux d'automatisation, les prototypes des expériences précédentes se trouvaient alors à au niveau d'automatisation dit « totale » puisque l'utilisateur n'avait pas véritablement la possibilité d'intervenir en cas d'erreur, bien que cela risquait de déclencher plus d'erreurs encore sur la suite de l'analyse. Ainsi, il y a un intérêt technique à permettre à l'utilisateur d'interrompre la reconnaissance afin de corriger une erreur dès qu'elle apparaît. De plus, on peut se demander si indépendamment de ce gain technique, il n'existerait pas un bénéfice supplémentaire lié au fonctionnement cognitif des utilisateurs. Lorsque les utilisateurs vérifient les interprétations automatiques sans avoir la possibilité d'interrompre le processus, ils doivent mémoriser les erreurs au fur et à mesure de l'analyse au risque, peut-être, d'en oublier.

Pour traiter cette problématique, après avoir présenté une série de travaux sur l'oubli des informations à court terme, sur le modèle du processeur humain qui apporte un éclairage quantitatif sur l'oubli et sur l'interférence, une expérience évaluant l'impact des interruptions de l'analyse par les utilisateurs est décrite et discutée.

## **1.1. L'oubli en mémoire à court terme**

L'oubli à court terme est un phénomène ayant fait l'objet de nombreux travaux de recherche. Déjà à la fin du XIX<sup>ème</sup> siècle, Ebbinghaus (1885) testait à travers ses propres performances la capacité de mémorisation de listes de syllabes. En une seule lecture, il parvenait à mémoriser en moyenne sept syllabes successives. Ce même chiffre sept inspirait à Miller (1956) sa célèbre revue de question résumant les principaux résultats expérimentaux portant sur le traitement et la mémorisation à court terme d'informations. Il décrivait notamment les expériences de Hayes (1952) et de Pollack (1952) visant à évaluer la mémorisation à court terme de différents types d'informations (chiffres binaires ou décimaux, lettres...) présentés à une cadence d'un item par seconde aux participants. Certains des items présentés, s'ils étaient regroupés, correspondaient à un ensemble déjà connu par les participants. Par

exemple, certaines suites de lettres pouvaient constituer des mots. Selon Miller (1956), bien que la quantité moyenne d'items mémorisés soit différente selon le type d'informations utilisées, la quantité moyenne de *chunks* (i.e. de groupes d'informations) mémorisée serait quant à elle approximativement constante et proche de sept.

Comme le souligne Cowan (2000), les théories plus récentes de Baddeley (1986) et de Richman, Staszewski et Simon (1995) indiquent que la limite de la mémorisation à court terme ne serait pas une limite de stockage, mais plutôt une limite de durée pendant laquelle un item peut être maintenu actif en mémoire à court terme sans autorépétition. Dans une revue détaillée de la littérature, Cowan (2000) défendait l'idée selon laquelle le nombre de *chunks* mémorisables à court terme serait en moyenne plus proche de quatre que de sept. Selon lui, la performance descendrait jusqu'à quatre lorsque l'autorépétition ainsi que la constitution de nouveaux *chunks* est impossible. Cela correspondrait donc à une mesure plus pure de la mémorisation à court terme.

## **1.2. Le modèle du processeur humain**

Le modèle du processeur humain est une tentative de caractérisation quantitative des capacités perceptives, motrices et cognitives humaines par l'agglomération des résultats de nombreuses études (Card, Moran, & Newell, 1986). Dans le contexte de la présente thèse, ce sont plus particulièrement les traitements perceptifs et cognitifs des stimuli visuels qui retiendront notre attention. Selon le modèle du processeur humain, l'oubli des *chunks* stockés en mémoire de travail lorsque les autorépétitions sont empêchées surviendrait en moyenne après 7 secondes (5 secondes au minimum et 226 secondes au maximum). L'oubli est caractérisé ici par la durée requise pour, en moyenne, n'avoir plus que 50% des *chunks* rappelés (*Half-life forgetting*).

Cette information très générale doit néanmoins être examinée avec une plus grande finesse car lorsque plusieurs *chunks* sont présentés successivement, cela tend à diminuer les performances de rappel à cause du phénomène d'interférence (Reitman, 1971; Nairne, 2002 pour une revue). Les travaux de Melton (1963) ont apporté des

informations plus précises sur ce point. Ils montraient que l'oubli d'un *chunk* apparaît d'autant plus rapidement que la liste de *chunks* à mémoriser est longue (dans l'expérience, la liste allait de un à cinq *chunks*). D'une manière plus générale, plus le nombre de *chunks* à rappeler est élevé, plus la durée de l'oubli de la moitié d'entre eux est courte. L'oubli d'un *chunk* lorsqu'il est présenté seul a lieu, selon le modèle du processeur humain, au bout de 73 secondes (226 secondes au maximum). Lorsque trois *chunks* sont présentés, l'oubli apparaît après 7 secondes (5 secondes au minimum et 34 secondes au maximum).

Il faut cependant préciser que les études sur lesquelles le modèle du processeur humain s'appuie ont été réalisées sur des listes de mots (Murdock, 1962), de nombres (Peterson & Peterson, 1959) ou de lettres (Peterson & Peterson, 1959; Melton, 1963) et donc pas spécifiquement sur du contenu visuo-spatial.

### **1.3. L'interférence en mémoire à court terme**

Une étude d'Elliott et Strawhorn (1976) montrait un effet de la similarité entre les informations stockées en mémoire et la tâche concurrente. L'interférence favorisant l'oubli serait plus importante lorsque la modalité de la tâche concurrente est la même que la modalité de présentation du stimulus à mémoriser (auditif vs. visuel dans cette expérience). Proctor et Fagnani (1978) ont aussi montré un effet de la similarité sur la mémorisation à court terme de triplets de consonnes avec une tâche distractive de Brown-Peterson. Les résultats de cette étude indiquent que lorsque la tâche distractive est dans la même modalité (auditive ou visuelle) que la tâche principale, la mémorisation des items est plus dégradée que lorsque les deux tâches sont présentées selon des modalités différentes. De la même manière, les performances à une tâche de reconnaissance d'odeurs sont diminuées lorsque d'autres odeurs sont présentées lors de la période de maintien en mémoire (Walk & Johns, 1984) alors qu'une tâche kinesthésique n'interfère pas avec une tâche concurrente de maintien en mémoire de nombres (Williams, Beaver, Spence, & Rundell, 1969).

L'interférence est aussi plus importante lorsque le type de stimuli est identique. Par exemple, le stockage des mots interfère plus avec le traitement d'autres mots qu'avec

la manipulation de numéros. C'est ce qu'ont notamment montré Smyth *et al.* (1988) avec une série de cinq expériences dans lesquelles les sujets devaient mémoriser des configurations de mouvements, de mots ou de positions spatiales avec une tâche distractive phonologique, motrice ou spatiale. Les items devaient être rappelés dans l'ordre selon un paradigme d'empan mnésique. Les interférences observées dans ces expériences n'apparaissent que lorsque la tâche principale et la tâche distractive mettaient en jeu le même type de stimulus (verbal, spatial ou moteur). D'autres travaux ont reproduit ce type de résultats avec des stimuli de nature différente. Par exemple, Klauer et Stegmaier, (1997) montraient qu'une tâche de localisation de sons interférait avec la mémorisation d'informations spatiales. Plus récemment, Lee, Lee et Tsai (2007) réalisaient le même type d'expérience, mais avec des listes de mots écrits. Là encore, la tâche de suppression phonologique durant la période de rétention avait un impact négatif sur la mémorisation à court terme.

Toutes ces études qui mettent en évidence un effet d'interférence n'intervenant strictement qu'en présence de stimuli de même nature ou portant sur la même modalité sont tout de même à nuancer. D'autres résultats conduisent à penser que l'interférence entre des stimuli différents serait réduite, mais pas nulle. C'est le cas par exemple des travaux de Vergauwe *et al.* (2009) qui ont mis en évidence une interférence entre une tâche de traitement et une tâche de stockage, y compris lorsque ces tâches renvoient l'une à un contenu spatial et l'autre à un contenu visuel. Ainsi, le traitement de stimuli visuels et spatiaux puiserait une part de ressources communes. De plus, une activité de mémorisation et une activité de traitement visuo-spatial interfèrent aussi entre eux, ce qui suggère là encore la sollicitation d'une part de ressources communes.

Sur un matériel visuo-spatial, Washburn et Astur (1998) ont observé dans une même expérience à la fois le phénomène d'oubli et celui d'interférence. Cette expérience portait sur une tâche de mise en correspondance d'informations visuo-spatiales. Plus le délai de rétention était long et moins la mise en correspondance était bien réalisée. Autrement dit, il y avait un oubli progressif de configurations visuelles stockées. De plus, cette étude montrait qu'un distracteur visuo-spatial (un curseur se déplace à

l'écran, générant un comportement de poursuite visuelle), empêchant l'autorépétition visuelle pendant le délai, accentuait ce phénomène d'oubli. Ce résultat renvoie à l'interprétation de Cowan (2000) pour qui la véritable mesure de l'oubli serait celle réalisée dans une condition d'interférence durant laquelle l'autorépétition a été empêchée.

## **1.4. Hypothèses**

Le fait que le phénomène d'oubli en mémoire à court terme apparaisse au bout de seulement quelques secondes (Card *et al.*, 1986) et que ce phénomène soit accentué en situation de double tâche interférente (Washburn & Astur, 1998) conduit à se demander si le format Séquentiel décrit dans la partie précédente ne comporterait pas de risque de gêner les utilisateurs. Avec le prototype sans bouton pause, les utilisateurs ne doivent pas seulement maintenir en mémoire la localisation des erreurs repérées pendant plusieurs dizaines de secondes (jusqu'à la fin de l'analyse), ils doivent en même temps continuer de chercher de nouvelles erreurs. En outre, la recherche de nouvelles erreurs peut être considérée comme une tâche visuo-spatiale concurrente à la tâche de maintien des erreurs déjà identifiées. Les participants repèrent des erreurs au fur et à mesure de l'analyse, mais ne peuvent pas faire d'autorépétitions visuo-spatiales à cause de la tâche de recherche de nouvelles erreurs. La concurrence entre le maintien d'une quantité de plus en plus importante d'informations visuo-spatiales en mémoire et l'identification de nouvelles erreurs devrait donc générer de l'oubli. Ainsi, l'hypothèse de la présente étude était que les participants qui peuvent interrompre l'analyse en pressant un bouton afin d'entourer les erreurs qu'ils repèrent au fur et à mesure devraient repérer plus d'erreurs que ceux ne pouvant pas interrompre l'analyse.

## 2. Méthode

### 2.1. Participants

Les participants de cette étude étaient 36 étudiants (10 hommes et 26 femmes) recrutés à partir de la base de volontaires du LOUSTIC. Le plus jeune avait 18 ans et le plus âgé avait 33 ans et 4 mois. Leur moyenne d'âge était de 24,4 ans ( $SD = 2,8$ ).

### 2.2. Matériel et procédure

Toutes les épreuves étaient réalisées sur un Tablet PC Asus Eee Slate 12.1. Chaque participant devait rechercher et entourer les erreurs d'interprétation du logiciel pour trois plans (voir figure 14) successifs dans une des deux conditions expérimentales : Sans Interruption (SI) et Avec Interruptions (AI). Dans la condition SI, le plan initial apparaissait immédiatement à l'écran et l'interprétation se construisait progressivement de manière visible par les participants, constituant ainsi une présentation séquentielle de l'interprétation superposée au plan initial. Les participants devaient attendre que la reconnaissance soit totalement terminée pour entourer les erreurs. La condition AI était identique à la précédente si ce n'est le fait que les participants avaient la possibilité d'interrompre l'analyse en cliquant sur un bouton à l'écran afin d'entourer des erreurs avant que la reconnaissance ne soit terminée.



Figure 14. Les trois plans interprétés que les participants devaient contrôler

Les plans étaient composés respectivement de 33, 45 et 44 symboles (portes, fenêtres coulissantes, fenêtres à battants, meubles). En l'absence d'interruption, la reconnaissance durait 109, 96 et 70 secondes. Ces plans étaient interprétés par le prototype et les participants avaient pour consigne d'entourer les erreurs d'interprétation commises par le logiciel (voir figure 15). Chaque participant contrôlait les interprétations des trois plans différents dans les mêmes conditions expérimentales. L'ordre de présentation des trois plans était contrebalancé à l'intérieur de chaque condition expérimentale.

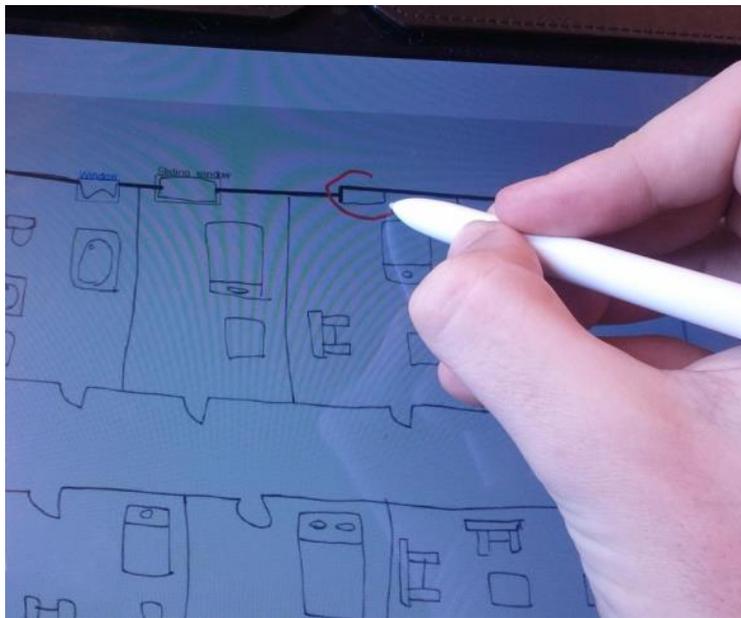


Figure 15. Participant entourant une erreur après avoir interrompu l'interprétation

## 3. Résultats

### 3.1. Avec Interruption vs Sans Interruption

Un test d'homoscédasticité de Levene ne révèle aucune différence significative entre les variances des distributions pour le nombre d'erreurs détectées ( $F(1, 34) = 0,073, p = 0,789$ ) entre le groupe SI ( $M = 15,89, SD = 3,306$ ) et le groupe AI ( $M = 14,94, SD = 3,718$ ). Une ANOVA conduite sur les performances des participants concernant la détection des erreurs ne révèle aucune différence significative,  $F(1, 34) = 0,649, MSE =$

8,028,  $p = 0,426$ . Le test de Levene sur la durée de la tâche révèle une différence significative ( $F(1, 34) = 5,700, p = 0,023$ ) entre les variances pour le groupe SI ( $M = 515,22, SD = 117,33$ ) et le groupe AI ( $M = 466,00, SD = 68,22$ ). Les différences éventuelles en termes de durée de la tâche ont donc été évaluées par le biais d'un test non paramétrique de Mann-Whitney qui ne révèle pas de différence significative,  $U = 134,5, p = 0,389$  (voir tableau 8).

Tableau 8. Durée de la tâche et erreurs détectées dans les groupes SI et AI

	Durée de la tâche		Erreurs détectées	
	$\underline{M}$	$\underline{SD}$	$\underline{M}$	$\underline{SD}$
SI	515,22	117,33	15,89	3,30
AI	466	68,22	14,94	3,71

### 3.2. Description des comportements d'interruptions

Les 18 participants du groupe AI ont effectué en moyenne 4,17 interruptions ( $SD = 2,83$ ) pendant la tâche (voir figure 16 pour la distribution des nombres d'interruptions).

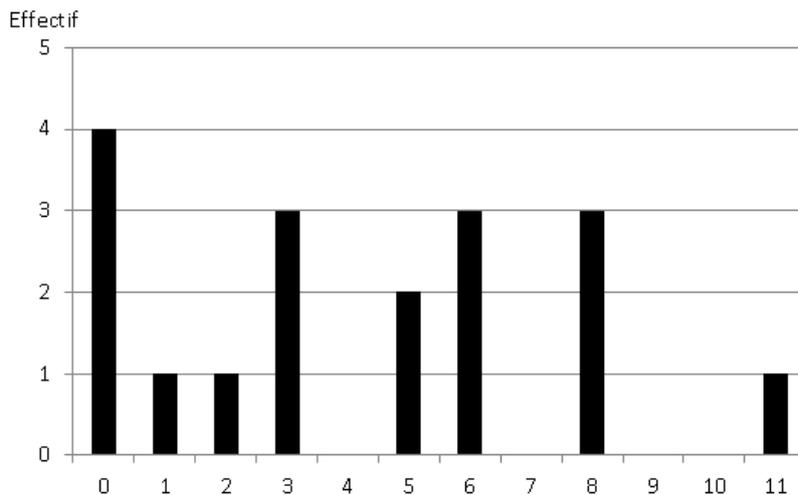


Figure 16. Distribution des nombres d'interruptions effectuées par les participants du groupe AI

Les analyses de corrélations révèlent une relation négative significative de la durée des contrôles postérieurs à la rétro-conversion avec la durée des interruptions ( $r(14) = -0,475, p = 0,046$ ) et avec le nombre d'interruptions ( $r(14) = -0,723, p < 0,001$ ). Plus les participants font d'interruptions et moins ils passent de temps à rechercher les erreurs après la fin de l'interprétation automatique.

### 3.3. Interruptions effectives

La figure 17 montre les moyennes et écarts-types des durées de la tâche en fonction du groupe (AI et SI) et de l'utilisation effective des interruptions. Les 4 participants du groupe AI n'ayant effectué aucune interruption sont regroupés sur les valeurs hautes de durées de la tâche. Si la présence à l'écran d'un bouton permettant d'interrompre l'analyse a un impact sur la performance des participants, cet impact devrait ne concerner en réalité que les participants ayant effectivement interrompu l'analyse au moins une fois.

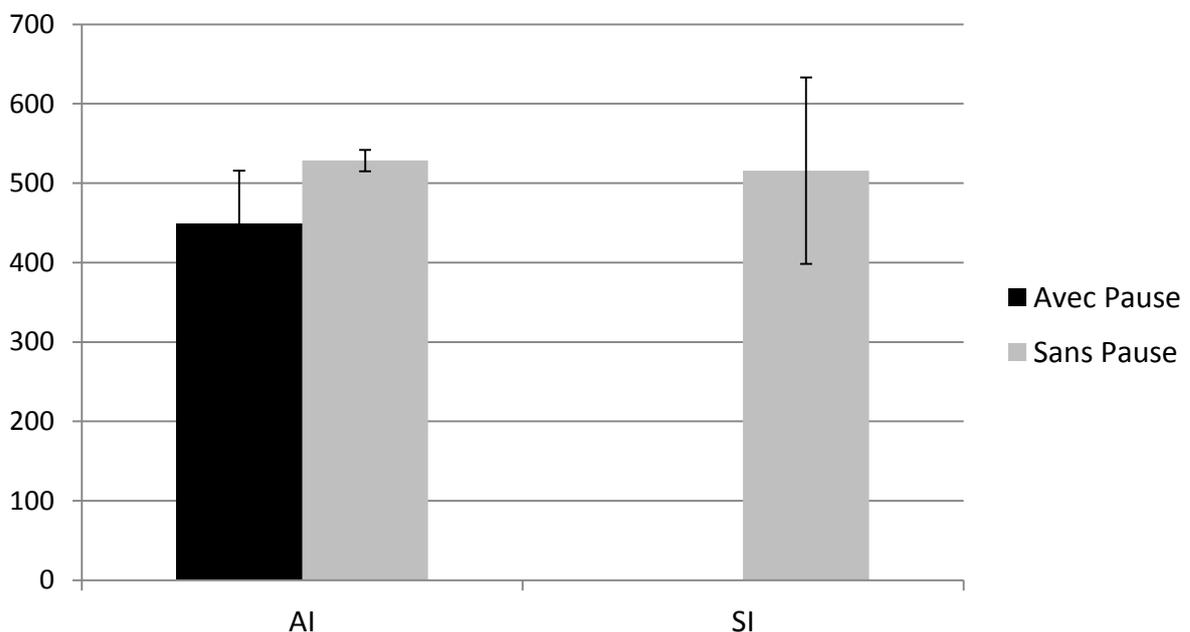


Figure 17. Moyennes et écarts-types des durées de la tâche en fonction du groupe expérimental (AI et SI) et de l'utilisation effective des interruptions

Les groupes expérimentaux SI et AI avaient été définis par la simple présence ou absence du bouton « pause » à l'écran durant la tâche. Un autre moyen de contraste a été utilisé, consistant à différencier les participants qui ont utilisé le bouton pause (groupe « avec pauses ») de ceux ne l'ayant pas utilisé (groupe « sans pause »). Le test d'homoscédasticité de Levene ne révèle pas de différence statistiquement significative entre les variances de scores de détection d'erreurs ( $F(1, 34) = 2,252, p = 0,143$ ) entre le groupe « avec pauses » ( $M = 15,64, SD = 1,985$ ) et le groupe « sans pause » ( $M = 15,27, SD = 4,233$ ). Une ANOVA conduite sur les performances des participants en détection d'erreurs ne révèle pas de différence statistiquement significative,  $F(1, 34) = 0,93, MSE = 1,172, p = 0,762$ . Un test de Levene sur les durées de la tâche ne révèle pas de différence significative ( $F(1, 34) = 2,475, p = 0,125$ ) entre les variances pour les groupes « avec pause » ( $M = 448,29, SD = 67,242$ ) et « sans pause » ( $M = 517,55, SD = 105,822$ ). Une ANOVA conduite sur la durée de la tâche révèle une différence significative entre les groupes,  $F(1, 34) = 4,747, MSE = 41040,244, p = 0,036$  (voir tableau 9). La durée moyenne de la tâche est significativement plus basse pour les participants ayant effectué des interruptions que pour ceux n'en ayant pas réalisé.

Tableau 9. Durée de la tâche et erreurs détectées avec et sans pauses

	Durée de la tâche		Erreurs détectées	
	<u>M</u>	<u>SD</u>	<u>M</u>	<u>SD</u>
Avec pauses	448,29	67,242	15,64	1,985
Sans pause	517,55	105,822	15,27	4,233

## 4. Discussion

L'hypothèse de l'étude présentée ici était que les participants à qui l'on permettrait d'interrompre la reconnaissance de symboles pour entourer les erreurs au fur et à mesure allaient en repérer plus que les autres. Cette hypothèse est invalidée de deux manières. Premièrement, il n'y a pas d'effet significatif lié à la possibilité ou non pour les participants d'interrompre l'analyse. En revanche, des résultats apparaissent lorsque les participants sont contrastés selon qu'ils aient ou non effectué des

interruptions. Deuxièmement, la différence significative observée ne concerne pas le nombre d'erreurs entourées mais la durée de la tâche. La durée moyenne de la tâche est significativement plus basse pour les participants qui ont effectué des interruptions de l'analyse que pour ceux qui ne l'ont pas fait. Ainsi, il n'y a pas d'effet de la simple présence du bouton pause à l'écran, mais l'utilisation de ce bouton a impacté la durée de la tâche. Ce résultat est appuyé par la corrélation négative entre le nombre d'interruptions et la durée des vérifications après la fin de l'interprétation.

Les participants qui n'ont pas fait d'interruptions ont entouré un nombre équivalent d'erreurs que ceux qui en ont fait. En revanche, ils les ont toutes entourées à la fin de l'analyse et non au fur et à mesure. Le fait que la différence entre les groupes concerne la durée de la tâche mais pas le nombre d'erreurs repérées laisse à penser que les participants n'ayant pas effectué d'interruption étaient conscients d'avoir peut-être oublié des erreurs. Ils auraient alors compensé ces oublis en prenant plus de temps pour réaliser les vérifications après l'analyse. Cette interprétation serait cohérente avec les résultats de l'étude de Washburn et Astur (1998) qui montrait un effet d'interférence entre un distracteur visuo-spatial et la mémorisation de configurations visuelles. Si les participants ont oublié l'emplacement de certaines erreurs à cause des interférences liées à la tâche, ils ont dû ré-explorer partiellement l'interprétation après la fin de l'analyse. Les participants n'ayant pas effectué d'interruption tendraient donc à explorer le plan deux fois : une première fois durant l'analyse, guidés par l'apparition séquentielle de l'interprétation et une seconde fois après la fin de l'analyse. Les participants ayant effectué des interruptions pour entourer les erreurs au fur et à mesure de l'analyse n'ont pas eu besoin de faire cette ré-exploration finale coûteuse en termes de temps. Ils ne contrôlaient donc qu'une seule fois l'interprétation.

On peut se demander pourquoi la différence de durées de la tâche entre les groupes apparaît significative lorsqu'on les distingue par l'utilisation effective des interruptions, mais pas lorsqu'on les distingue par la simple présence du bouton pause à l'écran. Avec les deux comparaisons, les différences vont dans le même sens, mais l'écart est plus important en tenant compte des interruptions effectives. Bien que nous ne l'ayons pas anticipé, ce résultat est facilement explicable. En effet, la présence du

bouton pause à l'écran n'apporte aucune aide aux participants s'ils ne l'utilisent pas. Les 4 participants du groupe AI n'ayant pas fait d'interruption auraient alors un profil de résultats ressemblant plus à ceux du groupe SI qu'à ceux de leur propre groupe, ce que l'on constate sur la figure 17. L'effet des interruptions serait donc dilué dans le groupe AI à cause des participants n'ayant pas utilisé cette fonctionnalité.

La principale limite de cette étude concerne le traitement des résultats. Le format de présentation (AI et SI) est une variable indépendante provoquée. Cela permet de faire l'hypothèse que les deux groupes sont équivalents en moyenne sur les caractéristiques individuelles non collectées pendant l'expérience comme par exemple le niveau de motivation pour la tâche ou encore l'état de fatigue. En revanche, l'utilisation ou non des interruptions est une variable indépendante partiellement invoquée. Certains participants n'utilisaient pas les interruptions parce que l'interface qu'ils manipulaient ne le permettait pas, mais d'autres avaient la possibilité de s'en servir ou non. L'expérience ne permet pas de déterminer quelles sont les raisons qui expliqueraient, parmi ces derniers, que certains ont choisi d'interrompre l'analyse et d'autres non. Il n'est donc pas impossible que le choix de distinguer les groupes sur la base de cette variable partiellement invoquée s'accompagne de variables confondues. Concrètement, si par exemple ce sont les participants les plus motivés pour la tâche qui ont fait le plus d'interruptions, alors la supériorité de leur performance pourrait tout aussi bien provenir de leur motivation que des interruptions elles-mêmes. Il convient donc de rester prudent quant aux interprétations réalisées sur ces résultats. Néanmoins, la manière dont se distribuent les durées de la tâche en fonction des groupes (voir figure 17), la corrélation négative entre la durée des corrections et le nombre d'interruptions, le faible nombre de participants du groupe AI à ne pas avoir fait d'interruption (4 au total) et le fait que la comparaison des groupes AI et SI aillent dans le même sens suggèrent tout de même que la cause de l'amélioration des performances est bien l'utilisation des interruptions plutôt qu'une variable confondue. En implémentant dans le prototype la possibilité d'interruption de l'analyse par les utilisateurs, le système passe du niveau « automatisation totale » au niveau « contrôle supervisé » sur l'échelle des niveaux d'automatisation de Kaber et Endsley (2003)

puisque l'humain a la possibilité d'intervenir dans la tâche lorsqu'il identifie des erreurs.

D'un point de vue applicatif, donner la possibilité aux utilisateurs d'interrompre l'analyse pour intervenir dès qu'ils le souhaitent comporte plusieurs avantages. Non seulement cela semble générer un gain de temps en amenant l'utilisateur à ne contrôler qu'une seule fois l'interprétation, mais cela peut aussi améliorer les performances du système puisque le logiciel final sera en mesure, lorsque l'utilisateur corrige une erreur, d'apprendre afin de ne pas reproduire cette erreur plus tard. Ainsi, plus une erreur est corrigée tôt et moins le logiciel en fera par la suite. Le prototype suivant, développé dans le cadre du projet MobiSketch, a donc intégré cette fonctionnalité d'interruption de l'analyse par l'utilisateur sous la forme d'un bouton pause. L'interaction entre l'utilisateur et le logiciel s'organise donc, dans une certaine mesure, comme une coopération puisque ces deux agents sont orientés vers l'objectif commun de générer une interprétation sans erreur et que pour ce faire, l'un et l'autre communiquent en temps réel et prennent en compte directement les sollicitations de l'autre pour ajuster leur propre comportement. En donnant la possibilité à l'utilisateur d'intervenir pendant l'analyse, le système passe au stade de contrôle supervisé sur l'échelle des niveaux d'automatisation de Kaber et Endsley (1997). Dans le modèle de Rasmussen (1983), la tâche de supervision de l'analyse par les participants est basée sur les règles. Autrement dit, l'apparition d'un signe spécifique à l'écran (une erreur) débloque une règle stockée en mémoire et déclenche sa mise en œuvre, *i.e.* la correction de l'erreur détectée.

# Étude 4 : Test utilisateur du prototype fonctionnel avec interruptions système

## 1. INTRODUCTION

L'expérience présentée dans la partie précédente de ce manuscrit montre que l'ajout dans le prototype d'une possibilité pour les utilisateurs d'interrompre l'analyse à n'importe quel moment, permettant donc de corriger les erreurs au fur et à mesure, peut constituer une aide. Cela leur évite de re-contrôler la totalité de l'interprétation après la fin de l'analyse à cause de l'oubli potentiel de certaines erreurs.

Comme cela a déjà été évoqué, le logiciel final sera capable d'apprendre tout au long de l'avancée dans la tâche grâce aux corrections effectuées par les utilisateurs. Du fait de cette logique de fonctionnement, le système ne devrait pas faire plusieurs fois la même erreur. Sur un plan très complexe comportant un grand nombre de symboles, corriger des erreurs tôt dans l'analyse pourrait ainsi éviter la propagation des erreurs en cascade (Ghorbel, Lemaitre, & Anquetil, 2012). Par exemple, si un symbole donné est présent à plusieurs endroits du plan et mal reconnu par le logiciel, grâce aux informations apportées par l'utilisateur, le logiciel va apprendre dès l'analyse de la première occurrence de ce symbole. Ainsi, il sera capable d'interpréter correctement les occurrences suivantes. Au total, il n'aura fallu qu'une seule intervention précoce de l'utilisateur pour éviter un ensemble d'erreurs. Sans intervention à ce moment, l'utilisateur aurait dû, à la fin de l'analyse, corriger un ensemble de mauvaises interprétations.

Bien que les prototypes testés dans les études précédentes proposent pour chaque symbole du plan une interprétation unique, ils accordent en réalité un score de

vraisemblance à cette interprétation (Ghorbel *et al.*, 2011). Ainsi, les utilisateurs allouent un niveau équivalent d'attention à la vérification de toutes les interprétations bien que le risque d'erreur ne soit pas le même sur chacune d'entre elles. On peut alors se demander si le fait de faire communiquer le système sur ses doutes quant à certaines interprétations pourrait améliorer les performances. Lorsque l'incertitude sur un symbole est importante, si le système s'arrête de lui-même pour demander à l'utilisateur de confirmer l'interprétation ou la corriger, cela devrait permettre de corriger les erreurs le plus tôt possible. Or, plus une erreur est corrigée tôt et plus le système peut en bénéficier par le biais de l'apprentissage.

Le test présenté dans cette partie n'a pas été construit sur la base d'une méthodologie expérimentale. Il s'agissait tout d'abord d'évaluer la fonctionnalité de sollicitations de l'utilisateur par le système. Cette évaluation est passée par des mesures comportementales et par un entretien pour obtenir des informations qualitatives. Il s'agissait aussi de confronter les participants à un prototype incluant l'ensemble des spécificités validées dans les études précédentes (plans intégrés, apparition séquentielle, interruptions de l'analyse par l'utilisateur) en plus des sollicitations par le système afin d'avoir un premier retour général sur l'interface. Ce retour devra être pris en compte pour améliorer l'interface y compris sur ses aspects graphiques.

## **2. METHODE**

### **2.1. Participants**

Dix-huit étudiants et jeunes actifs (3 femmes et 15 hommes) recrutés à partir de la base de volontaires du LOUSTIC ont participé à ce test. Le plus jeune avait 20 ans et 10 mois et le plus âgé avait 30 ans et 8 mois pour une moyenne de 24 ans et 9 mois ( $SD = 3,2$ ).

## 2.2. Matériel et procédure

Comme dans les études précédentes, la supervision de reconnaissance automatique était effectuée sur un Tablet PC Asus Eee Slate 12.1. Les participants devaient repérer et entourer les erreurs d'interprétation réalisées par le prototype lors de l'analyse de 3 plans successifs. Ces plans étaient identiques à ceux de l'expérience précédente. Chacun contenait 6 erreurs d'interprétations (voir figure 18). L'ordre de la présentation des plans était contrebalancé.



Figure 18. Les trois plans utilisés pour le test avec pour chacun, les six erreurs entourées en rouge

Les participants avaient la possibilité d'entourer les erreurs avec le stylet soit en interrompant l'analyse en cliquant sur un bouton « pause », soit après que la reconnaissance soit totalement terminée. De plus, lorsque le système accordait un degré de certitude trop faible à une interprétation, il interrompait l'analyse par lui-même, surlignait en bleu l'interprétation incertaine et faisait apparaître une fenêtre contenant la question « S'agit-il d'une erreur ? ». La fenêtre apparaissait toujours au même endroit, en plein centre de l'écran. Si les participants cliquaient sur « non », le surlignage bleu disparaissait et l'analyse reprenait à l'endroit où elle avait été interrompue. Si les participants cliquaient sur le bouton « oui », alors non seulement l'analyse reprenait là où elle s'était arrêtée, mais en plus le symbole qui était, selon les participants, mal interprété était automatiquement entouré en rouge. Ainsi, à la fin de l'interaction, toutes les erreurs d'interprétations devaient être entourées en rouge, soit par les participants avec le stylet, soit par le système de sollicitations supervisé par les participants (voir figure 19).

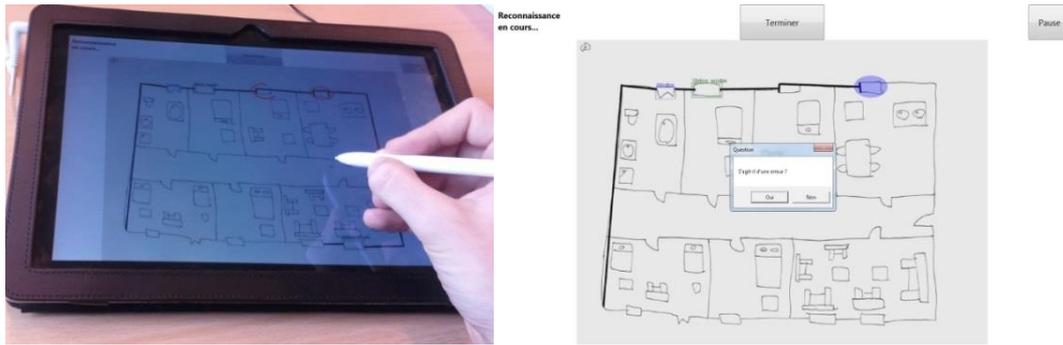


Figure 19. Le participant entoure une erreur avec le stylet (à gauche), le système sollicite l'utilisateur pour confirmer une interprétation incertaine (à droite)

Cependant, les sollicitations par le système ne renvoyaient pas toutes à des erreurs d'interprétations. En tout, le système sollicitait les participants 6 fois pour chacun des trois plans. Sur ces 6 sollicitations par plan, 3 concernaient des erreurs d'interprétation et 3 concernaient donc des interprétations justes ne devant donc pas être entourées. De plus, 3 erreurs par plans ne faisaient pas l'objet de sollicitation par le système.

Après avoir supervisé la reconnaissance des 3 plans, les participants étaient interrogés sous la forme d'un entretien semi-directif sur les qualités de l'interface. Un ensemble de 4 questions ouvertes étaient posées :

- « Qu'est-ce qu'on pourrait changer (ou ajouter, ou enlever) dans ce logiciel pour vous faciliter le repérage des erreurs ? »
- « Avez-vous utilisé le bouton « pause » ? »
- « Si oui, pensez-vous que cela vous a aidé / Sinon, pourquoi ? »
- « Lorsque le système était dans l'incertitude, il s'arrêtait pour vous demander s'il avait fait une erreur ou non. Est-ce que cela vous a plutôt aidé ou plutôt gêné ? »

# 3. RESULTATS

## 3.1. Données comportementales

### 3.1.1. Interruptions

Sur les 18 participants de l'échantillon, 7 n'ont pas interrompu l'analyse (voir figure 20). Cette moindre utilisation de la fonction d'interruption du système par rapport à l'expérience précédente pourrait provenir de l'ajout des sollicitations qui constituent aussi des interruptions en présence d'erreurs. La durée moyenne des interruptions du système par l'utilisateur était de 9,09 secondes ( $SD = 10,75$ ).

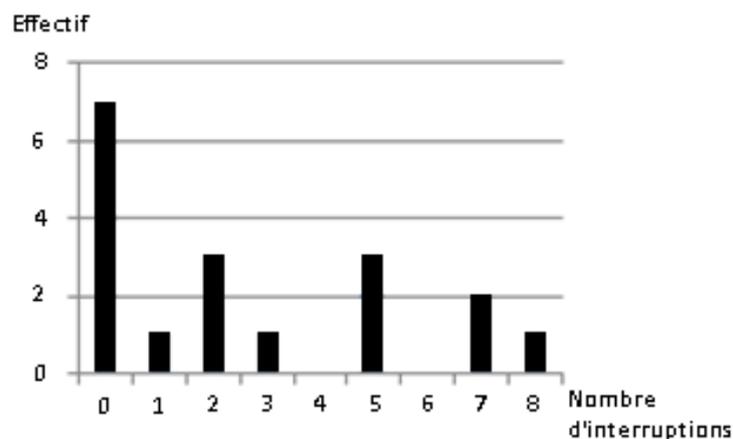


Figure 20. Effectifs en fonction du nombre d'interruptions

Pour affiner les analyses des données, les nombres d'interruptions ont été soumis à une analyse hiérarchique de clusters (Borgen & Barnett, 1987) afin de constituer des groupes de participants en fonction du nombre d'interruptions réalisées. La coupure de l'arbre hiérarchique permet d'obtenir une partition de deux classes, la première étant constituée des valeurs situées entre 1 et 3 alors que la seconde classe étant composée des valeurs allant de 5 à 8 (aucun participant n'ayant réalisé 4 interruptions). Les données de cette étude ont donc été traitées en distinguant les participants n'ayant fait aucune interruption, ceux ayant effectué 1, 2 ou 3 interruptions et ceux ayant fait entre 5 et 8 interruptions. Les 3 groupes ainsi

constitués comptaient respectivement 7, 5 et 6 participants. Les délais avant la première interruption en fonction des groupes sont présentés dans la figure 21.

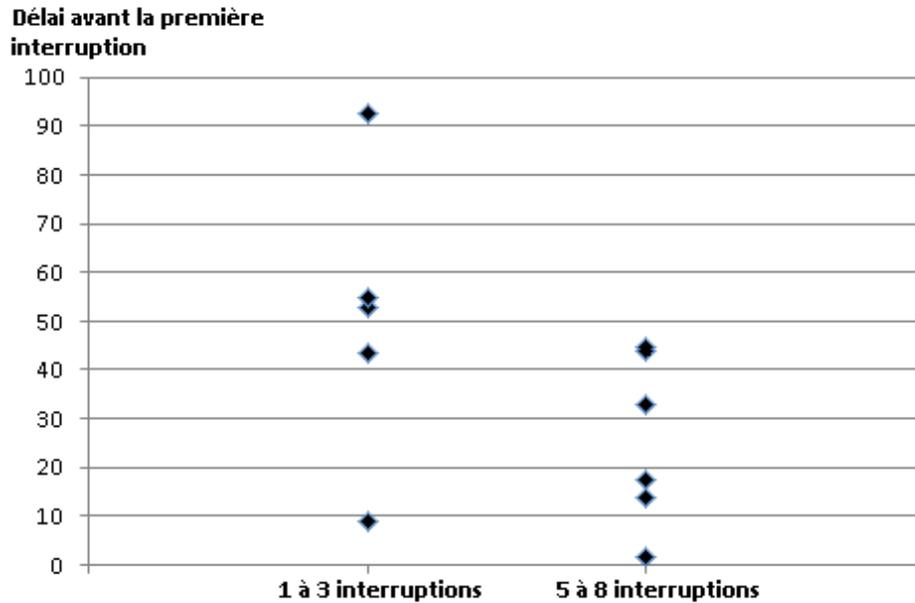


Figure 21. Délai avant la première interruption en fonction du groupe

Un test de Levene pour la moyenne des délais avant la première interruption révèle que l'hypothèse de l'équivalence des variances entre les groupes est acceptable ( $F(1, 9) = 0,273, p = 0,614$ ). Bien que la moyenne des délais avant la première interruption soit presque deux fois plus élevée dans le groupe « 1 à 3 interruptions » ( $M = 50,51$  ;  $SD = 29,83$ ) que dans le groupe « 5 à 8 interruptions » ( $M = 25,81$  ;  $SD = 17,53$ ), la différence testée par une ANOVA n'apparaît pas significative ( $F(1, 9) = 2,941, p = 0,121$ ). Le très faible effectif des groupes peut expliquer cette absence de différence.

Le test de Levene pour la durée moyenne des interruptions ne permet pas d'accepter l'hypothèse d'homoscédasticité ( $F(1, 9) = 4,715, p = 0,058$ ). C'est donc par le biais d'un test non paramétrique de Mann-Whitney que la comparaison des groupes a été effectuée. Bien que la durée moyenne des interruptions soit presque deux fois plus élevée dans le groupe « 1 à 3 interruptions » ( $M = 12,4$  ;  $SD = 15,69$ ) que dans le groupe « 5 à 8 interruptions » ( $M = 6,5$  ;  $SD = 3,21$ ), la différence testée par le U de Mann-Whitney n'apparaît pas significative ( $p = 1,00$ ). Cette absence de différence et

cette valeur de  $p$  aussi élevée peut s'expliquer à la fois par le faible nombre de participants et par le fait que le test de Mann-Whitney ne consiste pas à comparer des moyennes ou des variances mais la somme des rangs de chaque groupe (Fagerland & Sandvik, 2009). Ainsi, les différences de moyennes peuvent sembler importantes sans que cela n'apparaisse significativement différent à travers ce type de test non paramétrique.

Une analyse corrélationnelle a été conduite afin d'évaluer l'impact éventuel du nombre d'interruptions sur l'interaction en étant moins pénalisé par le faible nombre de participants. Cette analyse révèle un lien positif et significatif entre le délai d'apparition des interruptions et leur durée ( $r(14) = 0,223, p = 0,045$ ). Plus une interruption était effectuée tard dans l'interaction et plus elle durait longtemps. L'analyse révèle une autre corrélation significative mais négative entre le rang des interruptions et leur durée ( $r(14) = -0,278, p = 0,012$ ). Autrement dit, les premières interruptions effectuées par les participants pour un plan sont plus longues que les interruptions suivantes. En dépit de la non-significativité des comparaisons de groupes précédentes, ces deux résultats semblent indiquer que les participants ayant fait peu d'interruptions les auraient faits plutôt tardivement et plutôt longtemps par rapport à ceux en ayant fait beaucoup. Cela correspondrait à une stratégie consistant à n'interrompre le système que lorsque plusieurs erreurs se sont accumulées sur le plan. Ces interruptions seraient donc tardives, peu nombreuses et longues. À l'inverse, des participants interrompant le système à chaque erreur détectée interviendraient plus souvent, plus tôt et plus brièvement.

### 3.1.2. Durée de la tâche

La durée totale de la tâche est constituée de l'addition de la durée de l'analyse, des périodes d'interruption du système, des périodes de sollicitation par le système et des vérifications a posteriori (*i.e.* tout le temps qui s'écoule entre la fin de l'analyse et la validation finale de l'interprétation par l'utilisateur). En cumulant les données concernant les trois plans passés par les participants, on constate que le temps additionné des vérifications a posteriori, des sollicitations et des interruptions

représentent moins de la moitié du temps total de réalisation de la tâche (45,62%). Les participants passaient 6,5% de la durée de l'épreuve en état d'interruption, 17,16% en sollicitation et 22,51% à effectuer les vérifications a posteriori. Néanmoins, ces valeurs sont différentes selon la stratégie utilisée (voir figure 22).

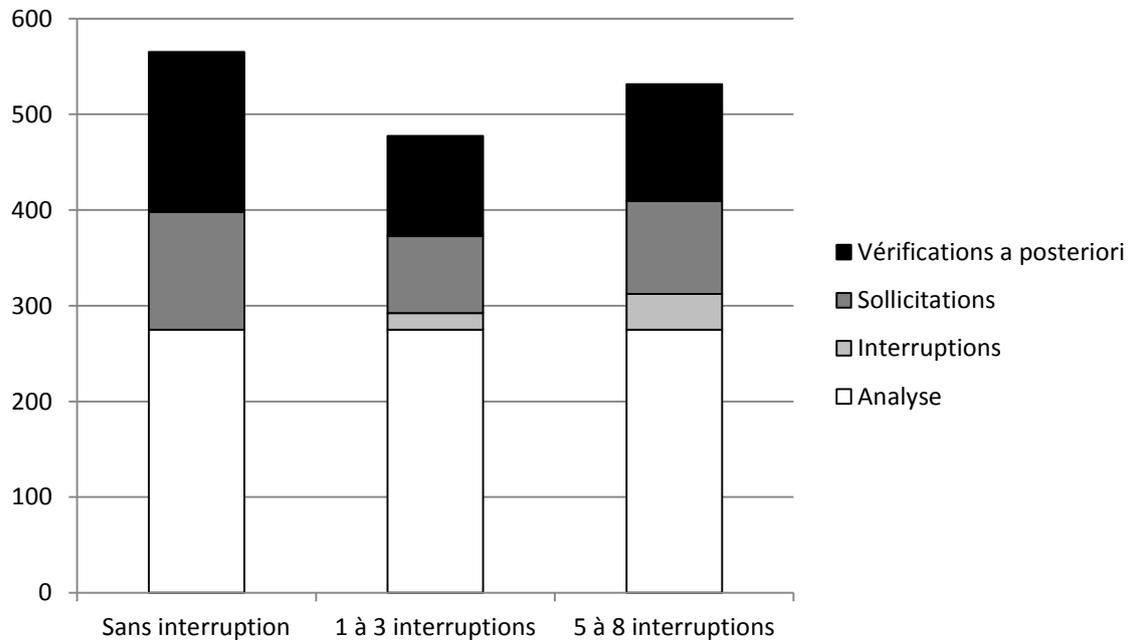


Figure 22. Durée moyenne en secondes de l'analyse, des interruptions, des sollicitations et des vérifications a posteriori pour les 3 plans cumulés

Une analyse de variances a été réalisée afin de vérifier l'impact éventuel de la stratégie (nombre d'interruptions effectuées) sur la durée des différentes étapes (vérification a posteriori, sollicitations et durée totale). L'homoscédasticité préalablement testée par la statistique de Levene révèle que l'hypothèse de l'égalité des variances est acceptable pour la durée totale ( $F(2, 15) = 0,601, p = 0,561$ ), la durée des sollicitations ( $F(2, 15) = 0,914, p = 0,422$ ) et la durée des vérifications a posteriori ( $F(2, 15) = 0,209, p = 0,813$ ). L'ANOVA ne permet pas de mettre en évidence d'effet de la stratégie sur la durée totale ( $F(2, 15) = 1,907, p = 0,183$ ), ni sur les vérifications a posteriori ( $F(2, 15) = 1,789, p = 0,201$ ). En revanche, elle révèle un effet significatif des stratégies sur la durée des sollicitations ( $F(2, 15) = 4,318, p = 0,033$ ). Les analyses de contrastes sur la variable dépendante « durée des sollicitations » montrent une différence significative entre les groupes « sans interruption » et « 1 à 3 interruptions » ( $t(15) = -2,869, p =$

0,012), mais pas entre les groupes « 1 à 3 interruptions » et « 5 à 8 interruptions » ( $t(15) = 1,092, p = 0,292$ ). Les participants dont la stratégie consistait à n'utiliser que peu d'interruptions ont donc répondu plus rapidement aux sollicitations du système. Ce résultat est discuté à la fin de ce chapitre.

### 3.1.3. Repérage des erreurs

À la fin de la tâche, une sauvegarde de l'image était réalisée automatiquement avec les symboles entourés soit avec le styler par les participants, soit automatiquement à travers une sollicitation. Puisqu'une erreur sur deux faisait l'objet d'une sollicitation, la correction de ces dernières avec et sans sollicitations a été comptabilisée de manière à prendre en compte cette distinction. Chaque participant était ainsi confronté en tout à 9 erreurs non signalées (3 pour chacun des 3 plans) et autant d'erreurs signalées par des sollicitations. Les participants ont repéré en moyenne 8,06 des 9 erreurs ayant fait l'objet d'une sollicitation ( $SD = 0,87$ ) alors qu'ils n'en ont repéré que 7,06 ( $SD = 1,76$ ) pour les erreurs non signalées par le système (voir figure 23). Les différences observées en termes de nombre d'erreurs repérées en fonction des stratégies sont très faibles (voir annexe 4).

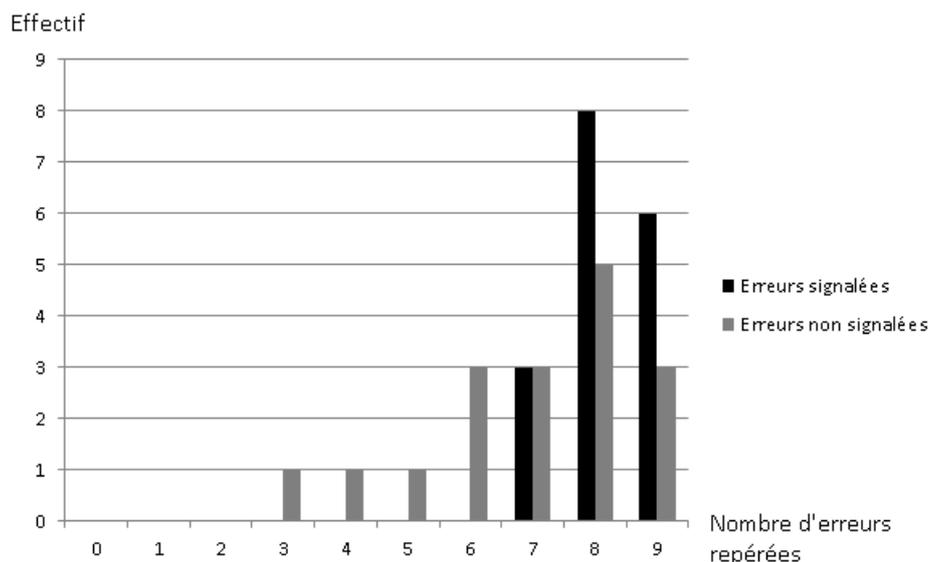


Figure 23. Distribution du nombre cumulé d'erreurs corrigées sur les 3 plans en distinguant les erreurs ayant fait l'objet d'une sollicitation de celles n'ayant pas fait l'objet de sollicitation

Un rapport de vraisemblance a été calculé pour évaluer l'impact des stratégies d'interruption sur le repérage des erreurs. Ce test n'a révélé aucun effet significatif sur le repérage des erreurs signalées par le système ( $LR(2, N = 18) = 0,170, p = 0,918$ ), ni sur celui des erreurs non signalées ( $LR(2, N = 18) = 0,283, p = 0,868$ ).

## **3.2. Données de l'entretien**

### **3.2.1. Interruptions du système**

A la question « Avez-vous utilisé le bouton « pause » ? », 11 participants ont répondu « oui » et 7 participants ont répondu « non ». Ce résultat est cohérent avec les observations des utilisations des interruptions à travers les fichiers log. Il était demandé aux participants d'expliquer pourquoi ils n'avaient pas utilisé les interruptions. L'ensemble de leurs réponses est présenté en annexe 5. Les commentaires ont été regroupés en fonction des idées auxquelles ils faisaient référence. Ce classement était réalisé en l'aveugle par trois expérimentateurs dans les catégories préalablement définies. Le taux de congruence des classements pour cette question était de 96%. Trois idées principales ressortaient de ces commentaires : la simplicité des plans (4 occurrences), la lenteur de l'analyse (1 occurrence) et le fait que les interruptions pourraient gêner la concentration (3 occurrences). Ainsi, la réponse la plus fréquente énoncée par les participants suggère que le plan était suffisamment simple pour ne pas avoir besoin des interruptions.

Parmi les 11 participants à avoir utilisé les interruptions, 9 ont prétendu que cela les avait aidé et les deux autres ont répondu que cela ne les avait ni gêné, ni aidé. Ainsi, aucun participant ayant utilisé les interruptions n'estime que cela l'ait gêné. Les deux participants ayant répondu que les interruptions ne les avaient ni aidé ni gêné ont justifié leur réponses de la manière suivante : « Cela oblige à quitter le plan des yeux pour mettre pause, on se perd » (renvoie à la catégorie « Gêne pour la concentration ») et « Je l'ai fait pour suivre au fur et à mesure, mais ça aurait été pareil de tout vérifier à la fin » (réponse ambiguë non catégorisée).

Concernant les commentaires des participants estimant avoir été aidés par les interruptions, ils ont été classés en deux catégories avec un taux de congruence de 92% : « Régulation de la vitesse de l'analyse » (2 occurrences) et « Difficulté de mémorisation de toutes les erreurs » (2 occurrences). Selon les participants, les interruptions les aideraient en leur fournissant la possibilité de réguler le rythme de l'analyse, mais aussi en leur permettant d'éviter l'oubli d'erreurs (voir annexe 6). Cette seconde raison est conforme aux interprétations des résultats proposées dans la partie précédente de ce manuscrit.

Pour résumer, lorsque les interruptions ne sont pas utilisées, c'est souvent parce que le plan est jugé suffisamment simple pour ne pas avoir besoin d'aide. En revanche, l'utilisation des interruptions permettrait de réguler le rythme de l'analyse lorsque celui-ci est trop rapide, mais aussi de ne pas oublier des erreurs précédemment détectées.

### 3.2.2. Sollicitations de l'utilisateur

La question suivante était posée aux participants durant l'entretien : « Lorsque le système était dans l'incertitude, il s'arrêtait pour vous demander s'il avait fait une erreur ou non. Est-ce que cela vous a plutôt aidé ou plutôt gêné ? ». À cette question, 10 participants ont répondu que cela les a plutôt aidés, 8 ont indiqué que cela les a plutôt gênés.

Les justifications des réponses pour les participants que les sollicitations ont gêné ont été classées en 6 catégories avec une congruence de 97% : « Sollicitations après plutôt que pendant » (2 occurrences), « fenêtre (boîte de dialogue) mal positionnée » (5 occurrences), « ne montre pas que des erreurs » (1 occurrence), « tâche simple » (1 occurrence) et « surlignage gênant » (1 occurrence) et « relâchement de l'attention » (1 occurrence). Les deux raisons principales qui semblent expliquer un jugement négatif sur les sollicitations sont donc le fait que les participants préféreraient qu'elles aient lieu après l'analyse plutôt que pendant et que le positionnement des fenêtres de sollicitations est jugé peu pertinent (voir annexe 7).

Concernant les participants prétendant que les sollicitations les ont aidé, les justifications des réponses ont été classées en 2 catégories avec une congruence de 100% : « Facilite la détection des erreurs » (5 occurrences) et « ralenti l'analyse » (1 occurrence). L'aide apportée par les sollicitations serait donc principalement une facilitation de la détection des erreurs (voir annexe 8).

Selon les participants interrogés, les sollicitations de l'utilisateur auraient donc pour principal avantage de faciliter la détection des erreurs mais pourrait notamment présenter l'inconvénient de parfois gêner la vérification à cause d'un mauvais emplacement de la fenêtre de dialogue.

### 3.2.3. Remarques générales

Durant l'entretien, la question suivante, ouverte et très générale, était posée aux participants : « Qu'est-ce qu'on pourrait changer (ou ajouter, ou enlever) dans ce logiciel pour vous faciliter le repérage des erreurs ? ». Les réponses des participants ont été réparties dans trois catégories, « visibilité des étiquetages des symboles » (4 occurrences), « ordre de l'analyse » (3 occurrences) et « trop rapide » (1 occurrence) avec un accord inter-juges de 96% (voir annexe 9).

Les deux éléments qu'il semble utile d'améliorer pour les participants sont l'ordre d'apparition des interprétations de symboles pour le rendre plus cohérent du point de vue de l'utilisateur et les caractéristiques graphiques de l'étiquetage des symboles.

## 4. DISCUSSION

La présente étude avait pour objectif de confronter les participants à une version mise à jour du prototype auquel avait été ajoutée une fonctionnalité de sollicitations des utilisateurs. Cette confrontation suivie d'un entretien devait fournir des informations concernant les problèmes d'utilisabilité de l'interface afin d'apporter des corrections à cette dernière.

Les résultats comportementaux semblent indiquer l'existence de trois principales stratégies qui sous-tendraient l'utilisation des interruptions. La première consisterait à

ne pas effectuer d'interruption. L'entretien révèle que les deux principales raisons qui motivent l'utilisation de cette stratégie sont la simplicité perçue des plans utilisés qui laissent penser que les interruptions ne sont pas utiles et le fait que l'utilisation du bouton « pause » risquerait de décentrer l'attention des participants de l'analyse en cours.

Parmi les utilisateurs des interruptions, certains prétendaient lors de l'entretien que cela leur permettait de réguler le rythme de l'analyse. Autrement dit, pour certains participants, les interruptions permettaient de contrôler des interprétations qui seraient apparues trop rapidement afin de pouvoir les vérifier au fur et à mesure de l'analyse. Cela pourrait indirectement expliquer pourquoi les participants n'ayant pas fait d'interruptions ont été plus lent pour répondre aux sollicitations. En effet, les sollicitations par le système constituaient aussi des interruptions de l'analyse. Celles-ci ont pu permettre aux participants de vérifier les interprétations qu'ils n'ont pas pu vérifier au fur et à mesure de l'analyse. Les sollicitations auraient alors servi à certains participants à réguler le rythme de l'analyse. La seconde stratégie est intermédiaire. Elle consiste à effectuer un faible nombre d'interruptions tardives. Lorsqu'un participant applique cette stratégie, il laisse s'accumuler quelques erreurs et interrompt alors l'analyse pour les entourer. Enfin, la troisième stratégie consiste à interrompre l'analyse au fur et à mesure de l'interaction pour entourer les erreurs détectées.

Cette interprétation des stratégies est renforcée par les résultats montrant que les participants ayant fait des interruptions tardives les ont faites plus longues et moins nombreuses. En revanche, les résultats ne permettent pas de mettre en évidence un impact de la stratégie (opérationnalisée par le nombre d'interruptions effectuées) sur la performance de détection d'erreurs.

La majorité des participants n'ayant pas utilisé la fonctionnalité d'interruption de l'analyse le justifiait par le fait que les plans traités étaient suffisamment simples pour ne pas en avoir besoin. Autrement dit, l'interruption pourrait être perçue comme une aide face à des plans plus complexes. L'aide apportée par cette fonctionnalité peut être de deux types : elle permet de réguler le rythme de l'analyse et d'entourer les

erreurs au fur et à mesure de l'analyse pour ne pas en oublier. Ces premiers éléments sont plutôt encourageants pour l'utilité de maintenir la fonction d'interruptions dans le prototype. Néanmoins, plusieurs participants n'ayant pas utilisé les interruptions ont déclaré ne pas l'avoir fait parce que cela aurait pu décentrer leur attention de l'analyse. Rappelons que pour interrompre l'analyse, les participants devaient cliquer sur un bouton qui se situait dans le coin haut/gauche de l'écran. Les commentaires des participants laissent penser qu'alors qu'ils contrôlent les interprétations au fur et à mesure en se laissant guider par l'analyse, interrompre le système les obligerait à quitter des yeux l'interprétation en cours pour cliquer sur le bouton.

Une solution a été proposée et implémentée dans le prototype pour régler ce problème : lorsque les participants cliquent où-que-ce-soit sur le plan, l'analyse s'interrompt. La décision de créer ce raccourci est cohérente avec le principe des actions minimales décrit par Bastien et Scapin (1993) car cela élimine une des actions à réaliser par l'utilisateur pour corriger une erreur. C'est aussi en accord avec le critère de flexibilité puisque cela permet aux utilisateurs de réaliser la même action de différentes manières (Bastien & Scapin, 1993). De plus, cela empêche la décentration de l'attention des participants qui, pour trouver du regard le bouton sur lequel cliquer perdraient le fil de la reconnaissance automatique. Ceux-ci auraient pris le risque de ne pas bénéficier de l'avantage que constitue le guidage résultant de la présentation séquentielle des informations. Enfin, ce raccourci est invisible et ne constitue donc en aucun cas une gêne pour les utilisateurs qui ne l'utiliseraient pas, comme le recommandent Molich et Nielsen (1990) ainsi que Nielsen (1994) avec le critère de flexibilité et efficacité de l'utilisation (*flexibility and efficiency of use*).

Concernant les sollicitations par le système, l'entretien révèle que 10 participants estiment qu'elles les ont aidé alors que 8 d'entre eux considèrent qu'elles les ont plutôt gêné. La majorité des participants dont le jugement était favorable pour cette fonctionnalité considéraient que celle-ci leur apportait une aide pour détecter les erreurs. Ce résultat est cohérent avec les performances des participants qui repéraient effectivement plus d'erreur signalées que d'erreurs non signalées. Ce dernier résultat est tout de même à considérer avec beaucoup de précautions car les erreurs signalées

et les erreurs non signalées étaient toujours les mêmes pour tous les participants. Autrement dit, il est possible qu'indépendamment de leur signalement éventuel, les erreurs ne soient pas toutes aussi faciles à identifier les unes que les autres. Les réponses des participants fournissent plusieurs pistes pour l'amélioration des sollicitations. Premièrement, cinq d'entre eux ont évoqué le fait que l'emplacement de la fenêtre dans laquelle apparaît la question « s'agit-il d'une erreur ? » était gênant. Le choix avait été fait de faire apparaître cette fenêtre à un endroit facile à trouver et stable : au centre de l'écran. Deux difficultés découlent de ce choix. La première est qu'il peut y avoir du contenu à traiter au centre de l'écran et dans ce cas, la fenêtre risque de cacher ce contenu. La seconde difficulté vient du fait que les participants doivent traiter deux informations qui sont potentiellement éloignées l'une de l'autre : la fenêtre et le symbole à contrôler. D'ailleurs, un participant a aussi évoqué le fait que surligner la zone à traiter pouvait rendre le symbole et son interprétation moins lisibles. Des améliorations ont été apportées pour régler à la fois le problème du surlignage et de l'emplacement de la fenêtre de sollicitation. Ces modifications ont consisté à définir pour chaque erreur l'emplacement de la fenêtre de sollicitation en fonction de l'endroit où se trouve l'interprétation à vérifier. Puisque la fenêtre contient des informations qui font référence à une information sur le plan, le fait de rapprocher ces deux sources visuelles devrait en faciliter la co-référenciation. Le surlignage de la zone à contrôler a été remplacé par un pointage de la fenêtre de sollicitation vers le symbole sous la forme d'une flèche (voir figure 24). Cette solution présente l'avantage de ne pas altérer la lisibilité des informations à contrôler. Les sollicitations système utilisées dans cette étude laissaient passer 50% des erreurs et produisait 50% de sollicitations sur des interprétations sans erreur. Cette précision est plutôt faible par rapport à ce qui est utilisé par exemple dans les expériences de Wiegmann, McCarley, Kramer et Wickens (2006) ou de Goh, Wiegmann et Madhavan (2005) avec un assistant automatique de détection d'une arme dans des images à rayons X de bagages. Vraisemblablement, l'augmentation de la précision des sollicitations dans les prototypes à venir devrait avoir un impact sur les performances. Non seulement cela pourrait améliorer la détection des erreurs, mais cela pourrait aussi accentuer le

phénomène de relâchement de l'attention évoqué par un participant qui parlait des sollicitations (« C'est une économie de réflexion pour nous, pas besoin de survoler minutieusement le plan »). Ce questionnement est approfondi dans la partie suivante de ce manuscrit.

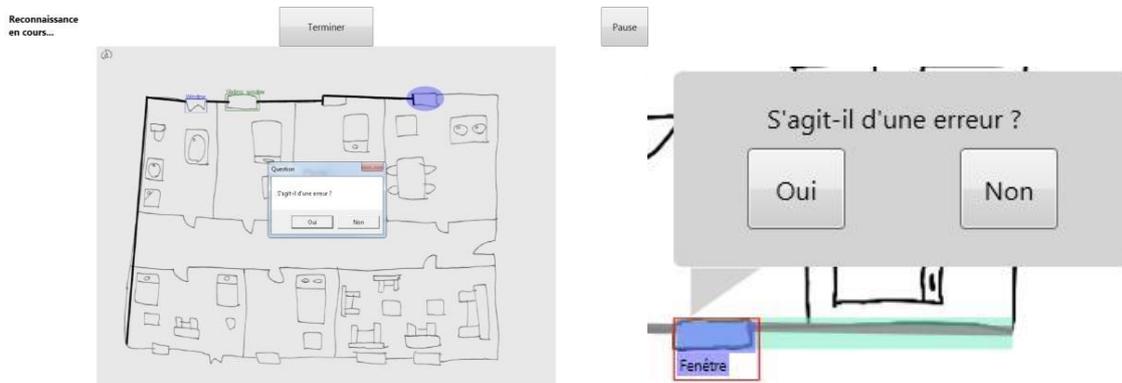


Figure 24. Sollicitations telles qu'elles étaient affichées lors du test (à gauche) et telles qu'elles ont été implémentées suite au test (à droite)

À la question portant sur les améliorations possibles du système, 4 participants ont mentionné le manque de visibilité des cadres utilisés pour spécifier qu'un symbole a été interprété. La couleur du cadre constituait un indice concernant l'interprétation réalisée puisque chaque interprétation possible était associée à une couleur différente. Comme le soulèvent les réponses des participants, conserver une couleur différente par interprétation pose le problème de diminuer la différence entre les couleurs utilisées, si bien qu'elles en deviennent plus difficiles à discriminer. Seules deux couleurs ont été retenues et chacune associée à une catégorie de symboles en accord avec le principe de groupement et de distinction entre les items de Bastien et Scapin (1993). Celui-ci indique que les informations qui doivent être interprétées comme appartenant à une même catégorie devraient être proches, soit spatialement, soit en termes de caractéristiques graphiques. Le bleu a été choisi pour désigner les ouvrants (portes et fenêtres) et le vert représente les meubles (canapé, table, etc ; voir figure 25).

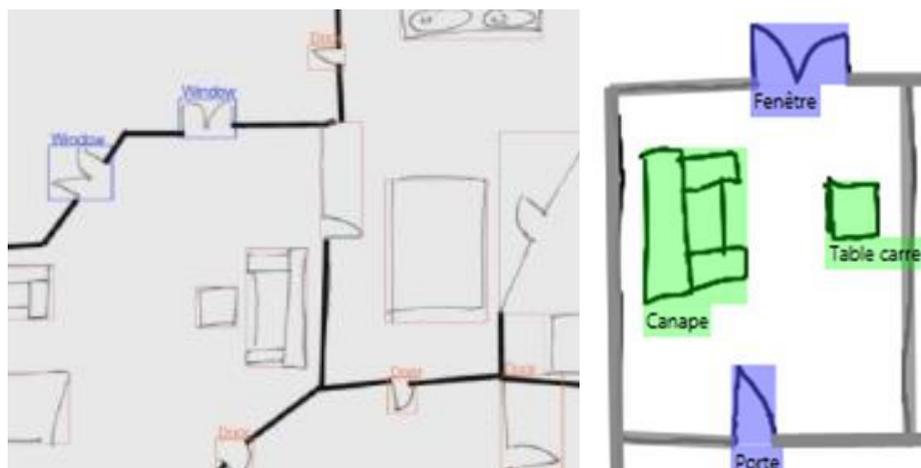


Figure 25. Graphismes des interprétations de symboles dans le prototype utilisé pour le test (à gauche) et dans la version améliorée du prototype (à droite)

À cette même question des améliorations globales éventuelles, 3 participants ont mentionné le problème de l'ordre dans lequel le plan est analysé qui selon eux ne serait pas logique. Puisqu'une part des performances à la tâche est liée au guidage attentionnel par le biais de l'apparition séquentielle de l'interprétation, un ordre plutôt intuitif d'apparition est vraisemblablement préférable. Les utilisateurs sont dans une situation dans laquelle leur niveau de conscience de l'état du système est faible (Endsley, 1995), ce qui les empêche d'anticiper et peut donc les ralentir. En fait, bien que non pertinent du point de vue de l'utilisateur, cet ordre correspond à la logique de fonctionnement du système de reconnaissance. Il est donc difficilement envisageable d'agir sur cet ordre pour le rendre plus pertinent. En revanche, la solution qui a été proposée est d'apporter des feedbacks aux utilisateurs pour communiquer sur les traitements effectués par le système et ainsi améliorer leur conscience de la situation. Cela a consisté en l'ajout d'un focus visuel sous la forme d'un halo de couleur signalant en permanence l'endroit que le système est en train d'analyser (voir figure 26). Ce faisant, le système indique à l'utilisateur quel sera le prochain symbole à interpréter. En plus d'améliorer la qualité du guidage attentionnel, cela devrait contribuer à augmenter la conscience que l'utilisateur a de l'état du système comme le préconise Endsley (1995), mais aussi Nielsen (1994) avec le critère de visibilité de l'état du système.

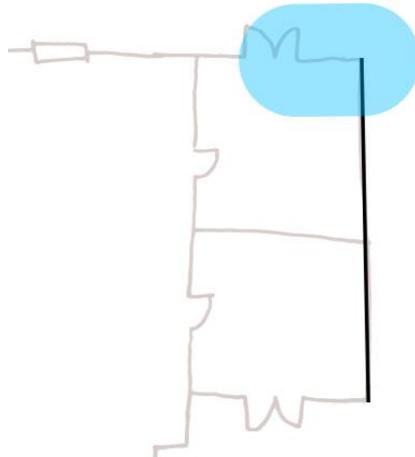


Figure 26. Focus de pré-guidage ajouté au prototype

Toutes les préconisations issues de ce test utilisateur ont été appliquées au prototype. Comme l'ont montré les études précédentes décrites dans ce manuscrit, la détection d'erreurs d'interprétation d'un plan n'est pas une tâche facile. Malgré les améliorations de l'interface, il n'est pas possible d'être sûr que les utilisateurs parviendront à repérer et à corriger toutes les erreurs, notamment pour des plans complexes. Les sollicitations de l'utilisateur par le système reposent sur l'idée que des erreurs signalées par le système seront plus probablement corrigées que les autres. Ainsi, cela serait un moyen supplémentaire d'aider les participants à corriger toutes les erreurs.

Le test présenté ici ne permettait pas d'étudier l'impact des sollicitations sur les performances de repérage des erreurs puisque les erreurs signalées n'étaient pas les mêmes que celles non signalées, cette variable n'ayant pas fait l'objet d'un contrebalancement. Une comparaison expérimentale a donc été effectuée pour combler cette lacune. De plus, un risque de relâchement de la vigilance à cause des sollicitations a été évoqué spontanément lors des entretiens (voir annexe 9). Cela suggère que les sollicitations pourraient faire l'objet d'une telle confiance par certains utilisateurs que cela pourrait générer un risque de diminution des contrôles. Cette dimension subjective qu'est la confiance dans le système a donc été prise en compte dans l'étude suivante.

# Étude 5 : Impact des sollicitations système sur la correction des erreurs<sup>4</sup>

## 1. INTRODUCTION

L'étude décrite dans la partie précédente de cette thèse a permis l'amélioration du prototype sur plusieurs aspects. D'abord, le système a été rendu plus flexible par l'ajout de raccourcis pour les interruptions. De plus, les choix graphiques des interprétations des symboles ont été affinés et un focus visuel a été implémenté afin de communiquer en permanence à l'utilisateur la zone du plan que le système est en train de traiter. Cette modification va dans le sens d'une amélioration de la conscience de la situation par l'utilisateur (Endsley, 1995). Le prototype intègre aussi une fonction de sollicitation de l'utilisateur lorsqu'une interprétation présente un risque d'erreur trop important. Cette communication du système sur ses propres doutes contribue également à augmenter la conscience de la situation par l'utilisateur. À l'instar du système de beautification décrit par Igarashi, Matsuoka, Kawachiya et Tanaka (1997), les sollicitations s'accompagnent de propositions d'interprétation alternatives dont seule la plus pertinente sera sélectionnée par l'utilisateur. De plus, l'apparition des sollicitations au fur et à mesure de l'analyse permet aux erreurs d'être corrigées plus tôt. Or, du fait des capacités d'apprentissage du système, une erreur corrigée pendant l'analyse tend à améliorer la fiabilité des interprétations suivantes (Ghorbel *et al.*, 2011). En proposant un ensemble de solutions possibles dont l'une doit être choisie par l'utilisateur, le système change de niveau d'automatisation. Il passe du « contrôle supervisé », dans lequel l'opérateur ne faisait que surveiller la machine pour

---

<sup>4</sup> Fleury, S., Jamet, E., Ghorbel, A., Lemaitre, A., & Anquetil, E. (soumis). Application of the Resources Model to the Supervision of an Automated Process

n'intervenir qu'en cas d'erreur, à un niveau que Kaber et Endsley (2003) ont appelé « prise de décision mixte ». Comme son nom l'indique, la « prise de décision mixte » est un niveau d'automatisation dans lequel les prises de décisions sont partagées. À ce niveau, le système fait un ensemble de propositions et l'utilisateur choisit la bonne réponse à apporter.

L'objectif de la présente étude est d'évaluer expérimentalement les conséquences des sollicitations sur le déroulement de l'interaction. Pour cette expérience, l'interface utilisée a été améliorée en suivant les recommandations issues de l'étude précédente, *i.e.* l'amélioration de l'aspect graphique des interprétations, l'ajout d'un focus de pré-guidage et le raccourci permettant l'interruption de l'analyse. Les sollicitations aussi ont été refondues. Elles apparaissent à proximité de l'interprétation concernée et la pointent par une flèche. De plus, la fiabilité des sollicitations a été augmentée puisque pour 6 erreurs, 4 faisaient l'objet de sollicitations. En effet, il s'agit d'un degré de fiabilité plus réaliste par rapport à ce qu'il devrait être possible d'obtenir à terme. Deux erreurs n'étaient donc pas signalées et 2 sollicitations portaient sur des interprétations qui n'étaient pas erronées. Comme le suggèrent les résultats de l'entretien de l'étude précédente, mais aussi l'expérience de Wiegmann *et al.* (2006) portant sur un assistant de détection d'arme à partir d'images de bagages obtenues rayon X, la confiance que les participants ont dans le système semble jouer un rôle important dans ce type de tâche. D'ailleurs, Kaber et Endsley (1997) décrivaient la sur-confiance dans le système comme un des symptômes de l'humain hors de la boucle. Cette dimension a donc été prise en compte dans la présente étude.

Enfin, les avancées dans le développement du système ont permis d'introduire dans le prototype utilisé la fonction de correction des erreurs. Les participants ne devaient donc plus seulement entourer les erreurs, mais il leur fallait effectuer la correction en sélectionnant la bonne interprétation. Avant la description de l'expérience, une revue de la littérature est proposée sur la détection de cibles avec assistance automatique, sur la confiance dans les systèmes automatisés ainsi que sur le modèle des ressources, (Wright, Fields, & Harrison, 1996). Ce dernier est centré sur la modélisation des interactions homme-machine à travers les ressources informationnelles distribuées.

## 1.1. Détection de cible avec assistance automatique

La tâche que les utilisateurs doivent réaliser pour superviser la reconnaissance automatique d'un plan consiste à vérifier les interprétations des différents symboles afin de détecter d'éventuelles erreurs. Ainsi, cette tâche présente certaines similitudes avec celle de détection d'une arme à partir d'images de bagages en désordre obtenues par rayon X ayant fait l'objet de plusieurs études. Dans celle conduite par Wiegmann *et al.* (2006), les participants devaient signaler pour chaque image présentée s'ils détectaient ou non la présence d'un couteau alors que 20% des images présentées contenaient effectivement un couteau. Certains participants devaient réaliser cette tâche sans aucune aide tandis que d'autres bénéficiaient d'une assistance automatique. Cette dernière, réalisée pour les besoins de l'expérience, consistait en une simulation d'un système de reconnaissance automatique de forme, capable d'identifier sur l'image la présence d'un couteau avec un taux de précision de 0,90 et un taux de fausses alarmes de 0,25. Selon la condition expérimentale des participants, l'assistance automatique pouvait se manifester de trois manières différentes : un message textuel avant l'apparition de l'image, un message textuel affiché en même temps que l'image ou enfin l'apparition d'un cercle sur l'image, entourant directement la zone dans laquelle se trouverait le couteau. Les résultats montraient que l'indice spatial (entourer directement la cible) améliore significativement la performance de détection des couteaux et qu'il s'agit, parmi les trois types d'assistances évalués, du plus efficace. Une expérience de Goh *et al.* (2005) allait dans le même sens en démontrant, sur le même type de tâche, une supériorité d'un indice direct (entourage de la cible potentielle) par rapport à un indice indirect signalant uniquement la présence supposée de la cible sur l'image sans précision quant à sa localisation.

Dans une recherche sur le même type de matériel, McCarley, Kramer, Wickens, Vidoni et Boot (2004) testaient notamment l'effet de l'entraînement sur les performances de participants novices pour cette tâche de détection de couteaux. Les résultats indiquaient que l'entraînement à cette tâche améliore les performances des

participants à la fois en termes de précision, de rapidité mais aussi de transfert (*i.e.* de repérage de couteaux de formes différentes). Ces résultats ont été confirmés par une seconde étude de Madhavan et Gonzalez (2006), dans laquelle le taux de sacs contenant des armes montait à 40%.

Dans l'étude de Wiegmann *et al.* (2006), les participants devaient spécifier le degré de confiance qu'ils accordaient à leur propre réponse concernant la présence d'une arme dans le bagage. Lorsque l'assistance automatique apportait un jugement congruent avec la réponse des participants, la confiance de ces derniers en leur propre réponse augmentait. À l'inverse, les suggestions non congruentes de l'assistant avaient tendance à diminuer la confiance des participants en leur propre réponse. Les participants accordaient donc un certain degré de confiance à l'aide apportée par l'assistance, puisque celle-ci pouvait les faire douter ou au contraire les conforter dans leur réponse. D'ailleurs, l'étude de McCarley, Wiegmann, Wickens et Kramer (2003), basée sur la même tâche, montrait que les participants avaient tendance à sous-évaluer le degré de précision de l'assistance automatique. L'effet d'une assistance automatique a aussi été évalué avec une tâche de contrôle de paramètres de vol d'un avion (Skitka, Mosier, & Burdick, 1999). Dans cette expérience, une assistance automatique tend à améliorer la précision des participants lorsque ces sollicitations sont pertinentes et à les diminuer lorsque ses interventions sont à mauvais escient. Les cas dans lesquels l'assistance a un impact négatif sur les performances sont considérés par les auteurs comme des erreurs de complaisance (Skitka *et al.*, 1999). Cela consisterait à accorder à l'assistance automatique une confiance excessive au regard de sa fiabilité. Le même type de résultat a été obtenu par Alberdi, Povyakalo, Strigini, Ayton, & Given-Wilson (2008) dans une étude portant sur une assistance automatique à la détection de lésion sur des mammographies. Là encore, les performances de détection par les participants augmentent lorsque l'assistance fournit des informations adéquates, mais diminuent lorsque l'assistance apporte des informations non pertinentes. Cependant, ni Skitka *et al.* (1999), ni Alberdi *et al.* (2008) n'ont effectué de mesure de la confiance afin de confirmer les interprétations de leurs résultats. Cette thématique de la confiance en le système est centrale lorsque l'on s'intéresse à

l'utilisation d'assistance automatique. Non seulement le degré de précision de l'assistance peut avoir un impact sur la confiance, mais cette dernière pourrait impacter à son tour la manière dont l'utilisateur va se comporter avec le système.

## 1.2. Le modèle des ressources

Le logiciel en cours de développement est d'une spécificité telle qu'il n'existe pas, à notre connaissance, de modèle théorique concernant ce type de système. Néanmoins, un modèle plus général traitant de la cognition comme d'un phénomène distribué, le modèle des ressources (Wright *et al.*, 1996; Fields, Wright, & Harrison, 1997; Wright, Fields, & Harrison, 2000), pourrait constituer un cadre d'analyse pertinent. La tâche réalisée par les participants dans l'expérience a donc été analysée dans une perspective de cognition distribuée, à travers le modèle des ressources. Ce modèle décrit le système constitué par l'humain en interaction avec la machine plutôt que de se centrer uniquement sur le fonctionnement cognitif de l'utilisateur. Le système ainsi décrit contient un certain nombre de ressources informationnelles qui peuvent être stockées dans la mémoire de l'utilisateur (*user-embedded*), dans la mémoire de la machine ou même sur un manuel papier (*device-embedded*).

### 1.2.1. Les ressources informationnelles

Wright *et al.* (1996) ont listé les différents types de ressources qui peuvent être contenus dans le système : les plans, l'état de l'interaction, les buts, les affordances, l'historique et les relations action-effet.

Le plan est une séquence d'actions, d'événements ou d'états. Il peut être interne, stocké en mémoire, ou externe, comme sous la forme d'une procédure écrite étape par étape. Le plan peut être disponible dès le début de la tâche sous forme écrite ou peut être construit par l'utilisateur durant l'interaction.

L'état de l'interaction, à un instant donné, correspond à l'ensemble des valeurs des variables pertinentes pour caractériser l'interaction. Par exemple, sur un site web, l'état correspond au fait d'être sur une page donnée du site. De plus, le site peut être

en train de charger des fichiers ou non. La transition d'un état à un autre peut être consécutive à une action de l'utilisateur, ou résulter d'une action de la machine elle-même. L'état peut être externe (par exemple, visible à l'écran) ou non. Ainsi, si l'utilisateur sait généralement sur quelle page d'un site web il se trouve, les chargements de fichiers qui ont lieu pendant qu'il navigue ne sont pas nécessairement signalés (état interne).

Les buts sont une description de l'état d'interaction à atteindre. Cette description peut être externe *i.e.* représentée à l'écran ou dans un manuel. Elle peut aussi être interne, *i.e.* stockée dans la mémoire de l'utilisateur. Malgré leurs ressemblances, les buts et le plan sont deux ressources différentes. Le plan est une procédure, une séquence d'événements qui aboutissent aux buts, tandis que les buts ne sont que la description de l'état visé. Par exemple, dans la tâche des Tours de Hanoï, les buts correspondent à la configuration finale que doit atteindre le joueur. Le plan renvoie quant à lui à l'ensemble des déplacements de disque que celui-ci doit effectuer pour atteindre ces buts.

Les affordances, dans le modèle des ressources, correspondent à l'ensemble des actions possibles que l'utilisateur peut réaliser pour un état donné de l'interaction. Ce sont les actions possibles qui sont identifiables par l'utilisateur. Par exemple, un bouton sur lequel est écrit ce qui se produit lorsque l'utilisateur le presse constitue une affordance.

L'historique est l'ensemble des actions, événements ou états, qui ont eu lieu depuis le début de l'interaction. Il peut être stocké dans la machine, permettant la fonction « *undo* » qui annule la dernière action réalisée. Il peut aussi être stocké dans la mémoire de l'utilisateur. L'historique ressemble au plan puisqu'il correspond à une séquence d'actions, d'événements ou d'états. Cependant, contrairement au plan, l'historique ne peut pas contenir de boucles conditionnelles puisqu'il décrit des actions, événements ou états qui ont eu lieu dans le passé.

Enfin, les relations action-effet sont l'ensemble des relations entre des actions de l'utilisateur et des changements dans l'état de l'interaction. Par exemple, dans beaucoup de programmes, l'action « cliquer sur la croix » a pour effet de « fermer la

fenêtre ». Les relations action-effet peuvent être externes (dans un manuel) ou internes (dans la mémoire de l'utilisateur).

### 1.2.2. Les stratégies d'interaction

Wright *et al.* (1996) ont défini une séquence d'interactions comme une succession d'étapes, *i.e.* de changements dans la configuration des ressources. Chaque étape est caractérisée par l'exécution de deux processus : la décision de choisir l'action suivante, adéquate au regard des ressources disponibles et la re-configuration (ou mise-à-jour) des ressources suite aux actions réalisées. Ainsi, le plan est mis à jour à chaque étape réalisée (Wright *et al.*, 1996). Les choix qui sont faits par les utilisateurs dépendent des stratégies d'interaction mises en place. Fields *et al.* (1997) ont défini six stratégies d'interaction principales au regard des ressources disponibles :

- Dans la stratégie intitulée « suivre le plan » (*plan following*), l'état de l'interaction constitue un indice permettant aux utilisateurs de savoir à quel niveau du plan ils se trouvent. Le plan les informe alors sur l'action suivante à réaliser. Les utilisateurs effectuent des allers et retours entre le plan et l'état de l'interaction afin d'identifier le niveau auquel ils se situent.
- La stratégie de planification (*planning strategy*) consiste à produire un plan qui peut ensuite être utilisé dans le cadre d'une stratégie « suivre le plan ». Cette planification est réalisée en coordonnant les buts, l'état de l'interaction, les relations action-effet et les affordances pour chaque étape du plan (Fields *et al.*, 1997; Wright *et al.*, 2000).
- Dans la stratégie de mise en correspondance sémantique (*semantic matching strategy*), les ressources utilisées sont les mêmes que dans la stratégie de planification. L'utilisateur décide de ce qu'il va faire en mettant en correspondance les relations action-effet avec les buts et vérifie si l'état d'interaction résultant satisfait bien les buts.

- L'interaction basée sur les informations affichées (*Display-based interaction*) est identique à la mise en correspondance sémantique, excepté le fait que les relations action-effet et les affordances sont externes.
- Dans l'exploration dirigée par le but (*goal-directed exploration*), l'utilisateur ne dispose pas de plan ; il procède par essai/erreur. Il essaie une action parmi les affordances disponibles et compare l'état de l'interaction avec les buts. Si ces derniers ne correspondent pas, il essaie alors une autre action parmi les affordances disponibles.
- Enfin, l'apprentissage par exploration (*learning by exploring*) peut conduire à créer un nouveau plan en observant l'historique d'une série d'actions dont le résultat est positif. Cela peut aussi étendre le répertoire de relations action-effet pour les futures mises en correspondance sémantiques (Fields *et al.*, 1997).

De nombreuses publications dans le domaine des interactions homme-machine s'appuient sur le modèle des ressources. Il s'agit d'un outil faisant figure de référence pour la modélisation de situations d'interactions, dans le but d'améliorer une interface (*e.g.* Doherty, Campos, & Harrison, 2008; Blandford & Furniss, 2006). Par exemple, il est appliqué pour la conception d'environnements virtuels (Smith & Hart, 2006), pour modéliser des environnements intelligents (Harrison, Massink, & Latella, 2009), pour la conception d'interfaces graphiques utilisateurs (Mehrnejad, Toreini, & Ghaemi Bafghi, 2011) ou encore pour la conception de technologies de soin (Rajkomar & Blandford, 2011).

### **1.3. Application du modèle des ressources à la reconnaissance de plans**

Dans la présente étude, il était demandé aux participants de superviser la reconnaissance automatique effectuée par notre prototype. Ils assistaient en temps réel à la reconnaissance des symboles et intervenaient pour corriger les erreurs d'interprétation lorsqu'ils en repéraient. L'objectif de cette recherche était d'étudier

l'impact des interruptions de l'interprétation par le système sur l'interaction dans la tâche de supervision de la reconnaissance automatique. Cette étude était aussi un moyen de mettre à l'épreuve la pertinence du modèle des ressources dans le contexte du développement d'un logiciel de reconnaissance de plans. Dans ce but, les caractéristiques de l'interaction avec les interruptions système ont été comparées avec celles de l'interaction sans interruption système.

Avec le prototype sans interruption système, l'activité consiste à surveiller le déroulement de la reconnaissance automatique et à vérifier au fur et à mesure les interprétations proposées pour chaque symbole par le prototype. Lorsque l'utilisateur localise une erreur d'interprétation, deux manières de la corriger s'offrent à lui. La première consiste à cliquer sur le bouton « pause » permettant ainsi d'interrompre la reconnaissance et à cliquer ensuite sur le symbole mal interprété pour faire apparaître les propositions de corrections. La seconde manière d'opérer consiste, durant la reconnaissance, à cliquer directement sur le symbole dont l'interprétation est erronée. Cette action a pour effet de bloquer le processus de reconnaissance et d'afficher les corrections possibles. L'utilisateur clique alors sur la bonne correction, le système l'applique avant de reprendre automatiquement son travail de reconnaissance.

Cette boucle peut être répétée pour chaque erreur identifiée jusqu'à la fin de l'interprétation automatique. L'objectif que les utilisateurs poursuivent est alors de parvenir à obtenir une interprétation du plan ne contenant aucune erreur. Ce but était expliqué dans la consigne donnée aux participants sous la forme d'un formulaire papier, disponible à la consultation pendant toute la durée de la tâche. Concrètement, la correspondance entre les symboles des plans d'architecture et leur sens était spécifiée dans la consigne. En termes de ressources informationnelles, cette page de consigne correspond aux buts. En effet, l'objectif des participants est de faire en sorte que l'ensemble des interprétations de symboles du plan soit conforme à ce qui est spécifié sur cette page. Dans cette tâche, les buts sont donc d'abord externes. Ils ne sont pas stockés dans la mémoire des utilisateurs au début de l'expérience. Il faut néanmoins préciser que les symboles utilisés dans cette étude n'avaient pas été choisis de manière arbitraire, mais qu'il s'agissait de représentations schématiques des objets

représentés. Il était donc possible pour les participants de deviner le sens d'un symbole donné, sans forcément avoir besoin de consulter la consigne pour cela.

Durant l'analyse, les affordances sont peu nombreuses puisque seules deux actions sont possibles pour l'utilisateur : interrompre la reconnaissance en cliquant sur le bouton « pause » et interrompre la reconnaissance en cliquant sur un symbole mal interprété pour voir s'afficher les propositions de correction. Ces relations action-effet sont expliquées dans la consigne. Cependant, si le bouton pause ne présente aucune ambiguïté quant à sa fonction, la possibilité d'interrompre le système en cliquant sur un symbole n'est pas véritablement une affordance. Pendant l'analyse par le logiciel, l'utilisateur doit détecter les erreurs éventuelles. Pour ce faire, il doit comparer l'état du système (les interprétations de symboles précédemment effectuées par le système) qui s'agrandit dynamiquement avec les buts présentés dans la consigne. Il doit ainsi faire face à un flux continu d'informations, ce qui implique que son travail de coordination soit réalisé rapidement pour identifier les erreurs. En coordonnant les buts avec l'état du système, l'utilisateur met en œuvre la stratégie de mise en correspondance sémantique. Lorsque l'utilisateur clique sur une interprétation pour la corriger, les nouvelles affordances consistent en la labélisation du symbole ciblé en cliquant sur un des boutons contenant chacun une interprétation différente. Autrement dit, l'utilisateur clique sur une des interprétations possibles et le système étiquette le symbole avec l'interprétation choisie. Ici, l'utilisateur doit coordonner ces affordances avec l'état du système (le symbole mal interprété) et les buts. Contrairement à la coordination mise en œuvre pour la détection des erreurs, celle effectuée pour la correction n'est pas soumise à un contexte de pression temporelle puisque l'analyse est en pause.

## **1.4. La confiance dans le système**

Le modèle des ressources vise à décrire la configuration des ressources informationnelles distribuées dans le système homme-machine. Cependant, ce modèle ne prend pas en compte les dimensions subjectives. Or, comme le suggère l'étude de McCarley *et al.* (2003), la confiance que l'utilisateur accorde à la machine pourrait être

une dimension dont il faudrait tenir compte dans la supervision de la reconnaissance automatique de plans.

Dans le domaine des interactions homme-machine, la confiance peut être définie comme « l'attitude d'un individu envers un agent qui va lui apporter de l'aide pour la réalisation de ses buts dans une situation caractérisée par de l'incertitude et de la vulnérabilité » (Lee & See, 2004, p. 51, notre traduction). Pour Fogg et Tseng (1999), « la confiance indique une croyance positive à propos de la fiabilité perçue d'une personne, d'un objet ou d'un processus » (p. 81, notre traduction). Dans la littérature en psychologie, les notions de confiance (*trust*) et de crédibilité (*credibility*) sont très liées puisque la crédibilité est définie comme une qualité perçue attribuée à une personne ou à un objet que l'on estime pouvoir croire (Fogg & Tseng, 1999).

La confiance que des utilisateurs accordent à un système automatique dépend de la manière dont se déroule l'interaction et en retour, cette confiance peut avoir un impact sur l'interaction. Il s'agit d'un enjeu majeur en ergonomie car lorsqu'un utilisateur n'a pas confiance dans un système automatique, il risque tout simplement de cesser de l'utiliser et de réaliser la tâche par lui-même (Lee & Moray, 1992).

Fogg et Tseng (1999) ont défini quatre principaux types de crédibilité :

- La crédibilité présumée est celle que les individus attribuent sur la base de suppositions très générales auxquelles ils adhèrent. Par exemple, beaucoup de gens acceptent l'idée que leurs amis leur disent la vérité, mais qu'un vendeur peut au contraire leur mentir.
- La crédibilité de réputation dépend des indicateurs tels que des titres officiels ou autres récompenses prestigieuses qui constituent des indices pour l'individu.
- La crédibilité de surface est attribuée par l'utilisateur après une simple inspection. Par exemple, c'est sur la base de la crédibilité de surface que des consommateurs jugent un livre selon sa couverture, ou un humain sur ses vêtements (Fogg & Tseng, 1999).

- La crédibilité basée sur l'expérience renvoie à la crédibilité que l'utilisateur attribue à un système à partir de sa propre expérience d'usage de ce système.

Avant d'avoir utilisé un système, les utilisateurs potentiels s'appuient donc exclusivement sur la crédibilité présumée, de réputation et de surface pour se faire une opinion. L'évaluation que les utilisateurs font de la crédibilité d'un système peut être plus ou moins en adéquation avec la précision réelle du système. La crédibilité peut ainsi être évaluée de manière juste, surestimée ou encore sous-estimée (voir tableau 10).

Tableau 10. Les évaluations de la crédibilité (Fogg & Tseng, 1999 ; notre traduction)

	L'utilisateur perçoit le produit comme crédible	L'utilisateur perçoit le produit comme non crédible
Le produit est crédible	Acceptation appropriée	Erreur d'incrédulité
Le produit n'est pas crédible	Erreur de naïveté	Rejet approprié

Souvent, la confiance ne se construit pas de manière consciente et active. Dans beaucoup de situations d'interaction, les individus ne réfléchissent pas activement au degré de confiance qu'ils accordent à la technologie qu'ils utilisent. Ils octroient alors une confiance « par défaut » (Hoffman, Johnson, Bradshaw, & Underbrink, 2013). Cependant, dans d'autres cas, la confiance se met en place de manière tout à fait consciente. Par exemple, un pilote qui a besoin de savoir si les conditions météorologiques seront mauvaises au cours du trajet qu'il a prévu d'emprunter va se poser la question de la confiance qu'il peut accorder à une prédiction météorologique forcément incertaine (Hoffman *et al.*, 2013).

La confiance ne renvoie pas seulement à deux entités, l'une faisant confiance à l'autre, mais aussi à une situation donnée ou à une tâche donnée. Un exemple présenté par Hoffman *et al.* (2013) illustre bien cette nécessaire contextualisation de la confiance : je peux certainement faire confiance à mon épouse pour me dire la vérité, mais si je

suis suspendu sur le rebord d'une falaise, je ne pourrais pas forcément lui faire confiance pour me remonter, car cela implique une force physique qu'elle pourrait ne pas avoir. De la même manière, la confiance dans les systèmes automatiques concerne une tâche précise, mais ne dit rien sur toutes les autres situations possibles.

Comme la confiance interpersonnelle, la confiance dans les systèmes est dynamique, se développe au fil des interactions, se renforce, ou au contraire se dégrade (Hoffman *et al.*, 2013; Lee & Moray, 1992; Khasawneh, Bowling, Jiang, Gramopadhye, & Melloy, 2003). D'une manière générale, le système tend à gagner en crédibilité lorsqu'il apporte des informations pertinentes, précises ou correctes. À l'inverse, il perd en crédibilité lorsqu'il apporte des informations fausses (Khasawneh *et al.*, 2003). Un système ayant perdu de la crédibilité peut en gagner à nouveau de deux manières : en transmettant pendant une période donnée des informations pertinentes ou en commettant toujours les mêmes erreurs, permettant ainsi à l'utilisateur de les anticiper (Fogg & Tseng, 1999). Les variations de la confiance dépendent aussi de la difficulté perçue de la tâche réalisée par le système. C'est ce que montrait une étude de Madhavan, Wiegmann et Lacson (2003), dans laquelle les participants devaient détecter une cible à l'écran. Dans cette expérience, les individus accordaient moins de confiance à un système qui faisait des erreurs sur des items qu'ils estimaient faciles, en comparaison à un système dont les erreurs portaient sur des items jugés difficiles. La confiance peut aussi dépendre de l'expertise des utilisateurs, de leur degré de compréhension du fonctionnement du système et de leurs besoins en informations (Fogg & Tseng, 1999).

Selon Lee et See (2004), lorsque le niveau de confiance dans l'automate correspond à son niveau effectif de fiabilité (voir figure 27), cela tend à induire une utilisation appropriée du système. En revanche, une confiance trop élevée par rapport à la fiabilité de l'automate peut conduire à une mauvaise utilisation, alors qu'une confiance trop faible peut être à l'origine d'un arrêt de l'utilisation. La sur-confiance est appelée « complaisance » par Miller et Parasuraman (2007). Elle apparaîtrait lorsque le niveau d'automatisation d'un système devient si élevé que l'opérateur à

l'illusion que la machine est infaillible. L'opérateur a alors tendance à diminuer son degré de vigilance dans la tâche.

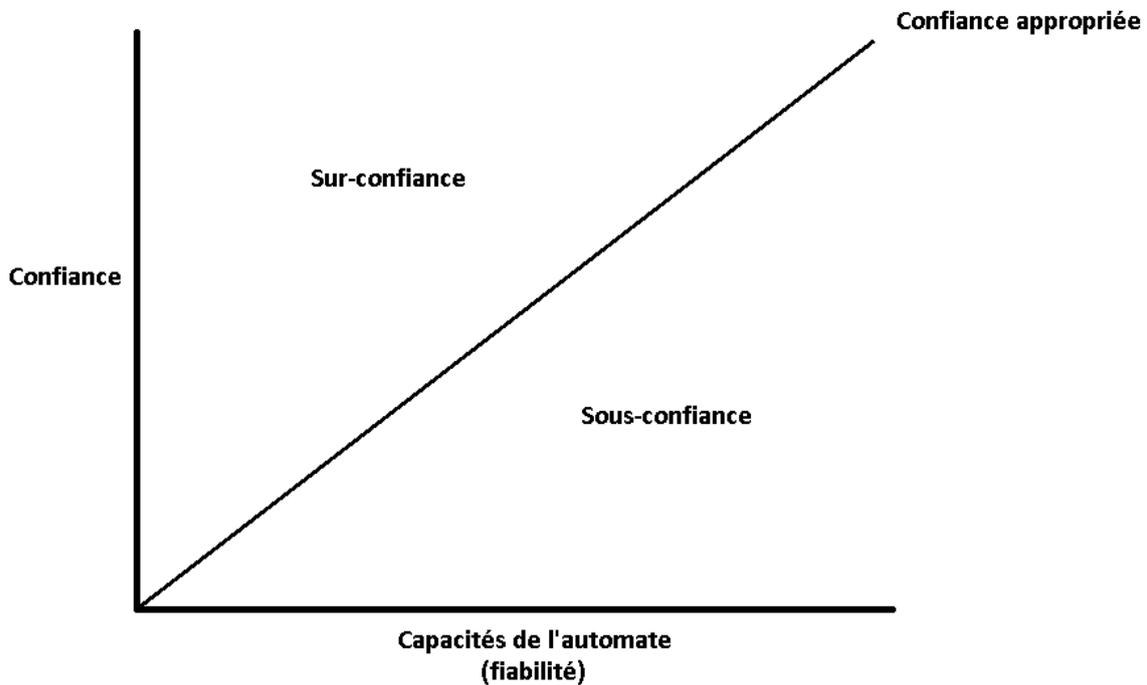


Figure 27. Lien entre la confiance et la fiabilité réelle du système (d'après Lee & See, 2004)

Ce désengagement attentionnel a été modélisé par Parasuraman et Manzey (2010) sous la forme d'une boucle impliquant à la fois la confiance, l'allocation des ressources attentionnelles et les feedbacks renvoyés par le système (voir figure 28). Selon ce modèle, l'opérateur alloue des ressources attentionnelles à l'ensemble des tâches non automatisées. En revanche, la quantité de ressources qu'il accorde aux tâches automatisées dépend du degré de confiance qu'il ressent à l'égard du système. Le résultat des actions sur l'ensemble de ces tâches se traduit sous la forme de feedbacks renvoyés par le système qui influencent la confiance de l'opérateur.

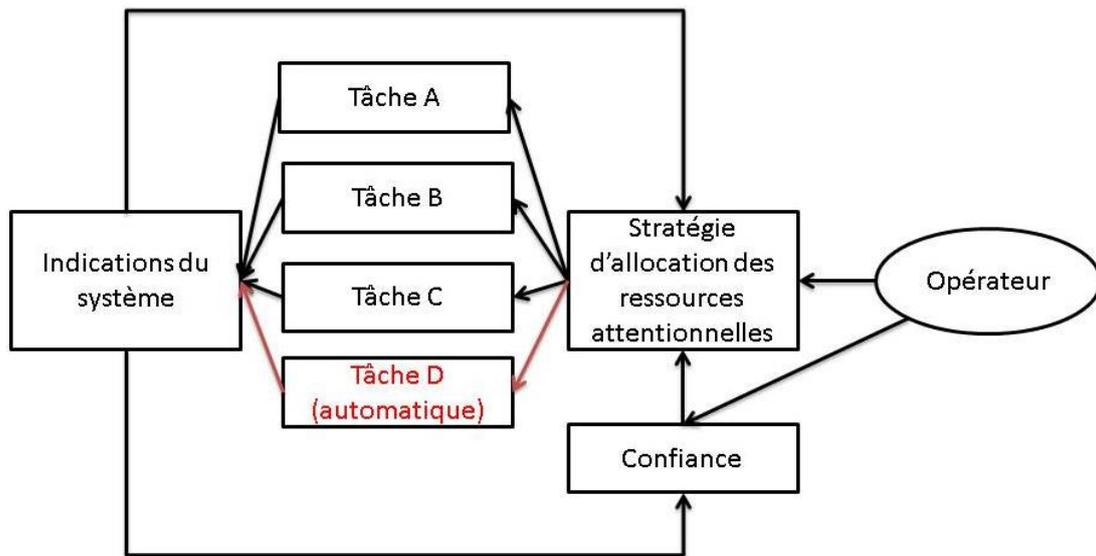


Figure 28. Modélisation de l'allocation des ressources attentionnelles en fonction de la confiance dans le système (d'après Parasuraman & Manzey, 2010)

Cet effet négatif de la sur-confiance a aussi été mis en évidence par des travaux portant sur un système d'alarme anticollision pour automobilistes (Abe, Itoh, & Tanaka, 2002). Dans cette recherche, les participants utilisaient un simulateur de conduite avec lequel ils devaient suivre une voiture qui ralentissait à un moment variable de la tâche. L'alarme était parfaitement fiable pour les premiers essais. Elle s'activait donc systématiquement aux moments des ralentissements. Au bout de quelques essais, des fausses alarmes apparaissaient (l'alarme se déclenchait alors que le véhicule ne ralentissait pas). Les résultats indiquent que les fausses alarmes augmentaient le temps de réaction des participants et diminuaient leur confiance dans le système (mesurée par questionnaire). L'absence d'alarme en présence d'un ralentissement augmente aussi le temps de réaction et diminue encore plus le niveau de confiance des participants.

Généralement, les concepteurs d'un système automatique cherchent à rendre la crédibilité de leur système la plus élevée possible. Pourtant, Lee et See (2004) conseillaient plutôt de concevoir le système de manière à ce que la confiance soit simplement appropriée, plutôt que trop élevée. Pour tendre vers un niveau de confiance approprié, plusieurs moyens existent. Il est par exemple possible de

communiquer à l'utilisateur des informations concernant les performances passées de l'automate, de présenter son mode de fonctionnement technique ou encore d'expliquer les limites de son champ d'application.

## 1.5. Hypothèses

Le modèle des ressources est un cadre d'analyse servant à décrire les ressources informationnelles réparties dans le système cognitif distribué que constituent la situation de travail, la machine et l'opérateur. Ainsi, ce modèle ne tient pas compte des variables subjectives alors que celles-ci pourraient avoir un impact sensible sur l'interaction. En effet, la confiance que les utilisateurs accordent au prototype pourrait être un élément déterminant (Jian, Bisantz, & Drury, 2000). Dans les exemples d'application précédents du modèle des ressources, les machines ne faisaient pas d'erreur (*e.g.* Fields, Wright, Marti, & Palmonari, 1998; Smith, Duke, & Wright, 1999; Wright *et al.*, 1996). En revanche, au sein de notre expérience, les sollicitations de l'utilisateur par le prototype étaient basées sur l'incertitude à propos des interprétations. Le système sollicitait l'utilisateur du fait d'un risque d'erreur. Lorsqu'il le faisait, cela pouvait être pour une erreur véritable ou non. De plus, des erreurs pouvaient exister dans l'interprétation sans faire pour autant l'objet de sollicitations par le prototype. Ce caractère faillible du système de sollicitations des utilisateurs en cas de risque d'erreurs donne de l'importance au degré de confiance que l'utilisateur accorde au prototype. Si l'utilisateur ne fait pas confiance à la machine, les sollicitations sont inutiles, car il va de toute manière re-contrôler la totalité de l'interprétation avec la même attention, sans tenir compte des suggestions. C'est ce que Fogg et Tseng (1999) appelaient l'erreur d'incrédulité (*incredulity error*). À l'inverse, si les utilisateurs accordent trop de confiance au prototype, ils risquent de négliger l'activité de contrôle et d'omettre des erreurs qui n'auraient pas fait l'objet de sollicitations. Dans ce cas, ils font une erreur de naïveté (*gullibility error*; Fogg & Tseng, 1999).

Pour trouver des erreurs d'interprétation, les participants doivent mettre en œuvre une stratégie de mise en correspondance sémantique consistant à décider de leurs

actions (quelle interprétation corriger) après avoir coordonné les relations action-effet avec les buts et l'état du système. Dans cette situation, les buts sont les interprétations justes décrites dans la consigne et l'état du système renvoie aux interprétations qui sont effectuées par le logiciel. Lorsque ces deux ressources informationnelles ne concordent pas, l'utilisateur doit réaliser une correction pour les faire correspondre. Par ailleurs, comme cela a été mentionné plus haut, le logiciel est capable de signaler à l'utilisateur quelles sont les interprétations qui sont sujettes à un important risque d'erreur. En indiquant successivement à l'utilisateur toutes les interprétations les plus incertaines, le logiciel lui fournit un plan, *i.e.* un ensemble de zones à vérifier dans un ordre donné. Ainsi, il permet à l'utilisateur de mettre en œuvre une seconde stratégie d'interaction, le suivi de plan, en parallèle avec la mise en correspondance sémantique. Cette stratégie de suivis de plan, plus focalisée sur les zones à risque, devrait favoriser la détection des erreurs ainsi soulignées par le logiciel par rapport à la stratégie de mise en correspondance sémantique seule. La première hypothèse de cette étude (hypothèse 1) était que les erreurs faisant l'objet d'une sollicitation par le prototype allaient être corrigées par une proportion plus importante de participants, en comparaison à une situation dans laquelle le prototype ne faisait pas de sollicitation.

Le fait que les buts soient externes pourrait causer des difficultés. Un traitement parallèle d'une consigne sur papier et de l'interprétation à l'écran à laquelle la consigne fait référence est une activité coûteuse sur le plan cognitif. Sweller et Chandler (1994) ainsi que Chandler et Sweller (1996) ont conduit une série d'expériences dans lesquelles ils demandaient à des participants d'apprendre le fonctionnement d'un logiciel grâce à un manuel, avant de réaliser une tâche spécifique sur ce logiciel. Un groupe de participants disposaient d'une version du manuel ne contenant que des informations textuelles et devaient interagir avec l'ordinateur pour apprendre. Un autre groupe de participants ne pouvait pas utiliser l'ordinateur durant la période d'apprentissage, mais disposait d'une version du manuel contenant, en plus des informations textuelles, des captures d'écrans du logiciel. Les participants qui devaient partager leur attention entre l'ordinateur et le manuel ont pris plus de temps

pour réaliser l'apprentissage, mais ils étaient aussi plus lents et moins précis dans la réalisation des exercices sur l'ordinateur. Dans notre expérience, l'activité de détection des erreurs présente des points communs avec la tâche décrite par Sweller et Chandler (1994). En effet, il s'agit de coordonner la consigne sur papier avec l'historique affiché à l'écran. Ainsi, le fait que la consigne contenant les buts soit présentée de manière externe sur papier peut constituer une gêne, entraînant une perte de temps plutôt qu'une diminution des performances de correction des erreurs. L'internalisation des buts devrait éliminer cette perte de temps au fur et à mesure de l'entraînement. Certains participants pourraient alors devenir suffisamment rapides pour contrôler correctement la totalité des interprétations réalisées par le prototype au fur et à mesure de leur apparition. Ainsi, la deuxième hypothèse de cette étude était que la durée de la tâche diminue progressivement au fil des essais, car du fait d'une coordination de plus en plus rapide (grâce à l'internalisation des buts), l'étape de correction nécessite de moins en moins de temps (hypothèse 2). Les deux hypothèses précédemment décrites sont basées sur la modélisation des interactions à travers le modèle des ressources. Si elles étaient validées, cela corroborerait l'idée selon laquelle le modèle des ressources est un outil pertinent pour la modélisation des interactions homme-machine dans lesquelles l'humain supervise les processus automatiques réalisés par la machine.

Néanmoins, le fait que le prototype puisse commettre des erreurs peut avoir un impact sur la confiance que les utilisateurs lui portent. Sur six sollicitations des utilisateurs pour chaque plan, quatre portent sur des erreurs et deux sur des symboles dont l'interprétation est pertinente. Les erreurs de la machine tendent à impacter négativement ce que Merritt et Ilgen (2008) appellent la confiance basée sur l'historique (*history-based-trust*). À l'inverse, l'étude de Wiegmann *et al.* (2006) démontrait que des opérateurs peuvent accorder une certaine crédibilité aux jugements de la machine qui leur suggère des réponses. Un phénomène de complaisance pourrait donc être induit par l'assistance automatique, comme le suggère l'étude de Skitka *et al.* (1999). La troisième hypothèse de cette étude (hypothèse 3) était que les utilisateurs travaillant avec le prototype doté de la

fonctionnalité de sollicitation allaient montrer un degré de confiance plus élevé en la machine. La quatrième hypothèse (hypothèse 4) était que ces utilisateurs devraient moins corriger les erreurs ne faisant pas l'objet de sollicitation lorsqu'ils sont dans le groupe avec sollicitations en comparaison avec ceux du groupe sans sollicitation. Les participants du groupe avec sollicitation, du fait de l'augmentation de leur confiance dans le système, devraient allouer moins de ressources attentionnelles au contrôle des erreurs n'ayant pas fait l'objet de sollicitation.

## **2. METHODE**

### **2.1. Participants**

Les participants de cette expérience étaient 48 étudiants et jeunes actifs (11 hommes et 37 femmes), recrutés à partir de la base de volontaires du LOUSTIC. Le plus jeune avait 18 ans et 9 mois alors que le plus âgé avait 44 ans et 2 mois. L'âge moyen était de 26,23 ans ( $SD = 5,93$ ).

### **2.2. Matériel**

Dans cette expérience, toutes les tâches ont été réalisées sur une tablette tactile Asus Eee Slate 12.1. Le prototype sur lequel les participants devaient travailler affichait en temps réel ses interprétations sous la forme d'une labélisation sous le symbole, tout en surlignant ce dernier en bleu si c'était un ouvrant et en vert si c'était un meuble. Chaque participant devait contrôler l'interprétation de quatre plans différents présentés successivement et était placé dans une des deux conditions expérimentales : Sans Interruption Système (SIS) et Avec Interruptions Système (AIS). Dans la condition SIS, le plan initial apparaissait à l'écran immédiatement et la reconnaissance automatique était effectuée par le prototype. Les participants avaient la possibilité d'interrompre l'interprétation pour corriger des erreurs avant la fin de l'analyse. La condition AIS était identique à la précédente, excepté le fait que lorsque le système était confronté à un symbole pour lequel l'interprétation présentait une incertitude

importante, il stoppait l'analyse et demandait à l'utilisateur de confirmer son interprétation (voir figure 29).

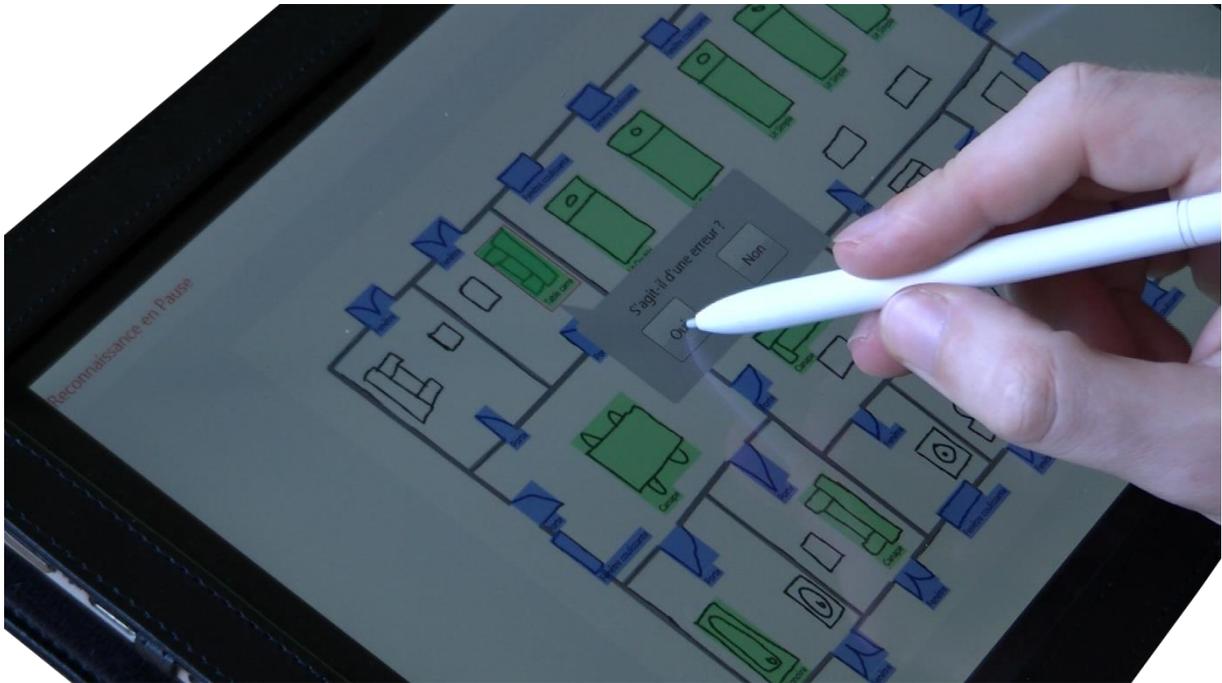


Figure 29. Le système attend que l'utilisateur confirme son interprétation

Si l'utilisateur lui signifiait qu'il s'agissait d'une erreur, le système proposait des corrections possibles. Le participant devait alors cliquer sur la bonne interprétation. Chaque plan utilisé était constitué de soixante symboles et de six erreurs d'interprétation. Dans la condition AIS, six sollicitations par le système avaient lieu dont quatre concernant des erreurs et deux des interprétations justes. Ainsi, deux erreurs de chaque plan ne faisaient pas l'objet d'une sollicitation.

## 2.3. Procédure

Un plan d'architecture était présenté à l'écran. Lorsque les participants étaient prêts, ils démarraient l'analyse par le prototype et avaient pour objectif de corriger les erreurs d'interprétation. Dans chacune des conditions expérimentales, quatre plans différents étaient ainsi présentés aux participants (voir figure 30). L'ordre de présentation de ces quatre plans était contrebalancé à l'intérieur de chaque groupe expérimental.



Figure 30. Plans d'architecture utilisés dans l'expérience

Avant le début de la tâche, il était demandé aux participants de lire entièrement la consigne leur expliquant leurs objectifs, les fonctionnalités du prototype et le sens des symboles utilisés (voir figure 31). Ensuite, ils devaient contrôler l'analyse et corriger les erreurs qu'ils détectaient. À la fin de cette tâche, il leur était demandé d'indiquer s'ils avaient ou non consulté la consigne pendant l'interaction avec le prototype. Enfin, les participants devaient répondre à un questionnaire mesurant leur confiance et leur méfiance dans le système, questionnaire construit à partir de l'échelle de Jian *et al.* (2000 ; voir annexe 10). Sur les douze items de l'échelle, les participants devaient spécifier leur degré d'accord sur une échelle de Likert en 7 points allant du pôle « Pas du tout d'accord » au pôle « Tout-à-fait d'accord ». Par exemple, les deux items suivants étaient proposés : « le système est trompeur », pour mesurer la méfiance et « le système est sûr », pour mesurer la confiance.

Symbole	Interprétation
	Fenêtre
	Porte
	Fenêtre coulissante
	Canapé
	Lit simple / Lit double
	Table avec chaises
	Lavabo simple
	Lavabo double
	Baignoire
	Table carrée
	Douche
	Toilette

Figure 31. Présentation des symboles et de leur interprétation

Les facteurs manipulés dans cette expérience étaient le format et le rang. Le format est une variable catégorielle intergroupe à deux modalités, AIS et SIS. Le rang est une variable ordinale intragroupe à quatre modalités correspondantes au rang de chaque plan parmi les 4 interprétés. Les variables dépendantes principales sont la durée de la tâche et la correction des erreurs. Le traitement des corrections d'erreurs a été réalisé en distinguant deux types d'erreurs : les quatre erreurs par plan faisant l'objet de sollicitations et les deux erreurs par plan ne faisant pas l'objet de sollicitation (dans la condition AIS). Du fait de cette distinction, le nombre d'erreurs pour chacune des catégories est faible. Dans ces conditions, la distribution des scores de correction d'erreurs ne permet pas des comparaisons de variances et de moyennes. Par conséquent, ces variables ont été recodées de manière binaire, le chiffre zéro signifiant qu'au moins une erreur de la catégorie n'a pas été corrigée et le chiffre un que toutes les erreurs de la catégorie ont été corrigées.

## 3. RESULTATS

### 3.1. Durée de la tâche

Les durées de la tâche sont un ensemble de valeurs continues, borné à gauche mais pas à droite. Par conséquent, afin d'évaluer l'effet de la variable intergroupe « format » et de la variable intragroupe « rang », nous avons modélisé la durée de la tâche selon une distribution gamma. Nous avons également tenu compte du caractère répété des mesures en utilisant la méthode d'estimation par équation d'estimation généralisée (Ballinger, 2004; Ghisletta & Spini, 2004).

L'interaction entre le format et le rang n'apparaît pas significative ( $Wald \chi^2(3, N = 191) = 5,534, p = 0,137$ ). En revanche, le test révèle un effet statistiquement significatif du rang ( $Wald \chi^2(3, N = 191) = 77,619, p < 0,001$ ) et du format ( $Wald \chi^2(1, N = 191) = 4238, p = 0,040$ ) sur la durée de la tâche (voir figure 32). Cet effet d'apprentissage valide la deuxième hypothèse de cette étude selon laquelle la tâche de correction des erreurs allait être réalisée de plus en plus rapidement au cours des essais.

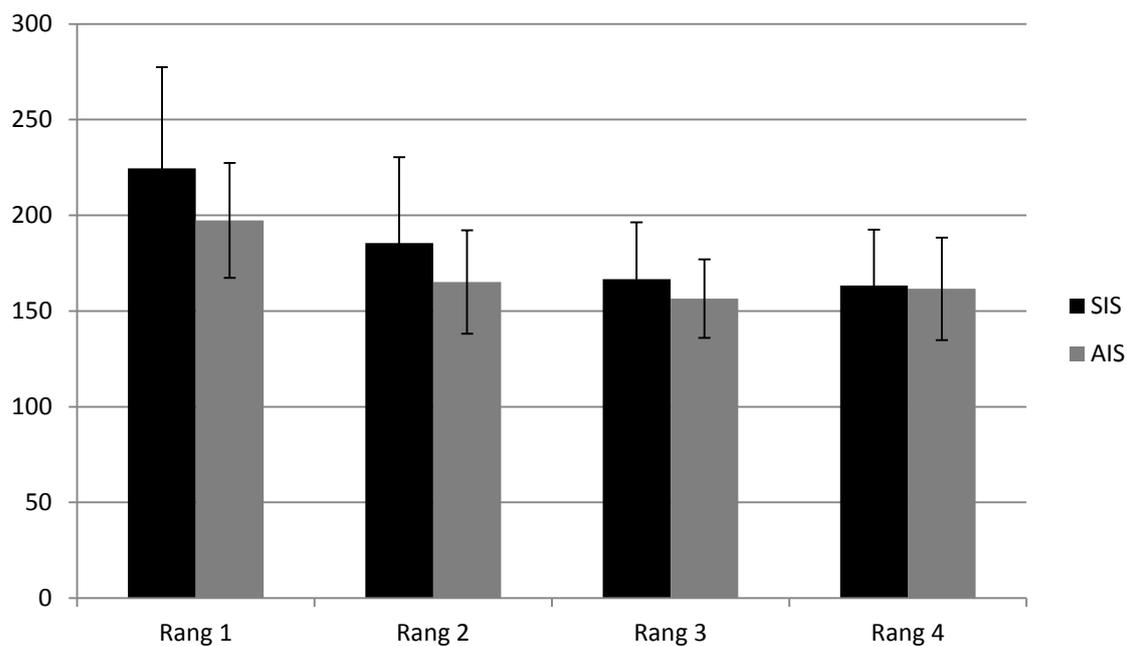


Figure 32. Moyennes et écarts-types pour la durée de la tâche dans chaque format et pour chaque rang

## 3.2. Erreurs

### 3.2.1. Erreurs signalées par le système

L'effet du format et du rang sur la proportion de correction totale des erreurs ayant fait l'objet de sollicitations de la part du système a été évalué en utilisant un modèle de régression logistique à effets aléatoires, tenant ainsi compte du caractère répété des données. L'interaction entre le format et le rang n'apparaît pas significative ( $Wald \chi^2(2, N = 191) = 1,321, p = 0,517$ ). En revanche, on constate un effet simple du rang ( $Wald \chi^2(3, N = 191) = 410,325, p < 0,001$ ) et du format ( $Wald \chi^2(1, N = 191) = 106,589, p < 0,001$ ) sur la proportion de correction totale des erreurs signalées (voir figure 33). Ce résultat valide la première hypothèse selon laquelle la détection des erreurs signalées par les sollicitations est améliorée.

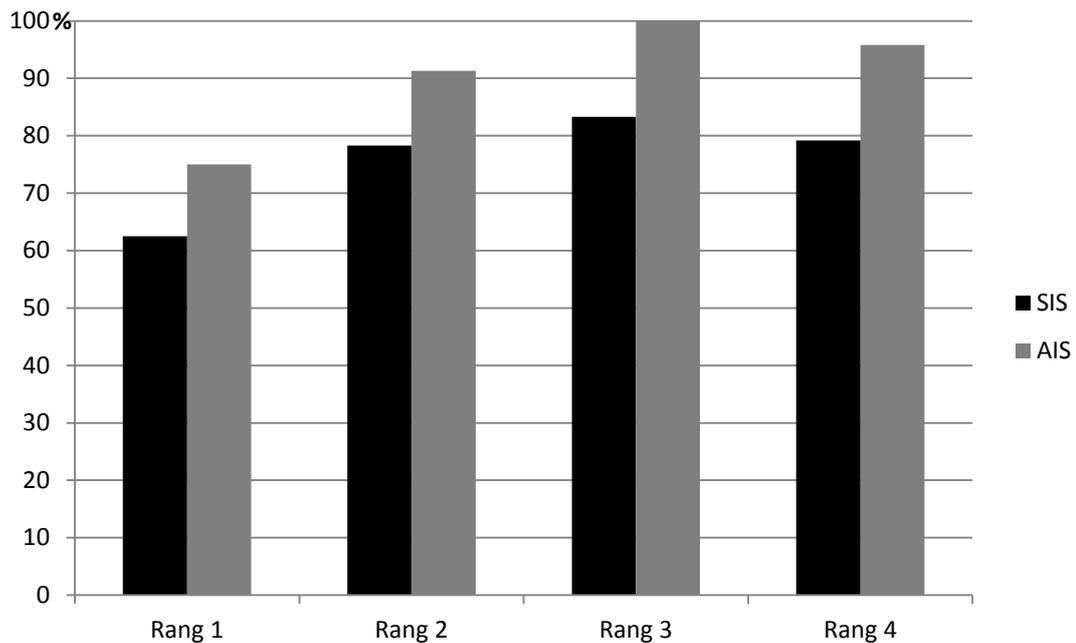


Figure 33. Pourcentage de participants ayant corrigé toutes les erreurs signalées par les sollicitations du système pour chacun des formats et des rangs

### 3.2.2. Erreurs non signalées par le système

Les effets du format et du rang sur la proportion de correction de la totalité des erreurs non signalées par les sollicitations du système ont été évalués à travers une régression logistique à mesures répétées. Celle-ci révèle un effet significatif du rang ( $Wald \chi^2(3, N = 191) = 270,145, p < 0,001$ ) et du format ( $Wald \chi^2(1, N = 191) = 101,186, p < 0,001$ ). De plus, l'interaction entre le format et le rang apparaît statistiquement significative ( $Wald \chi^2(2, N = 191) = 603,430, p < 0,001$ ). Sur la figure 34, on constate que la supériorité de la performance du groupe SIS n'apparaît qu'à partir du rang 2. Ce résultat est cohérent avec la quatrième hypothèse selon laquelle les sollicitations du système devraient générer une sur-confiance qui aurait pour conséquence de diminuer l'identification des erreurs non signalées.

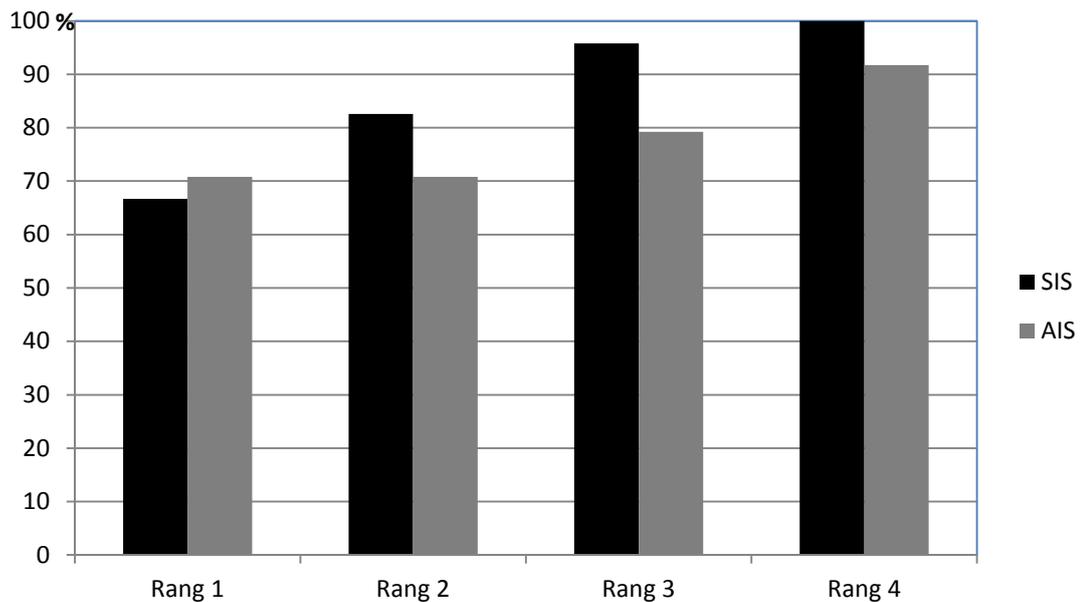


Figure 34. Pourcentage de participants ayant corrigé toutes les erreurs non signalées par les sollicitations du système pour chacun des formats et des rangs

La comparaison des proportions de participants ayant corrigé l'ensemble des erreurs (signalées ou non par le système) ne révèle aucune différence significative entre le groupe SIS (65,28%) et AIS (67,73%),  $Wald \chi^2(1, N = 191) = 0,064, p = 0,800$ . La variable format a un effet inverse sur les erreurs signalées par le système, en comparaison à son effet sur les erreurs non signalées. Ainsi, lorsque cette distinction entre les types d'erreurs n'est pas faite, aucune différence n'apparaît entre les groupes car les deux effets s'annulent.

### 3.3. Echelle de confiance dans les systèmes automatiques

Puisque les items utilisés pour mesurer la confiance et la méfiance ont été traduits pour cette expérience, une analyse factorielle confirmatoire a été réalisée sur les données afin de s'assurer que la même structure factorielle que celle de Jian *et al.* (2000) était retrouvée. Cette analyse a confirmé l'unidimensionnalité des items renvoyant à la confiance ( $\chi^2 = 20,15, p = 0,13, RMSEA = 0,1$ ) et ceux mesurant la méfiance ( $\chi^2 = 9,33, p = 0,41, RMSEA = 0,03$ ). De plus, la cohérence interne des items

évaluée par l'alpha de Cronbach est plutôt élevée pour la confiance ( $\alpha = 0,87$ ) et pour la méfiance ( $\alpha = 0,74$ ).

Le test de Levene sur la distribution des scores de méfiance révélait une homoscedasticité acceptable,  $F(1, 46) = 0,002, p = 0,966$ . Une ANOVA à un facteur ne montrait aucune différence statistiquement significative entre les groupes SIS et AIS,  $F(1, 46) = 0,961, MSE = 0,973, p = 0,332$  (voir tableau 11). Le test de Levene sur les scores de confiance confirmait lui aussi une homoscedasticité acceptable,  $F(1, 46) = 0,006, p = 0,936$ . Une ANOVA à un facteur montrait la supériorité statistiquement significative des scores de confiance dans le groupe AIS en comparaison à ceux obtenus dans le groupe SIS,  $F(1, 46) = 5,713, MSE = 5,685, p = 0,021$ . Ce résultat va dans le sens de la troisième hypothèse selon laquelle les participants du groupe AIS allaient accorder une plus grande confiance au système que les participants du groupe SIS.

Tableau 11. Moyennes et écarts-types des scores de confiance et de méfiance

	<b>Méfiance</b>		<b>Confiance</b>	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
SIS	3,94	1,02	3,76	,93
AIS	4,22	1,00	4,45	1,07

## 4. DISCUSSION

Un des objectifs théoriques de cette étude était d'évaluer le caractère prédictif du modèle des ressources pour une tâche de supervision de processus automatisés. Cette expérience visait également à estimer la contribution que pourrait apporter la prise en compte de la variable subjective « confiance » pour modéliser ce type de tâche. Dans cette optique, deux hypothèses découlant de l'application du modèle des ressources ont été mises à l'épreuve, ainsi que deux hypothèses basées sur la prise en compte de la confiance.

Les deux hypothèses basées sur le modèle des ressources sont validées par nos résultats. La première était que les participants du groupe AIS seraient meilleurs dans la correction des erreurs signalées par le système que ceux du groupe SIS. Notre expérience a montré qu'une part significativement plus élevée des participants du groupe AIS a corrigé toutes les erreurs signalées par rapport à ceux du groupe SIS. Ainsi, solliciter un utilisateur pour lui demander de vérifier une interprétation permet bien d'augmenter les chances que cette interprétation soit corrigée.

Dans la condition WIS, les participants utilisaient la stratégie de mise en correspondance sémantique consistant à coordonner l'état du système avec les buts afin de détecter les erreurs. Dans la condition IS, le logiciel fournit à l'utilisateur un plan en signalant au fur et à mesure de l'analyse les zones de l'interprétation comportant un risque d'erreur. Ce format favorise une stratégie de suivis de plan consistant à vérifier les interprétations faisant successivement l'objet de sollicitations par le logiciel. Cependant, la mise en œuvre de cette stratégie est réalisée en parallèle avec la stratégie de mise en correspondance sémantique puisque même dans la condition IS, certaines erreurs non signalées par le logiciel ont été corrigées par l'utilisateur. Ce résultat est aussi cohérent avec l'étude de Wiegmann *et al.* (2006), qui montrait une augmentation des performances de détection de cible lorsque le système suggère, avec une précision de 0,90, la présence ou l'absence de cette cible.

La deuxième hypothèse de cette étude était que la durée de la tâche allait diminuer au fil des essais. Cette hypothèse a été validée par les résultats de l'expérience. Il s'agit, là encore, de résultats cohérents avec le modèle des ressources. Pendant l'analyse, les participants reçoivent un flux d'informations à traiter (*i.e.* l'interprétation), qu'ils doivent coordonner avec les buts. Les buts étant externes, la coordination avec l'état du système se trouve ralentie. Ainsi, la phase de détection des erreurs est réalisée avec difficulté au début. L'internalisation progressive des buts peut améliorer la vitesse de la coordination. Notons que les symboles avaient été choisis de manière à ce que leur interprétation puisse être déduite dans une certaine mesure. L'internalisation des buts ne passe donc pas nécessairement par une consultation de la consigne pendant la tâche, bien que celle-ci ait été permise. Elle peut aussi provenir d'une déduction à la

première rencontre d'un symbole et au stockage en mémoire de cette information déduite, permettant une utilisation plus rapide lors d'une rencontre ultérieure avec ce symbole. En termes de prise en main du système, l'effet d'apprentissage observé est plutôt satisfaisant puisqu'en seulement 4 essais, la correction des erreurs est sensiblement améliorée et la durée de la tâche diminue en moyenne de 23%.

Le modèle des ressources était originellement destiné à permettre la modélisation des interactions homme-machine dans une perspective de cognition distribuée (Wright *et al.*, 1996; Wright *et al.*, 2000). Il a déjà été appliqué aux menus du tableur de Microsoft Excel pour servir d'exemple (Wright *et al.*, 1996), mais aussi à une tâche de contrôle du trafic aérien (Fields *et al.*, 1998) et même à la conception d'un environnement virtuel (Smith *et al.*, 1999). De plus, de nombreuses publications récentes décrivent des exemples d'utilisations du modèle des ressources (*e.g.* Harrison *et al.*, 2009; Rajkomar & Blandford, 2011; Mehrnejad *et al.*, 2011). Ici, l'application de ce modèle à une tâche de supervision d'interprétation automatique de plans est fructueuse du fait de la distribution des ressources informationnelles.

Le modèle des ressources constitue de ce fait un cadre d'analyse pertinent pour ce type de tâches, mais ne permet pas à lui seul d'expliquer l'ensemble des résultats obtenus. En effet, certains résultats nécessitent, pour être interprétés, la prise en compte de variables subjectives. Les performances de correction des erreurs non signalées sont plus faibles dans le groupe AIS que dans le groupe SIS. Cette différence n'apparaît qu'à partir du second plan traité. Cela suggère que les participants dans les deux conditions expérimentales débutent avec un niveau de confiance équivalent envers la machine et que cette confiance varie dans chacun des groupes selon la manière dont se déroule l'interaction. La différence de performance apparaissant au deuxième plan traité proviendrait d'une variation du niveau de confiance lui-même dépendant du degré de pertinence des sollicitations. Ce résultat est cohérent avec la description de la confiance dans les systèmes automatisés comme étant un phénomène dynamique en évolution permanente (Lee & Moray, 1992; Hoffman *et al.*, 2013).

De plus, la confiance dans la machine, mesurée par questionnaire après le traitement des quatre plans, est plus élevée chez les participants du groupe AIS que chez ceux du

groupe SIS. Il semble que la présence des sollicitations augmentent la confiance des utilisateurs au point de générer une erreur de naïveté (Fogg & Tseng, 1999). Ce degré de confiance plus élevé dans la condition AIS suggère que les participants ont l'impression que le prototype est capable d'indiquer efficacement les zones comportant un risque d'erreur sans en omettre. La naïveté des participants vis-à-vis de la machine affecte les stratégies d'interaction. Ainsi, à cause de cette confiance « naïve », ils auraient effectué moins de contrôles sur les interprétations ne faisant pas l'objet de sollicitation. D'ailleurs, 79% des participants du groupe SIS déclaraient avoir consulté la consigne pendant la tâche alors qu'ils ne sont que 46% dans le groupe AIS. Cette observation conforte l'idée selon laquelle les participants du groupe AIS ont eut tendance à se désengager de la tâche par rapport à ceux du groupe SIS. S'ils estimaient que le prototype était fiable dans ses sollicitations, ils avaient logiquement moins besoin de contrôler attentivement les interprétations et donc de revenir à la consigne. Selon les termes employés dans le modèle de Parasuraman et Manzey (2010), la confiance augmentée par les sollicitations aurait modifié les stratégies d'allocations de ressources des participants en les conduisant à diminuer leurs efforts sur la tâche de vérification des interprétations. Ce résultat est aussi cohérent avec les travaux de Miller et Parasuraman (2007). En effet, l'ajout des sollicitations peut être assimilé à une augmentation du niveau d'automatisation du système et cela génère une complaisance, *i.e.* une sur-confiance dans le système. Cette sur-confiance a alors pour conséquence une diminution de la vigilance. Le format AIS, au lieu de faciliter la supervision de l'analyse, semble faire basculer l'utilisateur vers une activité de traitement passif consistant à accepter ou rejeter les actions proposées par le système (Kaber & Endsley, 1997).

Wiegmann *et al.* (2006) ne présentaient pas de résultats dans lesquels ils auraient distingué les images sur lesquelles le système indique la cible des images sur lesquelles il ne l'indique pas. Pourtant, le système de détection de cible simulé dans cette étude n'est fiable qu'à 90%. De plus, les auteurs montraient que lorsque le jugement effectué par l'assistance automatique était congruent avec la réponse des participants, ceux-ci déclaraient un degré plus élevé de confiance dans leur propre réponse que lorsque ce

jugement était discordant. Cela démontre que les participants tenaient compte des suggestions de l'assistance automatique et lui faisaient confiance. Si le jugement d'une assistance automatique était créditée de confiance dans l'expérience de Wiegmann *et al.* (2006) comme dans la nôtre, il est vraisemblable que les participants ayant eut une assistance aient été moins précis dans la détection de cibles lorsque celles-ci ne faisaient pas l'objet d'un signalement. Du fait que Wiegmann *et al.* (2006) ne distinguaient pas les performances sur les images pour lesquelles les sollicitations étaient justes de celles pour lesquelles les sollicitations étaient absentes malgré la présence de la cible, il est possible qu'ils soient passés à côté d'un effet négatif des sollicitations sur certaines cibles. Ils rapportaient un effet positif global de l'assistance automatique sur l'identification des cibles, mais celui-ci pourrait provenir du taux relativement élevé de précision de l'assistant. Du fait de ce taux, l'effet positif de l'assistance dépasserait tant ses effets négatifs locaux que le résultat global serait positif.

Dans notre étude, l'absence de différence significative entre les deux conditions expérimentales pour la dimension subjective « méfiance » est surprenante puisque cette dimension est le contraire de la confiance qui elle, diffère entre les groupes. Ce résultat pourrait provenir des items utilisés qui, dans leur traduction française, ont pu étonner les participants. Par exemple, un item faisait référence aux « intentions » du système et un autre demandait aux participants d'évaluer si le système « se comportait de manière sournoise » (voir annexe 10). Faire référence à un logiciel avec ces termes habituellement plus volontiers utilisés pour parler d'êtres humains a peut-être impacté les réponses.

D'une manière générale, les résultats de notre étude sont cohérents avec les prédictions qui découlent de l'application du modèle des ressources. Cela confirme que ce modèle peut avoir un caractère prédictif et prescriptif pour une tâche de supervision de processus automatiques. Cependant, il est focalisé sur les informations distribuées dans le système d'interaction et ne tient pas compte des dimensions subjectives qui pourraient intervenir. Prendre en compte la confiance dans la machine est pourtant indispensable pour expliquer nos résultats. Le caractère prédictif du

modèle des ressources pourrait être amélioré s'il intégrait des variables subjectives. Certaines expériences et certains projets de conceptions utilisant le modèle des ressources pourraient donc conduire à des conclusions fausses ou à de mauvais choix de conception du fait de cette lacune.

Les résultats de cette étude permettent de faire des recommandations pour la conception d'interfaces de logiciels de DAO, de beautification, de reconnaissance automatique ou même de détection de cibles. Ils montrent que l'introduction des sollicitations de l'utilisateur peut comporter des avantages, mais aussi des risques. Cela peut permettre de gagner du temps et de mieux identifier certaines erreurs. D'un autre côté, cela peut aussi conduire à ne pas détecter d'autres erreurs. Concrètement, les logiciels qui intègrent cette fonctionnalité n'amélioreront pas forcément les performances de repérage des erreurs. Le bénéfice tiré de ce type d'assistance semble largement dépendant de sa précision. Il y a un équilibre à trouver, probablement spécifique à chaque interface et aux contraintes de précisions liées à la tâche. Trouver cet équilibre implique que les concepteurs testent différents degrés de précision des sollicitations pour trouver ce qui est le plus efficace pour leur produit. Par ailleurs, comme le suggèrent Lee et See (2004), un niveau de confiance plus approprié pourrait être obtenu en communiquant sur les performances réelles du système ou sur son mode de fonctionnement technique. Par exemple, avec le prototype utilisé, la consigne aurait pu inclure l'information selon laquelle en moyenne, un tiers des erreurs commises par le système ne faisaient pas l'objet de sollicitations. Ce type d'informations objectives pourrait maintenir les participants à un niveau de confiance pertinent au regard de la fiabilité réelle du système.

La principale limite de cette étude était que la comparaison était basée sur l'absence ou la présence des sollicitations. Pour confirmer nos interprétations, il faudrait faire varier le degré de pertinence des sollicitations afin de vérifier l'impact de cette variation à la fois sur la confiance et sur les performances. Des sollicitations plus pertinentes devraient améliorer la détection d'un plus grand nombre d'erreurs, mais cela devrait aussi augmenter la confiance dans la machine et donc réduire la détection des erreurs non signalées. À l'inverse, une diminution de la pertinence des

sollicitations devrait aider les participants pour la détection d'un plus faible nombre d'erreurs, mais cela devrait aussi éviter l'erreur de naïveté et ainsi ne pas diminuer leurs contrôles sur les autres interprétations. Dans l'expérience présentée ici, la pertinence des sollicitations n'était pas manipulée. Elle ne permet donc pas de confirmer en détail ces interprétations.

## **5. CONCLUSION**

Dans le cadre du projet MobiSketch, les résultats de cette expérience soulèvent autant de questions qu'ils apportent de réponses. Les spécificités techniques du projet donnent un intérêt supplémentaire aux sollicitations. En effet, puisque le système est capable d'apprendre au fur et à mesure de l'analyse, les sollicitations pourraient permettre de diminuer le nombre des erreurs réalisées. L'équilibre à trouver dans le degré de fiabilité des sollicitations devra faire l'objet d'études à venir, prenant en compte à la fois les variations de performance et de confiance. Une seconde piste de réflexion pourrait consister à éviter par des moyens détournés l'erreur de naïveté sans avoir à sacrifier une part de fiabilité des sollicitations. Par exemple, une information chiffrée pourrait être donnée à l'utilisateur sur la fiabilité des sollicitations afin d'impacter la crédibilité de réputation avant que ne s'installe la crédibilité basée sur l'expérience (Fogg & Tseng, 1999). Cela consisterait à afficher au début de l'interaction le taux d'erreurs ne faisant pas l'objet de sollicitation. Pour que la crédibilité basée sur l'expérience ne génère pas d'erreur de naïveté et puisque les informations non pertinentes tendent à diminuer la confiance (Khasawneh *et al.* 2003), une possibilité plus radicale pourrait consister à faire des sollicitations moins pertinentes au début de la tâche qu'à la fin pour que la confiance se maintienne à un niveau conduisant les participants à rester vigilants. Ces solutions éventuelles devraient être évaluées dans des travaux à venir avant d'être implémentées.

# Étude 6 : Évaluation du prototype auprès des utilisateurs finaux

## 1. Introduction

L'étude précédente a mis en évidence les avantages et les inconvénients des sollicitations de l'utilisateur par le système. D'une manière plus générale, les différentes études présentées dans ce manuscrit ont permis de répondre à plusieurs questions concernant l'ergonomie d'un système d'interprétation automatique de plan dans le but de faciliter le repérage et la correction des erreurs de reconnaissance.

Plus de 200 volontaires ont participé aux études précédentes et il n'aurait pas été possible de constituer un tel échantillon au sein d'une population de professionnels. Il a néanmoins été décidé de conclure ce travail par deux évaluations auprès de professionnels de l'architecture. La première évaluation était une enquête en ligne dans laquelle il était demandé aux participants de donner leur avis sur le système après avoir visionné une vidéo présentant ses fonctionnalités. Malheureusement, le faible nombre de réponses à cette enquête (9 participants) n'a pas permis de s'en satisfaire. C'est pourquoi des entretiens ont ensuite été réalisés afin d'obtenir des retours plus qualitatifs sur l'interface.

Puisque l'état d'avancement du prototype ne permettait pas de réaliser un véritable test utilisateur, cette seconde évaluation a consisté en une démonstration présentielle des fonctionnalités du système, suivie d'un entretien.

Cette partie se compose donc d'une présentation du *Technology Acceptance Model* (TAM) qui a été utilisé lors de l'enquête en ligne, suivie d'une description de l'enquête elle-même ainsi que de ses résultats, bien que la portée de ces derniers soit

sensiblement réduite à cause du très faible effectif. La seconde évaluation sous forme d'entretiens est ensuite présentée, et sera suivie par une conclusion.

## **1.1. Evaluation de l'intention d'usage : le TAM**

Le TAM est un modèle de l'acceptabilité des nouvelles technologies reposant principalement sur l'évaluation de l'utilité perçue et de la facilité d'utilisation perçue (Davis, 1989). L'acceptabilité peut être définie comme le fait d'être plus ou moins favorable à l'usage d'une technologie avant son implémentation (Schuitema, Steg, & Forward, 2010). Il est donc possible d'évaluer l'acceptabilité d'une technologie après l'avoir présentée à des futurs utilisateurs potentiels. L'acceptation, en revanche, correspond au fait d'être plus ou moins favorable à l'usage d'une technologie après son utilisation (Schuitema *et al.*, 2010). La principale différence entre l'acceptabilité et l'acceptation est donc temporelle, ces deux mesures se situant respectivement avant et après le moment auquel la technologie est rendue disponible pour les utilisateurs. Outre cette considération temporelle, l'acceptation renverrait à la volonté d'utiliser une technologie conformément à la fonction qui a guidé sa conception (Dillon & Morris, 1996).

Selon Davis (1993), il est possible d'améliorer l'adoption d'une technologie en modifiant les caractéristiques de cette dernière. L'usage réel peut en effet être prédit par des mesures antérieures de l'intention d'usage (Davis, Bagozzi, & Warshaw, 1989; Davis, 1993), mais aussi par des mesures d'utilité et de facilité d'usage perçue (Henderson & Divett, 2003). Les liens particuliers qu'entretiennent ensemble l'intention d'usage, l'utilité perçue et l'utilisabilité perçue sont décrits dans le TAM (voir figure 35).

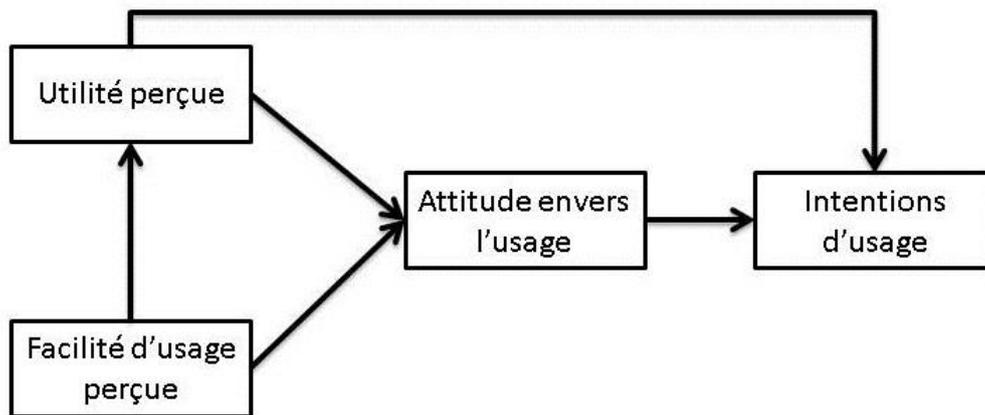


Figure 35. Le TAM, d'après Davis (1989)

Le TAM (Davis, 1989) est un modèle de référence dans le domaine de l'acceptabilité des nouvelles technologies. Bien qu'il ne soit pas le plus exhaustif quant aux dimensions prises en compte, il présente notamment l'avantage d'être parcimonieux (Venkatesh, Davis, & Morris, 2007; Ha & Stoel, 2009), ce qui le rend relativement peu coûteux à utiliser. Son second avantage concerne ses échelles qui sont à la fois très robustes et facilement généralisables (Venkatesh *et al.*, 2007). Du fait de ces qualités, ce modèle reste encore aujourd'hui très utilisé pour l'évaluation de l'acceptabilité de technologies comme par exemple un système de sécurité par puces RFID (Cheng, 2013), un système en ligne d'évaluation des performances d'étudiants (Muniandy, Ong, Phua, & Ong, 2011a; Muniandy, Ong, Phua, & Ong, 2011b) ou encore un magasin en ligne (Henderson & Divett, 2003). Toutefois, un certain nombre d'auteurs ont souligné les limites de ce modèle (voir par exemple Février, Gauduchea, Jamet, Rouxel, & Salembier, 2011). Par exemple, malgré la force des corrélations généralement mise en évidence avec le TAM, leurs poids varient de manière très importantes selon les études considérées, ce qui suggère que ce modèle est très incomplet (King & He, 2006).

La facilité d'usage et l'utilité perçues sont deux variables fondamentales dans la détermination de l'intention d'usage. La facilité d'usage perçue renvoie au « degré auquel une personne se situe dans la croyance qu'utiliser un système particulier va se faire sans effort » (Davis, 1989, p. 320, notre traduction). L'utilité perçue serait quant à

elle le « degré auquel une personne se situe dans la croyance selon laquelle l'utilisation d'un système particulier devrait augmenter ses performances au travail » (Davis, 1989, p. 320, notre traduction).

Dans l'échelle créée par Davis (1989), 6 items d'utilité perçue ont été retenus, ainsi que 6 items de facilité d'usage perçue. Pour chacun d'entre eux, les participants devaient estimer leur degré d'accord sur une échelle en 7 points. À titre d'illustration, la proposition « je trouverais ce système facile à utiliser » (notre traduction) correspondait à un exemple d'item mesurant la facilité d'utilisation perçue tandis que « utiliser ce système augmenterait ma productivité au travail » (notre traduction) correspondait à un exemple d'item mesurant l'utilité perçue.

## 2. Evaluation 1

L'objectif de la première évaluation était de mesurer l'utilité et l'utilisabilité perçue des différentes fonctions du système par des professionnels de l'architecture. L'état d'avancement du prototype ne permettant pas la réalisation d'un test qui inclurait l'ensemble des fonctionnalités, il a été décidé de réaliser une vidéo<sup>5</sup> présentant les fonctionnalités du système final à travers un cas d'usage spécifique : l'interprétation automatique, puis l'édition d'un plan d'architecture manuscrit réalisé sur support papier. Cette vidéo visait à servir de support de présentation du système. Elle était accompagnée d'items d'utilité et d'utilisabilité perçue portant sur les différentes fonctions présentées dans la vidéo. Cette approche devait permettre d'identifier les problèmes critiques, nécessitant prioritairement des corrections.

---

<sup>5</sup> Cette vidéo est disponible en ligne sur le site web du projet MobiSketch (<http://mobisketch.irisa.fr/index.php>) sous le nom « Vidéo des résultats du projet MobiSketch »

## 2.1. Méthode

### 2.1.1. Participants

Les participants à cette enquête étaient 9 volontaires dont 6 hommes et 3 femmes. Le plus jeune avait 22 ans et le plus âgé en avait 48, pour une moyenne d'âge de 27 ans ( $SD = 8,8$ ). On dénombrait parmi eux 3 étudiants en école d'architecture, 5 architectes et 1 conducteur de travaux. Ces participants ont été sélectionnés parce qu'il s'agissait de futurs utilisateurs potentiels du système développé dans le cadre du projet MobiSketch.

### 2.1.2. Matériel et procédure

Les participants étaient invités à se connecter sur une page web affichant à la fois une vidéo de présentation du système et des questions. La vidéo exposait un exemple d'usage possible du système qui consistait à dessiner à la main un plan sur papier, à le prendre en photo avec une tablette, à l'analyser et à l'éditer directement avec la tablette (voir figure 36). D'une durée de 2 minutes et 41 secondes, cette vidéo était accompagnée d'une voix-off décrivant l'ensemble des fonctionnalités.

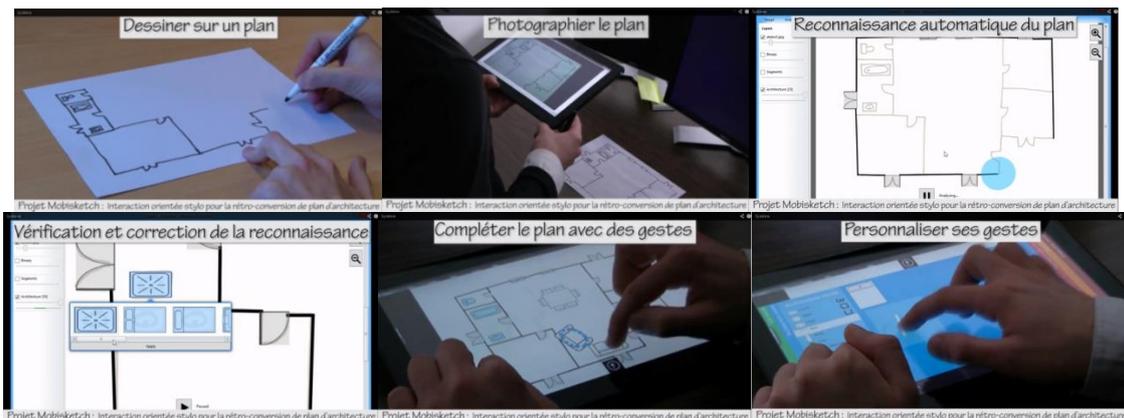


Figure 36. Captures d'écran des différentes étapes présentées dans la vidéo

Sous cette vidéo, les participants avaient accès à une série de questions (voir annexe 11) auxquelles ils devaient répondre par le biais d'échelles visuelles analogiques, ainsi qu'à travers des questions ouvertes. Ces questions portaient sur l'utilité, l'utilisabilité,

l'attitude et l'intention d'usage du système. Après avoir répondu à ces questions, ils parcouraient successivement les pages suivantes, dont chacune portait sur un aspect spécifique du système, à savoir la photographie d'un plan directement dans le logiciel, la reconnaissance automatique du plan, la correction des erreurs pendant l'analyse, l'édition du plan par des gestes et leur personnalisation ainsi que les sollicitations par le système.

Sur chacune des pages, une vidéo présentant uniquement la séquence portant sur la fonctionnalité considérée était affichée, ainsi que les mêmes questions d'utilisabilité et d'utilité perçue que sur la première page, mais faisant cette fois référence à la fonctionnalité concernée. Les items étaient des adaptations de ceux utilisés par Davis (1989). Par exemple, les items d'utilisabilité perçue étaient du type « apprendre à utiliser ce système serait facile pour moi » et les items d'utilité perçue était du type « je trouverais ce système utile dans mon travail ». Les participants devaient évaluer leur degré d'accord ou de désaccord avec les items sur des échelles visuelles analogiques. En plus de ces items, plusieurs questions plus ouvertes étaient posées comme par exemple « vous-même, que feriez-vous avec ce système ? ».

## **2.2. Résultats**

### **2.2.1. Utilité et utilisabilité perçues des fonctions du système**

La figure 37 présente les moyennes pour l'utilité et l'utilisabilité perçue pour chacune des fonctionnalités présentées. On constate sur ce graphique que l'évaluation de l'utilité perçue est systématiquement inférieure à l'évaluation de l'utilisabilité perçue. La moyenne la plus faible d'utilisabilité perçue concerne la fonctionnalité de sollicitation, qui est évaluée à un niveau de 74,8 sur l'échelle visuelle analogique codée en 100 points. L'ensemble des fonctions sont évaluées à des niveaux moyens situés entre 74,8 et 91,44, ce qui peut être considéré comme plutôt satisfaisant. En revanche, les évaluations de l'utilité sont plus contrastées. On constate notamment que l'évaluation moyenne de l'utilité globale du système, d'une valeur de 59,59, est plutôt

faible relativement aux autres scores. Le système serait donc jugé moyennement utile puisque 50 correspond à la moitié de l'échelle. Le fait que les interruptions soient évaluées si sévèrement sur le plan de l'utilité (54,22), tout comme les sollicitations (61,6), pourrait provenir en partie du fait que les participants n'ont pas effectué la tâche de supervision de l'analyse. Ils pourraient donc ne pas avoir perçu la difficulté que représente la détection des erreurs d'interprétation.

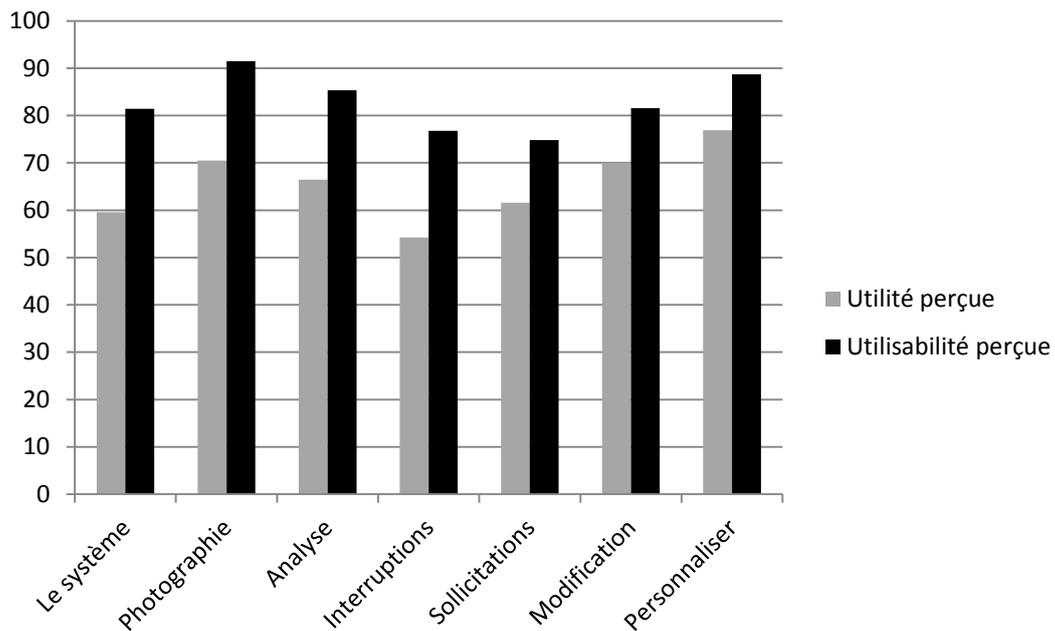


Figure 37. Moyenne des réponses pour l'utilité perçue et l'utilisabilité perçue de chacune des fonctionnalités présentées

Concernant le système en général, les items d'intention d'usage (« J'ai l'intention d'utiliser ce système ») et d'attitude envers l'usage (« Utiliser ce système est une bonne idée ») ont eux aussi donné lieu à des évaluations plutôt faibles, bien que positives (attitude moyenne = 66,7 ; intention d'usage moyenne = 64,22).

## 2.2.2. Questions d'usage

Malgré l'évolution technologique qui impacte à plusieurs niveaux le métier d'architecte, plusieurs auteurs considèrent que la première phase créative du début de la conception se fait toujours à la main, sur support papier (Manolopoulou, 2005; Bilda

& Gero, 2006). Afin de confirmer cette assertion, il était demandé aux participants de préciser s'ils travaillent effectivement de la sorte lorsqu'ils réalisent une activité de conception. À cette question, les 9 participants interrogés ont répondu « oui ». Lors de l'enquête, il était également demandé aux participants de préciser les logiciels qu'ils utilisaient dans le cadre de leur travail. Les réponses par ordre de fréquence étaient les suivantes : AutoCAD, ArchiCAD, Sketchup, suite Adobe (Photoshop, InDesign), suite Office, Artlantis et Maxwell Render. Les trois logiciels les plus cités servent directement au traitement de plans d'architecture.

AutoCAD architecture (Autodesk) est un logiciel de référence permettant la conception de plans d'architecture et la création de rendu 3D. ArchiCAD est quant à lui un logiciel de création de modèles en 3 dimensions ayant la spécificité d'inclure un grand nombre de fonctionnalités spécifiques au domaine de l'architecture (métrés, gestion des étages...), alors que Sketchup est un logiciel plus généraliste de création en 3D.

La totalité des réponses aux questions ouvertes posées est présentée en annexe 12. À la question « vous-même, que feriez-vous avec ce système ? », 4 personnes ont apporté une réponse. Pour deux d'entre elles, ce système pourrait servir lors des phases de réflexions avec les clients alors que les deux autres ont plutôt fait référence au fait de convertir directement une esquisse manuscrite, du support papier vers l'ordinateur.

Sur la même page, il était proposé aux participants de faire des remarques ou des commentaires sur le système. Les 4 réponses obtenues convergeaient vers l'idée selon laquelle le système devrait comporter des fonctions supplémentaires pour pouvoir mieux prendre en compte la complexité réelle des plans. Ainsi, les aspects à ajouter seraient la gestion des échelles, des différents types d'isolation pour les parois et le passage à la 3D. Le quatrième participant a répondu que ce type de système est inutile pour un professionnel car le programme serait trop simple par rapport à la complexité des plans utilisés.

Enfin, il était proposé aux participants de faire des remarques sur leur propre travail. Parmi les deux réponses collectées, la première a consisté à ré-évoquer la situation de conception avec le client qui pourrait être « plus conviviale », alors que la seconde a

insisté sur la nécessité d'étendre les possibilités techniques du système en lui permettant notamment de gérer les calques.

## 2.3. Discussion

L'objectif de cette enquête était de réaliser une évaluation des jugements subjectifs de professionnels sur le système développé dans le cadre du projet MobiSketch. Ces retours devaient permettre d'identifier des problèmes critiques du système afin de pouvoir les corriger. Compte tenu du faible nombre de participants à cette enquête, il convient de rester très prudent quant aux interprétations des résultats.

Les réponses aux items issus du TAM semblent indiquer que, du point de vue des professionnels interrogés, la principale qualité du système est sa facilité d'usage. En revanche, les autres aspects sur lesquels portaient les questions ont été évalués moins positivement, comme par exemple l'utilité ou l'intention d'usage du système. Les questions d'usage ont apporté un premier éclairage sur ces résultats. En effet, plusieurs participants ont souligné que le système, tel qu'il est actuellement conçu, ne permet pas de passer directement de l'esquisse papier au logiciel de DAO classique car il lui faudrait pour cela être en mesure de traiter la totalité des éléments qui constituent un plan, dont les échelles ou encore les caractéristiques des parois. Les prototypes actuels ne traitent en effet que des emplacements et orientations spatiales des différents ouvrants, mobiliers et parois, sans prise en compte des nombreuses caractéristiques de ces éléments. Dans l'état actuel, l'importation d'une esquisse interprétée directement dans un logiciel de DAO comme AutoCAD nécessiterait que l'utilisateur reprenne ensuite manuellement un certain nombre d'éléments comme les échelles ou même les angles. Dans ces conditions, il est difficile d'évaluer le gain de temps que représenterait l'analyse automatique du plan, trop superficielle, par rapport à une recopie du plan dans le logiciel de DAO, certes plus longue, mais débouchant directement sur une version véritablement exploitable du plan. Les résultats de cette enquête ont permis de mettre en évidence cette insuffisance importante du système, mais n'apportent pas d'éléments précis quant aux solutions qui permettraient de

résoudre ce problème. Plusieurs points ont été évoqués comme par exemple la gestion des calques ou des échelles, mais ces points restent à préciser.

## **3. Evaluation 2**

La seconde évaluation visait à compléter l'enquête en ligne ; insuffisante en termes d'effectif. Les résultats de l'évaluation 1 semblaient indiquer que, dans son état actuel, le prototype ne permettrait pas de réaliser de manière pertinente la conversion de l'esquisse dessinée sur papier en un format numérique utilisable via un logiciel de DAO. Cependant, plusieurs questions restent à éclaircir. Tout d'abord, les entretiens réalisés dans le cadre de l'évaluation 2 devaient permettre de clarifier les raisons qui empêchent l'application d'une reconnaissance d'esquisses papier permettant d'intégrer directement ces dernières dans un logiciel de DAO. Il s'agissait aussi de déterminer les améliorations qui permettraient de régler ce problème. Le deuxième objectif de l'évaluation 2 était de préciser l'idée selon laquelle le système pourrait être utile dans la relation client. Enfin, le troisième objectif était de dégager d'autres applications possibles du système dans le domaine de l'architecture.

Etant donné l'état d'avancement technique du prototype, celui-ci ne permettait pas encore de passer directement, depuis la même interface, de l'étape d'analyse d'un plan à celle d'édition d'un plan. Ainsi, il n'était pas envisageable de réaliser un véritable test utilisateur dans lequel les participants auraient utilisé le système. Il a donc été décidé de mettre en place une démonstration des différentes fonctionnalités pour chaque participant avant de procéder à un entretien portant sur les besoins et les contraintes des professionnels en rapport avec les possibilités techniques du système.

### **3.1. Méthode**

#### **3.1.1. Participants**

Les participants ont été sélectionnés dans la base de volontaires du LOUSTIC du fait de leur expertise des métiers de l'architecture. Il s'agissait de 4 hommes et 2 femmes de

22 à 63 ans pour une moyenne d'âge de 32 ans ( $SD = 17,6$ ). L'échantillon était composé de 4 étudiants en école d'architecture en fin de cursus (dont un en stage en agence au moment de l'entretien), d'un assistant architecte et d'un maître d'œuvre.

### 3.1.2. Matériel et procédure

La première partie du test consistait en une démonstration par l'expérimentateur des fonctionnalités du système. Les différentes étapes décrites dans la vidéo de l'enquête en ligne étaient réalisées devant les participants et expliquées oralement. Le plan d'architecture analysé était le même que celui de la vidéo. La tablette utilisée pour la démonstration était de type Samsung ATIV Smart PC pro avec un écran de 11.6 pouces et une définition de 1920 par 1080 pixels. Après la démonstration, les participants étaient invités à manipuler eux-mêmes le prototype. Enfin, ils étaient questionnés sur la pertinence du système pour leurs activités, sur son utilité et sa facilité d'usage, ainsi que sur ses applications possibles. L'ensemble de l'entretien était enregistré par l'intermédiaire d'un dictaphone.

## 3.2. Résultats

L'objectif de l'évaluation 2 était d'obtenir des informations concernant les applications possibles du système. Quatre applications principales ont émergé : créer un continuum papier/DAO, faire de l'esquisse de projet, de l'édition de plan en mobilité et faciliter la relation client. Les phrases des participants ont été réparties dans ces 4 catégories (certaines phrases ne renvoyaient à aucune de ces catégories). Cette répartition a été réalisée par deux observateurs différents avec un taux de congruence de 83,4%. Les phrases étant catégorisées de manière non congruente ont été écartées car considérées comme ambiguës.

### 3.2.1. Continuum papier/DAO

La réalisation d'un continuum entre le papier et la DAO correspond au cas d'usage qui était présenté dans la vidéo de l'évaluation 1. Elle renvoie à la situation de début de

projet dans laquelle l'architecte a réalisé des esquisses sur papier et souhaite reprendre ces esquisses sur un logiciel de DAO pour en faire des plans. Le système pourrait avoir pour fonction de réaliser automatiquement la transition entre l'esquisse dessinée sur papier et le logiciel de DAO, évitant ainsi la fastidieuse étape de recopie. L'ensemble des réponses concernant cette application sont présentées en annexe 13. Les commentaires des personnes interrogées indiquent que cette application serait, selon elles, utile. Par exemple, les deux extraits suivants vont dans ce sens :

- « S'il arrive à sortir un DWG<sup>6</sup> nickel d'un plan dessiné, c'est très intéressant. »
- « Il y a pas mal de bâtiments anciens pour lesquels on n'a que le plan papier. Là en projet, on fait de la réhabilitation et du coup, on a des grands plans à la main et on a tout repris sur ordi. Donc là, ça serait quand même plus rapide. »

Cependant, un certain scepticisme était exprimé quant à la capacité du logiciel à réaliser ce continuum (« là, ça me paraît difficilement exportable sur des logiciels comme AutoCAD, ArchiCAD ou autre. »).

Pour que le système puisse totalement remplacer la recopie, il faudrait qu'il soit capable de prendre en compte la totalité des informations techniques que gèrent les principaux logiciels de DAO (paramètres des murs, épaisseur, matière, isolations, *etc*). En particulier, il devrait être en mesure de gérer les échelles de manière précise. Si elles sont bien prises en compte, alors l'exportation peut s'avérer utile même s'il manque d'autres informations techniques, car celles-ci peuvent être ajoutées dans un second temps. Ainsi, concernant l'application d'interprétation automatique de plans papier pour éviter l'étape de recopie manuelle dans un logiciel de DAO, la priorité est l'ajout dans le système de la gestion précise des échelles.

### 3.2.2. Esquisse de projet

L'application du système à l'esquisse de projet consisterait d'abord à dessiner sur papier et à interpréter automatiquement le plan ainsi réalisé pour poursuivre un travail d'esquisse sur tablette. Les commentaires concernant cette application sont

---

<sup>6</sup> DWG est un format de fichier très répandu dans le domaine du DAO.

présentés en annexe 14. Deux principales idées se sont dégagées pour cette application.

La première consisterait à vectoriser les esquisses en début de projet. Cela permettrait des communications sur le projet à partir de supports plus nets et plus facilement éditables que des dessins manuscrits, puisque la vectorisation autorise à changer de taille sans perte de qualité. Ainsi, une esquisse d'intention pourrait évoluer dans le temps, en s'enrichissant sans devoir être redessinée à chaque modification importante.

La seconde idée consisterait à se servir du système comme d'un outil de dessin d'esquisses, permettant de bénéficier à la fois de la liberté d'utilisation du support papier et des fonctionnalités du support numérique. Concrètement, l'utilisateur pourrait faire une première esquisse sur papier, l'analyser et la faire évoluer progressivement sur l'interface.

### 3.2.3. Edition de plans en mobilité

Les remarques concernant l'édition de plans en mobilité sont présentées en annexe 15. Cette application renvoie aux commandes gestuelles sur la tablette permettant de faire des modifications sur un plan, des annotations depuis la tablette, tout en se trouvant sur un chantier. Selon les personnes interrogées, le système pourrait par exemple servir à faire des plans sur place, chez le client, sur un support avec lequel le résultat serait plus « propre » que celui souvent obtenu sur papier. En revanche, il semble peu envisageable qu'un architecte aille sur un chantier avec le plan complet sur tablette et effectue directement des modifications sur place. Les modifications sur le plan sont des traitements complexes qui ne seraient pas réalisés de cette manière. Un changement sur le plan peut en effet avoir un impact sur l'estimation des coûts, des quantités de matériaux, l'isolation, la lumière, *etc.* Tant que le système n'est pas capable de gérer ces différents aspects, l'utilisation en mobilité ne servirait qu'à effectuer de la prise de notes sur un plan (avec un système de calques), à anticiper des modifications à venir ou à créer des esquisses « propres » sur le terrain si les projets sont de faible envergure.

### 3.2.4. Relation client

L'application du système à la relation client consisterait à l'utiliser pour analyser un plan réalisé par le client lui-même ou en collaboration avec lui et/ou à effectuer les éditions sur un plan en temps réel en fonction de ses commentaires. Les remarques concernant cette fonctionnalité sont présentées en annexe 16. L'utilisation du système dans la relation client a été évoquée pour des vendeurs d'appartements, des décorateurs d'intérieur, des peintres en bâtiment et des architectes. Pour ce type d'applications, la simplicité du système par rapport aux logiciels habituels de DAO est une qualité. Cela permettrait de faire des propositions aux clients avec un support imagé, facile à comprendre et de modifier cette proposition en temps réel en fonction des commentaires. Dans l'extrait suivant, le participant donne un exemple d'usage :

- « Si on a rendez-vous à 16h avec un client, on arrive vers 15h, on commence à faire une esquisse, on la prend en photo, on la process et après, on modifie avec le client. »

L'intérêt pour cette application résiderait dans la simplicité d'utilisation du système. Il s'agirait d'une solution permettant de tester des idées à la fois rapidement et de manière plus séduisante pour le client en comparaison à ce que permet le support papier. L'utilisation de la tablette permet d'ailleurs de bénéficier de cette application de facilitation de la relation client en mobilité, par exemple sur le lieu des futurs travaux.

## 3.3. Discussion

L'objectif de l'évaluation 2 était de compléter la première approche réalisée par l'enquête en ligne en précisant les applications possibles du système et les points à améliorer pour permettre la mise en œuvre de ces applications. De plus, l'analyse du besoin réalisée par le partenaire industriel au début du projet nécessitait d'être complétée. En effet, au regard des dernières avancées techniques, il était possible de proposer aux participants des descriptions plus détaillées de fonctionnalités réalisables. Pour cela, 6 personnes (professionnels ou étudiants dans le domaine de

l'architecture) ont été rencontrés afin de leur présenter le système et d'obtenir des retours de leur part sur ses applications possibles.

L'application principale qui est envisagée depuis le début du projet est la possibilité de rétro-convertir une esquisse réalisée sur papier afin de l'importer directement dans un logiciel de DAO. Ce point a donc été largement abordé lors des entretiens. Par ailleurs, les réponses de l'évaluation 1 suggéraient une possibilité d'utilisation du système afin de servir de support dans la relation client. De plus, lors des entretiens, deux autres applications ont émergé : l'édition de plans en mobilité et l'esquisse de projets.

La rétro-conversion d'esquisses dans le but de les importer directement dans un logiciel de DAO est jugée à la fois utile et particulièrement difficile à mettre en place. Le principal obstacle serait la gestion des échelles. Sur AutoCAD par exemple, les dimensions de chaque élément sont spécifiées au moment de sa création. L'ajout d'un mur se fait par la définition d'un point de départ, d'une direction et d'une longueur avant de spécifier l'ensemble de ses caractéristiques annexes. Ce degré de précision spatial est difficile à envisager si le dessin est réalisé à la main. Pour être en mesure de créer un continuum entre le support papier et les logiciels classiques de DAO, le système devra être capable de régler ce problème en permettant au moins aux utilisateurs de spécifier les cotes.

Cependant, les capacités actuelles du système seraient suffisantes pour une utilisation dans le cadre des relations clients puisque cela implique principalement l'analyse automatique d'esquisses et les modifications à la volée du plan interprété. Enfin, l'application de travail sur les esquisses est aussi possible avec le système actuel. Cela consisterait à commencer par réaliser une première esquisse sur papier et à la rétro-convertir afin de poursuivre le travail de réflexion en passant par les modifications par commandes gestuelles.

Cette étude sur l'utilité et les applications du système souffre de plusieurs limites. Tout d'abord, seules 9 personnes ont répondu à l'enquête en ligne et 6 aux entretiens présentsiels. Ce faible nombre de réponses invite à la prudence quant aux conclusions que l'on peut tirer de cette étude. Par ailleurs, le prototype n'était pas suffisamment abouti pour réaliser un véritable test utilisateur. En effet, les évaluations de l'utilité et

de l'utilisabilité perçue étaient basées sur une démonstration alors qu'une utilisation réelle du système par les participants aurait certainement conduit à des remarques plus précises. Néanmoins, ces entretiens ont permis préciser les raisons qui ont conduit à un jugement modéré de l'intention d'usage lors de l'évaluation 1. Ainsi, aucun problème lié à l'utilisabilité n'a été évoqué par les participants. En revanche, c'est l'utilité du système au regard de ses différentes fonctionnalités qui a été critiqué. Or, une facilité d'usage élevée ne suffit pas à elle seule à générer des intentions d'usage. D'ailleurs, Davis (1989) a montré que bien que l'utilisabilité perçue soit un prédicteur de l'usage d'une technologie, l'utilité perçue en est un facteur encore plus important. Il est donc cohérent que dans notre étude, malgré une utilisabilité perçue plutôt élevée, l'intention d'usage du système se situe à un niveau modéré du fait du niveau moyen de l'utilité perçue.

# Conclusion générale<sup>7</sup>

## 1. Les niveaux d'automatisation au cours du projet

Tout au long du projet MobiSketch, le niveau d'automatisation des prototypes successifs a été progressivement modifié. Il est ainsi passé d'une automatisation maximale à un niveau appelé « prise de décision mixte », *i.e.* un partage des décisions entre l'homme et la machine (voir figure 38). Les prototypes utilisés dans les études 1 et 2 correspondaient à ce que Kaber et Endsley (1997) appelaient « automatisation totale ». Lorsque l'automatisation est totale, l'humain est totalement en dehors de la boucle de contrôle (Kaber & Endsley, 2003). En l'occurrence, avec les prototypes des études 1 et 2, l'analyse des plans se faisait entièrement sans intervention possible des utilisateurs, ces derniers ne pouvant vérifier les erreurs qu'une fois l'interprétation entièrement terminée. Dans l'étude 3, le niveau d'automatisation du prototype a été diminué pour passer de l'automatisation totale au contrôle supervisé. En permettant aux utilisateurs d'interrompre l'analyse pour intervenir dès qu'ils repéraient une erreur, le prototype de l'étude 3 opérait une réintroduction partielle des utilisateurs dans la boucle de contrôle. Dans les études 4 et 5, le partage des tâches entre l'humain et la machine a été de nouveau accentué. Dans ces études, le système sollicitait l'utilisateur lorsqu'il avait un doute sur une interprétation. Ainsi, non seulement l'utilisateur vérifiait les interprétations effectuées par le système, mais il prenait également part aux décisions d'interprétations pour les symboles les plus difficiles. Il ne s'agissait donc plus seulement d'une supervision de l'analyse, mais bien d'une forme de collaboration entre l'homme et la machine dans la prise de décision. Cet

---

<sup>7</sup> Les préconisations présentées dans cette partie ont fait l'objet d'un article:

Fleury, S., Jamet, E., Loup-Escande, E., Ghorbel, A., Lemaitre, A., & Anquetil, E. (in press). Towards Specifications for Automatic Recognition Software : An Example of a User-Centred Design. *Journal of Software Engineering and Applications*.

ajustement du niveau d'automatisation de la tâche est une démarche cohérente avec le caractère peu prévisible des erreurs. En effet, Sheridan et Verplank (1978) conseillaient d'adapter le degré d'automatisation des tâches à leur niveau de prévisibilité. Une tâche dont la prévisibilité est moyenne ne devrait être que partiellement automatisée.

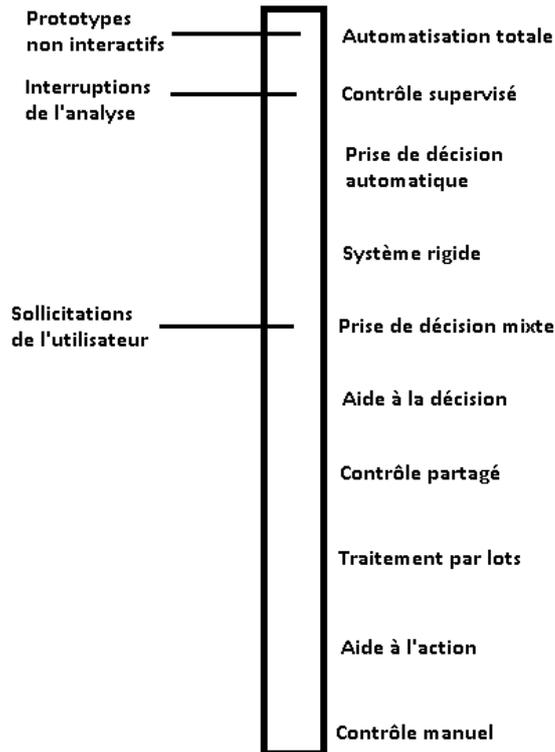


Figure 38. Évolution du niveau d'automatisation des prototypes au cours du projet MobiSketch, d'après l'échelle des niveaux d'automatisation de Kaber et Endsley (1997)

## 2. Préconisations

Cette thèse avait pour objectif de répondre à la problématique suivante : comment améliorer les performances du système homme-machine et diminuer les erreurs d'interprétations de symboles ? Le choix de cette problématique a été largement conforté par les résultats obtenus lors de la première étude puisque sur un plan très simple, seuls 32,5% des participants évalués sont parvenus à repérer l'ensemble des erreurs. L'amélioration des performances du système passe par la facilitation des

différentes étapes de traitement par lesquelles passe l'utilisateur durant la tâche. Ainsi, selon le modèle de Rasmussen (1983), cela consiste d'abord à faciliter l'extraction des caractéristiques, la reconnaissance et l'association été/tâche afin de favoriser la détection des erreurs. Cela implique à la fois d'optimiser la lisibilité des informations à l'écran, leur coréférenciation et leur contrôle exhaustif. Les deux étapes suivantes (règles stockées pour la tâche et pattern sensorimoteur automatisé) peuvent être facilitées en simplifiant la correction des erreurs par l'utilisateur. Par exemple, permettre à l'utilisateur d'interrompre l'analyse pour corriger des erreurs sans avoir à les maintenir en mémoire peut améliorer la mise en œuvre des corrections. Ce travail touchant à sa fin, un certain nombre de réponses concrètes à la problématique peut être proposé. Pour ce faire, la problématique générale a été décomposée en 5 questions plus précises :

- Comment présenter les informations à l'écran ?
- Faut-il permettre aux utilisateurs d'interrompre l'analyse par le système ?
- Quelle forme donner aux fonctions interactives ?
- Comment faciliter la vérification de l'interprétation ?
- Le système doit-il communiquer ses incertitudes aux utilisateurs ?

Les préconisations présentées ci-dessous ne s'appliquent pas uniquement à l'interprétation automatique de plans d'architecture. Elles concernent tous les logiciels d'interprétation de documents techniques (plans cadastraux, cartes routières, plans électriques, partitions musicales...) et de beautification qui impliquent un risque d'erreurs.

## **2.1. Comment présenter les informations à l'écran ?**

La première question abordée concerne la disposition à l'écran des informations que l'utilisateur doit traiter pour vérifier l'interprétation. L'étude 2 présentée dans ce manuscrit décrit 2 expériences dans lesquelles les participants devaient entourer des erreurs d'interprétation de plans dans 3 conditions expérimentales : avec une

interprétation séparée du plan, intégrée ou apparaissant de manière séquentielle. L'étude 2 a permis de mettre en évidence un effet positif de la superposition de l'interprétation. En effet, lorsque l'interprétation apparaissait de manière superposée au plan, les participants réalisaient la vérification plus rapidement que lorsque cette même interprétation apparaissait à côté du plan. Ce résultat indique qu'il serait préférable de superposer ces informations pour faciliter la corréférence des informations à comparer.

Cette même étude a permis de montrer qu'une présentation séquentielle de l'interprétation pouvait améliorer la performance de repérage des erreurs. Les participants du groupe Séquentiel étaient significativement plus nombreux à avoir repéré toutes les erreurs en comparaison aux participants du groupe Intégré. Ce résultat a été interprété comme une conséquence d'un guidage attentionnel qui serait généré par l'apparition séquentielle de l'interprétation. Le guidage permettrait d'aboutir à une exploration plus exhaustive du document à vérifier. Ainsi, plutôt que de considérer la reconnaissance automatique comme un traitement dont on ne présente à l'utilisateur que le résultat final, il est préférable d'afficher ce travail d'analyse à l'écran, au fur et à mesure qu'il se déroule. Cette modification n'exerce d'ailleurs aucune contrainte particulière puisque d'une part, les utilisateurs peuvent n'effectuer leur vérification qu'après la fin de l'analyse s'ils le souhaitent et d'autre part, ce temps d'analyse serait tout de même présent en l'absence de l'affichage séquentiel. Enfin, afficher en temps réel les interprétations effectuées par le système va dans le sens d'une amélioration de la conscience de la situation par les utilisateurs puisque, au fur et à mesure de la tâche, ceux-ci obtiennent ainsi des informations sur son état d'avancement, mais aussi sur la manière dont se déroulent les traitements effectués.

À la question de savoir comment présenter les informations à l'écran pour faciliter la correction des erreurs par les utilisateurs, les 2 expériences qui constituent l'étude 2 de ce travail permettent d'apporter la réponse suivante : pour améliorer l'efficacité des utilisateurs, il est préférable de présenter les informations à comparer de manière superposée plutôt que côte à côte. Le fait d'afficher les résultats de l'analyse en temps

réel, plutôt que de les cacher, peut également générer un guidage attentionnel permettant d'améliorer l'efficacité de la correction.

## 2.2. Faut-il permettre aux utilisateurs d'interrompre le système ?

L'étude 3 avait pour objectif d'évaluer l'impact des interruptions de l'analyse par l'utilisateur sur les performances. Dans la version « séquentielle » du prototype, les participants assistaient à l'analyse en temps réel mais n'avaient pas la possibilité d'intervenir en cas d'erreur. Autrement dit, lorsqu'ils repéraient des erreurs, ils devaient attendre la fin de l'analyse pour pouvoir les entourer. Une des versions testée dans l'étude 3 autorisait une interactivité puisque les participants pouvaient interrompre l'analyse lorsqu'ils le souhaitaient afin d'entourer des erreurs sans devoir attendre la fin de l'interprétation. Les résultats suggèrent que les interruptions de l'analyse permettent globalement de gagner du temps. Suite à cette étude, la fonction d'interruption de l'analyse par les utilisateurs a été intégrée dans le prototype (voir figure 39).

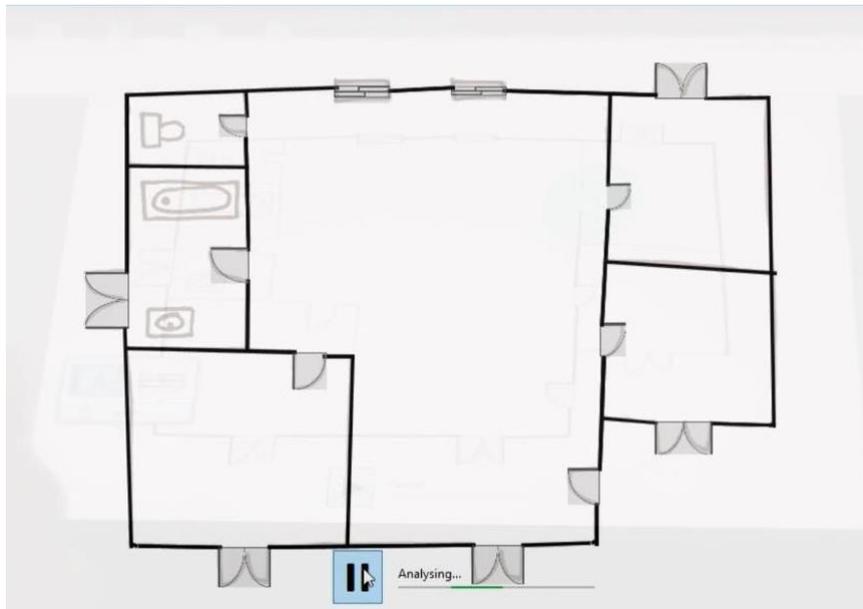


Figure 39. Analyse en cours sur une version de prototype incluant la fonction d'interruption de l'analyse

L'étude 3 permet donc de répondre à la question « faut-il permettre aux utilisateurs d'interrompre le système ? ». Il est effectivement pertinent de le permettre car lorsqu'ils le font, cela a tendance à leur faire gagner du temps en leur évitant d'oublier les erreurs repérées.

## **2.3. Quelle forme donner aux fonctions interactives ?**

Dans l'étude 4, les participants étaient soumis à un entretien semi-directif sur les qualités de l'interface après avoir réalisé avec cette dernière l'interprétation de 3 plans successifs. Le prototype utilisé avait la particularité d'intégrer la fonctionnalité de sollicitation des utilisateurs par le système en cas de risque d'erreur. Cette étude a permis de dégager plusieurs préconisations pour la conception de l'interface, mais aussi des éléments interprétatifs grâce aux réponses issues des entretiens.

Les résultats indiquent que les interruptions ainsi que les sollicitations sont toutes deux considérées par la plupart des participants comme apportant une aide plutôt qu'une gêne lors de la tâche. Néanmoins, certains participants ont évoqué un risque de diminution de l'attention sur l'analyse par l'utilisation des interruptions. Cette crainte pourrait s'expliquer par le fait que le bouton déclenchant l'interruption se situait à l'extérieur du plan à contrôler et que pour cliquer dessus, il fallait donc arrêter quelques instants de regarder le plan au risque de perdre le fil. Pour régler ce problème, nous suggérons d'implémenter un raccourci pour cette fonction d'interruption (*i.e.* qu'un simple clic n'importe où sur le plan déclenche l'interruption de l'analyse). Ainsi, les participants n'ont pas besoin de décentrer leur attention du plan pour interrompre l'analyse. De plus, cet ajout est totalement invisible pour un utilisateur qui ne souhaiterait pas s'en servir.

Au vu des commentaires portant sur les sollicitations, il a semblé que celles-ci pourraient être bien plus efficaces en apportant des améliorations relativement simples. Deux principaux points sont apparus dans les commentaires critiques sur les sollicitations : l'emplacement de la fenêtre de sollicitations ne s'adaptait pas à

l'emplacement du symbole ciblé et le surlignage du symbole ciblé nuisait à sa lisibilité. Ainsi, il est préférable de lier la zone d'apparition des sollicitations avec le symbole ciblé en la faisant apparaître le plus proche possible de ce dernier, afin d'améliorer la coréférenciation des informations. Il est d'ailleurs possible d'utiliser cette proximité pour remédier également au second élément gênant, à savoir le surlignage. En effet, il suffit alors de pointer le symbole qui fait l'objet d'une sollicitation par une flèche partant directement de la fenêtre (voir figure 40).

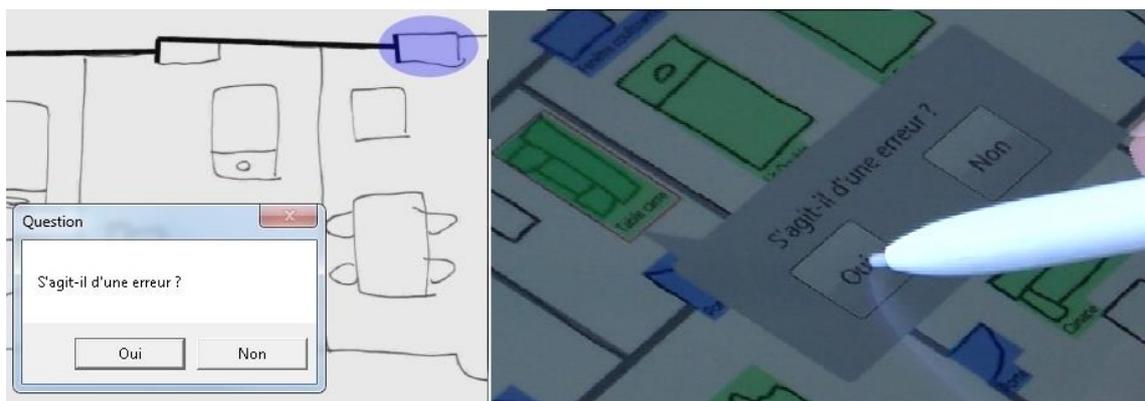


Figure 40. Présentations des sollicitations utilisées durant l'étude 4 (à gauche) et après leur amélioration dans le prototype (à droite)

Les fonctions interactives de ce logiciel sont les interruptions de l'analyse par l'utilisateur et les sollicitations de l'utilisateur par le système. L'étude 4 apporte des réponses quant à la forme qu'il est pertinent de donner à ces fonctions. Il serait ainsi préférable que les interruptions par l'utilisateur ne soient pas uniquement réalisées par l'utilisation d'un bouton situé à l'extérieur du plan, ce qui risquerait en effet de décentrer l'attention des utilisateurs. Sans nécessairement supprimer ce bouton, il est important de le compléter par une solution moins interférente avec la tâche de supervision, comme par exemple le fait de pouvoir cliquer directement sur le plan en interrompant ainsi l'analyse. Toujours afin d'éviter de décentrer l'attention des utilisateurs de la zone qu'ils sont en train de contrôler, des sollicitations du système devraient apparaître à proximité de l'endroit auquel ces sollicitations font référence, plutôt qu'à un endroit fixe.

## 2.4. Comment faciliter la vérification de l'interprétation ?

Lors de l'entretien semi-directif mené lors de l'étude 4, il était demandé aux participants de suggérer des modifications du logiciel permettant de faciliter la correction des erreurs. Concernant cette question, deux principales catégories de réponses ont émergé : la première consistait à dire que l'aspect graphique des sollicitations devait être amélioré et la seconde renvoyait à l'ordre d'apparition des interprétations, jugé non pertinent.

Le principe ergonomique de groupement et de distinction entre les items (Bastien & Scapin, 1993) repose sur le fait que les éléments qui sont visuellement proches dans l'espace, ou similaires en termes de forme ou de couleur, ont tendance à être interprétés comme appartenant à une même catégorie. Ainsi, il est important de bien choisir les emplacements des items, mais aussi leurs caractéristiques graphiques pour renseigner l'utilisateur sur les liens que ces items entretiennent entre eux. Dans un premier temps, en accord avec ce principe, il avait été décidé d'attribuer une couleur différente pour chaque interprétation. Les portes étaient donc encadrées en orange, les fenêtres en bleu, les fenêtres coulissantes en vert et les meubles en rouge. Dans les exemples traités, seuls ces 4 symboles différents étaient interprétés (les meubles n'étaient pas différenciés les uns des autres). En revanche, cette logique consistant à attribuer une couleur par symbole risquait de perdre de son intérêt lorsque le nombre de symboles différents à traiter allait augmenter. Plus ce nombre allait augmenter et plus il allait devenir difficile de discriminer les couleurs correspondantes. Afin de résoudre ce problème tout en restant cohérent avec le principe de groupement et de distinction entre les items (Bastien & Scapin, 1993), deux couleurs ont été utilisées (le bleu et le vert) pour renvoyer à deux catégories plus générales de symboles : les ouvrants (portes et fenêtres) et les meubles (voir figure 41). De plus, le simple encadrement des symboles a été remplacé par un surlignage permettant de mieux distinguer la couleur de chaque élément.

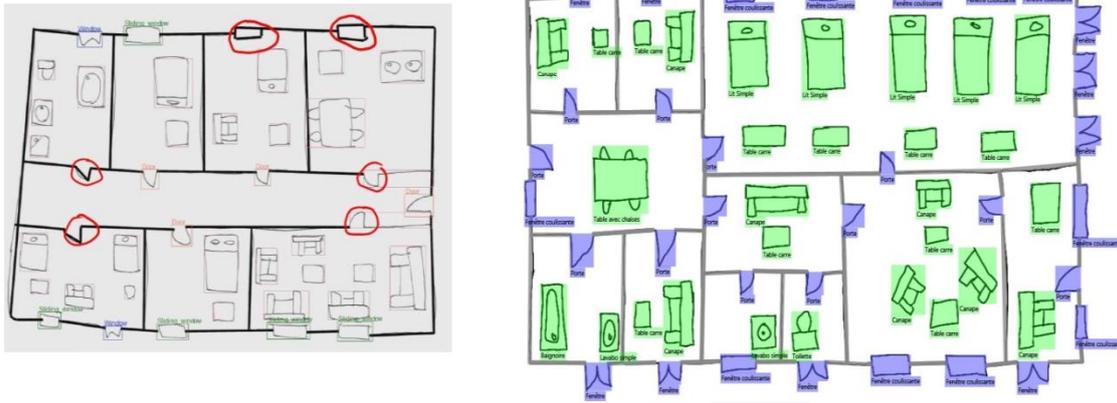


Figure 41. Capture d'écran d'un plan interprété utilisé dans l'étude 4 (à gauche) et d'un plan interprété après la prise en compte des préconisations (à droite)

L'autre commentaire récurrent des participants concernait l'ordre d'apparition des interprétations de symboles considéré comme spatialement incohérent. Il arrivait en effet que le système analyse successivement plusieurs symboles très proches les uns des autres et qu'il se mette soudainement à interpréter des symboles très éloignés sur le plan. Agir sur l'ordre de l'analyse serait techniquement très coûteux. La solution proposée a donc été d'envoyer en permanence à l'utilisateur des feedbacks sur les traitements effectués par le système. Ce choix s'inscrit dans une démarche d'amélioration de la conscience de la situation des utilisateurs (Endsley, 1995). Ces feedbacks continus se présentaient sous la forme d'un halo de couleur se déplaçant à l'écran et surlignant le symbole que le système était en train d'analyser (voir figure 42).

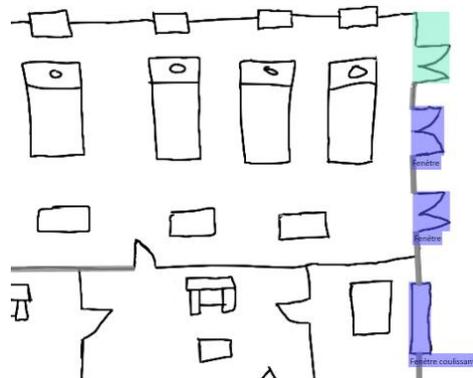


Figure 42. Exemple de focus signalant en temps réel à l'utilisateur la zone du plan traitée par le système (en bleu clair)

Pour faciliter la vérification des interprétations, il faut que les choix graphiques soient porteurs d'informations. Le meilleur niveau de catégorisation des symboles en fonction des couleurs dépend certainement beaucoup du type de document qui est traité, puisque celui-ci implique une variété de symboles plus ou moins grande. Chaque logiciel devrait donc nécessiter une réflexion spécifique en fonction du matériel qu'il sera amené à traiter. Quoi qu'il en soit, les heuristiques de Bastien et Scapin (1993), mais aussi celles de Nielsen (1994), semblent être des outils pertinents pour cette réflexion. Enfin, comme l'a préconisé Endsley (1995), il est important d'apporter à l'utilisateur un maximum de feedbacks sur l'état du système afin d'améliorer sa conscience de la situation et ainsi éviter les ralentissements de l'activité. Lorsque l'interprétation apparaît progressivement comme nous le recommandons, ces feedbacks peuvent prendre la forme d'un focus de couleur signalant à l'utilisateur le symbole sur lequel le système est en train de travailler.

## **2.5. Le système doit-il communiquer ses incertitudes aux utilisateurs ?**

Le fait de faire communiquer le système sur les risques d'erreurs devrait générer une amélioration de la conscience de la situation, se traduisant par une connaissance des difficultés d'interprétation rencontrées par le système. L'étude 5 avait pour objectif de tester l'impact des sollicitations de l'utilisateur par le système pour lui suggérer des risques d'erreurs. La tâche ne consistait plus seulement à repérer les erreurs puisque les participants devaient également réaliser la correction (voir figure 43).

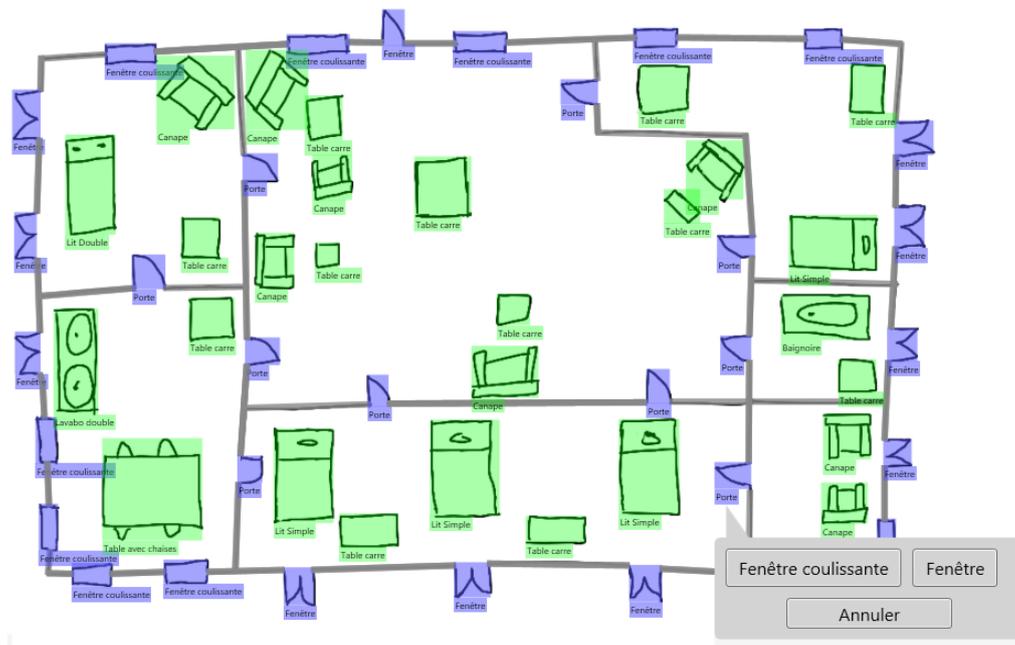


Figure 43. Capture d'écran d'un plan interprété avec affichage de propositions de corrections

Les erreurs qui font l'objet d'une sollicitation ont plus de chance d'être corrigées que des erreurs non signalées. Cependant, les sollicitations sont tout de même à utiliser avec parcimonie car elles ont tendance à impacter le niveau de confiance que les utilisateurs accordent au système. En signalant correctement des erreurs, le système peut ainsi conduire l'utilisateur à relâcher son attention allouée à la vérification des interprétations ne faisant pas l'objet de sollicitation. L'ajout des sollicitations peut donc à la fois améliorer les performances de correction des erreurs signalées tout en dégradant la correction des erreurs non signalées.

Ce phénomène de sur-confiance étant à l'origine d'un risque de diminution des performances, une solution pourrait être de communiquer aux utilisateurs les taux réels d'erreurs n'ayant pas fait l'objet de sollicitation dans les interactions précédentes. Ainsi, le niveau de confiance devrait se maintenir à un niveau plus adéquat, ce qui permettrait de bénéficier des effets positifs des sollicitations sans en subir pour autant les conséquences délétères.

### 3. Apports théoriques

Les apports théoriques de cette thèse sont pour certains généraux, liés au syndrome de l'humain hors de la boucle et pour d'autres, davantage spécifiques, renvoyant à des champs théoriques plus variés du fait de la diversité des modèles sur lesquels se sont appuyés les expériences. Ainsi, l'étude 2 a permis de montrer que le principe de contiguïté spatiale (Mayer, 2001) pouvait être étendu à la recherche d'erreurs d'interprétation du fait de la nécessaire coréférenciation des informations que cette tâche implique. Le rôle de la séquentialité pour guider l'attention des utilisateurs afin de rendre une vérification plus exhaustive a aussi été mis en évidence. L'étude 6 a quant à elle permis de montrer que la pertinence du modèle des ressources était sensiblement diminuée lorsque le fonctionnement du système homme-ordinateur qu'il décrit est impacté par le niveau de confiance que l'utilisateur accorde à la machine.

Le cadre théorique général de cette thèse était le syndrome de l'humain hors de la boucle décrit par Endsley (1995). En effet, dans le premier prototype utilisé, le plan était interprété de manière automatique et l'utilisateur ne pouvait rechercher les erreurs que dans un second temps, en comparant le plan avec son interprétation. Afin d'améliorer la conscience de la situation des utilisateurs, Endsley (1996) préconisait d'apporter aux utilisateurs le plus de feedbacks possible pendant les traitements. C'est aussi ce que conseillait Norman (1989) pour qui l'amélioration d'un système homme-ordinateur passe par une augmentation de la quantité et de la qualité des communications de l'ordinateur vers l'utilisateur. Cette recommandation a été appliquée de trois manières différentes et complémentaires : en affichant de manière séquentielle l'interprétation en temps réel, en signalant en permanence à l'utilisateur la zone du plan que le logiciel est en train d'analyser par un focus visuel et en montrant à l'utilisateur les parties du plan pour lesquelles l'interprétation est incertaine. De plus, selon Kaber et Endsley (1997), il est nécessaire d'ajuster le niveau d'automatisation du système, non pas en fonction de ce qui est techniquement réalisable, mais selon le niveau le plus adéquat pour optimiser les performances générales du système. Plus spécifiquement, une diminution du niveau d'automatisation tendrait à augmenter la

conscience de la situation des utilisateurs (Wickens, 2002). En passant d'une automatisation totale avec les premiers prototypes à une prise de décision mixte dans l'étude 6, le niveau d'automatisation du système a été diminué pour mieux correspondre aux caractéristiques des utilisateurs. L'ensemble de ces améliorations (augmentation des feedbacks, ajustement du niveau d'automatisation) a eu pour conséquence une nette amélioration des performances du système. En effet, seuls 32,5% des participants de l'étude 1 étaient parvenus à repérer l'ensemble des erreurs (sans avoir à effectuer de correction) sur un plan très simple alors que 67,73% de ceux de l'étude 5 ont corrigé l'ensemble des interprétations erronées sur des plans bien plus complexes.

## **4. Limites et perspectives**

La principale limite de ce travail concerne la population testée. Les 5 premières études ont porté sur une population novice dans le traitement des plans d'architecture. Les expériences réalisées portaient sur des aspects cognitifs supposés non spécifiques aux professionnels de l'architecture (effet de la contiguïté spatiale, mémoire à court terme). Ainsi, elles pouvaient être menées sur une population moins ciblée, ce qui rendait possible la constitution de groupes aux effectifs plus importants. L'étude 5, bien que portant aussi sur des novices, a constitué une première prise en compte de l'expertise en mettant en évidence les effets positifs de l'entraînement sur la précision et sur la durée de la tâche. Ensuite, l'étude 6 a porté sur une population entièrement constituée de futurs utilisateurs potentiels du système.

La seconde limite de cette thèse est son cadre théorique relativement hétérogène. En effet, les questionnements traités dans cette thèse ont été guidés en premier lieu par les applications. Les expérimentations ont permis d'apporter des réponses ne se trouvant pas dans la littérature. Cette démarche a pour conséquence une structure de manuscrit dans laquelle les aspects théoriques sont présentés de manières graduelles, au fur et à mesure, pour refléter leur mobilisation progressive au cours du projet.

Parmi les perspectives de recherche envisageables, un grand nombre de travaux pourraient être réalisés sur les sollicitations de l'utilisateur par le système. Trois principaux axes de recherches pourraient être envisagés. Le premier concerne les impacts des différents degrés de précision des sollicitations sur la détection des erreurs. Nous avons fait le choix d'un nombre de fausses alarmes identique au nombre d'erreurs ne faisant pas l'objet de sollicitation. Cependant, si on augmentait le seuil de vraisemblance à partir duquel une sollicitation est déclenchée, cela tendrait à diminuer le nombre de fausses alarmes et à augmenter le nombre d'erreurs non signalées. Une diminution de ce seuil aurait l'effet inverse. Un nombre trop élevé de fausses alarmes risquerait de rendre les sollicitations inutiles, voire gênantes, alors qu'un trop grand nombre d'erreurs non signalées pourrait conduire à un plus faible taux de correction. Le second axe consisterait à explorer les effets des sollicitations, selon leur degré de précision, sur la confiance dans le système, mais aussi l'impact de cette confiance sur les comportements d'interaction. Les solutions éventuelles qui permettraient d'améliorer l'adéquation de la confiance avec la fiabilité réelle du système nécessiteraient d'être évaluées. Enfin, les sollicitations utilisées dans nos expériences ne permettaient pas de rendre compte du caractère continu de l'incertitude. Le choix de signaler ou non un risque d'erreur fournit une information binaire. Améliorer la précision de ces feedbacks irait dans le sens d'une amélioration de la conscience de la situation. Il conviendrait alors d'évaluer des solutions qui consisteraient à afficher un degré de certitude du système, en utilisant par exemple une échelle de dégradé de couleurs pour chaque interprétation.

Les résultats présentés dans cette thèse ne s'appliquent pas uniquement à la reconnaissance de plans d'architectures. Concernant les systèmes d'analyse automatique de textes, les seules spécifications applicables sont celles qui renvoient à la fonctionnalité de sollicitation de l'utilisateur. La recherche d'erreurs d'analyse d'un texte passe elle aussi par la comparaison du texte initial avec son interprétation, mais il n'est pas envisageable de superposer ces deux sources d'informations pour les comparer. De plus, l'ordre de l'exploration visuelle d'un texte est suffisamment évident pour n'importe quel lecteur pour éviter les risques de ne pas contrôler

l'ensemble du document. En revanche, signaler à l'utilisateur les zones du texte caractérisées par un risque d'erreur élevé pourrait être possible. Dans ce cas, le risque de sur-confiance qui conduirait à ne pas suffisamment contrôler les zones ne faisant pas l'objet de sollicitation peut être présent. Les résultats des expériences présentées s'appliquent tous à l'ensemble des systèmes d'interprétation automatique de documents techniques structurés (partition musicale, carte routière, plan électrique, etc). En effet, la rétro-conversion de ces documents comporte deux principales constantes qui rendent applicables les spécifications proposées : elle est soumise à un risque d'erreurs et la recherche de ces erreurs par un utilisateur consiste à comparer la reconnaissance avec les symboles interprétés.

Pour terminer, soulignons la complémentarité de ce travail de thèse avec celui réalisé en parallèle au sein de l'IRISA (Ghorbel, 2012). Durant le projet, les résultats des travaux d'ergonomes orientaient la conception et en retour, les avancées techniques des partenaires chercheurs en informatique généraient de nouveaux questionnements en termes d'usages. Le caractère pluridisciplinaire du projet s'est notamment concrétisé à travers des communications scientifiques, communes aux deux équipes de recherche, dans lesquelles chacune des deux approches s'est trouvée enrichie par l'interaction avec la seconde. Pour traiter de thématiques de plus en plus complexes, ce type de collaboration interdisciplinaire devrait être amené à se développer à l'avenir.

# Bibliographie

- Abe, G., Itoh, M., & Tanaka, K. (2002). Dynamics of drivers' trust in warning systems. *17th IFAC World Congress, 5*, p. 2002. Barcelona, Spain.
- Abrams, R. A., & Christ, S. (2003). Motion onset captures attention. *Psychological Science 14*(5), 427-432.
- Ahmed, S., Liwicki, M., Weber, M., & Dengel, A. (2011). Improved Automatic Analysis of Architectural Floor Plans. *2011 International Conference on Document Analysis and Recognition*, (pp. 864-869).
- Ahmed, S., Weber, M., Langenhan, C., Dengel, A., & Petzold, F. (2013). Automatic Analysis and Sketch-Based Retrieval of Architectural Floor Plans. *Pattern Recognition Letters*.
- Ah-Soon, C., & Tombre, K. (1997). Variations on the Analysis of Architectural Drawings. *Proceedings of 4th International Conference on Document Analysis and Recognition*, (pp. 347-351). Ulm (Germany).
- Ah-Soon, C., & Tombre, K. (1998). Network-Based Recognition of Architectural Symbols. Dans A. Amin, D. Dori, P. Pudil, & H. Freeman (Éd.), *Proceedings of Joint IAPR Workshops SSPR'98 and SPR'98, 1451*, (pp. 252-261). Sydney, Australia.
- Ah-Soon, C., & Tombre, K. (2001). Architectural symbol recognition using a network of constraints. *Pattern Recognition Letters, 20*(2), 231-248.
- Alberdi, E., Povyakalo, A. A., Strigini, L., Ayton, P., & Given-Wilson, R. (2008). CAD in mammography: lesion-level versus case-level analysis of the effects of prompts on human decisions. *International Journal of Computer Assisted Radiology and Surgery, 3*(1-2), 115-122.
- Ayres, P., & Sweller, J. (2005). The Split-Attention Principle in Multimedia Learning. Dans R. E. Mayer, *The Cambridge Handbook of Multimedia Learning* (pp. 135-146). New York: Cambridge University Press.
- Baddeley, A. D. (1986). *Working memory*. Oxford: University Press.
- Baddeley, A. D. (1992, January 31). Working Memory. *Science, 255*(5044), 556-559.
- Baddeley, A. D. (2002). Is Working Memory Still Working ? *European Psychologist, 7*(2), 85-97.
- Bafna, S. (2008). How architectural drawings work - and what that implies for the role of representation in architecture. *The Journal of Architecture, 13*(5), 535-564.

- Bainbridge, D., & Bell, T. (2001). The Challenge of Optical Music Recognition. *Computers and the Humanities*, 35(2), 95-121.
- Bainbridge, L. (1983). Ironies of Automation. *Automatica*, 19(6), 775-779.
- Ballinger, G. A. (2004). Using Generalized Estimating Equations for longitudinal Data Analysis. *Organizational Research Methods*, 7(2), 127-150.
- Bastien, J. M., & Scapin, D. L. (1993). *Critères Ergonomiques pour l'évaluation d'interfaces Utilisateurs*. Technical report, INRIA.
- Bastien, J. M., & Scapin, D. L. (1993). *Ergonomics Criteria for the Evaluation of Human-Computer Interfaces*. Technical report, Institut National de Recherche en Informatique et en Automatique, Rocquencourt.
- Bauhoff, V., Huff, M., & Schawn, S. (2012). Distance Matters: Spatial Contiguity Effects as Trade-Off between Gazes Switches and Memory Load. *Applied Cognitive Psychology*, 26, 863-871.
- Berch, D. B., Krikorian, R., & Huba, E. M. (1998). The Corsi Block-Tapping Task: Methodological and Theoretical Considerations. *Brain and Cognition*, 38(3), 317-338.
- Bétrancourt, M., & Caro Dambreville, S. C. (1998). Intégrer des informations en escamots dans les textes techniques: quels effets sur les processus cognitifs? *Les hypermédias: approches cognitives et ergonomiques*, 157-173.
- Bétrancourt, M., & Tversky, B. (2000). Simple animations for organizing diagrams. *International Journal of Human Computer Studies*.
- Bétrancourt, M., Bisseret, A., & Faure, A. (2001). Sequential display of pictures and its effect on mental representations. Dans J.-F. Rouet, J. J. Levonen, & A. Biarreau (Éds.), *Multimedia learning: Cognitive and instructional issues* (pp. 112-118). The Netherland: Elsevier Science.
- Bilda, Z., & Gero, J. S. (2006). Reasoning with Internal and External Representations: A Case Study with Expert Architects. *Proceedings of CogSci/ICSS*, 6, (pp. 1020-1026).
- Bilda, Z., Gero, J. S., & Purcell, T. (2006). To sketch or not to sketch? That is the question. *Design Studies*, 27(5), 587-613.
- Bindemann, M. (2010). Scene and screen center biases early eye movements in scene viewing. *Vision Research*, 50, 2577-2587.
- Blandford, A., & Furniss, D. (2006). DiCoT: a methodology for applying Distributed Cognition to the design of team working systems. *Proc. DSVIS 2005*.
- Bobis, J., Sweller, J., & Cooper, M. (1993). Cognitive load effects in a primary-school geometry task. *Learning and Instruction*, 3(1), 1-21.

- Bodemer, D., Ploetzner, R., Feuerlein, I., & Spada, H. (2004). The active integration of information during learning with dynamic and interactive visualisations. *Learning and Instruction, 14*(3), 325-341.
- Borgen, F. H., & Barnett, D. C. (1987). Applying Cluster Analysis in Counseling Psychology Research. *Journal of Counseling Psychology, 34*(4), 456-468.
- Bruyer, R., & Scailquin, J.-C. (1998). The visuospatial sketchpad for mental images: Testing the multicomponent model of working memory. *Acta Psychologica, 98*(1), 17-36.
- Budd, M. A., Houtz, A., & Lambert, P. (2008). Comparison of nondominant- and dominant -hand performances on the Copy portion of the Rey Complex Figure Test (RCFT). *Journal of clinical and experimental neuropsychology, 30*(3), 380-386.
- Butcher, K. R. (2006). Learning From Text With Diagrams: Promoting Mental Model Development and Inference Generation. *Journal of Educational Psychology, 98*(1), 182-197.
- Cahill, H. E., & Hovland, C. I. (1960, March). The role of memory in the acquisition of concepts. *Journal of Experimental Psychology, 59*(3), 137-144.
- Card, S. K., Moran, T. P., & Newell, A. (1986). The Model Human Processor: An engineering model of human performance. Dans K. Boff (Éd.), *Handbook of perception and human performance* (pp. 1-35).
- Chandler, P., & Sweller, J. (1991). Cognitive Load Theory and the Format of Instruction. *Cognition and Instruction, 8*(4), 293-332.
- Chandler, P., & Sweller, J. (1992). The split-attention effect as a factor in the design of instruction. *British Journal of Educational Psychology, 62*(2), 233-246.
- Chandler, P., & Sweller, J. (1996). Cognitive Load While Learning to Use a Computer Program. *Applied Cognitive Psychology, 10*(2), 151-170.
- Chanquoy, L., & Alamargot, D. (2002). Mémoire de travail et rédaction de textes: évolution des modèles et bilan des premiers travaux. *L'Année psychologique, 102*(2), 363-398.
- Cheney, T. (1989, October 23). Robots milling around the office halls, pass man in glass booth with sign that reads: In Case of Emergency Break Glass. *The New Yorker*.
- Cheng, K.-M. (2013). An Evaluation Of RFID Door Security System At Taipei Arena Ice Land Based On Technology Acceptance Model. *International Journal of Management & Information Systems, 17*(2), 117-130.

- Chiang, Y. Y., Leyk, S., & Knoblock, C. A. (2013). Efficient and robust graphics recognition from historical maps. *Graphics Recognition. New Trends and Challenges* (pp. 25-35). Springer Berlin Heidelberg.
- Chiang, Y. Y., & Knoblock, C. A. (2011). Recognition of Multi-Oriented, Multi-Sized, and Curved Text. *2011 International Conference on Document Analysis and Recognition*, (pp. 1399-1403).
- Cierniak, G., Scheiter, K., & Gerjets, P. (2009). Explaining the split-attention effect: Is the reduction of extraneous cognitive load accompanied by an increase in germane cognitive load? *Computer in Human Behavior*, *25*(2), 315-324.
- Cooper, G. (1998). *Research into cognitive load theory and instructional design at UNSW*. Consulté le Août 2011, sur <http://dwb4.unl.edu/Diss/Cooper/UNSW.htm>
- Cowan, N. (2000). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*(1), 87-185.
- Craig, S. D., Gholson, B., & Driscoll, D. M. (2002). Animated Pedagogical Agents in Multimedia Educational Environments: Effects of Agent Properties, Picture Features, and Redundancy. *Journal of Educational Psychology*, *94*(2), 428-434.
- Darling, S., Della Sala, S., Logie, R. H., & Cantagallo, A. (2006). Neuropsychological evidence for separating components of visuo-spatial working memory. *Journal of Neurology*, *253*(2), 176-180.
- Davis, F. D. (1989). Perceived Usefulness, perceived ease of use, and user acceptance of information technology. *MIS quarterly*, *9*(3), 319-340.
- Davis, F. D. (1993). User acceptance of information technology: system characteristics, user perceptions and behavioral impacts. *Int. J. Man-Machine Studies*, *38*, 475-487.
- Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1989, August). User Acceptance of Computer Technology: A Comparison of Two Theoretical Models. *Management Science*, *35*(8), 982-1003.
- de Koning, B. B., Tabbers, H. K., Rikers, R. M., & Paas, F. (2009). Towards a Framework for Attention Cueing in Instructional Animations: Guidelines for Research and Design. *Educ Psychol Rev*, *21*(2), 113-140.
- Della Sala, S., Gray, C., Baddeley, A., Allamano, N., & Wilson, L. (1999). Pattern span: a tool for unwelding visuo-spatial memory. *Neuropsychologia*, *37*(10), 1189-1199.
- Dillon, A., & Morris, M. G. (1996). User acceptance of information technology: Theories and models. *Annual Review of Information Science and Technology*, *31*, 3-32.

- Do, E. L. (2002). Drawing marks, acts, and reacts: Toward a computational sketching interface for architectural design. *Artificial intelligence for Engineering Design, Analysis and Manufacturing*, 16(3), 149-171.
- Do, E. L., & Gross, M. D. (1996). Drawing as a Means to Design Reasoning. *AI and Design*.
- Do, E. L., & Gross, M. D. (2001). Thinking with diagrams in architectural design. *Artificial Intelligence Review*, 15(1), 135-149.
- Doherty, G., Campos, J., & Harrison, M. D. (2008). Resources for situated actions. Dans N. Graham, & P. Palanque (Éds.), *Interactive systems: Design, Specification and Verification* (Vol. 5136, pp. 194-207). Verlag: Springer.
- Dorr, M., Martinez, T., Gegenfurtner, K. R., & Barth, E. (2010). Variability of eye movements when viewing dynamic natural scenes. *Journal of vision*, 10(10), 1-17.
- Dunlap, R. D. (1997). *An assessment of visuo-spatial and phonological components of working memory*. Doctoral thesis, Texas Tech University.
- Ebbinghaus, H. (1885). *Memory: A Contribution to Experimental Psychology*. (H. A. Ruger, & C. E. Bussenius, Trads.) New York: Teachers College, Columbia University.
- Elliott, L. A., & Strawhorn, R. J. (1976). Interference in short-term memory for vocalization: Aural versus visual modality differences. *Journal of Experimental Psychology: Human Learning and Memory*, 2(6), 705-711.
- Elsen, C., Demaret, J. N., Yang, M. C., & Leclercq, P. (2012). Sketch-based interfaces for modeling and users' needs: Redefining connections. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, 26(3), 281-301.
- Endsley, M. R. (1995). Measurement of situation awareness in dynamic systems. *Human Factor*, 37(1), 65-84.
- Endsley, M. R. (1996). Automation and Situation Awareness. Dans R. Parasuraman, & M. Mouloua (Éds.), *Automation and human performance: Theory and applications* (pp. 163-181). Mahwah, NJ: Lawrence Erlbaum.
- Endsley, M. R. (1997). Situation Awareness, Automation & Free Flight. *FAA/Eurocontrol Air Traffic Management R&D Seminar*. Saclay, France.
- Endsley, M. R., & Kaber, D. B. (1999). Level of automation effects on performance, situation awareness and workload in a dynamic control task. *Ergonomics*, 42(3), 462-492.
- Erhel, S., & Jamet, E. (2006). Using pop-up windows to improve multimedia learning. *Journal of Computer Assisted Learning*, 22(2), 137-147.

- Erhel, S., & Jamet, E. (2011, June). How Can Positive Effects of Pop-up windows on Multimedia Learning be Explained? *Journal of Educational Multimedia and Hypermedia*, 20(2), 135-156.
- Fagerland, M. W., & Sandvik, L. (2009, February). The Wilcoxon-Mann-Whitney test under scrutiny. *Statistics in Medicine*, 28, 1487-1497.
- Ferber, J. (1999). *Multi-agent systems: an introduction to distributed artificial intelligence*. Addison-Wesley.
- Février, F., Gauducheau, N., Jamet, E., Rouxel, G., & Salembier, P. (2011). La prise en compte des affects dans le domaine des interactions homme-machine: quels modèles, quelles méthodes, quels bénéfices?. *Le travail humain*, 74(2), 183-201.
- Fields, B., Wright, P. C., & Harrison, M. D. (1997). Objectives, strategies and resources as design drivers. *Human Computer Interaction: INTERACT'97*, (pp. 14-18).
- Fields, B., Wright, P. C., Marti, P., & Palmonari, M. (1998). Air Traffic Control as a Distributed Cognitive System: a study of external representations. *ECCE9: Proceedings of the Ninth European Conference on Cognitive Ergonomics*, (pp. 85-90).
- Fitts, P. M. (1951). *Human engineering for an effective air-navigation and traffic-control system*. Oxford, England.
- Fleury, S., Ghorbel, A., Lemaitre, A., Anquetil, E., & Jamet, E. (2013). User-centred design of an interactive off-line handwritten architectural floor plan recognition. *International Conference on Document Analysis and Recognition (ICDAR'2013)*. Washington.
- Florax, M., & Ploetzner, R. (2010). What contributes to the split-attention effect? The role of text segmentation, picture labelling, and spatial proximity. *Learning and Instruction*, 20(3), 216-224.
- Flower, L., & Hayes, J. R. (1981). A Cognitive Process Theory of Writing. *College Composition and Communication*, 32(4), 365-387.
- Fogg, B. J., & Tseng, H. (1999). The elements of computer credibility. *CHI99*, (pp. 80-87).
- Franconeri, S. L., Simons, D. J., & Junge, J. A. (2004). Searching for stimulus-driven shifts of attention. *Psychonomics Bulletin & review*, 11(5), 876-881.
- Frenkel, S., & Bourdin, B. (2009, February). Verbal, visual, and spatio-sequential short-term memory: assessment of the storage capacities of children and teenagers with Down's syndrome. *Journal of Intellectual Disability Research*, 53(2), 152-160.
- Frick, R. W., & De Rose, A. (1986). The Suffix Effect and Preattentive Unit-formation in Visual Short-term Memory. *Canadian Journal of Psychology*, 40(2), 97-108.

- Ghisletta, P., & Spini, D. (2004). An Introduction to Generalized Estimating Equations and Application to Assess Selectivity Effects in a Longitudinal Study on Very Old Individuals. *Journal of Educational and Behavioral Statistics*, 29(4), 421-437.
- Ghorbel, A., Almaksour, A., Lemaitre, A., & Anquetil, E. (2011). Incremental learning for interactive sketch recognition. *Ninth IAPR International Workshop on Graphics RECognition*.
- Ghorbel, A. (2012). *Interprétation interactive de documents structurés: application à la rétroconversion de plans d'architecture manuscrits*. (Thèse de doctorat, Institut National des Sciences Appliquées, Rennes). Repéré à "[http://tel.archives-ouvertes.fr/docs/00/78/88/32/PDF/thA\\_se\\_Achraf\\_Ghorbel.pdf](http://tel.archives-ouvertes.fr/docs/00/78/88/32/PDF/thA_se_Achraf_Ghorbel.pdf)"
- Ghorbel, A., Anquetil, E., & Lemaitre, A. (2012). Optimization Analysis Base On A Breadth-First Exploration For A Structural Approach Of Sketches Interpretation. *10th IAPR International Workshop on Document Analysis Systems (DAS'12)*, (pp. 240-244). Gold Coast, Queensland: Australia.
- Ghorbel, A., Lemaitre, A., & Anquetil, E. (2012). Competitive hybrid exploration for off-line sketches structure recognition. *International Conference on Frontiers of Handwriting Recognition*. Bari, Italy.
- Ghorbel, A., Lemaitre, A., & Anquetil, E. (2012). Competitive hybrid exploration for off-line sketches structure recognition. *International Conference on Frontiers in Handwriting Recognition (ICFHR)*, (pp. 771-576). Bari, Italy.
- Ghorbel, A., Macé, S., Lemaitre, A., & Anquetil, E. (2011). Interactive competitive breadth-first exploration for sketch interpretation. *International Conference on Document Analysis and Recognition (ICDAR)*, (pp. 1195-1199).
- Ginns, P. (2006, December). Integrating information: A meta-analysis of the spatial contiguity and temporal contiguity effects. *Learning and Instruction*, 16(6), 511-525.
- Gobert, J. D. (1999). Expertise in the comprehension of architectural plans: knowledge acquisition and inference-making. Dans J. S. Gero, & B. Tversky (Éd.), *Visual And Spatial Reasoning In Design '99*. University of Sydney.
- Godjin, R., & Theeuwes, J. (2002, October 12). Oculomotor capture and Inhibition of Return: Evidence for an oculomotor suppression account of IOR. *Psychological Research*, 66(4), 234-246.
- Goh, J., Wiegmann, D. A., & Madhavan, P. (2005). Effects of Automation Failure in a Luggage Screening Task: A Comparison Between Direct and Indirect Cueing. *Proceedings of the Human Factors and Ergonomics Society 49th Annual Meeting*, (pp. 492-496).
- Ha, S., & Stoel, L. (2009). Consumer e-shopping acceptance: antecedents in a technology acceptance model. *Journal of Business Research*, 62(5), 565-571.

- Harrison, M. D., Massink, M., & Latella, D. (2009). Engineering Crowd Interaction within Smart Environments. *EICS'09*. Pittsburgh, Pennsylvania, USA.
- Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Human mental workload*, 1(3), 139-183.
- Hayes, J. R. (1952). Memory span for several vocabularies as a function of vocabulary size. *Quarterly Progress Report*, 338-352.
- Hayes, J. R., & Flower, L. S. (1980). Identifying the organization of writing processes. Dans L. W. Gregg, & E. R. Steinberg (Eds.), *Cognitive processes in writing*. Hillsdale, New Jersey: Lawrence Erlbaum.
- Hecker, R., & Mapperson, B. (1997). Dissociation of visual and spatial processing in working memory. *Neuropsychologia*, 35(5), 599-603.
- Henderson, R., & Divett, M. J. (2003). Perceived usefulness, ease of use and electronic supermarket use. *International Journal of Human-Computer Studies*, 59(3), 383-395.
- Hoffman, R. R., Johnson, M., Bradshaw, J. M., & Underbrink, A. (2013). Trust in Automation. *Intelligent Systems, IEEE*, 28(1), 84-88.
- Holsanova, J., Holmberg, N., & Holmqvist, K. (2009, December 23). Reading Information Graphics: The Role of Spatial Contiguity and Dual Attentional Guidance. *Applied Cognitive Psychology*, 29(3), 1215-1226.
- Hovland, C. I., & Weiss, W. (1953). Transmission of Information Concerning Concepts Through Positive and Negative Instances. *Journal of Experimental Psychology*, 45(3), 175-182.
- Igarashi, T., Matsuoka, S., Kawachiya, S., & Tanaka, H. (1997). Interactive beautification: A technique for rapid geometric design. *Proceedings of the 10th annual ACM symposium on User interface software and technology*, (pp. 105-114).
- ISO 9241-210. (2010). *Ergonomics of human-system interaction - Part 210: Human-centred design for interactive systems*. ISO.
- Jamet, E. (2006). Apprentissage à partir de documents électroniques illustrés: le rôle des présentations séquentielles. Dans A. Piolat, *Psychologie, Théories, Méthodes, Pratiques. Lire, Ecrire, Communiquer et Apprendre avec Internet* (pp. 557-572). Marseille: Solal.
- Jamet, E., & Arguel, A. (2008). La compréhension d'un document technique multimédia peut-elle être améliorée par une présentation séquentielle de son contenu ? *Le travail humain*, 71(3), 253-270.

- Jamet, E., Gavota, M., & Quaireau, C. (2008). Attention guiding in multimedia learning. *Learning and Instruction, 18*(2), 135-145.
- Jian, J.-Y., Bisantz, A. M., & Drury, C. G. (2000). Foundations for an empirically determined scale of trust in automated systems. *International Journal of Cognitive Ergonomics, 4*(1), 53-71.
- Johnson, C. I., & Mayer, R. E. (2012). An Eye Movement Analysis of the Spatial Contiguity Effect in Multimedia Learning. *Journal of Experimental Psychology: Applied, 18*(2), 178-191.
- Jones, D., Farrand, P., Stuart, G., & Morris, N. (1995). Functional Equivalence of Verbal and Spatial Information in Serial Short-Term Memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(4), 1008-1018.
- Just, M. A., & Carpenter, P. A. (1992). A Capacity Theory of Comprehension: Individual Differences in Working Memory. *Psychological Review, 99*(1), 122-149.
- Kaber, D. B., & Endsley, M. R. (1997). Out-of-the-Loop Performance Problems and the Use of Intermediate Levels of Automation for Improved Control System Functioning and Safety. *Process Safety Progress, 16*(3), 126-131.
- Kaber, D. B., & Endsley, M. R. (2003). The effects to level of automation and adaptative automation on human performance, situation awareness and workload in a dynamic control task. *Theoretical Issues in Ergonomics Science, 5*(2), 113-153.
- Kalyuga, S., Chandler, P., & Sweller, J. (1999). Managing Split-attention and Redundancy in Multimedia Instruction. *Applied Cognitive Psychology, 13*(4), 351-371.
- Kellogg, R. (1996). A model of working memory in writing. Dans M. Levy, & S. Ransdell (Éds.), *The science of writing: Theories, methods, individual differences, and applications* (pp. 57-72). Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Kester, L., Kirschner, P. A., & Van Merriënboer, J. G. (2005). The management of cognitive load during complex cognitive skill acquisition by means of computer-simulated problem solving. *British Journal of Educational Psychology, 75*(1), 71-85.
- Khasawneh, M. T., Bowling, S. R., Jiang, X., Gramopadhye, A. K., & Melloy, B. J. (2003). A model for predicting human trust in automated systems. *Proceedings of the 8th Annual International Conference on Industrial Engineering - Theory, Applications and Practice*, (pp. 216-222). Las Vegas.
- Kim, S., & Astion, M. (2000). Patterns of image comparison using compare and contrast feature in Urinalysis Tutor. *British Journal of Educational Technology, 31*(4), 349-357.

- King, W. R., & He, J. (2006). A meta-analysis of the technology acceptance model. *Information & Management*, 43(6), 740-755.
- Klauer, K. C., & Stegmaier, R. (1997). Interference in Immediate Spatial Memory: Shifts of Spatial Attention or Central-executive Involvement. *The Quarterly Journal of Experimental Psychology*, 50(1), 79-99.
- Koutamanis, A., & Mitossi, V. (1993). Architectural computer vision: Automated recognition of architectural drawings. *Design and decision support systems in architecture*.
- Kroft, P., & Wickens, C. D. (2001). *The Display of Multiple Geographical Data Bases: Implications of Visual Attention*. Technical Report, University of Illinois, Institute of Aviation.
- Larkin, J. H., & Simon, H. A. (1987). Why a Diagram is (Sometimes) Worth Ten Thousand Words. *Cognitive Science*, 11(1), 65-99.
- Lecourtois, C. (2005). Architecturologie appliquée à une sémiotique de l'esquisse architecturale. *SCAN'O5, Séminaire de conception architecturale numérique*.
- Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors*, 46(1), 50-80.
- Lee, J., & Moray, N. (1992). Trust, control strategies and allocation of function in human-machine systems. *Ergonomics*, 35(10), 1245-1270.
- Lee, Y. S., Lee, H. M., & Tsai, S. H. (2007). Effects of post-cue interval on intentional forgetting. *British Journal of Psychology*, 98(2), 257-272.
- Lin, L., & Atkinson, R. K. (2011). Using animations and visual cueing to support learning of scientific concepts and processes. *Computers & Education*, 56(3), 650-658.
- Lladós, J., Valveny, E., Sanchez, G., & Martí, E. (2002). Symbol recognition: Current Advances and Perspectives. In D. Blostein, & Y.-B. Kwon, *Graphics Recognition. Algorithms and Applications* (Vol. 4, pp. 104-128). Kingston, Canada: Springer Berlin Heidelberg.
- Logie, R. H. (1986). Visuo-Spatial Processing in Working Memory. *The Quarterly Journal of Experimental Psychology*, 38(2), 229-247.
- Logie, R. H. (1995). *Visuo-spatial Working Memory*. Hove, UK.
- Lu, T., Tai, C.-L., Su, F., & Cai, S. (2005). A new recognition model for electronic architectural drawings. *Computer-aided Design*, 37(10), 1053-1069.
- Ludwig, C. J., Ranson, A., & Gilchrist, I. D. (2008). Oculomotor capture by transient events: A comparison of abrupt onsets, offsets, motion, and flicker. *Journal of Vision*, 8(14), 1-16.

- Lyn, F., & Dulaney, R. (2009). A Case for Drawing. *Enquiry/The ARCC Journal of Architectural Research*, 6(1), 23-30.
- Macé, S., Locteau, H., Valveny, E., & Tabbone, S. (2010). A system to detect rooms in architectural floor plan images. *Proceedings of the 9th IAPR International Workshop on Document Analysis Systems*, (pp. 167-174).
- Madhavan, P., & Gonzalez, C. (2006). Effects of sensitivity, criterion shifts, and subjective confidence on the development of automaticity in airline luggage screening. *Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting*, (pp. 334-338).
- Madhavan, P., Wiegmann, D. A., & Lacson, F. C. (2003). Automation failures on tasks easily performed by operators undermines trust in automated aids. *Proceedings of the 47th Annual Meeting of the Human Factors and Ergonomics Society*. Santa Monica.
- Mani, K., & Johnson-Laird, P. N. (1982). The mental representation of spatial descriptions. *Memory & Cognition*, 10(2), 181-187.
- Manolopoulou, Y. (2005). Unformed drawing: notes, sketches, and diagrams. *The Journal of Architecture*, 10(5), 517-525.
- Maycock, B., Liu, G., & Klein, R. M. (2009). Where to Begin? Eye-Movement When Drawing. *Journal of Research Practice*, 5(2).
- Mayer, R. E. (1989). Systematic Thinking Fostered by Illustrations in Scientific Text. *Journal of Educational Psychology*, 81(2), 240-246.
- Mayer, R. E. (2001). *Multimedia Learning*. New York: Cambridge University Press.
- Mayer, R. E., & Chandler, P. (2001, June). When Learning Is Just a Click Away: Does Simple User Interaction Foster Deeper Understanding of Multimedia Messages? *Journal of Educational Psychology*, 93(2), 390-397.
- Mayer, R. E., & Gallini, J. K. (1990). When Is an Illustration Worth Ten Thousand Words? *Journal of Educational Psychology*, 82(4), 715-726.
- Mayer, R. E., Steinhoff, K., Bower, G., & Mars, R. (1995). A generative theory of textbook design: Using annotated illustrations to foster meaningful learning of science text. *Educational Technology Research and Development*, 43(1), 31-44.
- McCarley, J. S., Kramer, A. F., Wickens, C. D., Vidoni, E. D., & Boot, W. R. (2004). Visual Skills in Airport-Security Screening. *Psychological Science*, 15(5), 302-306.
- McCarley, J. S., Wiegmann, D. A., Wickens, C. D., & Kramer, A. F. (2003). Effects of age on utilization and perceived reliability of an automated decision-making aid for luggage screening. *Proceedings of the 47th Annual Meeting of the Human Factors and Ergonomics Society*. Santa Monica.

- McCutchen, D. (1996). A capacity theory of writing: working memory in text composition. *Educational Psychology Review*, 8, 299-325.
- Mehrnejad, M., Toreini, E., & Ghaemi Bafghi, A. (2011). Security Analyzing and Designing GUI with the Resources Model. *8th International ISC Conference on Information Security and Cryptology*, (pp. 14-15).
- Melton, A. W. (1963). Implications of short-term memory for a general theory of memory. *Journal of Verbal Learning and Verbal Behavior*, 2, 1-21.
- Merritt, S. M., & Ilgen, D. R. (2008, April 1). Not All Trust Is Created Equal: Dispositional and History-Based Trust in Human-Automation Interactions. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 50(194), 194-210.
- Miller, C., & Parasuraman, R. (2007). Designing For Flexible Interaction Between Humans and Automation: Delegation Interfaces for Supervisory Control. *Human Factors*, 49(1), 57-75.
- Miller, G. A. (1956). The Magical Number Seven, Plus or Minus Two Some Limits on Our Capacity for Processing Information. *Psychological Review*, 101(2), 343-352.
- Molich, R., & Nielsen, J. (1990). Improving a Human-Computer Dialogue. *Communication of the ACM*. 33, pp. 338-348. Edgar H. Sibley Panel Editor.
- Morel, G., Chauvin, C., Rossi, A., & Berruet, P. (2008). Concevoir des Systèmes Sociotechniques Complexes Résilients et Reconfigurables pour Garantir un Niveau de Sécurité Optimal. *ERGO IA 2008*. Bidart/Biarritz, France.
- Moreno, R., & Mayer, R. E. (1999). Cognitive Principles of Multimedia Learning: The Role of Modality and Contiguity. *Journal of Educational Psychology*, 91(2), 358-368.
- Muniandy, B., Ong, M. Y., Phua, K. K., & Ong, S. L. (2011a, September). Assessing Key Performance Indicators Monitoring System (KPI-MS) of a university using Technology Acceptance Model. *International Journal of Social Science and Humanity*, 1(3), 171-176.
- Muniandy, B., Ong, M. Y., Phua, K. K., & Ong, S. L. (2011b). User Acceptance of a Key Performance Indicator Monitoring System (KPI-MS) in Higher Education: An Application of the Technology Acceptance Model. *2nd International Conference on Education and Management Technology*. 13, pp. 179-184. Singapore: IACSIT Press.
- Murdock, B. B. (1962). The Serial Position Effect of Free Recall. *Journal of Experimental Psychology*, 64(5), 482-488.

- Mwangi, W., & Sweller, J. (1998). Learning to Solve Compare Word Problems: The Effect of Example Format and Generating Self-Explanations. *Cognition and Instruction*, 16(2), 173-199.
- Nairne, J. S. (2002). Remembering Over the Short-Term: The Case Against the Standard Model. *Annual Review of Psychology*, 53(1), 53-81.
- Nickles, G. M., Melloy, B. J., & Gramopadhye, A. K. (2003). A comparison of three levels of training designed to promote systematic search behavior in visual inspection. *International Journal of Industrial Ergonomics*, 32, 331-339.
- Nielsen, J. (1994). Enhancing the Explanatory Power of Usability Heuristics. *Proceedings of the SIGCHI conference on Human factors in computing systems: celebrating interdependence*, (pp. 152-158). Boston.
- Norman, D. A. (1989). *The problem of automation: Inappropriate feedback and interaction not overautomation*. La Jolla: San Diego, CA: Institute for Cognitive Science, U.C.
- Ogier, J.-M., Mullot, R., Labiche, J., & Lecourtier, Y. (1995). Interprétation de documents par cycles "perceptifs" de construction d'objets cohérents. Application aux données cadastrales. *Traitement du signal*, 12(6), 627-637.
- Olive, T., & Piolat, A. (2005). Le rôle de la mémoire de travail dans la production écrite de textes. *Psychologie française*, 50(3), 373-390.
- Orbay, G., & Kara, L. B. (2011). Beautification of design sketches using trainable stroke clustering and curve fitting. *IEEE Transactions on Visualization and Computer Graphics*, 17(5), 694-708.
- Ouyang, T. Y., & Davis, R. (2009). A visual approach to sketched symbol recognition. *Proceedings of the 21st international joint conference on artificial intelligence* (pp. 1463-1468). Morgan Kaufman Publisher Inc.
- Ozcelik, E., Arsalan-Ari, I., & Cagiltay, K. (2010). Why does signaling enhance multimedia learning? Evidence from eye movements. *Computers in Human Behavior*, 26(1), 110-117.
- Ozcelik, E., Karakus, T., Kursun, E., & Cagiltay, K. (2009). An eye-tracking study of how color coding affects multimedia learning. *Computer & Education*, 53(2), 445-453.
- Parasuraman, R., & Manzey, D. H. (2010). Complacency and bias in human use of automation: An attentional integration. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 52(3), 381-410.
- Parasuraman, R., & Riley, V. (1997, June). Humans and Automation: Use, Misuse, Disuse, Abuse. *Human Factors*, 39(2), 230-253.

- Pearson, D. G., & Sahraie, A. (2003). Oculomotor control and the maintenance. *The Quarterly Journal of Experimental Psychology*, 56A(7), 1089-1111.
- Peterson, L. R., & Peterson, M. J. (1959). Short-Term Retention of Individual Verbal Items. *Journal of Experimental Psychology*, 58(3), 193-198.
- Pickering, S. J., Gathercole, S. E., Melanle, H., & Lloyd, S. A. (2001). Development of memory for pattern and path: Further evidence for the fractionation of visuo-spatial memory. *The Quarterly Journal of Experimental Psychology*, 54A(2), 397-420.
- Pollack, I. (1952). *The assimilation of sequentially-encoded information*. Human Resources Research Laboratories, Air Research and Development Command, Bolling Air Force Base.
- Proctor, R. W., & Fagnani, C. A. (1978). Effects of distractor-stimulus modality in the Brown-Peterson distractor task. *Journal of Experimental Psychology: Human Learning and Memory*, 4(6), 676-684.
- Pugin, L. (2006). Optical music recognition of early typographic prints using hidden Markov models. *Proceedings of the 7th International Conference on Music Information Retrieval*, (pp. 53-56).
- Purnell, K. N., Solman, R. T., & Sweller, J. (1991). The effects of technical illustrations on cognitive load. *Instructional Science*, 20(5-6), 443-462.
- Quinn, J. G. (1994). Towards a clarification of spatial processing. *The Quarterly Journal of Experimental Psychology*, 47A(2), 465-480.
- Quinn, J. G., & Ralston, G. E. (1986). Movement and Attention in Visual Working Memory. *The Quarterly Journal of Experimental Psychology*, 38(4), 689-703.
- Rajkomar, A., & Blandford, A. (2011). Distributed cognition for evaluating healthcare technology. *Proceedings of the 25th BCS Conference on Human-Computer Interaction* (pp. 341-350). British Computer Society.
- Raphael, C., & Wang, J. (2011). New Approaches to Optical Music Recognition. *Proceedings of the 12th International Conference on Music Information Retrieval*, (pp. 305-310). Miami, USA.
- Rasmussen, J. (1979). *On the structure of knowledge - a morphology of mental models in a man-machine system context*. Riso National Laboratory, Denmark.
- Rasmussen, J. (1983, May). Skills, Rules, and Knowledge; Signals, Signs, and Symbols, and Other Distinctions in Human Performance Models. *IEEE Transaction on Systems, Man, and Cybernetics*, 13(3), 257-266.
- Rasmussen, J. (1986). A framework for Cognitive Task Analysis. Dans E. Hollnagel, G. Mancini, & D. Woods (Éds.), *Intelligent Decision Support Systems in Process Environment*. Berlin: Springer Verlag.

- Raynaud, D. (1999). Le schème: opérateur de la conception architecturale. *Intellectica*, 29(2), 35-69.
- Reason, J. (1993). *L'erreur humaine*. (J.-M. Hoc, Trad.) Paris: Presses Universitaires de France.
- Rebelo, A., Fujinaga, I., Paszkiewicz, F., Marcal, A. R., Guedes, C., & Cardoso, J. S. (2012). Optical music recognition: state-of-the-art and open issues. *International Journal of Multimedia Information Retrieval*, 1(3), 173-190.
- Reitman, J. S. (1971). Mechanisms of Forgetting in Short-Term Memory. *Cognitive Psychology*, 2(2), 185-195.
- Richman, H. B., Staszewski, J. J., & Simon, H. A. (1995). Simulation of Expert Memory Using EPAM IV. *Psychological Review*, 102(2), 305-330.
- Sarter, N. B., Woods, D. D., & Billings, C. E. (1997). Automation surprises. Dans G. Salvendy (Éd.), *Handbook of Human Factors & Ergonomics, second edition*. Wiley.
- Schmidt-Weigand, F., Kohnert, A., & Glowalla, U. (2010, March 16). Explaining the Modality and Contiguity Effects: New Insights From Investigating Students' Viewing Behavior. *Applied Cognitive Psychology*, 24, 226-237.
- Schuitema, G., Steg, L., & Forward, S. (2010). Explaining differences in acceptability before and acceptance after the implementation of a congestion charge in Stockholm. *Transportation Research*, 44(2), 99-109.
- Schütz, A. C., Braun, D. I., & Gegenfurtner, K. R. (2011). Eye movements and perception: A selective review. *Journal of Vision*, 11(5), 1-30.
- Senyapili, B., & Basa, Y. (2006). The Shifting Tides of Academe: Oscillation between Hand and Computer in Architectural Education. *International Journal of Technology and Design Education*, 16(3), 273-283.
- Sezgin, T. M., Stahovich, T., & Davis, R. (2007). Sketch Based Interfaces: Early Processing for Sketch Understanding. *Proceedings of the ACM SIGGRAPH 2006 COURSES*. New York.
- Sheridan, T. B., & Verplank, W. L. (1978). *Human and computer control of undersea teleoperators*. Massachusetts Institute of Technology, Office of Naval Research, Arlington.
- Skitka, L. J., Mosier, K. L., & Burdick, M. (1999). Does automation bias decision-making?. *International Journal of Human-Computer Studies*, 51(5), 991-1006.
- Smith, S. P., & Hart, J. (2006). Evaluating distributed cognitive resources for wayfinding in a desktop virtual environment. *1st IEEE Symposium on 3D User Interfaces*. Alexandria.

- Smith, S., Duke, D., & Wright, P. (1999). Using the Resources Model in Virtual Environment Design. *Workshop on User Centered Design and Implementation of Virtual Environments*, (pp. 57-72).
- Smyth, M. M. (1996). Interference with Rehearsal in Spatial Working Memory in the Absence of Eye Movements. *The Quarterly Journal of Experimental Psychology*, 49A(4), 940-949.
- Smyth, M. M., & Pendleton, L. R. (1989). Working Memory for Movements. *The Quarterly Journal of Experimental Psychology*, 41A(2), 235-250.
- Smyth, M. M., Pearson, A., & Pendleton, L. R. (1988). Movement and Working Memory: Patterns and Positions in Space. *The Quarterly Journal of Experimental Psychology*, 40A(3), 497-514.
- Suwa, M., & Tversky, B. (2002). External representations contribute to the dynamic construction of ideas. *Diagrammatic representation and inference*, pp. 149-160.
- Swanson, H. L., & Berninger, V. W. (1996). Individual Differences in Children's Working Memory and Writing Skill. *Journal of Experimental Child Psychology*, 63(54), 358-385.
- Sweller, J. (1988). Cognitive Load During Problem Solving: Effects on Learning. *Cognitive Science*, 12(2), 257-285.
- Sweller, J., & Chandler, P. (1994). Why some material is difficult to learn. *Cognition and instruction*, 12(3), 185-233.
- Sweller, J., Chandler, P., Tierney, P., & Cooper, M. (1990). Cognitive Load as a Factor in the Structuring of Technical Material. *Journal of Experimental Psychology: General*, 119(2), 176-192.
- Tarmizi, R. A., & Sweller, J. (1988). Guidance During Mathematical Problem Solving. *Journal of Experimental Psychology*, 80(4), 424-436.
- Theeuwes, J., Kramer, A. F., Hahn, S., & Irwin, D. E. (1998). Our eyes do not always go where we want them to go: Capture of the Eyes by New Objects. *Psychological Science*, 9(5), 379-385.
- Tindall-Ford, S., Chandler, P., & Sweller, J. (1997). When Two Sensory Modes Are Better Than One. *Journal of Experimental Psychology: Applied*, 3(4), 257-287.
- Torrance, M., & Galbraith, D. (2006). The processing demands of writing. Dans C. MacArthur, S. Graham, & J. Fitzgerlad (Éds.), *Handbook of Writing Research* (pp. 67-80). New York: Guilford Publications.
- Tricot, A. (1998). Charge cognitive et apprentissage. Une présentation des travaux de John Sweller. *Revue de Psychologie de L'Education*, 3, 37-64.

- Tversky, B., & Suwa, M. (2009). Thinking with sketches. Dans A. B. Markman, & A. B. Wood, *Tools for innovation: The science behind the practical methods that drive new ideas* (pp. 75-84). New York, US: Oxford University Press.
- Unwin, S. (2007). Analysing architecture through drawing. *Building Research & information*, 35(1), 101-110.
- Valveny, E., & Marti, E. (1999). Application of Deformable Template Matching to Symbol Recognition in Hand-written Architectural Drawings. *Proceedings of 5th International Conference on Document Analysis and Recognition*, (pp. 483-486). Bangalore (India).
- Venkatesh, V., Davis, F. D., & Morris, M. G. (2007). Dead or alive? The development, trajectory and future of technology adoption research. *Journal of the association for information systems*, 8(4), 267-286.
- Vergauwe, E., Barrouiller, P., & Camos, V. (2009). Visual and Spatial Working Memory Are Not That Dissociated After All: A Time-Based Resource-Sharing Account. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35(4), 1012-1028.
- Walk, H. A., & Johns, E. E. (1984). Interference and facilitation in short-term memory for odors. *Perception & Psychophysics*, 36(6), 508-514.
- Washburn, D. A., & Astur, R. S. (1998, March). Nonverbal working memory of humans and monkeys: Rehearsal in the sketchpad? *Memory & Cognition*, 26(2), 277-286.
- Wickens, C. D. (2002). Situation awareness and workload in aviation. *Current Directions in Psychological Science*, 11(4), 128-133.
- Wickens, C. D., & Carswell, C. M. (1995). The proximity compatibility principle: Its psychological foundation and relevance to display design. *Human Factors*, 37(3), 473-494.
- Wiegmann, D., McCarley, J. S., Kramer, A. F., & Wickens, C. D. (2006, August). Age and Automation Interact to Influence Performance of a Simulated Luggage Screening Task. *Aviation, Space, and Environmental Medicine*, 77(8), 825-831.
- Williams, H. L., Beaver, W. S., Spence, M. T., & Rundell, O. H. (1969). Digital and Kinesthetic Memory with Interpolated Information Processing. *Journal of Experimental Psychology*, 80(3), 530-536.
- Wilson, S., Bekker, M., Johnson, H., & Johnson, P. (1996). Costs and Benefits of User Involvement in Design: Practitioners' Views. *Proceedings of HCI on People and Computers XI*, (pp. 221-240). London.
- Wright, P. C., Fields, B., & Harrison, M. D. (1996). Distributed information resources: A new approach to interaction modelling. *Proceedings of ECCE8: Eighth European Conference on Cognitive Ergonomics*, (pp. 10-13).

Wright, P. C., Fields, R. E., & Harrison, M. D. (2000). Analysing human-computer interaction as distributed cognition: the resources model. *Human Computer Interaction*, 15(1), 1-42.

# Annexes

## Annexe 1 : Corrélations entre les scores aux tests d'empan visuel et spatial avec les scores aux épreuves de correction, recopie et complétion de plan

		Score de correction	Score de recopie	Score de complétion
Empan visuel	Corrélation de Pearson	0,005	0,393*	0,245
	Sig. (bilatéral)	0,976	0,012	0,127
	N	40	40	40
Empan spatial	Corrélation de Pearson	0,202	-0,008	0,223
	Sig. (bilatéral)	0,212	0,959	0,166
	N	40	40	40

\* $p < 0,05$

## **Annexe 2 : Corrélations entre les scores aux tests d'empan visuel et spatial avec la durée des épreuves de correction, recopie et complétion de plan**

		Durée de la correction	Durée de la recopie	Durée de la complétion
Empan visuel	Corrélation de Pearson	0,090	0,003	0,251
	Sig. (bilatéral)	0,581	0,985	0,159
	N	40	38	33
Empan spatial	Corrélation de Pearson	0,264	0,069	0,170
	Sig. (bilatéral)	0,099	0,680	0,345
	N	40	38	33

## **Annexe 3 : Corrélations entre les scores aux tests d'empan visuel et spatial avec le nombre d'Allers et Retours Visuels (ARV) effectués lors des épreuves de correction, recopie et complétion de plan**

		Nombre d'ARV pendant la correction	Nombre d'ARV pendant la recopie	Nombre d'ARV pendant la complétion
Empan visuel	Corrélation de Pearson	-0,076	0,033	0,338
	Sig. (bilatéral)	0,641	0,843	0,055
	N	40	38	33
Empan spatial	Corrélation de Pearson	0,161	0,228	0,242
	Sig. (bilatéral)	0,322	0,168	0,176
	N	40	38	33

## Annexe 4. Nombre d'erreurs repérées en fonction de la stratégie

	Erreurs signalées		Erreurs non signalées	
	M	SD	M	SD
Sans interruption	7,86	0,90	7	2,08
1 à 3 interruptions	8,4	0,55	7	1,58
5 à 8 interruptions	8	1,1	7,17	1,83

# Annexe 5. Ensemble des commentaires des participants justifiant leur non utilisation des interruptions classés par catégories

Catégories	Commentaires
Simplicité des plans	Si les plans avaient été plus complexes, je l'aurais utilisé
	Il n'y avait pas beaucoup d'erreurs, donc pas besoin
	Il n'y avait pas beaucoup d'erreurs à mémoriser. Je l'aurai utilisé s'il y en avait eu plus
	En plus, pas besoin de s'arrêter parce que le plan était simple
Lenteur de l'analyse	Le système était assez lent. S'il avait été plus rapide, je l'aurais utilisé
Gêne pour la concentration	Permet de faire tranquillement une pré-exploration avant de tout re-vérifier méthodiquement à la fin
	Je cherchais à suivre le cheminement jusqu'au bout pour rester concentré
	Il était plus facile de suivre ce qu'il fait sans intervenir.

## **Annexe 6. Ensemble des commentaires des participants précisant pourquoi l'utilisation des interruptions les a aidé**

Catégories	Commentaires
Réguler la vitesse de l'analyse	Pour ne pas perdre le fil
	Certaines pièces étaient grandes donc cela demandait à être plus attentif. J'ai fait des pauses dans ces moments-là
Difficulté de mémorisation de toutes les erreurs	Cela évite d'oublier des erreurs
	Réduit les risques d'oubli

## **Annexe 7. Ensemble des commentaires des participants estimant que les sollicitations ne les ont pas aidé classés par catégories**

Catégories	Commentaires
Sollicitations après plutôt que pendant	Je préférerais voir à la fin
	Je préfère voir ça à la fin
Fenêtre mal positionnées	La fenêtre qui arrive parfois au-dessus de l'erreur
	La fenêtre apparaissait parfois au-dessus des erreurs
	La fenêtre cache parfois les erreurs
	Déplacer la fenêtre
	Le message d'erreur est par-dessus l'erreur
Ne montre pas que des erreurs	Car souvent, ça n'était pas une erreur
Tâche simple	Pas besoin, on y arrive tout seul
Surlignage gênant	Le surlignage de l'erreur cache la proposition du système
Relâchement de l'attention	C'est une économie de réflexion pour nous, pas besoin de survoler minutieusement le plan

## **Annexe 8. Ensemble des commentaires des participants estimant que les sollicitations les ont aidé classés par catégories**

Catégories	Commentaires
Facilite la détection des erreurs	C'est une économie de réflexion pour nous, pas besoin de survoler minutieusement le plan
	Guide vers des erreurs discrètes
	Permet d'avoir des exemples d'erreurs
	Comme ça, on n'a pas besoin de voir l'erreur
	Il montre des erreurs que je n'aurais pas forcément vues
Ralenti l'analyse	Il allait vite, donc je n'avais pas le temps de suivre, donc c'était bien

## Annexe 9. Commentaires des participants sur l'interface regroupés par catégories

Catégories	Commentaires
Visibilité des cadres	Les traits sont peu visibles, trop fins
	Les encadrements sont trop fins
	Peu de différence entre les différents encadrements
	Les cadres ne sont pas assez saillants
Ordre de l'analyse	L'ordre d'apparition n'était pas logique
	La progression du logiciel n'est pas logique
	L'ordre d'apparition n'est pas logique
Trop rapide	C'est un peu trop rapide

# **Annexe 10. Items de l'échelle de confiance dans les systèmes automatisés (Jian, Bisantz, & Drury, 2000 ; notre traduction)**

Items de méfiance :

- Le système est trompeur
- Le système se comporte de manière sournoise
- Je me méfie des intentions, des actions ou des résultats du système
- Je me méfie du système
- Les actions du système auront un résultat nuisible ou préjudiciable

Items de confiance :

- Je suis confiant dans le système
- Le système assure une sécurité
- Le système est intègre
- Le système est sûr
- Le système est fiable
- Je peux faire confiance au système
- Je suis familier avec le système

# **Annexe 11. Items du questionnaire en ligne (d'après Davis, 1989)**

Items de facilité d'usage perçue :

- Apprendre à utiliser cette fonction serait facile pour moi
- Je trouverais facile d'utiliser cette fonction
- Je serais capable d'utiliser cette fonction

Items d'utilité perçue

- Je trouverais cette fonction utile dans mon travail
- Utiliser cette fonction dans mon travail me permettrait d'accomplir mes tâches plus rapidement
- Utiliser cette fonction rendrait mon travail plus facile

## Annexe 12. Réponses des participants aux questions ouvertes

Questions	Réponses
Vous-même, que feriez-vous avec ce système ?	Je pourrais réaliser les esquisses de projet de mes clients sous leurs yeux en suivant leurs commentaires après avoir pris en photo le plan de l'existant.
	Sortir des plans propres à partir d'un brouillon fait main.
	Je dessinerais mes plans d'esquisse rapidement, pour les photographier et les récupérer en version ordinateur.
	De l'esquisse de projet pour de la communication avec les clients.
Si vous le souhaitez, vous pouvez faire des commentaires ou des remarques sur le système	Prévoir peut être un outil de mise à l'échelle aussi intuitif que les autres fonctions de votre logiciel. Je vous fais confiance !
	Idée intéressante mais programme grand public (usage inutile pour un professionnel). La complexité des « modes de construire » actuelle ne peut être retranscrite avec un programme aussi simple.
	Le projet est chouette mais souvent les architectes montent leur 3D à partir de leurs plans et là le logiciel est un logiciel 2D uniquement (en opposition avec Archicad par exemple) et pas compatible avec un logiciel de 3D (type 3DS Max, sketshup...)
	On pourrait ajouter des systèmes d'isolation pour les parois, avec un système de dessin reconnaissable pour chaque type d'isolation
Si vous le souhaitez, vous pouvez faire des remarques ou commentaires sur votre	La phase de création/réflexion sur un projet est le moment le plus agréable

travail	intellectuellement dans mon travail. Cet outil pourrait rendre le rendez-vous où l'on discute du projet avec le client potentiel plus convivial peut être ! En tout cas, c'est une chouette proposition !
	Très bon projet, les points importants pour moi vont être la possibilité de modifier et ajouter des murs, d'avoir sa propre bibliothèque d'objets et de pouvoir gérer les calques, avec une différenciation entre calques de notation et calques plan.

# Annexe 13. Réponses concernant l'application « Continuum papier/DAO »

Si on peut prendre en photo un plan qu'on a fait vite à la main avec les isolants, les fenêtres et tout et que lui après il reconnaît de manière précise ce qu'on a dessiné, qu'il comprenne que c'est un isolant... S'il arrive à sortir un DWG nickel d'un plan dessiné, c'est très intéressant.

C'est vrai que si on a l'échelle, après on peut passer du papier au logiciel et on peut l'enrichir avec. Une fois qu'on a l'esquisse, on enrichit avec tous les détails. Si on n'a pas d'échelle, on s'en sert comme ça mais après on fait autre chose. Ça obligerait à tout le refaire.

Là c'est rapide (la rétro-conversion). Sur AutoCAD, il faut faire beaucoup d'étapes, c'est long. Après, on pourrait cliquer sur une baie, mettre des cotes et qu'elle se mette directement au format demandé.

Il y a pas mal de bâtiments anciens pour lesquels on n'a que le plan papier. Là en projet, on fait de la réhabilitation et du coup, on a des grands plans à la main et on a tout repris sur ordi. Donc là, ça serait quand même plus rapide.

Mais après, s'il fait la prise de photo et qu'il faut juste remettre les cotes, c'est déjà plus rapide que de tout redessiner.

Pour que ça puisse remplacer la recopie manuelle dans un logiciel, il faudrait par exemple la possibilité d'éditer un mur et de lui attribuer par exemple « mur en parpaing de telle épaisseur », « isolation extérieure ou intérieure », des choses comme ça. On rentre directement les infos techniques, la norme technique. Moi ce qui me semble compliqué, c'est pour les échelles. Il faut dessiner à l'échelle à la main. Sur AutoCAD, on dessine un trait et ensuite on lui attribue une épaisseur, un motif. (...) Cette précision serait difficile à avoir à la main.

Les fichiers qui sortiront de ce logiciel seront très bien pour la partie conception et réflexion, mais c'est difficilement exportable après tant que les problèmes d'échelle et de correspondances de traits ne sont pas réglés. Là, ça me paraît difficilement exportable sur des logiciels comme AutoCAD, ArchiCAD ou autre. Pour un travail de réflexion, de présentation c'est bien parce que c'est rapide et clair, ça permet d'avoir un dessin lisse.

## Annexe 14. Réponses concernant l'application « esquisse de projet »

Moi je m'en servais plus pour de l'esquisse de projet. Du coup, l'échelle va être donnée par la taille des portes, le lavabo, la salle d'eau... Cela peut être pratique (...) pour vérifier que les pièces que l'on a dessiné sont bien aux normes, par exemple pour une chambre il y a des surfaces minimum.

Cela peut servir à se rendre compte si le projet est viable. Parce que lorsque l'on dessine à la main, on peut ne pas faire gaffe à certaines choses. Par exemple on peut avoir dessiné un mur de 7 mètres et en face un mur de 6 mètres.

Les fichiers qui sortiront de ce logiciel seront très bien pour la partie conception et réflexion, mais c'est difficilement exportable après tant que les problèmes d'échelle et de correspondances de traits ne sont pas réglés. Là, ça me paraît difficilement exportable sur des logiciels comme AutoCAD, ArchiCAD ou autre. Pour un travail de réflexion, de présentation c'est bien parce que c'est rapide et clair, ça permet d'avoir un dessin lisse.

Juste pour la conception et la réflexion, avoir la possibilité de faire un croquis et de le mettre en vectoriel pour ensuite le tirer sur des feuilles, enfin sur des mises en pages, alors qu'il reste un croquis sans prétention d'être un plan, c'est bien.

Le fait de pouvoir bouger les meubles, c'est bien parce que ça permet d'essayer, puis finalement on bouge. À la main, il faudrait effacer donc là, c'est quand même plus rapide et facile. Mais ça serait juste pour la phase de conception parce que si on n'a pas d'échelle, on laisserait ça pour tout refaire sur autre chose.

Il pourrait faire ça (vectorisation) sur un croquis sans forcément retranscrire l'info en mémoire, mais juste reproduire le trait. Pour qu'on puisse retraiter pour les mises en page. En tant qu'étudiant, ce genre d'outil est pratique parce que lorsqu'on dessine, on peut faire un croquis d'intention, rapidement, juste des formes géométriques très simples et on ne le fait pas forcément très grand. Le fait de pouvoir le numériser, on a vraiment le choix et on a vraiment un produit numérique pour la mise en page. On peut l'agrandir, peut-être le retoucher en mettant des aplats de couleur pour souligner une zone. Comme c'est vectoriel, on peut l'agrandir, c'est plus net. Comme le logiciel peut reconnaître les traits, il faudrait pouvoir changer ces traits là et en faire par exemple un trait pointillé ou d'une couleur ou d'une épaisseur donnée. Comme ça, à partir d'un dessin tout bête, on peut révéler une idée ou quelque chose plus simplement et plus rapidement. Je pense qu'il n'y a pas beaucoup de choses à faire pour arriver à ce résultat. En complément de ce type de plan, ça peut être très bien parce qu'autant on fait un plan d'intérieur qui montre un peu les espaces, mais vu que ce n'est pas un logiciel de 3D, il faudrait pouvoir montrer un croquis qui donne un peu la figure générale du projet ou pouvoir présenter une idée forte du projet.

## **Annexe 15. Réponses concernant l'application « Edition de plan en mobilité »**

Pourvoir visiter une baraque, quand on n'a pas les plans et pouvoir dessiner soi-même directement sur la tablette, ça peut être pas mal quand on a un métré à faire ou ce genre de chose. Bon après c'est rare de ne pas trouver les plans d'une maison aussi, mais ça peut être intéressant de dessiner sur le terrain.

Des modifications sur le plan directement sur le chantier ? Je ne pense pas. C'est tout un processus les modifications à apporter si le chantier a commencé. Un gars tout seul ne va pas changer le plan sur le chantier, c'est tellement lourd... Il faut obligatoirement passer par le logiciel de base.

C'est bien parce que pour nous qui allons sur le terrain, on n'y va pas forcément. On nous ramène des plans, des fois c'est un peu brouillon pour nous qui travaillons sur ordinateur et quand on fait des plans qui ont été faits sur le terrain et qui ne sont pas toujours clairs. Là c'est sympa, cela permet sur le terrain de faire directement les modifications.

## Annexe 16. Réponses concernant l'application « relation client »

Ça a une autre utilité pour les vendeurs d'appartements. Vous achetez un appartement, vous avez vos meubles, vous allez vous dire « où est-ce que je vais les mettre ? Comment je vais pouvoir disposer les meubles dans l'appartement ? » Par exemple, j'ai une grande table de salon, un grand canapé, j'ai un canapé énorme, ou un truc comme ça, donc ça permet de réussir à dire « je pourrais disposer cela comme ça ou comme ça », ou alors par contre, vu le mobilier que j'ai, l'appartement ne convient pas, il faudrait peut-être une pièce un peu plus grande ou quelque chose comme ça. (...) Vous faites appel à un décorateur, vous lui laissez carte blanche pour décorer votre appartement que ce soit mobilier ou autres accessoires, même des appliques murales, des luminaires... Donc un décorateur, un vendeur d'appartement... C'est rapide à faire pour un client. Il a sa tablette, il a le type d'appartement, voilà, vous pouvez mettre votre table ici, votre salon là, même l'aménagement d'une cuisine, des choses comme ça...

ça pourrait être bien d'avoir la possibilité d'avoir une notion de lumière dans une pièce. Par exemple, cette pièce est totalement aveugle (montre sur le plan). Cela permettrait de montrer à un client quand cela ne marche pas bien. Parce que nous on le voit très vite la notion de lumière mais un client ne le perçoit pas. Par exemple sous forme d'un calque avec en noir ce qui va être sans lumière et en jaune ce qui va être bien éclairé. Mais quelque chose d'assez terne pour que l'on puisse voir correctement le plan et ne pas avoir l'information tout le temps affichée. Dans la relation client, souvent, les personnes ont des idées préconçues. Si on leur dit qu'ici, une fenêtre va tout changer, il va dire : une fenêtre ça va rajouter 1500 euros et du coup, on peut leur montrer l'intérêt d'ajouter une fenêtre.

Si on a rendez-vous à 16h avec un client, on arrive vers 15h, on commence à faire une esquisse, on la prend en photo, on la process et après, on modifie avec le client. Cela permettrait de montrer sur le plan directement tout en restant dans le vrai pour ne pas faire d'erreur.

J'aime bien la partie où l'on peut modifier tout ce qu'on veut. Parce que c'est une problématique quand on est en visite, le client va nous dire, si je veux une porte fenêtre ici, qu'est-ce que cela va donner. Du coup on montre sur le plan et on peut montrer les erreurs et pourquoi cela n'est pas possible. Les gens ne comprennent pas si on ne leur fait pas un petit dessin.

Généralement, nous on fait le plan en entier, on voit le client, si cela ne lui plait pas, on fait la modification. Là, avec la tablette, on pourrait faire la modification directement et cela serait plus intéressant. Quand on reçoit le client, c'est sur un ordi fixe. On fait les modifications comme ça et ça n'est pas très pratique. Aller sur le terrain et voir ça directement avec lui c'est plus intéressant.

Si on fait des plans comme ça, ou alors pour montrer à des clients si jamais on fait des maisons individuelles, là ça peut être bien parce qu'on lui dit « vous voulez votre chambre là, une fenêtre ici, etc.

Il y a parfois des clients qui dessinent mais... ça va dépendre des clients en fait, des fois il y en a qui ne se sentent pas à l'aise à dessiner et d'autres qui ont déjà réfléchi à ce qu'ils veulent. Ça leur permettrait d'être plus en confiance : « je peux montrer mes idées ». Cela permettrait plus d'échange et de confiance dans les premiers rendez-vous avec les clients. Avant de signer, ça peut permettre d'établir une bonne relation.

Ça peut être utile aussi à certains corps de métiers, par exemple en peinture. Quand vous avez acheté un appartement, vous faites vos choix, couleurs, papiers, des choses comme ça. Ça peut être transférable aussi. On sort un petit plan comme ça pour le client et qui peut être donné aussi à l'ouvrier sur place. On remet après les références, les choses comme ça et c'est propre. En plus il y a une trace derrière. Ou alors aux fournisseurs. Les entreprises dans des showrooms pour faire choisir leur clients, nous on utilisait un fournisseur qui gérait de recevoir tous nos clients. Quand vous faites 40 appartements, vous avez 40 clients et chacun a ses choix, donc tout ça il faut le gérer. Le revêtement de sol... Vous faites des choix de sol, après avec votre principe de symboles c'est facile. L'ouvrier sait que dans telle pièce, il va y avoir telle couleur, tel matériaux. Après il faut que ça puisse être imprimé et donné à l'ouvrier directement, il a ça dans sa cabane ou son local de chantier et pour lui c'est facile.

# **Index des Illustrations**

Figure 1. “Robots milling around the office halls, pass man in glass booth with sign that reads: In Case of Emergency Break Glass” (Cheney, 1989).....	9
Figure 2. Version simplifiée du modèle des 3 niveaux de performances des opérateurs (d’après Rasmussen, 1983).....	15
Figure 3. Type de tâche en fonction de la prévisibilité et du degré d’automatisation (d’après Sheridan & Verplank, 1978) .....	18
Figure 4. Production de textes et mémoire de travail (Kellogg, 1996) .....	29
Figure 5. Captures d’écran du plan dessiné par un participant (à gauche) et du plan « interprété » (à droite) sur lequel le participant a entouré les erreurs .....	33
Figure 6. Captures d’écran des épreuves électroniques des blocs de Corsi (à gauche) et des motifs visuels (à droite).....	34
Figure 7. Plan du LOUSTIC complété par un participant (gauche), modèle sur papier (centre) et plan « converti » (droite).....	35
Figure 8. Exemples de productions des participants pour les épreuves de complétion (gauche), de recopie (centre) et de correction (droite) .....	36
Figure 9. Effectif (en ordonnée) en fonction du score à la tâche de complétion (en abscisse).....	37
Figure 10. Effectif (en ordonnée) en fonction du score à l’épreuve de recopie (en abscisse).....	38
Figure 11. Effectif (en ordonnée) en fonction du score à l’épreuve de correction (en abscisse).....	38
Figure 12. Durée de la tâche (à gauche) et nombre d’allers et retours modèle/cible (à droite) en fonction du nombre d’erreurs repérées (en abscisse).....	39
Figure 13. Formats Séparé (en haut), Intégré (en bas à gauche) et Séquentiel (en bas à droite) .....	56
Figure 14. Les trois plans interprétés que les participants devaient contrôler.....	75
Figure 15. Participant entourant une erreur après avoir interrompu l’interprétation.....	76
Figure 16. Distribution des nombres d’interruptions effectuées par les participants du groupe AI.....	77
Figure 17. Moyennes et écarts-types des durées de la tâche en fonction du groupe expérimental (AI et SI) et de l’utilisation effective des interruptions .....	78
Figure 18. Les trois plans utilisés pour le test avec pour chacun, les six erreurs entourées en rouge .....	85
Figure 19. Le participant entoure une erreur avec le stylet (à gauche), le système sollicite l’utilisateur pour confirmer une interprétation incertaine (à droite) .....	86
Figure 20. Effectifs en fonction du nombre d’interruptions.....	87
Figure 21. Délai avant la première interruption en fonction du groupe .....	88
Figure 22. Durée moyenne en secondes de l’analyse, des interruptions, des sollicitations et des vérifications a posteriori pour les 3 plans cumulés .....	90
Figure 23. Distribution du nombre cumulé d’erreurs corrigées sur les 3 plans en distinguant les erreurs ayant fait l’objet d’une sollicitation de celles n’ayant pas fait l’objet de sollicitation .....	91

Figure 24. Sollicitations telles qu'elles étaient affichées lors du test (à gauche) et telles qu'elles ont été implémentées suite au test (à droite).....	98
Figure 25. Graphismes des interprétations de symboles dans le prototype utilisé pour le test (à gauche) et dans la version améliorée du prototype (à droite) .....	99
Figure 26. Focus de pré-guidage ajouté au prototype .....	100
Figure 27. Lien entre la confiance et la fiabilité réelle du système (d'après Lee & See, 2004).....	114
Figure 28. Modélisation de l'allocation des ressources attentionnelles en fonction de la confiance dans le système (d'après Parasuraman & Manzey, 2010) .....	115
Figure 29. Le système attend que l'utilisateur confirme son interprétation .....	120
Figure 30. Plans d'architecture utilisés dans l'expérience .....	121
Figure 31. Présentation des symboles et de leur interprétation .....	121
Figure 32. Moyennes et écarts-types pour la durée de la tâche dans chaque format et pour chaque rang .....	123
Figure 33. Pourcentage de participants ayant corrigé toutes les erreurs signalées par les sollicitations du système pour chacun des formats et des rangs.....	124
Figure 34. Pourcentage de participants ayant corrigé toutes les erreurs non signalées par les sollicitations du système pour chacun des formats et des rangs .....	125
Figure 35. Le TAM, d'après Davis (1989).....	135
Figure 36. Captures d'écrans des différentes étapes présentées dans la vidéo.....	137
Figure 37. Moyenne des réponses pour l'utilité perçue et l'utilisabilité perçue de chacune des fonctionnalités présentées .....	139
Figure 38. Évolution du niveau d'automatisation des prototypes au cours du projet MobiSketch, d'après l'échelle des niveaux d'automatisation de Kaber et Endsley (1997) .....	150
Figure 39. Analyse en cours sur une version de prototype incluant la fonction d'interruption de l'analyse.....	153
Figure 40. Présentations des sollicitations utilisées durant l'étude 4 (à gauche) et après leur amélioration dans le prototype (à droite) .....	155
Figure 41. Capture d'écran d'un plan interprété utilisé dans l'étude 4 (à gauche) et d'un plan interprété après la prise en compte des préconisations (à droite) .....	157
Figure 42. Exemple de focus signalant en temps réel à l'utilisateur la zone du plan traitée par le système (en bleu clair).....	157
Figure 43. Capture d'écran d'un plan interprété avec affichage de propositions de corrections .....	159

# **Index des Tableaux**

Tableau 1. Les niveaux d'automatisations (Kaber & Endsley, 1997; Endsley & Kaber, 1999).....	23
Tableau 2. Moyennes et écart-types des durées de la tâche dans les groupes Séparé et Intégré.....	58
Tableau 3. Proportion de participants ayant repéré toutes les erreurs dans les groupes Séparé, Intégré et Séquentiel .....	59
Tableau 4. Moyennes et écart-types pour la durée de la tâche dans les groupes Séparé, Intégré et Séquentiel .....	63
Tableau 5. Proportion de participants ayant repéré toutes les erreurs .....	64
Tableau 6. Moyennes et écart-type des temps de fixations cumulés sur les erreurs et en dehors des zones d'erreur .....	65
Tableau 7. Proportion de participants ayant contrôlé l'ensemble des erreurs .....	66
Tableau 8. Durée de la tâche et erreurs détectées dans les groupes SI et AI .....	77
Tableau 9. Durée de la tâche et erreurs détectées avec et sans pauses .....	79
Tableau 10. Les évaluations de la crédibilité (Fogg & Tseng, 1999 ; notre traduction) .....	112
Tableau 11. Moyennes et écarts-types des scores de confiance et de méfiance.....	126

---

## LE ROLE DE L'UTILISATEUR DANS LES SYSTEMES DE TRAITEMENTS AUTOMATIQUES

---

### Résumé

Actuellement, l'interprétation automatique de symboles peut être appliquée à de nombreux types de documents techniques (partitions musicales, plans électriques, etc). L'automatisation de ce type de tâche peut permettre un gain de temps pour des professionnels (Lu, Tai, Su, & Cai, 2005), mais elle peut aussi générer une perte de conscience de la situation par les opérateurs (Endsley, 1995). De plus, ces traitements impliquent un risque d'erreurs d'interprétation des symboles (Orbay & Kara, 2011). Il est donc crucial qu'un humain intervienne pour corriger ces erreurs. L'objectif de cette thèse était d'améliorer la compréhension des interactions homme-machine prenant place dans ce type de tâche afin de déterminer comment améliorer la performance du système. Plusieurs études ont ainsi été réalisées dans le cadre d'une démarche de conception centrée-utilisateur (ISO 9241-210, 2010) d'un logiciel d'interprétation automatique de plans d'architecture. Ces études se sont appuyées sur un cadre théorique varié, puisant à la fois dans la psychologie cognitive, les interactions homme-machine et l'ergonomie des documents multimédias. Elles ont mis en évidence des résultats débouchant sur des préconisations pour la conception de systèmes de reconnaissance. Ainsi, l'affichage de l'interprétation superposée au document interprété facilite la coréférenciation des informations à contrôler. De plus, permettre à l'utilisateur d'assister en temps réel à l'analyse et d'intervenir au fur et à mesure améliore les performances. Enfin, le signalement par la machine des risques d'erreurs est bénéfique, mais cela peut générer une réduction de la vigilance due à un phénomène de sur-confiance.

### Mots clés

Conception centrée-utilisateur, interaction homme-machine, ergonomie, interprétation automatique, supervision

---

## THE ROLE OF THE USER IN AUTOMATIC PROCESSING SYSTEMS

---

### Abstract

*Automatic symbol recognition can currently be applied to a wide range of technical documents (music scores, electrical diagrams, etc.). Automating this kind of task can help some professionals save time (Lu et al., 2005), but it can also generate a loss of situation awareness among operators (Endsley, 1995). Furthermore, as these systems sometimes interpret symbols incorrectly (Lu, Tai, Su, & Cai, 2005; Orbay & Kara, 2011), humans need to be able to intervene in the process to correct any errors as and when they arise. The objective of this thesis was to improve our understanding of the human-machine interactions elicited by this kind of task, in order to determine how best to enhance system performance. We conducted a series of studies during the development of a user-centred design (ISO 9241-210, 2010) of automatic interpretation software for architectural plans. These studies were set within a heterogeneous theoretical framework bringing together cognitive psychology, human-machine interactions and the ergonomics of the multimedia documents. The results they yielded led to several recommendations for designing recognition systems. For instance, superimposing the interpretation on the interpreted document makes it easier to cross-check the information. Furthermore, allowing users to monitor the interpretation process in real time and interrupt it at any point increases performances. Finally, while the highlighting of potential errors can also be beneficial, it can also make users over-confident and thus less watchful.*

### Keywords

*User-centred design, human-computer interaction, ergonomics, automatic interpretation, supervision*